

The American Economic Review

VOLUME XLVIII

JUNE 1958

NUMBER THREE

THE COST OF CAPITAL, CORPORATION FINANCE AND THE THEORY OF INVESTMENT

By FRANCO MODIGLIANI AND MERTON H. MILLER*

What is the "cost of capital" to a firm in a world in which funds are used to acquire assets whose yields are uncertain; and in which capital can be obtained by many different media, ranging from pure debt instruments, representing money-fixed claims, to pure equity issues, giving holders only the right to a pro-rata share in the uncertain venture? This question has vexed at least three classes of economists: (1) the corporation finance specialist concerned with the techniques of financing firms so as to ensure their survival and growth; (2) the managerial economist concerned with capital budgeting; and (3) the economic theorist concerned with explaining investment behavior at both the micro and macro levels.¹

In much of his formal analysis, the economic theorist at least has tended to side-step the essence of this cost-of-capital problem by proceeding as though physical assets—like bonds—could be regarded as yielding known, sure streams. Given this assumption, the theorist has concluded that the cost of capital to the owners of a firm is simply the rate of interest on bonds; and has derived the familiar proposition that the firm, acting rationally, will tend to push investment to the point

* The authors are, respectively, professor and associate professor of economics in the Graduate School of Industrial Administration, Carnegie Institute of Technology. This article is a revised version of a paper delivered at the annual meeting of the Econometric Society, December 1956. The authors express thanks for the comments and suggestions made at that time by the discussants of the paper, Evsey Domar, Robert Eisner and John Lintner, and subsequently by James Duesenberry. They are also greatly indebted to many of their present and former colleagues and students at Carnegie Tech who served so often and with such remarkable patience as a critical forum for the ideas here presented.

¹ The literature bearing on the cost-of-capital problem is far too extensive for listing here. Numerous references to it will be found throughout the paper though we make no claim to completeness. One phase of the problem which we do not consider explicitly, but which has a considerable literature of its own is the relation between the cost of capital and public utility rates. For a recent summary of the "cost-of-capital theory" of rate regulation and a brief discussion of some of its implications, the reader may refer to H. M. Somers [20].

where the marginal yield on physical assets is equal to the market rate of interest.² This proposition can be shown to follow from either of two criteria of rational decision-making which are equivalent under certainty, namely (1) the maximization of profits and (2) the maximization of market value.

According to the first criterion, a physical asset is worth acquiring if it will increase the net profit of the owners of the firm. But net profit will increase only if the expected rate of return, or yield, of the asset exceeds the rate of interest. According to the second criterion, an asset is worth acquiring if it increases the value of the owners' equity, *i.e.*, if it adds more to the market value of the firm than the costs of acquisition. But what the asset adds is given by capitalizing the stream it generates at the market rate of interest, and this capitalized value will exceed its cost if and only if the yield of the asset exceeds the rate of interest. Note that, under either formulation, the cost of capital is equal to the rate of interest on bonds, regardless of whether the funds are acquired through debt instruments or through new issues of common stock. Indeed, in a world of sure returns, the distinction between debt and equity funds reduces largely to one of terminology.

It must be acknowledged that some attempt is usually made in this type of analysis to allow for the existence of uncertainty. This attempt typically takes the form of superimposing on the results of the certainty analysis the notion of a "risk discount" to be subtracted from the expected yield (or a "risk premium" to be added to the market rate of interest). Investment decisions are then supposed to be based on a comparison of this "risk adjusted" or "certainty equivalent" yield with the market rate of interest.³ No satisfactory explanation has yet been provided, however, as to what determines the size of the risk discount and how it varies in response to changes in other variables.

Considered as a convenient approximation, the model of the firm constructed via this certainty—or certainty-equivalent—approach has admittedly been useful in dealing with some of the grosser aspects of the processes of capital accumulation and economic fluctuations. Such a model underlies, for example, the familiar Keynesian aggregate investment function in which aggregate investment is written as a function of the rate of interest—the same riskless rate of interest which appears later in the system in the liquidity-preference equation. Yet few would maintain that this approximation is adequate. At the macroeconomic level there are ample grounds for doubting that the rate of interest has

² Or, more accurately, to the marginal cost of borrowed funds since it is customary, at least in advanced analysis, to draw the supply curve of borrowed funds to the firm as a rising one. For an advanced treatment of the certainty case, see F. and V. Lutz [13].

³ The classic examples of the certainty-equivalent approach are found in J. R. Hicks [8] and O. Lange [11].

as large and as direct an influence on the rate of investment as this analysis would lead us to believe. At the microeconomic level the certainty model has little descriptive value and provides no real guidance to the finance specialist or managerial economist whose main problems cannot be treated in a framework which deals so cavalierly with uncertainty and ignores all forms of financing other than debt issues.⁴

Only recently have economists begun to face up seriously to the problem of the cost of capital *cum* risk. In the process they have found their interests and endeavors merging with those of the finance specialist and the managerial economist who have lived with the problem longer and more intimately. In this joint search to establish the principles which govern rational investment and financial policy in a world of uncertainty two main lines of attack can be discerned. These lines represent, in effect, attempts to extrapolate to the world of uncertainty each of the two criteria—profit maximization and market value maximization—which were seen to have equivalent implications in the special case of certainty. With the recognition of uncertainty this equivalence vanishes. In fact, the profit maximization criterion is no longer even well defined. Under uncertainty there corresponds to each decision of the firm not a unique profit outcome, but a plurality of mutually exclusive outcomes which can at best be described by a subjective probability distribution. The profit outcome, in short, has become a random variable and as such its maximization no longer has an operational meaning. Nor can this difficulty generally be disposed of by using the mathematical expectation of profits as the variable to be maximized. For decisions which affect the expected value will also tend to affect the dispersion and other characteristics of the distribution of outcomes. In particular, the use of debt rather than equity funds to finance a given venture may well increase the expected return to the owners, but only at the cost of increased dispersion of the outcomes.

Under these conditions the profit outcomes of alternative investment and financing decisions can be compared and ranked only in terms of a *subjective* “utility function” of the owners which weighs the expected yield against other characteristics of the distribution. Accordingly, the extrapolation of the profit maximization criterion of the certainty model has tended to evolve into utility maximization, sometimes explicitly, more frequently in a qualitative and heuristic form.⁵

The utility approach undoubtedly represents an advance over the certainty or certainty-equivalent approach. It does at least permit us

⁴ Those who have taken a “case-method” course in finance in recent years will recall in this connection the famous Liguigas case of Hunt and Williams, [9, pp. 193–96] a case which is often used to introduce the student to the cost-of-capital problem and to poke a bit of fun at the economist’s certainty-model.

⁵ For an attempt at a rigorous explicit development of this line of attack, see F. Modigliani and M. Zeman [14].

to explore (within limits) some of the implications of different financing arrangements, and it does give some meaning to the "cost" of different types of funds. However, because the cost of capital has become an essentially subjective concept, the utility approach has serious drawbacks for normative as well as analytical purposes. How, for example, is management to ascertain the risk preferences of its stockholders and to compromise among their tastes? And how can the economist build a meaningful investment function in the face of the fact that any given investment opportunity might or might not be worth exploiting depending on precisely who happen to be the owners of the firm at the moment?

Fortunately, these questions do not have to be answered; for the alternative approach, based on market value maximization, can provide the basis for an operational definition of the cost of capital and a workable theory of investment. Under this approach any investment project and its concomitant financing plan must pass only the following test: Will the project, as financed, raise the market value of the firm's shares? If so, it is worth undertaking; if not, its return is less than the marginal cost of capital to the firm. Note that such a test is entirely independent of the tastes of the current owners, since market prices will reflect not only their preferences but those of all potential owners as well. If any current stockholder disagrees with management and the market over the valuation of the project, he is free to sell out and reinvest elsewhere, but will still benefit from the capital appreciation resulting from management's decision.

The potential advantages of the market-value approach have long been appreciated; yet analytical results have been meager. What appears to be keeping this line of development from achieving its promise is largely the lack of an adequate theory of the effect of financial structure on market valuations, and of how these effects can be inferred from objective market data. It is with the development of such a theory and of its implications for the cost-of-capital problem that we shall be concerned in this paper.

Our procedure will be to develop in Section I the basic theory itself and to give some brief account of its empirical relevance. In Section II, we show how the theory can be used to answer the cost-of-capital question and how it permits us to develop a theory of investment of the firm under conditions of uncertainty. Throughout these sections the approach is essentially a partial-equilibrium one focusing on the firm and "industry." Accordingly, the "prices" of certain income streams will be treated as constant and given from outside the model, just as in the standard Marshallian analysis of the firm and industry the prices of all inputs and of all other products are taken as given. We have chosen to focus at this level rather than on the economy as a whole because it

is at the level of the firm and the industry that the interests of the various specialists concerned with the cost-of-capital problem come most closely together. Although the emphasis has thus been placed on partial-equilibrium analysis, the results obtained also provide the essential building blocks for a general equilibrium model which shows how those prices which are here taken as given, are themselves determined. For reasons of space, however, and because the material is of interest in its own right, the presentation of the general equilibrium model which rounds out the analysis must be deferred to a subsequent paper.

I. *The Valuation of Securities, Leverage, and the Cost of Capital*

A. *The Capitalization Rate for Uncertain Streams*

As a starting point, consider an economy in which all physical assets are owned by corporations. For the moment, assume that these corporations can finance their assets by issuing common stock only; the introduction of bond issues, or their equivalent, as a source of corporate funds is postponed until the next part of this section.

The physical assets held by each firm will yield to the owners of the firm—its stockholders—a stream of “profits” over time; but the elements of this series need not be constant and in any event are uncertain. This stream of income, and hence the stream accruing to any share of common stock, will be regarded as extending indefinitely into the future. We assume, however, that the mean value of the stream over time, or average profit per unit of time, is finite and represents a random variable subject to a (subjective) probability distribution. We shall refer to the average value over time of the stream accruing to a given share as the return of that share; and to the mathematical expectation of this average as the expected return of the share.⁶ Although individual investors may have different views as to the shape of the probability distri-

⁶ These propositions can be restated analytically as follows: The assets of the i th firm generate a stream:

$$X_i(1), X_i(2) \cdots X_i(T)$$

whose elements are random variables subject to the joint probability distribution:

$$\chi_i[X_i(1), X_i(2) \cdots X_i(t)].$$

The return to the i th firm is defined as:

$$X_i = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T X_i(t).$$

X_i is itself a random variable with a probability distribution $\Phi_i(X_i)$ whose form is determined uniquely by χ_i . The expected return $\bar{X}_i$ is defined as $\bar{X}_i = E(X_i) = \int X_i \Phi_i(X_i) dX_i$. If N_i is the number of shares outstanding, the return of the i th share is $x_i = (1/N_i)X_i$ with probability distribution $\phi_i(x_i)dx_i = \Phi_i(N_i x_i)d(N_i x_i)$ and expected value $\bar{x}_i = (1/N_i)\bar{X}_i$.

bution of the return of any share, we shall assume for simplicity that they are at least in agreement as to the expected return.⁷

This way of characterizing uncertain streams merits brief comment. Notice first that the stream is a stream of profits, not dividends. As will become clear later, as long as management is presumed to be acting in the best interests of the stockholders, retained earnings can be regarded as equivalent to a fully subscribed, pre-emptive issue of common stock. Hence, for present purposes, the division of the stream between cash dividends and retained earnings in any period is a mere detail. Notice also that the uncertainty attaches to the mean value over time of the stream of profits and should not be confused with variability over time of the successive elements of the stream. That variability and uncertainty are two totally different concepts should be clear from the fact that the elements of a stream can be variable even though known with certainty. It can be shown, furthermore, that whether the elements of a stream are sure or uncertain, the effect of variability per se on the valuation of the stream is at best a second-order one which can safely be neglected for our purposes (and indeed most others too).⁸

The next assumption plays a strategic role in the rest of the analysis. We shall assume that firms can be divided into "equivalent return" classes such that the return on the shares issued by any firm in any given class is proportional to (and hence perfectly correlated with) the return on the shares issued by any other firm in the same class. This assumption implies that the various shares within the same class differ, at most, by a "scale factor." Accordingly, if we adjust for the difference in scale, by taking the *ratio* of the return to the expected return, the probability distribution of that ratio is identical for all shares in the class. It follows that all relevant properties of a share are uniquely characterized by specifying (1) the class to which it belongs and (2) its expected return.

The significance of this assumption is that it permits us to classify firms into groups within which the shares of different firms are "homogeneous," that is, perfect substitutes for one another. We have, thus, an analogue to the familiar concept of the industry in which it is the commodity produced by the firms that is taken as homogeneous. To complete this analogy with Marshallian price theory, we shall assume in the

⁷ To deal adequately with refinements such as differences among investors in estimates of expected returns would require extensive discussion of the theory of portfolio selection. Brief references to these and related topics will be made in the succeeding article on the general equilibrium model.

⁸ The reader may convince himself of this by asking how much he would be willing to rebate to his employer for the privilege of receiving his annual salary in equal monthly installments rather than in irregular amounts over the year. See also J. M. Keynes [10, esp. pp. 53–54].

analysis to follow that the shares concerned are traded in perfect markets under conditions of atomistic competition.⁹

From our definition of homogeneous classes of stock it follows that in equilibrium in a perfect capital market the price per dollar's worth of expected return must be the same for all shares of any given class. Or, equivalently, in any given class the price of every share must be proportional to its expected return. Let us denote this factor of proportionality for any class, say the k th class, by $1/\rho_k$. Then if p_j denotes the price and $\bar{x}_j$ is the expected return per share of the j th firm in class k , we must have:

$$(1) \quad p_j = \frac{1}{\rho_k} \bar{x}_j;$$

or, equivalently,

$$(2) \quad \frac{\bar{x}_j}{p_j} = \rho_k \text{ a constant for all firms } j \text{ in class } k.$$

The constants ρ_k (one for each of the k classes) can be given several economic interpretations: (a) From (2) we see that each ρ_k is the expected rate of return of any share in class k . (b) From (1) $1/\rho_k$ is the price which an investor has to pay for a dollar's worth of expected return in the class k . (c) Again from (1), by analogy with the terminology for perpetual bonds, ρ_k can be regarded as the market rate of capitalization for the expected value of the uncertain streams of the kind generated by the k th class of firms.¹⁰

B. Debt Financing and Its Effects on Security Prices

Having developed an apparatus for dealing with uncertain streams we can now approach the heart of the cost-of-capital problem by dropping the assumption that firms cannot issue bonds. The introduction of debt-financing changes the market for shares in a very fundamental way. Because firms may have different proportions of debt in their capi-

⁹ Just what our classes of stocks contain and how the different classes can be identified by outside observers are empirical questions to which we shall return later. For the present, it is sufficient to observe: (1) Our concept of a class, while not identical to that of the industry is at least closely related to it. Certainly the basic characteristics of the probability distributions of the returns on assets will depend to a significant extent on the product sold and the technology used. (2) What are the appropriate class boundaries will depend on the particular problem being studied. An economist concerned with general tendencies in the market, for example, might well be prepared to work with far wider classes than would be appropriate for an investor planning his portfolio, or a firm planning its financial strategy.

¹⁰ We cannot, on the basis of the assumptions so far, make any statements about the relationship or spread between the various ρ 's or capitalization rates. Before we could do so we would have to make further specific assumptions about the way investors believe the probability distributions vary from class to class, as well as assumptions about investors' preferences as between the characteristics of different distributions.

tal structure, shares of different companies, even in the same class, can give rise to different probability distributions of returns. In the language of finance, the shares will be subject to different degrees of financial risk or "leverage" and hence they will no longer be perfect substitutes for one another.

To exhibit the mechanism determining the relative prices of shares under these conditions, we make the following two assumptions about the nature of bonds and the bond market, though they are actually stronger than is necessary and will be relaxed later: (1) All bonds (including any debts issued by households for the purpose of carrying shares) are assumed to yield a constant income per unit of time, and this income is regarded as certain by all traders regardless of the issuer. (2) Bonds, like stocks, are traded in a perfect market, where the term perfect is to be taken in its usual sense as implying that any two commodities which are perfect substitutes for each other must sell, in equilibrium, at the same price. It follows from assumption (1) that all bonds are in fact perfect substitutes up to a scale factor. It follows from assumption (2) that they must all sell at the same price per dollar's worth of return, or what amounts to the same thing must yield the same rate of return. This rate of return will be denoted by r and referred to as the rate of interest or, equivalently, as the capitalization rate for sure streams. We now can derive the following two basic propositions with respect to the valuation of securities in companies with different capital structures:

Proposition I. Consider any company j and let $\bar{X}_j$ stand as before for the expected return on the assets owned by the company (that is, its expected profit before deduction of interest). Denote by D_j the market value of the debts of the company; by S_j the market value of its common shares; and by $V_j \equiv S_j + D_j$ the market value of all its securities or, as we shall say, the market value of the firm. Then, our Proposition I asserts that we must have in equilibrium:

$$(3) \quad V_j \equiv (S_j + D_j) = \bar{X}_j / \rho_k, \text{ for any firm } j \text{ in class } k.$$

That is, the market value of any firm is independent of its capital structure and is given by capitalizing its expected return at the rate ρ_k appropriate to its class.

This proposition can be stated in an equivalent way in terms of the firm's "average cost of capital," $\bar{X}_j / V_j$, which is the ratio of its expected return to the market value of all its securities. Our proposition then is:

$$(4) \quad \frac{\bar{X}_j}{(S_j + D_j)} \equiv \frac{\bar{X}_j}{V_j} = \rho_k, \text{ for any firm } j, \text{ in class } k.$$

That is, the average cost of capital to any firm is completely independent of

its capital structure and is equal to the capitalization rate of a pure equity stream of its class.

To establish Proposition I we will show that as long as the relations (3) or (4) do not hold between any pair of firms in a class, arbitrage will take place and restore the stated equalities. We use the term arbitrage advisedly. For if Proposition I did not hold, an investor could buy and sell stocks and bonds in such a way as to exchange one income stream for another stream, identical in all relevant respects but selling at a lower price. The exchange would therefore be advantageous to the investor quite independently of his attitudes toward risk.¹¹ As investors exploit these arbitrage opportunities, the value of the overpriced shares will fall and that of the underpriced shares will rise, thereby tending to eliminate the discrepancy between the market values of the firms.

By way of proof, consider two firms in the same class and assume for simplicity only, that the expected return, X , is the same for both firms. Let company 1 be financed entirely with common stock while company 2 has some debt in its capital structure. Suppose first the value of the levered firm, V_2 , to be larger than that of the unlevered one, V_1 . Consider an investor holding s_2 dollars' worth of the shares of company 2, representing a fraction α of the total outstanding stock, S_2 . The return from this portfolio, denoted by Y_2 , will be a fraction α of the income available for the stockholders of company 2, which is equal to the total return X_2 less the interest charge, rD_2 . Since under our assumption of homogeneity, the anticipated total return of company 2, X_2 , is, under all circumstances, the same as the anticipated total return to company 1, X_1 , we can hereafter replace X_2 and X_1 by a common symbol X . Hence, the return from the initial portfolio can be written as:

$$(5) \quad Y_2 = \alpha(X - rD_2).$$

Now suppose the investor sold his αS_2 worth of company 2 shares and acquired instead an amount $s_1 = \alpha(S_2 + D_2)$ of the shares of company 1. He could do so by utilizing the amount αS_2 realized from the sale of his initial holding and borrowing an additional amount αD_2 on his own credit, pledging his new holdings in company 1 as a collateral. He would thus secure for himself a fraction $s_1/S_1 = \alpha(S_2 + D_2)/S_1$ of the shares and earnings of company 1. Making proper allowance for the interest payments on his personal debt αD_2 , the return from the new portfolio, Y_1 , is given by:

¹¹ In the language of the theory of choice, the exchanges are movements from inefficient points in the interior to efficient points on the boundary of the investor's opportunity set; and not movements between efficient points along the boundary. Hence for this part of the analysis nothing is involved in the way of specific assumptions about investor attitudes or behavior other than that investors behave consistently and prefer more income to less income, *ceteris paribus*.

$$(6) \quad Y_1 = \frac{\alpha(S_2 + D_2)}{S_1} X - r\alpha D_2 = \alpha \frac{V_2}{V_1} X - r\alpha D_2.$$

Comparing (5) with (6) we see that as long as $V_2 > V_1$ we must have $Y_1 > Y_2$, so that it pays owners of company 2's shares to sell their holdings, thereby depressing S_2 and hence V_2 ; and to acquire shares of company 1, thereby raising S_1 and thus V_1 . We conclude therefore that levered companies cannot command a premium over unlevered companies because investors have the opportunity of putting the equivalent leverage into their portfolio directly by borrowing on personal account.

Consider now the other possibility, namely that the market value of the levered company V_2 is less than V_1 . Suppose an investor holds initially an amount s_1 of shares of company 1, representing a fraction α of the total outstanding stock, S_1 . His return from this holding is:

$$Y_1 = \frac{s_1}{S_1} X = \alpha X.$$

Suppose he were to exchange this initial holding for another portfolio, also worth s_1 , but consisting of s_2 dollars of stock of company 2 and of d dollars of bonds, where s_2 and d are given by:

$$(7) \quad s_2 = \frac{S_2}{V_2} s_1, \quad d = \frac{D_2}{V_2} s_1.$$

In other words the new portfolio is to consist of stock of company 2 and of bonds in the proportions S_2/V_2 and D_2/V_2 , respectively. The return from the stock in the new portfolio will be a fraction s_2/S_2 of the total return to stockholders of company 2, which is $(X - rD_2)$, and the return from the bonds will be rd . Making use of (7), the total return from the portfolio, Y_2 , can be expressed as follows:

$$Y_2 = \frac{s_2}{S_2} (X - rD_2) + rd = \frac{s_1}{V_2} (X - rD_2) + r \frac{D_2}{V_2} s_1 = \frac{s_1}{V_2} X = \alpha \frac{S_1}{V_2} X$$

(since $s_1 = \alpha S_1$). Comparing Y_2 with Y_1 we see that, if $V_2 < S_1 \equiv V_1$, then Y_2 will exceed Y_1 . Hence it pays the holders of company 1's shares to sell these holdings and replace them with a mixed portfolio containing an appropriate fraction of the shares of company 2.

The acquisition of a mixed portfolio of stock of a levered company j and of bonds in the proportion S_j/V_j and D_j/V_j respectively, may be regarded as an operation which "undoes" the leverage, giving access to an appropriate fraction of the unlevered return X_j . It is this possibility of undoing leverage which prevents the value of levered firms from being consistently less than those of unlevered firms, or more generally prevents the average cost of capital $\bar{X}_j/V_j$ from being systematically higher for levered than for nonlevered companies in the same class.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

Since we have already shown that arbitrage will also prevent V_2 from being larger than V_1 , we can conclude that in equilibrium we must have $V_2 = V_1$, as stated in Proposition I.

Proposition II. From Proposition I we can derive the following proposition concerning the rate of return on common stock in companies whose capital structure includes some debt: the expected rate of return or yield, i , on the stock of any company j belonging to the k th class is a linear function of leverage as follows:

$$(8) \quad i_j = \rho_k + (\rho_k - r)D_j/S_j.$$

That is, *the expected yield of a share of stock is equal to the appropriate capitalization rate ρ_k for a pure equity stream in the class, plus a premium related to financial risk equal to the debt-to-equity ratio times the spread between ρ_k and r .* Or equivalently, the market price of any share of stock is given by capitalizing its expected return at the continuously variable rate i_j of (8).¹²

A number of writers have stated close equivalents of our Proposition I although by appealing to intuition rather than by attempting a proof and only to insist immediately that the results were not applicable to the actual capital markets.¹³ Proposition II, however, so far as we have been able to discover is new.¹⁴ To establish it we first note that, by definition, the expected rate of return, i , is given by:

$$(9) \quad i_j \equiv \frac{\bar{X}_j - rD_j}{S_j}.$$

From Proposition I, equation (3), we know that:

$$\bar{X}_j = \rho_k(S_j + D_j).$$

Substituting in (9) and simplifying, we obtain equation (8).

¹² To illustrate, suppose $\bar{X} = 1000$, $D = 4000$, $r = 5$ per cent and $\rho_k = 10$ per cent. These values imply that $V = 10,000$ and $S = 6000$ by virtue of Proposition I. The expected yield or rate of return per share is then:

$$i = \frac{1000 - 200}{6000} = .1 + (.1 - .05) \frac{4000}{6000} = 13\frac{1}{3} \text{ per cent.}$$

¹³ See, for example, J. B. Williams [21, esp. pp. 72-73]; David Durand [3]; and W. A. Morton [15]. None of these writers describe in any detail the mechanism which is supposed to keep the average cost of capital constant under changes in capital structure. They seem, however, to be visualizing the equilibrating mechanism in terms of switches by investors between stocks and bonds as the yields of each get out of line with their "riskiness." This is an argument quite different from the pure arbitrage mechanism underlying our proof, and the difference is crucial. Regarding Proposition I as resting on investors' attitudes toward risk leads inevitably to a misunderstanding of many factors influencing relative yields such as, for example, limitations on the portfolio composition of financial institutions. See below, esp. Section I.D.

¹⁴ Morton does make reference to a linear yield function but only " . . . for the sake of simplicity and because the particular function used makes no essential difference in my conclusions" [15, p. 443, note 2].

C. *Some Qualifications and Extensions of the Basic Propositions*

The methods and results developed so far can be extended in a number of useful directions, of which we shall consider here only three: (1) allowing for a corporate profits tax under which interest payments are deductible; (2) recognizing the existence of a multiplicity of bonds and interest rates; and (3) acknowledging the presence of market imperfections which might interfere with the process of arbitrage. The first two will be examined briefly in this section with some further attention given to the tax problem in Section II. Market imperfections will be discussed in Part D of this section in the course of a comparison of our results with those of received doctrines in the field of finance.

Effects of the Present Method of Taxing Corporations. The deduction of interest in computing taxable corporate profits will prevent the arbitrage process from making the value of all firms in a given class proportional to the expected returns generated by their physical assets. Instead, it can be shown (by the same type of proof used for the original version of Proposition I) that the market values of firms in each class must be proportional in equilibrium to their expected return net of taxes (that is, to the sum of the interest paid and expected net stockholder income). This means we must replace each $\bar{X}_j$ in the original versions of Propositions I and II with a new variable $\bar{X}_j^\tau$ representing the total income net of taxes generated by the firm:

$$(10) \quad \bar{X}_j^\tau \equiv (\bar{X}_j - rD_j)(1 - \tau) + rD_j \equiv \bar{\pi}_j^\tau + rD_j,$$

where $\bar{\pi}_j^\tau$ represents the expected net income accruing to the common stockholders and τ stands for the average rate of corporate income tax.¹⁵

After making these substitutions, the propositions, when adjusted for taxes, continue to have the same form as their originals. That is, Proposition I becomes:

$$(11) \quad \frac{\bar{X}_j^\tau}{V_j} = \rho_k^\tau, \text{ for any firm in class } k,$$

and Proposition II becomes

$$(12) \quad i_j \equiv \frac{\bar{\pi}_j^\tau}{S_j} = \rho_j^\tau + (\rho_k^\tau - r)D_j/S_j$$

where ρ_k^τ is the capitalization rate for income net of taxes in class k .

Although the form of the propositions is unaffected, certain interpretations must be changed. In particular, the after-tax capitalization rate

¹⁵ For simplicity, we shall ignore throughout the tiny element of progression in our present corporate tax and treat τ as a constant independent of $(X_j - rD_j)$.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

ρ_k^r can no longer be identified with the "average cost of capital" which is $\rho_k = \bar{X}_j / V_j$. The difference between ρ_k^r and the "true" average cost of capital, as we shall see, is a matter of some relevance in connection with investment planning within the firm (Section II). For the description of market behavior, however, which is our immediate concern here, the distinction is not essential. To simplify presentation, therefore, and to preserve continuity with the terminology in the standard literature we shall continue in this section to refer to ρ_k^r as the average cost of capital, though strictly speaking this identification is correct only in the absence of taxes.

Effects of a Plurality of Bonds and Interest Rates. In existing capital markets we find not one, but a whole family of interest rates varying with maturity, with the technical provisions of the loan and, what is most relevant for present purposes, with the financial condition of the borrower.¹⁶ Economic theory and market experience both suggest that the yields demanded by lenders tend to increase with the debt-equity ratio of the borrowing firm (or individual). If so, and if we can assume as a first approximation that this yield curve, $r = r(D/S)$, whatever its precise form, is the same for all borrowers, then we can readily extend our propositions to the case of a rising supply curve for borrowed funds.¹⁷

Proposition I is actually unaffected in form and interpretation by the fact that the rate of interest may rise with leverage; while the average cost of *borrowed* funds will tend to increase as debt rises, the average cost of funds from *all* sources will still be independent of leverage (apart from the tax effect). This conclusion follows directly from the ability of those who engage in arbitrage to undo the leverage in any financial structure by acquiring an appropriately mixed portfolio of bonds and stocks. Because of this ability, the ratio of earnings (*before* interest charges) to market value—*i.e.*, the average cost of capital from all

¹⁶ We shall not consider here the extension of the analysis to encompass the time structure of interest rates. Although some of the problems posed by the time structure can be handled within our comparative statics framework, an adequate discussion would require a separate paper.

¹⁷ We can also develop a theory of bond valuation along lines essentially parallel to those followed for the case of shares. We conjecture that the curve of bond yields as a function of leverage will turn out to be a nonlinear one in contrast to the linear function of leverage developed for common shares. However, we would also expect that the rate of increase in the yield on new issues would not be substantial in practice. This relatively slow rise would reflect the fact that interest rate increases by themselves can never be completely satisfactory to creditors as compensation for their increased risk. Such increases may simply serve to raise r so high relative to ρ that they become self-defeating by giving rise to a situation in which even normal fluctuations in earnings may force the company into bankruptcy. The difficulty of borrowing more, therefore, tends to show up in the usual case not so much in higher rates as in the form of increasingly stringent restrictions imposed on the company's management and finances by the creditors; and ultimately in a complete inability to obtain new borrowed funds, at least from the institutional investors who normally set the standards in the market for bonds.

sources—must be the same for all firms in a given class.¹⁸ In other words, the increased cost of borrowed funds as leverage increases will tend to be offset by a corresponding reduction in the yield of common stock. This seemingly paradoxical result will be examined more closely below in connection with Proposition II.

A significant modification of Proposition I would be required only if the yield curve $r=r(D/S)$ were different for different borrowers, as might happen if creditors had marked preferences for the securities of a particular class of debtors. If, for example, corporations as a class were able to borrow at lower rates than individuals having equivalent personal leverage, then the average cost of capital to corporations might fall slightly, as leverage increased over some range, in reflection of this differential. In evaluating this possibility, however, remember that the relevant interest rate for our arbitrage operators is the rate on brokers' loans and, historically, that rate has not been noticeably higher than representative corporate rates.¹⁹ The operations of holding companies and investment trusts which can borrow on terms comparable to operating companies represent still another force which could be expected to wipe out any marked or prolonged advantages from holding levered stocks.²⁰

Although Proposition I remains unaffected as long as the yield curve is the same for all borrowers, the relation between common stock yields and leverage will no longer be the strictly linear one given by the original Proposition II. If r increases with leverage, the yield i will still tend to

¹⁸ One normally minor qualification might be noted. Once we relax the assumption that all bonds have certain yields, our arbitrage operator faces the danger of something comparable to "gambler's ruin." That is, there is always the possibility that an otherwise sound concern—one whose long-run expected income is greater than its interest liability—might be forced into liquidation as a result of a run of temporary losses. Since reorganization generally involves costs, and because the operation of the firm may be hampered during the period of reorganization with lasting unfavorable effects on earnings prospects, we might perhaps expect heavily levered companies to sell at a slight discount relative to less heavily indebted companies of the same class.

¹⁹ Under normal conditions, moreover, a substantial part of the arbitrage process could be expected to take the form, not of having the arbitrage operators go into debt on personal account to put the required leverage into their portfolios, but simply of having them reduce the amount of corporate bonds they already hold when they acquire underpriced unlevered stock. Margin requirements are also somewhat less of an obstacle to maintaining any desired degree of leverage in a portfolio than might be thought at first glance. Leverage could be largely restored in the face of higher margin requirements by switching to stocks having more leverage at the corporate level.

²⁰ An extreme form of inequality between borrowing and lending rates occurs, of course, in the case of preferred stocks, which can not be directly issued by individuals on personal account. Here again, however, we would expect that the operations of investment corporations plus the ability of arbitrage operators to sell off their holdings of preferred stocks would act to prevent the emergence of any substantial premiums (for this reason) on capital structures containing preferred stocks. Nor are preferred stocks so far removed from bonds as to make it impossible for arbitrage operators to approximate closely the risk and leverage of a corporate preferred stock by incurring a somewhat smaller debt on personal account.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

rise as D/S increases, but at a decreasing rather than a constant rate. Beyond some high level of leverage, depending on the exact form of the interest function, the yield may even start to fall.²¹ The relation between i and D/S could conceivably take the form indicated by the curve MD

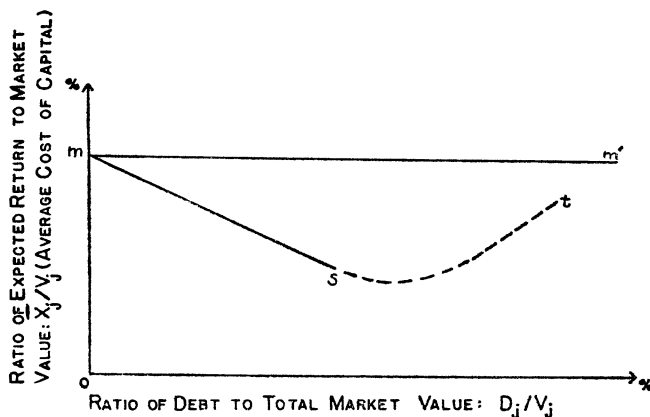


FIGURE 1

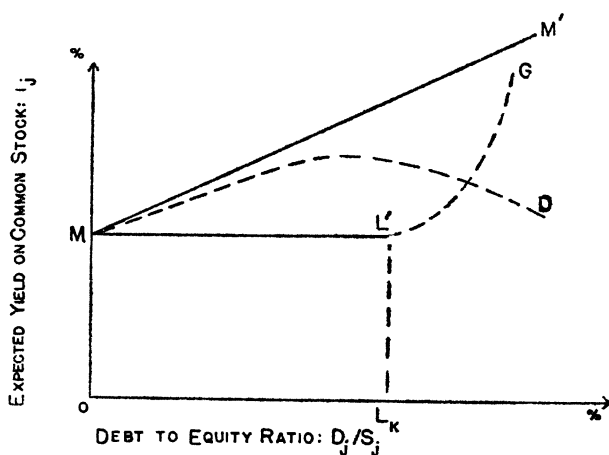


FIGURE 2

in Figure 2, although in practice the curvature would be much less pronounced. By contrast, with a constant rate of interest, the relation would be linear throughout as shown by line MM' , Figure 2.

The downward sloping part of the curve MD perhaps requires some

²¹ Since new lenders are unlikely to permit this much leverage (*cf.* note 17), this range of the curve is likely to be occupied by companies whose earnings prospects have fallen substantially since the time when their debts were issued.

comment since it may be hard to imagine why investors, other than those who like lotteries, would purchase stocks in this range. Remember, however, that the yield curve of Proposition II is a consequence of the more fundamental Proposition I. Should the demand by the risk-lovers prove insufficient to keep the market to the peculiar yield-curve MD , this demand would be reinforced by the action of arbitrage operators. The latter would find it profitable to own a pro-rata share of the firm as a whole by holding its stock *and* bonds, the lower yield of the shares being thus offset by the higher return on bonds.

D. *The Relation of Propositions I and II to Current Doctrines*

The propositions we have developed with respect to the valuation of firms and shares appear to be substantially at variance with current doctrines in the field of finance. The main differences between our view and the current view are summarized graphically in Figures 1 and 2. Our Proposition I [equation (4)] asserts that the average cost of capital, $\bar{X}_j/V_j$, is a constant for all firms j in class k , independently of their financial structure. This implies that, if we were to take a sample of firms in a given class, and if for each firm we were to plot the ratio of expected return to market value against some measure of leverage or financial structure, the points would tend to fall on a horizontal straight line with intercept $\rho_k r$, like the solid line mm' in Figure 1.²² From Proposition I we derived Proposition II [equation (8)] which, taking the simplest version with r constant, asserts that, for all firms in a class, the relation between the yield on common stock and financial structure, measured by D_j/S_j , will approximate a straight line with slope $(\rho_k r - r)$ and intercept $\rho_k r$. This relationship is shown as the solid line MM' in Figure 2, to which reference has been made earlier.²³

By contrast, the conventional view among finance specialists appears to start from the proposition that, other things equal, the earnings-price ratio (or its reciprocal, the times-earnings multiplier) of a firm's common stock will normally be only slightly affected by "moderate" amounts of debt in the firm's capital structure.²⁴ Translated into our no-

²² In Figure 1 the measure of leverage used is D_j/V_j (the ratio of debt to market value) rather than D_j/S_j (the ratio of debt to equity), the concept used in the analytical development. The D_j/V_j measure is introduced at this point because it simplifies comparison and contrast of our view with the traditional position.

²³ The line MM' in Figure 2 has been drawn with a positive slope on the assumption that $\rho_k r > r$, a condition which will normally obtain. Our Proposition II as given in equation (8) would continue to be valid, of course, even in the unlikely event that $\rho_k r < r$, but the slope of MM' would be negative.

²⁴ See, e.g., Graham and Dodd [6, pp. 464-66]. Without doing violence to this position, we can bring out its implications more sharply by ignoring the qualification and treating the yield as a virtual constant over the relevant range. See in this connection the discussion in Durand [3, esp. pp. 225-37] of what he calls the "net income method" of valuation.

tation, it asserts that for any firm j in the class k ,

$$(13) \quad \frac{\bar{X}_j^r - rD_j}{S_j} \equiv \frac{\bar{\pi}_j^r}{S_j} = i_k^*, \text{ a constant for } \frac{D_j}{S_j} \leq L_k$$

or, equivalently,

$$(14) \quad S_j = \bar{\pi}_j^r / i_k^*.$$

Here i_k^* represents the capitalization rate or earnings-price ratio on the common stock and L_k denotes some amount of leverage regarded as the maximum "reasonable" amount for firms of the class k . This assumed relationship between yield and leverage is the horizontal solid line ML' of Figure 2. Beyond L' , the yield will presumably rise sharply as the market discounts "excessive" trading on the equity. This possibility of a rising range for high leverages is indicated by the broken-line segment $L'G$ in the figure.²⁵

If the value of shares were really given by (14) then the over-all market value of the firm must be:

$$(16) \quad V_j \equiv S_j + D_j = \frac{\bar{X}_j^r - rD_j}{i_k^*} + D_j = \frac{\bar{X}_j^r}{i_k^*} + \frac{(i_k^* - r)D_j}{i_k^*}.$$

That is, for any given level of expected total returns after taxes ($\bar{X}_j^r$) and assuming, as seems natural, that $i_k^* > r$, the value of the firm must tend to *rise* with debt;²⁶ whereas our Proposition I asserts that the value of the firm is completely independent of the capital structure. Another way of contrasting our position with the traditional one is in terms of the cost of capital. Solving (16) for $\bar{X}_j^r/V_j$ yields:

$$(17) \quad \bar{X}_j^r/V_j = i_k^* - (i_k^* - r)D_j/V_j.$$

According to this equation, the average cost of capital is not independent of capital structure as we have argued, but should tend to *fall* with increasing leverage, at least within the relevant range of moderate debt ratios, as shown by the line ms in Figure 1. Or to put it in more familiar terms, debt-financing should be "cheaper" than equity-financing if not carried too far.

When we also allow for the possibility of a rising range of stock yields for large values of leverage, we obtain a U-shaped curve like mst in

²⁵ To make it easier to see some of the implications of this hypothesis as well as to prepare the ground for later statistical testing, it will be helpful to assume that the notion of a critical limit on leverage beyond which yields rise rapidly, can be epitomized by a quadratic relation of the form:

$$(15) \quad \bar{\pi}_j^r/S_j = i_k^* + \beta(D_j/S_j) + \alpha(D_j/S_j)^2, \quad \alpha > 0.$$

²⁶ For a typical discussion of how a promoter can, supposedly, increase the market value of a firm by recourse to debt issues, see W. J. Eiteman [4, esp. pp. 11-13].

Figure 1.²⁷ That a yield-curve for stocks of the form $ML'G$ in Figure 2 implies a U-shaped cost-of-capital curve has, of course, been recognized by many writers. A natural further step has been to suggest that the capital structure corresponding to the trough of the U is an "optimal capital structure" towards which management ought to strive in the best interests of the stockholders.²⁸ According to our model, by contrast, no such optimal structure exists—all structures being equivalent from the point of view of the cost of capital.

Although the falling, or at least U-shaped, cost-of-capital function is in one form or another the dominant view in the literature, the ultimate rationale of that view is by no means clear. The crucial element in the position—that the expected earnings-price ratio of the stock is largely unaffected by leverage up to some conventional limit—is rarely even regarded as something which requires explanation. It is usually simply taken for granted or it is merely asserted that this is the way the market behaves.²⁹ To the extent that the constant earnings-price ratio has a rationale at all we suspect that it reflects in most cases the feeling that moderate amounts of debt in "sound" corporations do not really add very much to the "riskiness" of the stock. Since the extra risk is slight, it seems natural to suppose that firms will not have to pay noticeably higher yields in order to induce investors to hold the stock.³⁰

A more sophisticated line of argument has been advanced by David Durand [3, pp. 231-33]. He suggests that because insurance companies and certain other important institutional investors are restricted to debt securities, nonfinancial corporations are able to borrow from them at interest rates which are lower than would be required to compensate

²⁷ The U-shaped nature of the cost-of-capital curve can be exhibited explicitly if the yield curve for shares as a function of leverage can be approximated by equation (15) of footnote 25. From that equation, multiplying both sides by S_j we obtain: $\bar{\pi}_j r = \bar{X}_j r - r D_j = i_k^* S_j + \beta D_j + \alpha D_j^2 / S_j$ or, adding and subtracting $i_k^* D_j$ from the right-hand side and collecting terms,

$$(18) \quad \bar{X}_j r = i_k^* (S_j + D_j) + (\beta + r - i_k^*) D_j + \alpha D_j^2 / S_j.$$

Dividing (18) by V_j gives an expression for the cost of capital:

$$(19) \quad \bar{X}_j r / V_j = i_k^* - (i_k^* - r - \beta) D_j / V_j + \alpha D_j^2 / S_j V_j = i_k^* - (i_k^* - r - \beta) D_j / V_j + \alpha (D_j / V_j)^2 / (1 - D_j / V_j)$$

which is clearly U-shaped since α is supposed to be positive.

²⁸ For a typical statement see S. M. Robbins [16, p. 307]. See also Graham and Dodd [6, pp. 468-74].

²⁹ See e.g., Graham and Dodd [6, p. 466].

³⁰ A typical statement is the following by Guthmann and Dougall [7, p. 245]: "Theoretically it might be argued that the increased hazard from using bonds and preferred stocks would counterbalance this additional income and so prevent the common stock from being more attractive than when it had a lower return but fewer prior obligations. In practice, the extra earnings from 'trading on the equity' are often regarded by investors as more than sufficient to serve as a 'premium for risk' when the proportions of the several securities are judiciously mixed."

creditors in a free market. Thus, while he would presumably agree with our conclusions that stockholders could not gain from leverage in an unconstrained market, he concludes that they can gain under present institutional arrangements. This gain would arise by virtue of the "safety superpremium" which lenders are willing to pay corporations for the privilege of lending.³¹

The defective link in both the traditional and the Durand version of the argument lies in the confusion between investors' subjective risk preferences and their objective market opportunities. Our Propositions I and II, as noted earlier, do not depend for their validity on any assumption about individual risk preferences. Nor do they involve any assertion as to what is an adequate compensation to investors for assuming a given degree of risk. They rely merely on the fact that a given commodity cannot consistently sell at more than one price in the market; or more precisely that the price of a commodity representing a "bundle" of two other commodities cannot be consistently different from the weighted average of the prices of the two components (the weights being equal to the proportion of the two commodities in the bundle).

An analogy may be helpful at this point. The relations between $1/\rho_k$, the price per dollar of an unlevered stream in class k ; $1/r$, the price per dollar of a sure stream, and $1/i_j$, the price per dollar of a levered stream j , in the k th class, are essentially the same as those between, respectively, the price of whole milk, the price of butter fat, and the price of milk which has been thinned out by skimming off some of the butter fat. Our Proposition I states that a firm cannot reduce the cost of capital—i.e., increase the market value of the stream it generates—by securing part of its capital through the sale of bonds, even though debt money appears to be cheaper. This assertion is equivalent to the proposition that, under perfect markets, a dairy farmer cannot in general earn more for the milk he produces by skimming some of the butter fat and selling it separately, even though butter fat per unit weight, sells for more than whole milk. The advantage from skimming the milk rather than selling whole milk would be purely illusory; for what would be gained from selling the high-priced butter fat would be lost in selling the low-priced residue of thinned milk. Similarly our Proposition II—that the price per dollar of a levered stream falls as leverage increases—is an ex-

³¹ Like Durand, Morton [15] contends "that the actual market deviates from [Proposition I] by giving a changing over-all cost of money at different points of the [leverage] scale" (p. 443, note 2, inserts ours), but the basis for this contention is nowhere clearly stated. Judging by the great emphasis given to the lack of mobility of investment funds between stocks and bonds and to the psychological and institutional pressures toward debt portfolios (see pp. 444-51 and especially his discussion of the optimal capital structure on p. 453) he would seem to be taking a position very similar to that of Durand above.

act analogue of the statement that the price per gallon of thinned milk falls continuously as more butter fat is skimmed off.³²

It is clear that this last assertion is true as long as butter fat is worth more per unit weight than whole milk, and it holds even if, for many consumers, taking a little cream out of the milk (adding a little leverage to the stock) does not detract noticeably from the taste (does not add noticeably to the risk). Furthermore the argument remains valid even in the face of institutional limitations of the type envisaged by Durand. For suppose that a large fraction of the population habitually dines in restaurants which are required by law to serve only cream in lieu of milk (entrust their savings to institutional investors who can only buy bonds). To be sure the price of butter fat will then tend to be higher in relation to that of skimmed milk than in the absence such restrictions (the rate of interest will tend to be lower), and this will benefit people who eat at home and who like skim milk (who manage their own portfolio and are able and willing to take risk). But it will still be the case that a farmer cannot gain by skimming some of the butter fat and selling it separately (firm cannot reduce the cost of capital by recourse to borrowed funds).³³

Our propositions can be regarded as the extension of the classical theory of markets to the particular case of the capital markets. Those who hold the current view—whether they realize it or not—must as-

³² Let M denote the quantity of whole milk, B/M the proportion of butter fat in the whole milk, and let p_M , p_B and p_α denote, respectively, the price per unit weight of whole milk, butter fat and thinned milk from which a fraction α of the butter fat has been skimmed off. We then have the fundamental perfect market relation:

$$(a) \quad p_\alpha(M - \alpha B) + p_B \alpha B = p_M M, \quad 0 \leq \alpha \leq 1,$$

stating that total receipts will be the same amount $p_M M$, independently of the amount αB of butter fat that may have been sold separately. Since p_M corresponds to $1/\rho$, p_B to $1/r$, p_α to $1/i$, M to $\bar{X}$ and αB to rD , (a) is equivalent to Proposition I, $S + D = \bar{X}/\rho$. From (a) we derive:

$$(b) \quad p_\alpha = p_M \frac{M}{M - \alpha B} - p_B \frac{\alpha B}{M - \alpha B}$$

which gives the price of thinned milk as an explicit function of the proportion of butter fat skimmed off; the function decreasing as long as $p_B > p_M$. From (a) also follows:

$$(c) \quad 1/p_\alpha = 1/p_M + (1/p_M - 1/p_B) \frac{p_B \alpha B}{p_\alpha (M - \alpha B)}$$

which is the exact analogue of Proposition II, as given by (8).

³³ The reader who likes parables will find that the analogy with interrelated commodity markets can be pushed a good deal farther than we have done in the text. For instance, the effect of changes in the market rate of interest on the over-all cost of capital is the same as the effect of a change in the price of butter on the price of whole milk. Similarly, just as the relation between the prices of skim milk and butter fat influences the kind of cows that will be reared, so the relation between i and r influences the kind of ventures that will be undertaken. If people like butter we shall have Guernseys; if they are willing to pay a high price for safety, this will encourage ventures which promise smaller but less uncertain streams per dollar of physical assets.

sume not merely that there are lags and frictions in the equilibrating process—a feeling we certainly share,³⁴ claiming for our propositions only that they describe the central tendency around which observations will scatter—but also that there are large and *systematic* imperfections in the market which permanently bias the outcome. This is an assumption that economists, at any rate, will instinctively eye with some skepticism.

In any event, whether such prolonged, systematic departures from equilibrium really exist or whether our propositions are better descriptions of long-run market behavior can be settled only by empirical research. Before going on to the theory of investment it may be helpful, therefore, to look at the evidence.

E. Some Preliminary Evidence on the Basic Propositions

Unfortunately the evidence which has been assembled so far is amazingly skimpy. Indeed, we have been able to locate only two recent studies—and these of rather limited scope—which were designed to throw light on the issue. Pending the results of more comprehensive tests which we hope will soon be available, we shall review briefly such evidence as is provided by the two studies in question: (1) an analysis of the relation between security yields and financial structure for some 43 large electric utilities by F. B. Allen [1], and (2) a parallel (unpublished) study by Robert Smith [19], for 42 oil companies designed to test whether Allen's rather striking results would be found in an industry with very different characteristics.³⁵ The Allen study is based on average figures for the years 1947 and 1948, while the Smith study relates to the single year 1953.

The Effect of Leverage on the Cost of Capital. According to the received view, as shown in equation (17) the average cost of capital, $\bar{X}^*/V$, should decline linearly with leverage as measured by the ratio D/V , at least through most of the relevant range.³⁶ According to Proposition I, the average cost of capital within a given class k should tend to have the same value ρ_k^* independently of the degree of leverage. A simple test

³⁴ Several specific examples of the failure of the arbitrage mechanism can be found in Graham and Dodd [6, e.g., pp. 646–48]. The price discrepancy described on pp. 646–47 is particularly curious since it persists even today despite the fact that a whole generation of security analysts has been brought up on this book!

³⁵ We wish to express our thanks to both writers for making available to us some of their original worksheets. In addition to these recent studies there is a frequently cited (but apparently seldom read) study by the Federal Communications Commission in 1938 [22] which purports to show the existence of an optimal capital structure or range of structures (in the sense defined above) for public utilities in the 1930's. By current standards for statistical investigations, however, this study cannot be regarded as having any real evidential value for the problem at hand.

³⁶ We shall simplify our notation in this section by dropping the subscript j used to denote a particular firm wherever this will not lead to confusion.

of the merits of the two alternative hypotheses can thus be carried out by correlating $\bar{X}'/V$ with D/V . If the traditional view is correct, the correlation should be significantly negative; if our view represents a better approximation to reality, then the correlation should not be significantly different from zero.

Both studies provide information about the average value of D —the market value of bonds and preferred stock—and of V —the market value of all securities.³⁷ From these data we can readily compute the ratio D/V and this ratio (expressed as a percentage) is represented by the symbol d in the regression equations below. The measurement of the variable $\bar{X}'/V$, however, presents serious difficulties. Strictly speaking, the numerator should measure the expected returns net of taxes, but this is a variable on which no direct information is available. As an approximation, we have followed both authors and used (1) the average value of actual net returns in 1947 and 1948 for Allen's utilities; and (2) actual net returns in 1953 for Smith's oil companies. Net return is defined in both cases as the sum of interest, preferred dividends and stockholders' income net of corporate income taxes. Although this approximation to expected returns is undoubtedly very crude, there is no reason to believe that it will systematically bias the test in so far as the sign of the regression coefficient is concerned. The roughness of the approximation, however, will tend to make for a wide scatter. Also contributing to the scatter is the crudeness of the industrial classification, since especially within the sample of oil companies, the assumption that all the firms belong to the same class in our sense, is at best only approximately valid.

Denoting by x our approximation to $\bar{X}'/V$ (expressed, like d , as a percentage), the results of the tests are as follows:

$$\text{Electric Utilities } x = 5.3 + .006d \quad r = .12 \\ (\pm .008)$$

$$\text{Oil Companies } x = 8.5 + .006d \quad r = .04. \\ (\pm .024)$$

The data underlying these equations are also shown in scatter diagram form in Figures 3 and 4.

The results of these tests are clearly favorable to our hypothesis.

³⁷ Note that for purposes of this test preferred stocks, since they represent an *expected* fixed obligation, are properly classified with bonds even though the tax status of preferred dividends is different from that of interest payments and even though preferred dividends are really fixed only as to their maximum in any year. Some difficulty of classification does arise in the case of convertible preferred stocks (and convertible bonds) selling at a substantial premium, but fortunately very few such issues were involved for the companies included in the two studies. Smith included bank loans and certain other short-term obligations (at book values) in his data on oil company debts and this treatment is perhaps open to some question. However, the amounts involved were relatively small and check computations showed that their elimination would lead to only minor differences in the test results.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

283

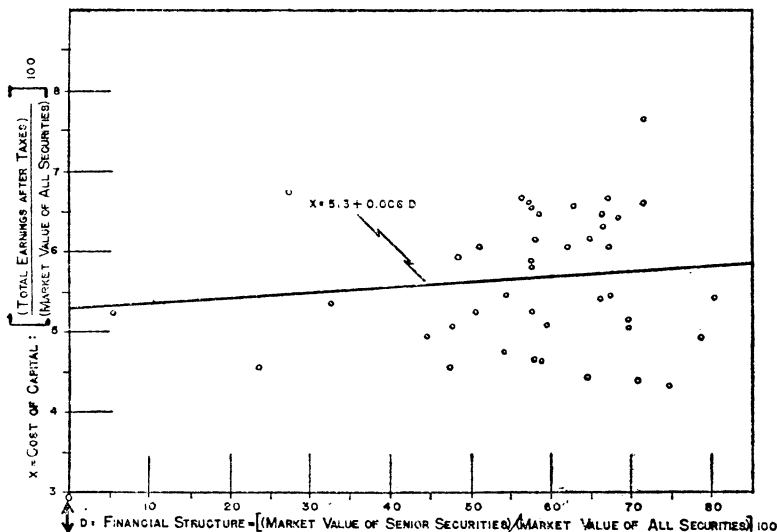


FIGURE 3. COST OF CAPITAL IN RELATION TO FINANCIAL STRUCTURE FOR 43 ELECTRIC UTILITIES, 1947-48

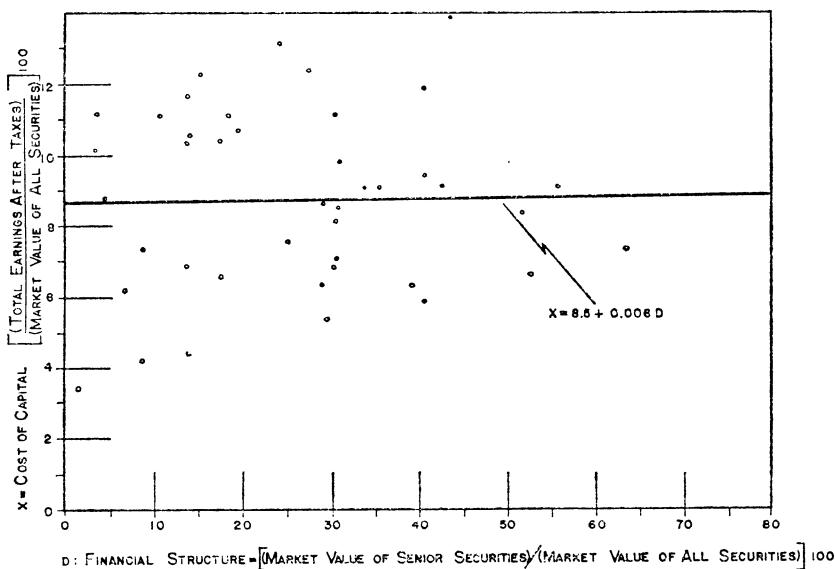


FIGURE 4. COST OF CAPITAL IN RELATION TO FINANCIAL STRUCTURE FOR 42 OIL COMPANIES, 1953

Both correlation coefficients are very close to zero and not statistically significant. Furthermore, the implications of the traditional view fail to be supported even with respect to the sign of the correlation. The data in short provide no evidence of any tendency for the cost of capital to fall as the debt ratio increases.³⁸

It should also be apparent from the scatter diagrams that there is no hint of a curvilinear, U-shaped, relation of the kind which is widely believed to hold between the cost of capital and leverage. This graphical impression was confirmed by statistical tests which showed that for both industries the curvature was not significantly different from zero, its sign actually being opposite to that hypothesized.³⁹

Note also that according to our model, the constant terms of the regression equations are measures of ρ_k^r , the capitalization rates for unlevered streams and hence the average cost of capital in the classes in question. The estimates of 8.5 per cent for the oil companies as against 5.3 per cent for electric utilities appear to accord well with a priori expectations, both in absolute value and relative spread.

The Effect of Leverage on Common Stock Yields. According to our Proposition II—see equation 12 and Figure 2—the expected yield on common stock, $\bar{\pi}^r/S$, in any given class, should tend to increase with leverage as measured by the ratio D/S . The relation should tend to be linear and with positive slope through most of the relevant range (as in the curve MM' of Figure 2), though it might tend to flatten out if we move

³⁸ It may be argued that a test of the kind used is biased against the traditional view. The fact that both sides of the regression equation are divided by the variable V which may be subject to random variation might tend to impart a positive bias to the correlation. As a check on the results presented in the text, we have, therefore, carried out a supplementary test based on equation (16). This equation shows that, if the traditional view is correct, the market value of a company should, for given $\bar{X}^r$, increase with debt through most of the relevant range; according to our model the market value should be uncorrelated with D , given $\bar{X}^r$. Because of wide variations in the size of the firms included in our samples, all variables must be divided by a suitable scale factor in order to avoid spurious results in carrying out a test of equation (16). The factor we have used is the book value of the firm denoted by A . The hypothesis tested thus takes the specific form:

$$V/A = a + b(\bar{X}^r/A) + c(D/A)$$

and the numerator of the ratio $\bar{X}^r/A$ is again approximated by actual net returns. The partial correlation between V/A and D/A should now be positive according to the traditional view and zero according to our model. Although division by A should, if anything, bias the results in favor of the traditional hypothesis, the partial correlation turns out to be only .03 for the oil companies and $-.28$ for the electric utilities. Neither of these coefficients is significantly different from zero and the larger one even has the wrong sign.

³⁹ The tests consisted of fitting to the data the equation (19) of footnote 27. As shown there, it follows from the U-shaped hypothesis that the coefficient α of the variable $(D/V)^2/(1-D/V)$, denoted hereafter by d^* , should be significant and positive. The following regression equations and partials were obtained:

$$\text{Electric Utilities } x = 5.0 + .017d - .003d^*; r_{xd^*} = -.15$$

$$\text{Oil Companies } x = 8.0 + .05d - .03d^*; r_{xd^*} = -.14.$$

far enough to the right (as in the curve MD'), to the extent that high leverage tends to drive up the cost of senior capital. According to the conventional view, the yield curve as a function of leverage should be a horizontal straight line (like ML') through most of the relevant range; far enough to the right, the yield may tend to rise at an increasing rate. Here again, a straight-forward correlation—in this case between $\bar{\pi}^r/S$ and D/S —can provide a test of the two positions. If our view is correct, the correlation should be significantly positive; if the traditional view is correct, the correlation should be negligible.

Subject to the same qualifications noted above in connection with $\bar{X}^r$, we can approximate $\bar{\pi}^r$ by actual stockholder net income.⁴⁰ Letting z denote in each case the approximation to $\bar{\pi}^r/S$ (expressed as a percentage) and letting h denote the ratio D/S (also in percentage terms) the following results are obtained:

$$\text{Electric Utilities } z = 6.6 + .017h \quad r = .53 \\ (+.004)$$

$$\text{Oil Companies } z = 8.9 + .051h \quad r = .53. \\ (\pm .012)$$

These results are shown in scatter diagram form in Figures 5 and 6.

Here again the implications of our analysis seem to be borne out by the data. Both correlation coefficients are positive and highly significant when account is taken of the substantial sample size. Furthermore, the estimates of the coefficients of the equations seem to accord reasonably well with our hypothesis. According to equation (12) the constant term should be the value of ρ_k^r for the given class while the slope should be $(\rho_k^r - r)$. From the test of Proposition I we have seen that for the oil companies the mean value of ρ_k^r could be estimated at around 8.7. Since the average yield of senior capital during the period covered was in the order of $3\frac{1}{2}$ per cent, we should expect a constant term of about 8.7 per cent and a slope of just over 5 per cent. These values closely approximate the regression estimates of 8.9 per cent and 5.1 per cent respectively. For the electric utilities, the yield of senior capital was also on the order of $3\frac{1}{2}$ per cent during the test years, but since the estimate of the mean value of ρ_k^r from the test of Proposition I was 5.6 per cent,

⁴⁰ As indicated earlier, Smith's data were for the single year 1953. Since the use of a single year's profits as a measure of expected profits might be open to objection we collected profit data for 1952 for the same companies and based the computation of $\bar{\pi}^r/S$ on the average of the two years. The value of $\bar{\pi}^r/S$ was obtained from the formula:

$$\left(\text{net earnings in 1952} \cdot \frac{\text{assets in '53}}{\text{assets in '52}} + \text{net earnings in '1953} \right) \frac{1}{2} \\ \div (\text{average market value of common stock in '53}).$$

The asset adjustment was introduced as rough allowance for the effects of possible growth in the size of the firm. It might be added that the correlation computed with $\bar{\pi}^r/S$ based on net profits in 1953 alone was found to be only slightly smaller, namely .50.

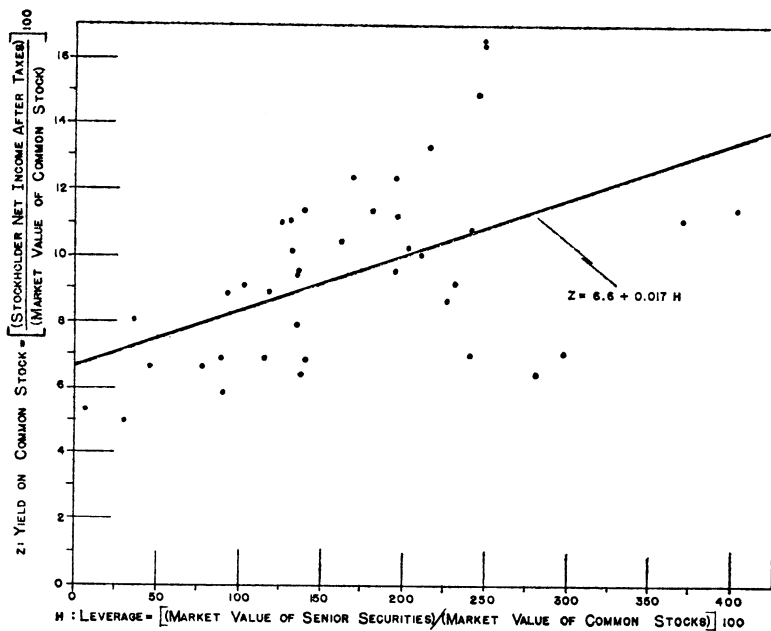


FIGURE 5. YIELD ON COMMON STOCK IN RELATION TO LEVERAGE FOR 43 ELECTRIC UTILITIES, 1947-48

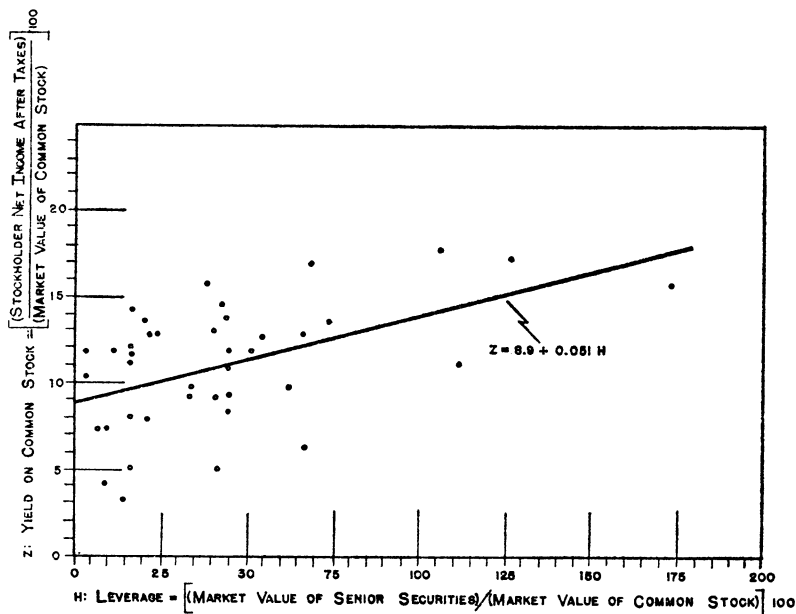


FIGURE 6. YIELD ON COMMON STOCK IN RELATION TO LEVERAGE FOR 42 OIL COMPANIES, 1952-53

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

287

the slope should be just above 2 per cent. The actual regression estimate for the slope of 1.7 per cent is thus somewhat low, but still within one standard error of its theoretical value. Because of this underestimate of the slope and because of the large mean value of leverage ($\bar{h}=160$ per cent) the regression estimate of the constant term, 6.6 per cent, is somewhat high, although not significantly different from the value of 5.6 per cent obtained in the test of Proposition I.

When we add a square term to the above equations to test for the presence and direction of curvature we obtain the following estimates:

$$\text{Electric Utilities } z = 4.6 + .004h - .007h^2$$

$$\text{Oil Companies } z = 8.5 + .072h - .016h^2.$$

For both cases the curvature is negative. In fact, for the electric utilities, where the observations cover a wider range of leverage ratios, the negative coefficient of the square term is actually significant at the 5 per cent level. Negative curvature, as we have seen, runs directly counter to the traditional hypothesis, whereas it can be readily accounted for by our model in terms of rising cost of borrowed funds.⁴¹

In summary, the empirical evidence we have reviewed seems to be broadly consistent with our model and largely inconsistent with traditional views. Needless to say much more extensive testing will be required before we can firmly conclude that our theory describes market behavior. Caution is indicated especially with regard to our test of Proposition II, partly because of possible statistical pitfalls⁴² and partly because not all the factors that might have a systematic effect on stock yields have been considered. In particular, no attempt was made to test the possible influence of the dividend pay-out ratio whose role has tended to receive a great deal of attention in current research and thinking. There are two reasons for this omission. First, our main objective has been to assess the *prima facie* tenability of *our* model, and in this model, based as it is on rational behavior by investors, dividends per se play no role. Second, in a world in which the policy of dividend stabilization is widespread, there is no simple way of disentangling the true effect of dividend payments on stock prices from their apparent effect,

⁴¹ That the yield of senior capital tended to rise for utilities as leverage increased is clearly shown in several of the scatter diagrams presented in the published version of Allen's study. This significant negative curvature between stock yields and leverage for utilities may be partly responsible for the fact, previously noted, that the constant in the linear regression is somewhat higher and the slope somewhat lower than implied by equation (12). Note also in connection with the estimate of ρ_k^7 that the introduction of the quadratic term reduces the constant considerably, pushing it in fact below the *a priori* expectation of 5.6, though the difference is again not statistically significant.

⁴² In our test, *e.g.*, the two variables z and h are both ratios with S appearing in the denominator, which may tend to impart a positive bias to the correlation (*cf.* note 38). Attempts were made to develop alternative tests, but although various possibilities were explored, we have so far been unable to find satisfactory alternatives.

the latter reflecting only the role of dividends as a proxy measure of long-term earning anticipations.⁴³ The difficulties just mentioned are further compounded by possible interrelations between dividend policy and leverage.⁴⁴

II. *Implications of the Analysis for the Theory of Investment*

A. *Capital Structure and Investment Policy*

On the basis of our propositions with respect to cost of capital and financial structure (and for the moment neglecting taxes), we can derive the following simple rule for optimal investment policy by the firm:

Proposition III. If a firm in class k is acting in the best interest of the stockholders at the time of the decision, it will exploit an investment opportunity if and only if the rate of return on the investment, say ρ^* , is as large as or larger than ρ_k . That is, *the cut-off point for investment in the firm will in all cases be ρ_k and will be completely unaffected by the type of security used to finance the investment.* Equivalently, we may say that regardless of the financing used, the marginal cost of capital to a firm is equal to the average cost of capital, which is in turn equal to the capitalization rate for an unlevered stream in the class to which the firm belongs.⁴⁵

To establish this result we will consider the three major financing alternatives open to the firm—bonds, retained earnings, and common stock issues—and show that in each case an investment is worth undertaking if, and only if, $\rho^* \geq \rho_k$.⁴⁶

Consider first the case of an investment financed by the sale of bonds. We know from Proposition I that the market value of the firm before the investment was undertaken was:⁴⁷

$$(20) \quad V_0 = \bar{X}_0 / \rho_k$$

⁴³ We suggest that failure to appreciate this difficulty is responsible for many fallacious, or at least unwarranted, conclusions about the role of dividends.

⁴⁴ In the sample of electric utilities, there is a substantial negative correlation between yields and pay-out ratios, but also between pay-out ratios and leverage, suggesting that either the association of yields and leverage or of yields and pay-out ratios may be (at least partly) spurious. These difficulties however do not arise in the case of the oil industry sample. A preliminary analysis indicates that there is here no significant relation between leverage and pay-out ratios and also no significant correlation (either gross or partial) between yields and pay-out ratios.

⁴⁵ The analysis developed in this paper is essentially a comparative-statics, not a dynamic analysis. This note of caution applies with special force to Proposition III. Such problems as those posed by expected changes in r and in ρ_k over time will not be treated here. Although they are in principle amenable to analysis within the general framework we have laid out, such an undertaking is sufficiently complex to deserve separate treatment. Cf. note 17.

⁴⁶ The extension of the proof to other types of financing, such as the sale of preferred stock or the issuance of stock rights is straightforward.

⁴⁷ Since no confusion is likely to arise, we have again, for simplicity, eliminated the subscripts identifying the firm in the equations to follow. Except for ρ_k , the subscripts now refer to time periods.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

and that the value of the common stock was:

$$(21) \quad S_0 = V_0 - D_0.$$

If now the firm borrows I dollars to finance an investment yielding ρ^* its market value will become:

$$(22) \quad V_1 = \frac{\bar{X}_0 + \rho^* I}{\rho_k} = V_0 + \frac{\rho^* I}{\rho_k}$$

and the value of its common stock will be:

$$(23) \quad S_1 = V_1 - (D_0 + I) = V_0 + \frac{\rho^* I}{\rho_k} - D_0 - I$$

or using equation 21,

$$(24) \quad S_1 = S_0 + \frac{\rho^* I}{\rho_k} - I.$$

Hence $S_1 \gtrless S_0$ as $\rho^* \gtrless \rho_k$.⁴⁸

To illustrate, suppose the capitalization rate for uncertain streams in the k th class is 10 per cent and the rate of interest is 4 per cent. Then if a given company had an expected income of 1,000 and if it were financed entirely by common stock we know from Proposition I that the market value of its stock would be 10,000. Assume now that the managers of the firm discover an investment opportunity which will require an outlay of 100 and which is expected to yield 8 per cent. At first sight this might appear to be a profitable opportunity since the expected return is double the interest cost. If, however, the management borrows the necessary 100 at 4 per cent, the total expected income of the company rises to 1,008 and the market value of the firm to 10,080. But the firm now will have 100 of bonds in its capital structure so that, paradoxically, the market value of the stock must actually be reduced from 10,000 to 9,980 as a consequence of this apparently profitable investment. Or, to put it another way, the gains from being able to tap cheap, borrowed funds are more than offset for the stockholders by the market's discounting of the stock for the added leverage assumed.

Consider next the case of retained earnings. Suppose that in the course of its operations the firm acquired I dollars of cash (without impairing

⁴⁸ In the case of bond-financing the rate of interest on bonds does not enter explicitly into the decision (assuming the firm borrows at the market rate of interest). This is true, moreover, given the conditions outlined in Section I.C, even though interest rates may be an increasing function of debt outstanding. To the extent that the firm borrowed at a rate other than the market rate the two I 's in equation (24) would no longer be identical and an additional gain or loss, as the case might be, would accrue to the shareholders. It might also be noted in passing that permitting the two I 's in (24) to take on different values provides a simple method for introducing underwriting expenses into the analysis.

the earning power of its assets). If the cash is distributed as a dividend to the stockholders their wealth W_0 , after the distribution will be:

$$(25) \quad W_0 = S_0 + I = \frac{\bar{X}_0}{\rho_k} - D_0 + I$$

where $\bar{X}_0$ represents the expected return from the assets exclusive of the amount I in question. If however the funds are retained by the company and used to finance new assets whose expected rate of return is ρ^* , then the stockholders' wealth would become:

$$(26) \quad W_1 = S_1 = \frac{\bar{X}_0 + \rho^* I}{\rho_k} - D_0 = S_0 + \frac{\rho^* I}{\rho_k}.$$

Clearly $W_1 \geq W_0$ as $\rho^* \geq \rho_k$ so that an investment financed by retained earnings raises the net worth of the owners if and only if $\rho^* > \rho_k$.⁴⁹

Consider finally, the case of common-stock financing. Let P_0 denote the current market price per share of stock and assume, for simplicity, that this price reflects currently expected earnings only, that is, it does not reflect any future increase in earnings as a result of the investment under consideration.⁵⁰ Then if N is the original number of shares, the price per share is:

$$(27) \quad P_0 = S_0 / N$$

and the number of new shares, M , needed to finance an investment of I dollars is given by:

$$(28) \quad M = \frac{I}{P_0}.$$

As a result of the investment the market value of the stock becomes:

$$S_1 = \frac{\bar{X}_0 + \rho^* I}{\rho_k} - D_0 = S_0 + \frac{\rho^* I}{\rho_k} = NP_0 + \frac{\rho^* I}{\rho_k}$$

and the price per share:

$$(29) \quad P_1 = \frac{S_1}{N + M} = \frac{1}{N + M} \left[NP_0 + \frac{\rho^* I}{\rho_k} \right].$$

⁴⁹ The conclusion that ρ_k is the cut-off point for investments financed from internal funds applies not only to undistributed net profits, but to depreciation allowances (and even to the funds represented by the current sale value of any asset or collection of assets). Since the owners can earn ρ_k by investing funds elsewhere in the class, partial or total liquidating distributions should be made whenever the firm cannot achieve a marginal internal rate of return equal to ρ_k .

⁵⁰ If we assumed that the market price of the stock did reflect the expected higher future earnings (as would be the case if our original set of assumptions above were strictly followed) the analysis would differ slightly in detail, but not in essentials. The cut-off point for new investment would still be ρ_k , but where $\rho^* > \rho_k$ the gain to the original owners would be larger than if the stock price were based on the pre-investment expectations only.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

Since by equation (28), $I = MP_0$, we can add MP_0 and subtract I from the quantity in bracket, obtaining:

$$(30) \quad \begin{aligned} P_1 &= \frac{1}{N+M} \left[(N+M)P_0 + \frac{\rho^* - \rho_k}{\rho_k} I \right] \\ &= P_0 + \frac{1}{N+M} \frac{\rho^* - \rho_k}{\rho_k} I > P_0 \text{ if,} \end{aligned}$$

and only if, $\rho^* > \rho_k$.

Thus an investment financed by common stock is advantageous to the current stockholders if and only if its yield exceeds the capitalization rate ρ_k .

Once again a numerical example may help to illustrate the result and make it clear why the relevant cut-off rate is ρ_k and not the current yield on common stock, i . Suppose that ρ_k is 10 per cent, r is 4 per cent, that the original expected income of our company is 1,000 and that management has the opportunity of investing 100 having an expected yield of 12 per cent. If the original capital structure is 50 per cent debt and 50 per cent equity, and 1,000 shares of stock are initially outstanding, then, by Proposition I, the market value of the common stock must be 5,000 or 5 per share. Furthermore, since the interest bill is $.04 \times 5,000 = 200$, the yield on common stock is $800/5,000 = 16$ per cent. It may then appear that financing the additional investment of 100 by issuing 20 shares to outsiders at 5 per share would dilute the equity of the original owners since the 100 promises to yield 12 per cent whereas the common stock is currently yielding 16 per cent. Actually, however, the income of the company would rise to 1,012; the value of the firm to 10,120; and the value of the common stock to 5,120. Since there are now 1,020 shares, each would be worth 5.02 and the wealth of the original stockholders would thus have been increased. What has happened is that the dilution in expected earnings per share (from .80 to .796) has been more than offset, in its effect upon the market price of the shares, by the decrease in leverage.

Our conclusion is, once again, at variance with conventional views,⁵¹ so much so as to be easily misinterpreted. Read hastily, Proposition III seems to imply that the capital structure of a firm is a matter of indifference; and that, consequently, one of the core problems of corporate finance—the problem of the optimal capital structure for a firm—is no problem at all. It may be helpful, therefore, to clear up such possible misunderstandings.

⁵¹ In the matter of investment policy under uncertainty there is no single position which represents "accepted" doctrine. For a sample of current formulations, all very different from ours, see Joel Dean [2, esp. Ch. 3], M. Gordon and E. Shapiro [5], and Harry Roberts [17].

B. *Proposition III and Financial Planning by Firms*

Misinterpretation of the scope of Proposition III can be avoided by remembering that this Proposition tells us only that the type of instrument used to finance an investment is irrelevant to the question of whether or not the investment is worth while. This does not mean that the owners (or the managers) have no grounds whatever for preferring one financing plan to another; or that there are no other policy or technical issues in finance at the level of the firm.

That grounds for preferring one type of financial structure to another will still exist within the framework of our model can readily be seen for the case of common-stock financing. In general, except for something like a widely publicized oil-strike, we would expect the market to place very heavy weight on current and recent past earnings in forming expectations as to future returns. Hence, if the owners of a firm discovered a major investment opportunity which they felt would yield much more than ρ_k , they might well prefer not to finance it via common stock at the then ruling price, because this price may fail to capitalize the new venture. A better course would be a pre-emptive issue of stock (and in this connection it should be remembered that stockholders are free to borrow and buy). Another possibility would be to finance the project initially with debt. Once the project had reflected itself in increased actual earnings, the debt could be retired either with an equity issue at much better prices or through retained earnings. Still another possibility along the same lines might be to combine the two steps by means of a convertible debenture or preferred stock, perhaps with a progressively declining conversion rate. Even such a double-stage financing plan may possibly be regarded as yielding too large a share to outsiders since the new stockholders are, in effect, being given an interest in any similar opportunities the firm may discover in the future. If there is a reasonable prospect that even larger opportunities may arise in the near future and if there is some danger that borrowing now would preclude more borrowing later, the owners might find their interests best protected by splitting off the current opportunity into a separate subsidiary with independent financing. Clearly the problems involved in making the crucial estimates and in planning the optimal financial strategy are by no means trivial, even though they should have no bearing on the basic decision to invest (as long as $\rho^* \geq \rho_k$).⁵²

Another reason why the alternatives in financial plans may not be a matter of indifference arises from the fact that managers are concerned

⁵² Nor can we rule out the possibility that the existing owners, if unable to use a financing plan which protects their interest, may actually prefer to pass up an otherwise profitable venture rather than give outsiders an "excessive" share of the business. It is presumably in situations of this kind that we could justifiably speak of a shortage of "equity capital," though this kind of market imperfection is likely to be of significance only for small or new firms.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

with more than simply furthering the interest of the owners. Such other objectives of the management—which need not be necessarily in conflict with those of the owners—are much more likely to be served by some types of financing arrangements than others. In many forms of borrowing agreements, for example, creditors are able to stipulate terms which the current management may regard as infringing on its prerogatives or restricting its freedom to maneuver. The creditors might even be able to insist on having a direct voice in the formation of policy.⁵³ To the extent, therefore, that financial policies have these implications for the management of the firm, something like the utility approach described in the introductory section becomes relevant to financial (as opposed to investment) decision-making. It is, however, the utility functions of the managers per se and not of the owners that are now involved.⁵⁴

In summary, many of the specific considerations which bulk so large in traditional discussions of corporate finance can readily be superimposed on our simple framework without forcing any drastic (and certainly no systematic) alteration of the conclusion which is our principal concern, namely that for investment decisions, the marginal cost of capital is ρ_k .

C. The Effect of the Corporate Income Tax on Investment Decisions

In Section I it was shown that when an unintegrated corporate income tax is introduced, the original version of our Proposition I,

$$\bar{X}/V = \rho_k = \text{a constant}$$

must be rewritten as:

$$(11) \quad \frac{(\bar{X} - rD)(1 - \tau) + rD}{V} \equiv \frac{\bar{X}\tau}{V} = \rho_k\tau = \text{a constant}.$$

Throughout Section I we found it convenient to refer to $\bar{X}\tau/V$ as the cost of capital. The appropriate measure of the cost of capital relevant

⁵³ Similar considerations are involved in the matter of dividend policy. Even though the stockholders may be indifferent as to payout policy as long as investment policy is optimal, the management need not be so. Retained earnings involve far fewer threats to control than any of the alternative sources of funds and, of course, involve no underwriting expense or risk. But against these advantages management must balance the fact that sharp changes in dividend rates, which heavy reliance on retained earnings might imply, may give the impression that a firm's finances are being poorly managed, with consequent threats to the control and professional standing of the management.

⁵⁴ In principle, at least, this introduction of management's risk preferences with respect to financing methods would do much to reconcile the apparent conflict between Proposition III and such empirical findings as those of Modigliani and Zeman [14] on the close relation between interest rates and the ratio of new debt to new equity issues; or of John Lintner [12] on the considerable stability in target and actual dividend-payout ratios.

to investment decisions, however, is the ratio of the expected return *before taxes* to the market value, *i.e.*, $\bar{X}/V$. From (11) above we find:

$$(31) \quad \frac{\bar{X}}{V} = \frac{\rho_k^r - \tau_r(D/V)}{1 - \tau} = \frac{\rho_k^r}{1 - \tau} \left[1 - \frac{\tau r D}{\rho_k^r V} \right],$$

which shows that the cost of capital now depends on the debt ratio, decreasing, as D/V rises, at the constant rate $\tau r/(1-\tau)$.⁵⁵ Thus, with a corporate income tax under which interest is a deductible expense, gains can accrue to stockholders from having debt in the capital structure, even when capital markets are perfect. The gains however are small, as can be seen from (31), and as will be shown more explicitly below.

From (31) we can develop the tax-adjusted counterpart of Proposition III by interpreting the term D/V in that equation as the proportion of debt used in any additional financing of V dollars. For example, in the case where the financing is entirely by new common stock, $D=0$ and the required rate of return ρ_k^S on a venture so financed becomes:

$$(32) \quad \rho_k^S = \frac{\rho_k^r}{1 - \tau}.$$

For the other extreme of pure debt financing $D=V$ and the required rate of return, ρ_k^D , becomes:

$$(33) \quad \rho_k^D = \frac{\rho_k^r}{1 - \tau} \left[1 - \tau \frac{r}{\rho_k^r} \right] = \rho_k^S \left[1 - \tau \frac{r}{\rho_k^r} \right] = \rho_k^S - \frac{\tau}{1 - \tau} r. \text{ } ^{56}$$

For investments financed out of retained earnings, the problem of defining the required rate of return is more difficult since it involves a comparison of the tax consequences to the individual stockholder of receiving a dividend versus having a capital gain. Depending on the time of realization, a capital gain produced by retained earnings may be taxed either at ordinary income tax rates, 50 per cent of these rates, 25 per

⁵⁵ Equation (31) is amenable, in principle, to statistical tests similar to those described in Section I.E. However we have not made any systematic attempt to carry out such tests so far, because neither the Allen nor the Smith study provides the required information. Actually, Smith's data included a very crude estimate of tax liability, and, using this estimate, we did in fact obtain a negative relation between $\bar{X}/V$ and D/V . However, the correlation ($-.28$) turned out to be significant only at about the 10 per cent level. While this result is not conclusive, it should be remembered that, according to our theory, the slope of the regression equation should be in any event quite small. In fact, with a value of τ in the order of .5, and values of ρ_k^r and r in the order of 8.5 and 3.5 per cent respectively (*cf.* Section I.E) an increase in D/V from 0 to 60 per cent (which is, approximately, the range of variation of this variable in the sample) should tend to reduce the average cost of capital only from about 17 to about 15 per cent.

⁵⁶ This conclusion does not extend to preferred stocks even though they have been classed with debt issues previously. Since preferred dividends except for a portion of those of public utilities are not in general deductible from the corporate tax, the cut-off point for new financing via preferred stock is exactly the same as that for common stock.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT

cent, or zero, if held till death. The rate on any dividends received in the event of a distribution will also be a variable depending on the amount of other income received by the stockholder, and with the added complications introduced by the current dividend-credit provisions. If we assume that the managers proceed on the basis of reasonable estimates as to the average values of the relevant tax rates for the owners, then the required return for retained earnings ρ_k^R can be shown to be:

$$(34) \quad \rho_k^R = \rho_k^* \frac{1}{1 - \tau} \frac{1 - \tau_d}{1 - \tau_g} = \frac{1 - \tau_d}{1 - \tau_g} \rho_k^*$$

where τ_d is the assumed rate of personal income tax on dividends and τ_g is the assumed rate of tax on capital gains.

A numerical illustration may perhaps be helpful in clarifying the relationship between these required rates of return. If we take the following round numbers as representative order-of-magnitude values under present conditions: an after-tax capitalization rate ρ_k^* of 10 per cent, a rate of interest on bonds of 4 per cent, a corporate tax rate of 50 per cent, a marginal personal income tax rate on dividends of 40 per cent (corresponding to an income of about \$25,000 on a joint return), and a capital gains rate of 20 per cent (one-half the marginal rate on dividends), then the required rates of return would be: (1) 20 per cent for investments financed entirely by issuance of new common shares; (2) 16 per cent for investments financed entirely by new debt; and (3) 15 per cent for investments financed wholly from internal funds.

These results would seem to have considerable significance for current discussions of the effect of the corporate income tax on financial policy and on investment. Although we cannot explore the implications of the results in any detail here, we should at least like to call attention to the remarkably small difference between the "cost" of equity funds and debt funds. With the numerical values assumed, equity money turned out to be only 25 per cent more expensive than debt money, rather than something on the order of 5 times as expensive as is commonly supposed to be the case.⁵⁷ The reason for the wide difference is that the traditional

⁵⁷ See *e.g.*, D. T. Smith [18]. It should also be pointed out that our tax system acts in other ways to reduce the gains from debt financing. Heavy reliance on debt in the capital structure, for example, commits a company to paying out a substantial proportion of its income in the form of interest payments taxable to the owners under the personal income tax. A debt-free company, by contrast, can reinvest in the business all of its (smaller) net income and to this extent subject the owners only to the low capital gains rate (or possibly no tax at all by virtue of the loophole at death). Thus, we should expect a high degree of leverage to be of value to the owners, even in the case of closely held corporations, primarily in cases where their firm was not expected to have much need for additional funds to expand assets and earnings in the future. To the extent that opportunities for growth were available, as they presumably would be for most successful corporations, the interest of the stockholders would tend to be better served by a structure which permitted maximum use of retained earnings.

view starts from the position that debt funds are several times cheaper than equity funds even in the absence of taxes, with taxes serving simply to magnify the cost ratio in proportion to the corporate rate. By contrast, in our model in which the repercussions of debt financing on the value of shares are taken into account, the *only* difference in cost is that due to the tax effect, and its magnitude is simply the tax on the "grossed up" interest payment. Not only is this magnitude likely to be small but our analysis yields the further paradoxical implication that the stockholders' gain from, and hence incentive to use, debt financing is actually smaller the lower the rate of interest. In the extreme case where the firm could borrow for practically nothing, the advantage of debt financing would also be practically nothing.

III. Conclusion

With the development of Proposition III the main objectives we outlined in our introductory discussion have been reached. We have in our Propositions I and II at least the foundations of a theory of the valuation of firms and shares in a world of uncertainty. We have shown, moreover, how this theory can lead to an operational definition of the cost of capital and how that concept can be used in turn as a basis for rational investment decision-making within the firm. Needless to say, however, much remains to be done before the cost of capital can be put away on the shelf among the solved problems. Our approach has been that of static, partial equilibrium analysis. It has assumed among other things a state of atomistic competition in the capital markets and an ease of access to those markets which only a relatively small (though important) group of firms even come close to possessing. These and other drastic simplifications have been necessary in order to come to grips with the problem at all. Having served their purpose they can now be relaxed in the direction of greater realism and relevance, a task in which we hope others interested in this area will wish to share.

REFERENCES

1. F. B. ALLEN, "Does Going into Debt Lower the 'Cost of Capital'?", *Analysis Jour.*, Aug. 1954, 10, 57-61.
2. J. DEAN, *Capital Budgeting*. New York 1951.
3. D. DURAND, "Costs of Debt and Equity Funds for Business: Trends and Problems of Measurement" in Nat. Bur. Econ. Research, *Conference on Research in Business Finance*. New York 1952, pp. 215-47.
4. W. J. EITEMAN, "Financial Aspects of Promotion," in *Essays on Business Finance* by M. W. Waterford and W. J. Eiteman. Ann Arbor, Mich. 1952, pp. 1-17.
5. M. J. GORDON and E. SHAPIRO, "Capital Equipment Analysis: The Required Rate of Profit," *Manag. Sci.*, Oct. 1956, 3, 102-10.

MODIGLIANI AND MILLER: THEORY OF INVESTMENT 297

6. B. GRAHAM and L. DODD, *Security Analysis*, 3rd ed. New York 1951.
7. G. GUTHMANN and H. E. DOUGALL, *Corporate Financial Policy*, 3rd ed. New York 1955.
8. J. R. HICKS, *Value and Capital*, 2nd ed. Oxford 1946.
9. P. HUNT and M. WILLIAMS, *Case Problems in Finance*, rev. ed. Homewood, Ill. 1954.
10. J. M. KEYNES, *The General Theory of Employment, Interest and Money*. New York 1936.
11. O. LANGE, *Price Flexibility and Employment*. Bloomington, Ind. 1944.
12. J. LINTNER, "Distribution of Incomes of Corporations among Dividends, Retained Earnings and Taxes," *Am. Econ. Rev.*, May 1956, 46, 97-113.
13. F. LUTZ and V. LUTZ, *The Theory of Investment of the Firm*. Princeton 1951.
14. F. MODIGLIANI and M. ZEMAN, "The Effect of the Availability of Funds, and the Terms Thereof, on Business Investment" in Nat. Bur. Econ. Research, *Conference on Research in Business Finance*. New York 1952, pp. 263-309.
15. W. A. MORTON, "The Structure of the Capital Market and the Price of Money," *Am. Econ. Rev.*, May 1954, 44, 440-54.
16. S. M. ROBBINS, *Managing Securities*. Boston 1954.
17. H. V. ROBERTS, "Current Problems in the Economics of Capital Budgeting," *Jour. Bus.*, 1957, 30 (1), 12-16.
18. D. T. SMITH, *Effects of Taxation on Corporate Financial Policy*. Boston 1952.
19. R. SMITH, "Cost of Capital in the Oil Industry," (hectograph). Pittsburgh: Carnegie Inst. Tech. 1955.
20. H. M. SOMERS, "'Cost of Money' as the Determinant of Public Utility Rates," *Buffalo Law Rev.*, Spring 1955, 4, 1-28.
21. J. B. WILLIAMS, *The Theory of Investment Value*. Cambridge, Mass. 1938.
22. U. S. Federal Communications Commission, *The Problem of the "Rate of Return" in Public Utility Regulation*. Washington 1938.



The Effect of the Firm's Capital Structure on the Systematic Risk of Common Stocks

Robert S. Hamada

The Journal of Finance, Vol. 27, No. 2, Papers and Proceedings of the Thirtieth Annual Meeting of the American Finance Association, New Orleans, Louisiana, December 27-29, 1971. (May, 1972), pp. 435-452.

Stable URL:

<http://links.jstor.org/sici?sici=0022-1082%28197205%2927%3A2%3C435%3ATEOTFC%3E2.0.CO%3B2-3>

The Journal of Finance is currently published by American Finance Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/afina.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

THE EFFECT OF THE FIRM'S CAPITAL STRUCTURE ON THE SYSTEMATIC RISK OF COMMON STOCKS

ROBERT S. HAMADA*

I. INTRODUCTION

ONLY RECENTLY has there been an interest in relating the issues historically associated with corporation finance to those historically associated with investment and portfolio analyses. In fact, rigorous theoretical attempts in this direction were made only since the capital asset pricing model of Sharpe [13], Lintner [6], and Mossin [11], itself an extension of the Markowitz [7] portfolio theory. This study is one of the first empirical works consciously attempting to show and test the relationships between the two fields. In addition, differences in the observed systematic or nondiversifiable risk of common stocks, β , have never really been analyzed before by investigating some of the underlying differences in the firms.

In the capital asset pricing model, it was demonstrated that the efficient set of portfolios to any individual investor will always be some combination of lending at the risk-free rate and the "market portfolio," or borrowing at the risk-free rate and the "market portfolio." At the same time, the Modigliani and Miller (MM) propositions [9, 10] on the effect of corporate leverage are well known to the students of corporation finance. In order for their propositions to hold, personal leverage is required to be a perfect substitute for corporate leverage. If this is true, then corporate borrowing could substitute for personal borrowing in the capital asset pricing model as well.

Both in the pricing model and the MM theory, borrowing, from whatever source, while maintaining a fixed amount of equity, increases the risk to the investor. Therefore, in the mean-standard deviation version of the capital asset pricing model, the covariance of the asset's rate of return with the market portfolio's rate of return (which measures the nondiversifiable risk of the asset—the proxy β will be used to measure this) should be greater for the stock of a firm with a higher debt-equity ratio than for the stock of another firm in the same risk-class with a lower debt-equity ratio.¹

This study, then, has a number of purposes. First, we shall attempt to link empirically corporation finance issues with portfolio and security analyses through the effect of a firm's leverage on the systematic risk of its common

* Graduate School of Business, University of Chicago, currently visiting at the Graduate School of Business Administration, University of Washington. The research assistance of Christine Thomas and Leon Tsao is gratefully acknowledged. This paper has benefited from the comments made at the Finance Workshop at the University of Chicago, and especially those made by Eugene Fama. Remaining errors are due solely to the author.

1. This very quick summary of the theoretical relationship between what is known as corporation finance and the modern investment and portfolio analyses centered around the capital asset pricing model is more thoroughly presented in [5], along with the necessary assumptions required for this relationship.

stock. Then, we shall attempt to test the MM theory, or at least provide another piece of evidence on this long-standing controversial issue. This test will not rely on an explicit valuation model, such as the MM study of the electric utility industry [8] and the Brown study of the railroad industry [2]. A procedure using systematic risk measures (β s) has been worked out in this paper for this purpose.

If the MM theory is validated by this procedure, then the final purpose of this study is to demonstrate a method for estimating the cost of capital of individual firms to be used by them for scale-changing or nondiversifying investment projects. The primary component of any firm's cost of capital is the capitalization rate for the firm if the firm had no debt and preferred stock in its capital structure. Since most firms do have fixed commitment obligations, this capitalization rate (we shall call it $E(R_A)$; MM denote it ρ^r) is unobservable. But if the MM theory and the capital asset pricing model are correct, then it is possible to estimate $E(R_A)$ from the systematic risk approach for individual firms, even if these firms are members of a one-firm risk-class.²

With this statement of the purposes for this study, we shall, in Section II, discuss the alternative general procedures that are possible for estimating the effect of leverage on systematic risk and select the most feasible ones. The results are presented in Section III. And finally, tests of the MM versus the traditional theories of corporation finance are presented in Section IV.

II. SOME POSSIBLE PROCEDURES AND THE SELECTED ESTIMATING RELATIONSHIPS

There are at least four general procedures that can be used to estimate the effect of the firm's capital structure on the systematic risk of common stocks. The first is the MM valuation model approach. By estimating ρ^r with an explicit valuation model as they have for the electric utility industry, it is possible to relate this ρ^r with the use of the capital asset pricing model to a nonleveraged systematic risk measure, ${}_A\beta$. Then the difference between the observed common stock's systematic risk (which we shall denote ${}_B\beta$) and ${}_A\beta$ would be due solely to leverage. But the difficulties of this approach for all firms are many.

The MM valuation model approach requires the specification, in advance, of risk-classes. All firms in a risk-class are then assumed to have the same ρ^r —the capitalization rate for an all-common equity firm. Unfortunately, there must be enough firms in a risk-class so that a cross-section analysis will yield statistically significant coefficients. There may not be many more risk-classes (with enough observations) now that the electric utility and railroad industries have been studied. In addition, the MM approach requires estimating expected asset earnings and estimating the capitalized growth potential implicit in stock prices. If it is possible to consider growth and expected earnings without having

2. It is, in fact, this last purpose of making applicable and practical some of the implications of the capital asset pricing model for corporation finance issues that provided the initial motivation for this paper. In this context, if one is familiar with the fair rate of return literature for regulated utilities, for example, an industry where debt is so prevalent, adjusting correctly for leverage is not frequently done and can be very critical.

to specify their exact magnitude at a specific point in time, considerable difficulty and possible measurement errors will be avoided.

The second approach is to run a regression between the observed systematic risk of a stock and a number of accounting and leverage variables in an attempt to explain this observed systematic risk. Unfortunately, without a theory, we do not know which variables to include and which variables to exclude and whether the relationship is linear, multiplicative, exponential, curvilinear, etc. Therefore, this method will also not be used.

A third approach is to measure the systematic risk before and after a new debt issue. The difference can then be attributed to the debt issue directly. An attractive feature of this procedure is that a good estimate of the market value of the incremental debt issue can be obtained. A number of disadvantages, unfortunately, are associated with this direct approach. The difference in the systematic risk may be due not only to the additional debt, but also to the reason the debt was issued. It may be used to finance a new investment project, in which case the project's characteristics will also be reflected in the new systematic risk measure. In addition, the new debt issue may have been anticipated by the market if the firm had some long-run target leverage ratio which this issue will help maintain; conversely, the market may not fully consider the new debt issue if it believes the increase in leverage is only temporary. For these reasons, this seemingly attractive procedure will not be employed.

The last approach, which will be used in this study, is to assume the validity of the MM theory from the outset. Then the observed rate of return of a stock can be adjusted to what *it would have been* over the same time period had the firm no debt and preferred stock in its capital structure. The difference between the observed systematic risk, ${}_B\beta$, and the systematic risk for this adjusted rate of return time series, ${}_A\beta$, can be attributed to leverage, if the MM theory is correct. The final step, then, is to test the MM theory.

To discuss this more specifically, consider the following relationship for the dollar return to the common shareholder from period $t - 1$ to t :

$$(X - I)_t(1 - \tau)_t - p_t + \Delta G_t = d_t + cg_t \quad (1)$$

where X_t represents earnings before taxes, interest, and preferred dividends and is assumed to be unaffected by fixed commitment obligations; I_t represents interest and other fixed charges paid during the period; τ is the corporation income tax rate; p_t is the preferred dividends paid; ΔG_t represents the change in capitalized growth over the period; and d_t and cg_t are common shareholder dividends and capital gains during the period, respectively.

Equation (1) relates the corporation finance types of variables with the market holding period return important to the investors. The first term on the left-hand-side of (1) is profits after taxes and after interest which is the earnings the common and preferred shareholders receive on their investment for the period. Subtracting out p_t leaves us with the earnings the common shareholder would receive from currently-held assets.

To this must be added any change in capitalized growth since we are trying to explain the common shareholder's market holding period dollar return. ΔG_t

must be added for growth firms to the current period's profits from existing assets since capitalized growth opportunities of the firm—future earnings from new assets over and above the firm's cost of capital which are already reflected in the stock price at $(t - 1)$ —should change over the period and would accrue to the common shareholder. Assuming shareholders at the start of the period estimated these growth opportunities on average correctly, the expected value of ΔG_t would not be zero, but should be positive. For example, consider growth opportunities five years from now which yield more than the going rate of return and are reflected in today's stock price. These growth opportunities will become one year closer to fruition at time t than at time $t - 1$ so that their present value would become larger. ΔG_t then represents this increase in the present value of these future opportunities simply because it is now four years away rather than five.³

Since the systematic risk of a common stock is:

$${}_B\beta = \frac{\text{cov}(R_{B_t}, R_{M_t})}{\sigma^2(R_{M_t})} \quad (2)$$

where R_{B_t} is the common shareholder's rate of return and R_{M_t} is the rate of return on the market portfolio, then substitution of (1) into (2) yields:

$${}_B\beta = \frac{\text{cov} \left[\frac{(X - I)(1 - \tau)_t - p_t + \Delta G_t}{S_{B_{t-1}}}, R_{M_t} \right]}{\sigma^2(R_{M_t})} \quad (2a)$$

where $S_{B_{t-1}}$ denotes the market value of the common stock at the beginning of the period.

The systematic risk for the same firm over the same period *if* there were no debt and preferred stock in its capital structure is:

$$\begin{aligned} {}_A\beta &= \frac{\text{cov}(R_{A_t}, R_{M_t})}{\sigma^2(R_{M_t})} \\ &= \frac{\text{cov} \left[\frac{X(1 - \tau)_t + \Delta G_t}{S_{A_{t-1}}}, R_{M_t} \right]}{\sigma^2(R_{M_t})} \end{aligned} \quad (3)$$

where R_{A_t} and $S_{A_{t-1}}$ represent the rate of return and the market value, respectively, to the common shareholder if the firm had no debt and preferred stock. From (3), we can obtain:

$${}_A\beta S_{A_{t-1}} = \frac{\text{cov}[X(1 - \tau)_t + \Delta G_t, R_{M_t}]}{\sigma^2(R_{M_t})} \quad (3a)$$

3. Continual awareness of the difficulties of estimating capitalized growth, or changes in growth, especially in conjunction with leverage considerations, for purposes such as valuation or cost of capital is a characteristic common to students of corporation finance. This is the reason for the emphasis on growth in this paper and for presenting a method to neutralize for differences in growth when comparing rates of return.

Next, by expanding and rearranging (2a), we have:

$${}_B\beta S_{B,t-1} = \frac{\text{cov}[X(1-\tau)_t + \Delta G_t, R_{M_t}]}{\sigma^2(R_{M_t})} - \frac{\text{cov}[I(1-\tau)_t, R_{M_t}]}{\sigma^2(R_{M_t})} - \frac{\text{cov}(p_t, R_{M_t})}{\sigma^2(R_{M_t})} \quad (2b)$$

If we assume as an empirical approximation that interest and preferred dividends have negligible covariance with the market, at least relative to the (pure equity) common stock's covariance, then substitution of the LHS of (3a) into the RHS of (2b) yields:⁴

$${}_B\beta S_{B,t-1} = {}_A\beta S_{A,t-1} \quad (4)$$

or

$${}_A\beta = \left(\frac{S_B}{S_A} \right)_{t-1} {}_B\beta \quad (4a)$$

Because $S_{A,t-1}$, the market value of common stock *if* the firm had no debt and preferred stock, is not observable since most firms do have debt and/or preferred stock, a theory is required in order to measure what this quantity *would have been* at $t-1$. The MM theory [10] will be employed for this purpose, that is:

$$S_{A,t-1} = (V - \tau D)_{t-1}. \quad (5)$$

Equation (5) indicates that if the Federal government tax subsidy for debt financing, τD , where D is the market value of debt, is subtracted from the observed market value of the firm, V_{t-1} (where V_{t-1} is the sum of S_B , D and the observed market value of preferred), then the market value of an unleveraged firm is obtained. Underlying (5) is the assumption that the firm is near its target leverage ratio so that no more or no less debt subsidy is capitalized already into the observed stock price. The conditions under which this MM relationship hold are discussed carefully in [4].

It is at this point that problems in obtaining satisfactory estimates of ${}_A\beta$ develop, since (4) theoretically holds only for the next period. As a practical matter, the accepted, and seemingly acceptable, method of obtaining estimates of a stock's systematic risk, ${}_B\beta$, is to run a least squares regression between a stock's and market portfolio's *historical* rates of return. Using past data for ${}_B\beta$, it is not clear which *period's* ratio of market values to apply in (4a) to estimate the firm's systematic risk, ${}_A\beta$. There would be no problem if the market value ratios of debt to equity and preferred stock to equity remained relatively stable over the past for each firm, but a cursory look at these data reveals that this is not true for the large majority of firms in our sample. Should we use the market value ratio required in (4a) that was observed at the start of our regression period, at the end of our regression period, or some kind of average over the period? In addition, since these different observed ratios will give us different estimates for ${}_A\beta$, it is not clear, without some criterion, how we should select from among the various estimates.

4. This general method of arriving at (4) was suggested by the comments of William Sharpe, one of the discussants of this paper at the annual meeting. A much more cumbersome and less general derivation of (4) was in the earlier version.

It is for this purpose—to obtain a standard—that a more cumbersome and more data demanding approach to obtain estimates of ${}_A\beta$ is suggested. Given the large fluctuations in market leverage ratios, intuitively it would appear that the firm's risk is more stable than the common stock's risk. In that event, a leverage-free rate of return time series for each firm should be derived and the market model applied to this time series directly. In this manner, the beta coefficient would give us a *direct* estimate of ${}_A\beta$ which can then be used as a criterion to determine if any of the market value ratios discussed above can be applied to (4a) successfully.

For this purpose, the "would-have-been" rate of return for the common stock if the firm had no debt and preferred is:

$$R_{A_t} = \frac{X_t(1 - \tau)_t + \Delta G_t}{S_{A_{t-1}}}. \quad (6)$$

The numerator of (6) can be rearranged to be:

$$X_t(1 - \tau)_t + \Delta G_t \equiv [(X - I)_t(1 - \tau)_t - p_t + \Delta G_t] + p_t + I_t(1 - \tau)_t.$$

Substituting (1):

$$X_t(1 - \tau)_t + \Delta G_t = [d_t + cg_t] + p_t + I_t(1 - \tau)_t.$$

Therefore, (6) can be written as:

$$R_{A_t} = \frac{d_t + cg_t + p_t + I_t(1 - \tau)_t}{S_{A_{t-1}}}. \quad (7)$$

Since $S_{A_{t-1}}$ is unobservable for the firms with leverage, the MM theory, equation (5), will be employed; then:

$$R_{A_t} = \frac{d_t + cg_t + p_t + I_t(1 - \tau)_t}{(V - \tau D)_{t-1}}. \quad (8)$$

The observed rate of return on the common stock is, of course:

$$R_{B_t} = \frac{(X - I)_t(1 - \tau)_t - p_t + \Delta G_t}{S_{B_{t-1}}} = \frac{d_t + cg_t}{S_{B_{t-1}}}. \quad (9)$$

Equation (8) is the rate of return to the common shareholder of the same firm and over the same period of time as (9). However, in (8) there are the underlying assumptions that the firm never had any debt and preferred stock and that the MM theory is correct; (9) incorporates the exact amount of debt and preferred stock that the firm actually did have over this time period and no leverage assumption is being made. Both (8) and (9) are now in forms where they can be measured with available data. One can note that it is unnecessary to estimate the change in growth, or earnings from current assets, since these should be captured in the market holding period return, $d_t + cg_t$.

Using CRSP data for (9) and both CRSP and Compustat data for the components of (8), a time series of yearly R_{A_t} and R_{B_t} for $t = 1948-1967$ were derived for 304 different firms. These 304 firms represent an exhaustive sample of the firms with complete data on both tapes for all the years.

A number of "market model" [1, 12] variants were then applied to these data. For each of the 304 firms, the following regressions were run:

$$R_{Ait} = {}_A\alpha_i + {}_A\beta_i R_{Mt} + {}_A\epsilon_{it} \quad (10a)$$

$$R_{Bit} = {}_B\alpha_i + {}_B\beta_i R_{Mt} + {}_B\epsilon_{it} \quad (10b)$$

$$\ln(1 + R_{Ait}) = {}_{AC}\alpha_i + {}_{AC}\beta_i \ln(1 + R_{Mt}) + {}_{AC}\epsilon_{it} \quad (10c)$$

$$\ln(1 + R_{Bit}) = {}_{BC}\alpha_i + {}_{BC}\beta_i \ln(1 + R_{Mt}) + {}_{BC}\epsilon_{it} \quad (10d)$$

$$i = 1, 2, \dots, 304$$

$$t = 1948-1967$$

where R_{Mt} is the observed NYSE arithmetic stock market rate of return with dividends reinvested, α_i and β_i are constants for each firm-regression, and the usual conditions are assumed for the properties of the disturbance terms, ϵ_{it} . Equations (10c) and (10d) are the continuously-compounded rate of return versions of (10a) and (10b), respectively.⁵

III. THE RESULTS

An abbreviated table of the regression results for each of the four variants, equations (10a)-(10d), summarized across the 304 firms is shown in Table 1.

The first column designated "mean" is the average of the statistic (indicated by the rows) over all 304 firms. Therefore, the mean ${}_A\hat{\alpha}$ of 0.0221 is the intercept term of equation (10a) averaged over 304 different firm-regressions. The second and third columns give the deviation measures indicated, of the 304 point estimates of, say, ${}_A\hat{\alpha}$. The mean standard error of estimate in the last column is the average over 304 firms of the individual standard errors of estimate.

The major conclusion drawn from Table 1 is the following mean β comparisons:

$${}_B\hat{\beta} > {}_A\hat{\beta}, \text{ i.e., } 0.9190 > 0.7030$$

$${}_{BC}\hat{\beta} > {}_{AC}\hat{\beta}, \text{ i.e., } 0.9183 > 0.7263.$$

The directional results of these betas, assuming the validity of the MM theory, are not imperceptible and clearly are not negligible differences from the investor's point of view. This is obtained in spite of all the measurement and data problems associated with estimating a time series of the RHS of (8) for

5. Because the R_{Mt} used in equations (10) is defined as the observed stock market return, and since adjusting for capital structure is the major purpose of this exercise, it was decided that the same four regressions should be replicated on a leverage-adjusted stock market rate of return. The major reason for this additional adjustment is the belief that the rates of return over time and their relationship with the market are more stable when we can abstract from all changes in leverage and get at the underlying risk of all firms.

For the 221 firms (out of the total 304) whose fiscal years coincide with the calendar year, average values for the components of the RHS of (8) were obtained for each year so that R_{Mt} could be adjusted in the same way as for the individual firms—a yearly time series of stock market rates of return, if all the firms on the NYSE had no debt and no preferred in their capital structure, was derived. The results, when using this adjusted market portfolio rate of return time series, were not very different from the results of equations (10), and so will not be reported here separately.

TABLE 1
SUMMARY RESULTS OVER 304 FIRMS OF EQUATIONS (10a)-(10d)

	Mean	Mean Absolute Deviation*	Standard Deviation	Mean Standard Error of Estimate
${}_A\hat{\alpha}$	0.0221	0.0431	0.0537	0.0558
${}_A\hat{\beta}$	0.7030	0.2660	0.3485	0.2130
${}_A\hat{R}^2$	0.3799	0.1577	0.1896	
${}_A\hat{\rho}$	0.0314			
${}_B\hat{\alpha}$	0.0187	0.0571	0.0714	0.0720
${}_B\hat{\beta}$	0.9190	0.3550	0.4478	0.2746
${}_B\hat{R}^2$	0.3864	0.1578	0.1905	
${}_B\hat{\rho}$	0.0281			
${}_{AC}\hat{\alpha}$	0.0058	0.0427	0.0535	0.0461
${}_{AC}\hat{\beta}$	0.7263	0.2700	0.3442	0.2081
${}_{AC}\hat{R}^2$	0.3933	0.1586	0.1909	
${}_{AC}\hat{\rho}$	0.0268			
${}_{BC}\hat{\alpha}$	-0.0052	0.0580	0.0729	0.0574
${}_{BC}\hat{\beta}$	0.9183	0.3426	0.4216	0.2591
${}_{BC}\hat{R}^2$	0.4012	0.1602	0.1922	
${}_{BC}\hat{\rho}$	0.0262			

* Defined as: $\frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$, where $N = 304$. $\hat{\rho}$ = first order serial correlation coefficient.

each firm. One of the reasons for the “traditional” theory position on leverage is precisely this point—that small and reasonable amounts of leverage cannot be discerned by the market. In fact, if the MM theory is correct, leverage has explained as much as, roughly, 21 to 24 per cent of the value of the mean β .

We can also note that if the covariance between the asset and market rates of return, as well as the market variance, was constant over time, then the systematic risk from the market model is related to the expected rate of return by the capital asset pricing model. That is:

$$E(R_{At}) = R_{Ft} + {}_A\beta[E(R_{Mt}) - R_{Ft}] \tag{11a}$$

$$E(R_{Bt}) = R_{Ft} + {}_B\beta[E(R_{Mt}) - R_{Ft}] \tag{11b}$$

Equation (11a) indicates the relationship between the expected rate of return for the common stock shareholder of a debt-free and preferred-free firm, to the systematic risk, ${}_A\beta$, as obtained in regressions (10a) or (10c). The LHS of (11a) is the important $\rho\tau$ for the MM cost of capital. The MM theory [9, 10] also predicts that shareholder expected yield must be higher (for the same real firm) when the firm has debt than when it does not. Financial risk is greater, therefore, shareholders require more expected return. Thus, $E(R_{Bt})$ must be greater than $E(R_{At})$. In order for this MM prediction to be true, from (11a) and (11b) it can be observed that ${}_B\beta$ must be greater than ${}_A\beta$, which is what we obtained.

Using the results underlying Table 1, namely the firm and stock betas, as the

criterion for selecting among the possible observed market value ratios that can be used, if any, for (4), the following cross-section regressions were run:

$$(\beta_B)_i = a_1 + b_1 \left(\frac{S_A}{S_B} \beta_A \right)_i + u_{1i} \quad i = 1, 2, \dots, 102 \quad (12a)$$

$$(\beta_{BC})_i = a_2 + b_2 \left(\frac{S_A}{S_B} \beta_{AC} \right)_i + u_{2i} \quad i = 1, 2, \dots, 102 \quad (12b)$$

$$(\beta_A)_i = a_3 + b_3 \left(\frac{S_B}{S_A} \beta_B \right)_i + u_{3i} \quad i = 1, 2, \dots, 102 \quad (13a)$$

$$(\beta_{AC})_i = a_4 + b_4 \left(\frac{S_B}{S_A} \beta_{BC} \right)_i + u_{4i} \quad i = 1, 2, \dots, 102 \quad (13b)$$

Because the preferred stock market values were not as reliable as debt, only the 102 firms (out of 304) that did not have preferred in any of the years were used. The test for the adequacy of this alternative approach, equation (4), to adjust the systematic risk of common stocks for the underlying firm's capital structure, is whether the intercept term, a , is equal to zero, and the slope coefficient, b , is equal to one in the above regressions (as well as, of course, a high R^2)—these requirements are implied by (4). The results of this test would also indicate whether future "market model" studies that only use common stock rates of return without adjusting, or even noting, for the firm's debt-equity ratio will be adequate. The total firm's systematic risk may be stable (as long as the firm stays in the same risk-class), whereas the common stock's systematic risk may not be stable merely because of unanticipated capital structure changes—the data underlying Table 3 indicate that there were very few firms which did not have major changes in their capital structure over the twenty years studied.

The results of these regressions, when using the average S_A and average S_B over the twenty years for each firm, are shown in the first column panel of Table 2. These regressions were then replicated twice, first using the December 31, 1947 values of S_{A1} and S_{B1} instead of the twenty-year average for each firm, and then substituting the December 31, 1966 values of S_{A1} and S_{B1} for the 1947 values. These results are in the second and third panels of Table 2.⁶

From the first panel of Table 2, it appears that this alternative approach via (4a) for adjusting the systematic risk for the firm's leverage is quite

6. The point should be made that we are not merely regressing a variable on itself in (12) and (13). (12a) and (12b) can be interpreted as correlating the β_B obtained from (10b) and (10d)—the LHS variable in (12a) and (12b)—against the β_B obtained from rearranging (4)—the RHS variable in (12a) and (12b)—to determine whether the use of (4) is as good a means of obtaining β_B as the direct way via the equations (10). We would be regressing a variable on itself only if the β_A were calculated using (4a), and then the β_A thus obtained, inserted into (12a) and (12b).

Instead, we are obtaining β_A using the MM model in *each* of the twenty years so that a leverage-adjusted 20 year time series of R_{A1} is derived. Of course, if there were no data nor measurement problems, and if the debt-to-equity ratio were perfectly stable over this twenty year period for each firm, then we should obtain perfect correlation in (12a) and (12b), with $a = 0$ and $b = 1$, as (4) would be an identity.

TABLE 2
RESULTS FOR THE EQUATIONS (12a), (12b), (13a), AND (13b)*

	Using 20-Year Average for $\left(\frac{S_A}{S_B}\right)_i$				Using 1947 Value for $\left(\frac{S_A}{S_B}\right)_i$				Using 1966 Value for $\left(\frac{S_A}{S_B}\right)_i$			
	a	b	R^2		a	b	R^2		a	b	R^2	
Eq. (12a)	-0.022 (0.021) constant suppressed	1.062 (0.021) 1.042 (0.009)	0.962 0.962		0.150 (0.048) constant suppressed	0.842 (0.045) 0.966 (0.021)	0.781 0.781		0.085 (0.041) constant suppressed	0.905 (0.038) 0.976 (0.017)	0.849 0.849	
Eq. (12b)	-0.003 (0.013) constant suppressed	1.016 (0.013) 1.014 (0.005)	0.984 0.984		0.159 (0.047) constant suppressed	0.816 (0.044) 0.952 (0.019)	0.773 0.773		0.124 (0.037) constant suppressed	0.843 (0.034) 0.947 (0.015)	0.859 0.859	
	Using 20-Year Average for $\left(\frac{S_B}{S_A}\right)_i$				Using 1947 Value for $\left(\frac{S_B}{S_A}\right)_i$				Using 1966 Value for $\left(\frac{S_B}{S_A}\right)_i$			
	a	b	R^2		a	b	R^2		a	b	R^2	
Eq. (13a)	0.030 (0.016) constant suppressed	0.931 (0.017) 0.960 (0.007)	0.969 0.969		0.112 (0.028) constant suppressed	0.843 (0.030) 0.948 (0.015)	0.888 0.888		0.080 (0.027) constant suppressed	0.898 (0.030) 0.976 (0.014)	0.902 0.902	
Eq. (13b)	0.007 (0.010) constant suppressed	0.979 (0.011) 1.004 (0.012)	0.988 0.911		0.119 (0.026) constant suppressed	0.852 (0.028) 0.967 (0.013)	0.902 0.902		0.063 (0.026) constant suppressed	0.942 (0.029) 1.005 (0.012)	0.911 0.911	

* Standard error in parentheses.

satisfactory (at least with respect to our sample of firms and years) only if long-run averages of S_A and S_B are used. The second and third panels indicate that the equations (8) and (10) procedure is markedly superior when only one year's market value ratio is used as the adjustment factor. The annual debt-to-equity ratio is much too unstable for this latter procedure.

Thus, when forecasting systematic risk is the primary objective—for example, for portfolio decisions or for estimating the firm's cost of capital to apply to prospective projects—a long-run forecasted leverage adjustment is required. Assuming the firm's risk is more stable than the common stock's risk,⁷ and if there is some reason to believe that a better forecast of the firm's future leverage can be obtained than using simply a past year's (or an average of past years') leverage, it should be possible to improve the usual extrapolation forecast of a stock's systematic risk by forecasting the total firm's systematic risk first, and then using the independent leverage estimate as an adjustment.

IV. TESTS OF THE MM VS. TRADITIONAL THEORIES OF CORPORATION FINANCE

To determine if the difference, ${}_B\beta - {}_A\beta$, found in this study is indeed the correct effect of leverage, some confirmation of the MM theory (since it was assumed to be correct up to this point) from the systematic risk approach is needed. Since a direct test by this approach seems impossible, an indirect, inferential test is suggested.

The MM theory [9, 10] predicts that for firms in the same risk-class, the capitalization rate if all the firms were financed with only common equity, $E(R_A)$, would be the same—regardless of the actual amount of debt and preferred each individual firm had. This would imply, from (11a), that if $E(R_A)$ must be the same for all firms in a risk-class, so must ${}_A\beta$. And if these firms had different ratios of fixed commitment obligations to common equity, this difference in financial risk would cause their observed ${}_B\beta$ s to be different.

The major competing theory of corporation finance is what is now known as the “traditional theory,” which has contrary implications. This theory predicts that the capitalization rate for common equity, $E(R_B)$, (sometimes called the required or expected stock yield, or expected earnings-price ratio) is constant, as debt is increased, up to some critical leverage point (this point being a function of gambler's ruin and bankruptcy costs).⁸ The clear implication of this constant, horizontal, equity yield (or their initial downward sloping cost of capital curve) is that changes in market or covariability risk are assumed not to be discernible to the shareholders as debt is increased. Then the traditional theory is saying that the ${}_B\beta$ s, a measure of this covariability risk, would be the same for all firms in a given risk-class irregardless of differences in leverage, as long as the critical leverage point is not reached.

Since there will always be unavoidable errors in estimating the β 's of indi-

7. A faint, but possible, empirical indication of this point may be obtained from Table 1. The ratio of the mean point estimate to the mean standard error of estimate is less for the firm β than for the stock β in both the discrete and continuously compounded cases.

8. This interpretation of the traditional theory can be found in [9, especially their figure 2, page 275, and their equation (13) and footnote 24 where reference is made to Durand and Graham and Dodd].

TABLE 3
INDUSTRY MARKET VALUE RATIOS OF PREFERRED STOCK (P) AND DEBT (D) TO COMMON STOCK (S)

Industry Number	Industry	Number of Firms	Mean* ROM** ROCR***	P/S	D/S	P + D S
20	Food and Kindred Products	30	0.00 0.00 0.00	0.22 1.18 2.52	0.81 0.00 0.00	1.04 0.00 0.00
28	Chemicals and Allied Products	30	0.00 0.00 0.00	0.07 0.51 1.54	0.25 0.00 0.00	0.33 0.00 0.00
29	Petroleum and Coal Products	18	0.00 0.00 0.00	0.06 0.26 0.83	0.22 0.00 0.00	0.27 0.03 0.00
33	Primary Metals	21	0.00 0.00 0.00	0.14 1.31 4.69	0.54 0.00 0.00	0.68 0.00 0.00
35	Machinery, except Electrical	28	0.00 0.00 0.00	0.07 0.49 1.28	0.33 0.00 0.00	0.40 0.00 0.00

Capital Structure and Systematic Risk

447

TABLE 3 (Continued)

Industry Number	Industry	Number of Firms	P/S	D/S	P + D	
					S	S
36	Electrical Machinery & Equipment	13	Mean	0.06	0.35	0.41
			ROM	0.00	0.00	0.01
			ROCR	0.00	0.00	0.00
37	Transportation Equipment	24	Mean	0.08	0.38	0.47
			ROM	0.00	0.00	0.00
			ROCR	0.00	0.00	0.00
49	Utilities	27	Mean	0.25	1.03	1.28
			ROM	0.00	0.49	0.52
			ROCR	0.00	0.12	0.12
53	Dep't Stores, Order Houses & Vending Mach. Operators	17	Mean	0.13	0.49	0.62
			ROM	0.00	0.01	0.01
			ROCR	0.00	0.00	0.00

* "Mean" refers to the average ratio over 20 years and over all firms in the industry.

** "Range of Means" (ROM) refers to the lowest firm's mean (over 20 years) ratio and the highest firm's mean (over 20 years) ratio in the industry.

*** "Range of Company Ranges" (ROCR) refers to the lowest and highest ratio in the industry, regardless of the year.

vidual firms and in specifying a risk-class, we would not expect to find a set of firms with identical systematic risk. But by specifying reasonable a priori risk-classes, if the individual firms had closer or less scattered $_{A}\beta$ s than $_{B}\beta$ s, then this would support the MM theory and contradict the traditional theory. If, instead, the $_{B}\beta$ s were not discernibly more diverse than the $_{A}\beta$ s, and the leverage ratio differed considerably among firms, then this would indicate support for the traditional theory.⁹

In order to test this implication, risk-classes must be first specified. The SEC two-digit industry classification was used for this purpose. Requiring enough firms for statistical reasons in any given industry, nine risk-classes were specified that had at least 13 firms; these nine classes are listed in Table 3 with their various leverage ratios.¹⁰ It is clear from this table that our first requirement is met—that there is a considerable range of leverage ratios among firms in a risk-class and also over the twenty-year period.

Three tests will be performed to distinguish between the MM and traditional theories. The first is simply to calculate the standard deviation of the unbiased β estimates in a risk-class. The second is a chi-square test of the distribution of β 's in an industry compared to the distribution of the β 's in the total sample. Finally, an analysis of variance test on the estimated variance of the β 's between industries, as opposed to within industries, is performed. In all tests, only the point estimate of β (which should be unbiased) for each stock and firm is used.¹¹

The first test is reported in Table 4. If we compare the standard deviation of $_{AC}\beta$ with the standard deviation of $_{BC}\beta$ by industries (or risk-classes), we can note that $\sigma(_{AC}\beta)$ is less than $\sigma(_{BC}\beta)$ for eight out of the nine classes. The probability of obtaining this is only 0.0195, given a 50% probability that $\sigma(_{AC}\beta)$ can be larger or smaller than $\sigma(_{BC}\beta)$. These results indicate that the systematic risk of the firms in a given risk-class, if they were all financed only with common equity, is much less diverse than their observed stock's systematic risk. This supports the MM theory, at least in contrast to the traditional theory.¹²

9. The traditional theory also implies that $E(R_A)$ is equal to $E(R_B)$ for all firms. Unfortunately, we do not have a functional relationship between these traditional theory capitalization rates and the measured β s of this study. Clearly, since the $_{A}\beta$ s were obtained assuming the validity of the MM theory, they would not be applicable for the traditional theory. In fact, no relationship between the $_{A}\beta$ and $_{B}\beta$ for a given firm, or for firms in a given risk-class, can be specified as was done for the capitalization rates.

10. The tenth largest industry had only eight firms. For our purpose of testing the uniformity of firm β s relative to stock β s within a risk-class, the use of the two-digit industry classification as a proxy does not seem as critical as, for instance, its use for the purpose of performing an MM valuation model study [8] wherein the ρ^T must be pre-specified to be exactly the same for all firms in the industry.

11. Since these β s are estimated in the market model regressions with error, precise testing should incorporate the errors in the β estimation. Unfortunately, to do this is extremely difficult and more importantly, requires the normality assumption for the market model disturbance term. Since there is considerable evidence that is contrary to this required assumption [see 3], our tests will ignore the β measurement error entirely. But ignoring this is partially corrected in our first and third tests since means and variances of these point estimate β s must be calculated, and this procedure will "average out" the individual measurement errors by the factor $1/N$.

12. Of course, there could always be another theory, as yet not formulated, which could be even

TABLE 4
 MEAN AND STANDARD DEVIATION OF INDUSTRY β 's

Industry Number	Industry	Number of Firms		${}_A\beta$	${}_B\beta$	${}_{AO}\beta$	${}_{BO}\beta$
20	Food & Kindred Products	30	Mean β	0.515	0.815	0.528	0.806
			$\sigma(\beta)$	0.232	0.448	0.227	0.424
28	Chemicals & Allied Products	30	Mean β	0.747	0.928	0.785	0.946
			$\sigma(\beta)$	0.237	0.391	0.216	0.329
29	Petroleum & Coal Products	18	Mean β	0.633	0.747	0.656	0.756
			$\sigma(\beta)$	0.144	0.188	0.148	0.176
33	Primary Metals	21	Mean β	1.036	1.399	1.106	1.436
			$\sigma(\beta)$	0.223	0.272	0.197	0.268
35	Machinery, except Electrical	28	Mean β	0.878	1.037	0.917	1.068
			$\sigma(\beta)$	0.262	0.240	0.271	0.259
36	Electrical Machinery and Equipment	13	Mean β	0.940	1.234	0.951	1.164
			$\sigma(\beta)$	0.320	0.505	0.283	0.363
37	Transportation Equipment	24	Mean β	0.860	1.062	0.875	1.048
			$\sigma(\beta)$	0.225	0.313	0.225	0.289
49	Utilities	27	Mean β	0.160	0.255	0.166	0.254
			$\sigma(\beta)$	0.086	0.133	0.098	0.147
53	Department Stores, etc.	17	Mean β	0.652	0.901	0.692	0.923
			$\sigma(\beta)$	0.187	0.282	0.198	0.279

Our second test, the chi-square test, requires us to rank our 300 ${}_A\beta$ s into ten equal categories, each with 30 ${}_A\beta$ s (four miscellaneous firms were taken out randomly). By noting the value of the highest and lowest ${}_A\beta$ for each of the ten categories, a distribution of the number of ${}_A\beta$ s in each category, by risk-class, can be obtained. This was then repeated for the other three betas. To test whether the distribution for each of the four β 's and for each of the risk-classes follows the expected uniform distribution, a chi-square test was performed.¹³

Even with just casual inspection of these distributions of the betas by risk-class, it is clear that two industries, primary metals and utilities, are so highly skewed that they greatly exaggerate our results.¹⁴ Eliminating these

more strongly supported than the MM theory. If we compare $\sigma({}_A\beta)$ to $\sigma({}_B\beta)$ by risk-classes in Table 4, precisely the same results are obtained as those reported above for the continuously-compounded betas.

13. By risk-classes, seven of the nine chi-square values of ${}_A\beta$ are larger than those of ${}_B\beta$, as are eight out of nine for the continuously-compounded betas. This would occur by chance with probabilities of 0.0898 and 0.0195, respectively, if there were a 50% chance that either the firm or stock chi-square value could be larger. Nevertheless, if we inspect the individual chi-square values by risk-class, we note that most of them are large so that the probabilities of obtaining these values are highly unlikely. For all four β s, the distributions for most of the risk-classes are nonuniform.

14. Primary metals have extremely large betas; utilities have extremely small betas.

two industries, and also two miscellaneous firms so that an even 250 firms are in the sample, new upper and lower values of the β 's were obtained for each of the ten class intervals and for each of the four β 's.

In Table 5, the chi-square values are presented; for the total of all risk-classes, the probability of obtaining a chi-square value less than 120.63 is over 99.95% (for $A\beta$), whereas the probability of obtaining a chi-square value less than 99.75 is between 99.5% and 99.9% (for $B\beta$). More sharply contrasting results are obtained when $AC\beta$ is compared to $BC\beta$. For $AC\beta$, the probability of obtaining less than 128.47 is over 99.95%, whereas for $BC\beta$, the probability of obtaining less than 78.65 is only 90.0%. By abstracting from financial risk, the underlying systematic risk is much less scattered when grouped into risk-classes than when leverage is assumed not to affect the systematic risk. The null hypothesis that the β 's in a risk-class come from the same distribution as all β 's is rejected for $AC\beta$, but not for $BC\beta$ (at the 90% level). Although this, in itself, does not tell us *how* a risk-class differs from the total market, an inspection of the distributions of the betas by risk-class underlying Table 5 does indicate more clustering of the $AC\beta$ s than the $BC\beta$ s so that the MM theory is again favored over the traditional theory.

The analysis of variance test is our last comparison of the implications of the two theories. The ratio of the estimated variance between industries to the estimated variance within the industries (the F-statistic) when the seven

TABLE 5
CHI-SQUARE RESULTS FOR ALL β 'S AND ALL INDUSTRIES
(EXCEPT UTILITIES AND PRIMARY METALS)

Industry		$A\beta$	$B\beta$	$AC\beta$	$BC\beta$
Food and Kindred	Chi-Square	18.67	11.33	26.00	9.33
	$P\{\chi^2 < \}^* =$	95-97.5%	70-75%	99.5-99.9%	50-60%
Chemicals	Chi-Square	9.33	10.67	12.00	7.33
	$P\{\chi^2 < \} =$	50-60%	60-70%	75-80%	30-40%
Petroleum	Chi-Square	17.56	25.33	18.67	22.00
	$P\{\chi^2 < \} =$	95-97.5%	99.5-99.9%	95-97.5%	99-99.5%
Machinery	Chi-Square	19.14	12.00	24.86	9.14
	$P\{\chi^2 < \} =$	97.5-98%	75-80%	99.5-99.9%	50-60%
Electrical Machinery	Chi-Square	13.92	7.77	12.38	9.31
	$P\{\chi^2 < \} =$	80-90%	40-50%	80-90%	50-60%
Transportation Equipment	Chi-Square	15.17	16.83	13.50	6.83
	$P\{\chi^2 < \} =$	90-95%	90-95%	80-90%	30-40%
Dep't Stores	Chi-Square	14.18	3.59	14.18	3.59
	$P\{\chi^2 < \} =$	80-90%	5-10%	80-90%	5-10%
Miscellaneous	Chi-Square	12.67	12.22	6.89	11.11
	$P\{\chi^2 < \} =$	80-90%	80-90%	30-40%	70-75%
Total	Chi-Square	120.63	99.75	128.47	78.65
	$P\{\chi^2 < \} =$	over 99.95%	99.5-99.90%	over 99.95%	90.0%

* Example: $P\{\chi^2 < 18.67\} = 95-97.5\%$ for 9 degrees of freedom.

industries are considered (again, the two obviously skewed industries, primary metals and utilities, were eliminated) is less for ${}_B\beta$ ($F = 3.90$) than for ${}_A\beta$ ($F = 9.99$), and less for ${}_{BC}\beta$ ($F = 4.18$) than for ${}_{AC}\beta$ ($F = 10.83$). The probability of obtaining these F-statistics for ${}_A\beta$ and ${}_{AC}\beta$ is less than 0.001, but for ${}_B\beta$ and ${}_{BC}\beta$ greater than or equal to 0.001. These results are consistent with the results obtained from our two previous tests. The MM theory is more compatible with the data than the traditional theory.¹⁵

V. CONCLUSIONS

This study attempted to tie together some of the notions associated with the field of corporation finance with those associated with security and portfolio analyses. Specifically, if the MM corporate tax leverage propositions are correct, then approximately 21 to 24% of the observed systematic risk of common stocks (when averaged over 304 firms) can be explained merely by the added financial risk taken on by the underlying firm with its use of debt and preferred stock. Corporate leverage does count considerably.

To determine whether the MM theory is correct, a number of tests on a contrasting implication of the MM and "traditional" theories of corporation finance were performed. The data confirmed MM's position, at least vis-à-vis our interpretation of the traditional theory's position. This should provide another piece of evidence on this controversial topic.

Finally, if the MM theory and the capital asset pricing model are correct, and if the adjustments made in equations (8) or (4a) result in accurate measures of the systematic risk of a leverage-free firm, the possibility is greater, without resorting to a fullblown risk-class study of the type MM did for the electric utility industry [8], of estimating the cost of capital for individual firms.

REFERENCES

1. M. Blume. "Portfolio Theory: A Step Toward Its Practical Application," *Journal of Business* 43 (April, 1970), 152-173.
2. P. Brown. "Some Aspects of Valuation in the Railroad Industry." Unpublished Ph.D. dissertation, Graduate School of Business, University of Chicago, 1968.
3. E. Fama. "The Behavior of Stock Market Prices," *Journal of Business* 38 (January, 1965), 34-105.
4. E. Fama, and M. Miller. *The Theory of Finance*. Chapter 4, Holt, Rinehart and Winston, 1972.
5. R. Hamada. "Portfolio Analysis, Market Equilibrium and Corporation Finance," *Journal of Finance* (March, 1969), 13-31.
6. J. Lintner. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics* (February, 1965), 13-37.
7. H. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons, Inc., 1959.
8. M. Miller, and F. Modigliani. "Some Estimates of the Cost of Capital to the Electric Utility Industry, 1954-57," *American Economic Review* (June, 1966), 333-91.

15. All of our tests, it should be emphasized, although consistent, are only inferential. Aside from assuming that the two-digit SEC industry classification is a good proxy for risk-classes and that the errors in estimating the individual β s can be safely ignored, the tests rely on the two theories exhausting all the reasonable theories on leverage. But there is always the use of another line of reasoning. If the results of the MM electric utility study [8] are correct, and if these results can be generalized to all firms and to all risk-classes, then it can be claimed that the MM theory is universally valid. Then our result in Section III does indicate the correct effect of the firm's capital structure on the systematic risk of common stocks.

9. F. Modigliani, and M. Miller. "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review* (June, 1958), 261-97.
10. ———. "Corporate Income Taxes and the Cost of Capital: A Correction," *American Economic Review* (June, 1963), 433-43.
11. J. Mossin. "Equilibrium in a Capital Asset Market," *Econometrica* (October, 1966), 768-83.
12. W. Sharpe. "A Simplified Model for Portfolio Analysis," *Management Science* (January, 1963), 277-93.
13. ———. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance* (September, 1964), 425-42.

SESSION TOPIC: STOCK MARKET PRICE BEHAVIOR

SESSION CHAIRMAN: BURTON G. MALKIEL

EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*

EUGENE F. FAMA**

I. INTRODUCTION

THE PRIMARY ROLE of the capital market is allocation of ownership of the economy's capital stock. In general terms, the ideal is a market in which prices provide accurate signals for resource allocation: that is, a market in which firms can make production-investment decisions, and investors can choose among the securities that represent ownership of firms' activities under the assumption that security prices at any time "fully reflect" all available information. A market in which prices always "fully reflect" available information is called "efficient."

This paper reviews the theoretical and empirical literature on the efficient markets model. After a discussion of the theory, empirical work concerned with the adjustment of security prices to three relevant information subsets is considered. First, *weak form* tests, in which the information set is just historical prices, are discussed. Then *semi-strong form* tests, in which the concern is whether prices efficiently adjust to other information that is obviously publicly available (e.g., announcements of annual earnings, stock splits, etc.) are considered. Finally, *strong form* tests concerned with whether given investors or groups have monopolistic access to any information relevant for price formation are reviewed.¹ We shall conclude that, with but a few exceptions, the efficient markets model stands up well.

Though we proceed from theory to empirical work, to keep the proper historical perspective we should note to a large extent the empirical work in this area preceded the development of the theory. The theory is presented first here in order to more easily judge which of the empirical results are most relevant from the viewpoint of the theory. The empirical work itself, however, will then be reviewed in more or less historical sequence.

Finally, the perceptive reader will surely recognize instances in this paper where relevant studies are not specifically discussed. In such cases my apologies should be taken for granted. The area is so bountiful that some such injustices are unavoidable. But the primary goal here will have been accomplished if a coherent picture of the main lines of the work on efficient markets is presented, along with an accurate picture of the current state of the arts.

* Research on this project was supported by a grant from the National Science Foundation. I am indebted to Arthur Laffer, Robert Aliber, Ray Ball, Michael Jensen, James Lorie, Merton Miller, Charles Nelson, Richard Roll, William Taylor, and Ross Watts for their helpful comments.

** University of Chicago—Joint Session with the Econometric Society.

1. The distinction between weak and strong form tests was first suggested by Harry Roberts.

II. THE THEORY OF EFFICIENT MARKETS

A. *Expected Return or "Fair Game" Models*

The definitional statement that in an efficient market prices "fully reflect" available information is so general that it has no empirically testable implications. To make the model testable, the process of price formation must be specified in more detail. In essence we must define somewhat more exactly what is meant by the term "fully reflect."

One possibility would be to posit that equilibrium prices (or expected returns) on securities are generated as in the "two parameter" Sharpe [40]-Lintner [24, 25] world. In general, however, the theoretical models and especially the empirical tests of capital market efficiency have not been this specific. Most of the available work is based only on the assumption that the conditions of market equilibrium can (somehow) be stated in terms of expected returns. In general terms, like the two parameter model such theories would posit that conditional on some relevant information set, the equilibrium expected return on a security is a function of its "risk." And different theories would differ primarily in how "risk" is defined.

All members of the class of such "expected return theories" can, however, be described notationally as follows:

$$E(\tilde{p}_{j,t+1}|\Phi_t) = [1 + E(\tilde{r}_{j,t+1}|\Phi_t)]p_{jt}, \quad (1)$$

where E is the expected value operator; p_{jt} is the price of security j at time t ; $p_{j,t+1}$ is its price at $t + 1$ (with reinvestment of any intermediate cash income from the security); $r_{j,t+1}$ is the one-period percentage return $(p_{j,t+1} - p_{jt})/p_{jt}$; Φ_t is a general symbol for whatever set of information is assumed to be "fully reflected" in the price at t ; and the tildes indicate that $p_{j,t+1}$ and $r_{j,t+1}$ are random variables at t .

The value of the equilibrium expected return $E(\tilde{r}_{j,t+1}|\Phi_t)$ projected on the basis of the information Φ_t would be determined from the particular expected return theory at hand. The conditional expectation notation of (1) is meant to imply, however, that whatever expected return model is assumed to apply, the information in Φ_t is fully utilized in determining equilibrium expected returns. And this is the sense in which Φ_t is "fully reflected" in the formation of the price p_{jt} .

But we should note right off that, simple as it is, the assumption that the conditions of market equilibrium can be stated in terms of expected returns elevates the purely mathematical concept of expected value to a status not necessarily implied by the general notion of market efficiency. The expected value is just one of many possible summary measures of a distribution of returns, and market efficiency per se (i.e., the general notion that prices "fully reflect" available information) does not imbue it with any special importance. Thus, the results of tests based on this assumption depend to some extent on its validity as well as on the efficiency of the market. But some such assumption is the unavoidable price one must pay to give the theory of efficient markets empirical content.

The assumptions that the conditions of market equilibrium can be stated

in terms of expected returns and that equilibrium expected returns are formed on the basis of (and thus “fully reflect”) the information set Φ_t have a major empirical implication—they rule out the possibility of trading systems based only on information in Φ_t that have expected profits or returns in excess of equilibrium expected profits or returns. Thus let

$$x_{j,t+1} = p_{j,t+1} - E(p_{j,t+1}|\Phi_t). \quad (2)$$

Then

$$E(\tilde{x}_{j,t+1}|\Phi_t) = 0 \quad (3)$$

which, *by definition*, says that the sequence $\{x_{jt}\}$ is a “fair game” with respect to the information sequence $\{\Phi_t\}$. Or, equivalently, let

$$z_{j,t+1} = r_{j,t+1} - E(\tilde{r}_{j,t+1}|\Phi_t), \quad (4)$$

then

$$E(\tilde{z}_{j,t+1}|\Phi_t) = 0, \quad (5)$$

so that the sequence $\{z_{jt}\}$ is also a “fair game” with respect to the information sequence $\{\Phi_t\}$.

In economic terms, $x_{j,t+1}$ is the excess market value of security j at time $t + 1$: it is the difference between the observed price and the expected value of the price that was projected at t on the basis of the information Φ_t . And similarly, $z_{j,t+1}$ is the return at $t + 1$ in excess of the equilibrium expected return projected at t . Let

$$\alpha(\Phi_t) = [\alpha_1(\Phi_t), \alpha_2(\Phi_t), \dots, \alpha_n(\Phi_t)]$$

be any trading system based on Φ_t which tells the investor the amounts $\alpha_j(\Phi_t)$ of funds available at t that are to be invested in each of the n available securities. The total excess market value at $t + 1$ that will be generated by such a system is

$$V_{t+1} = \sum_{j=1}^n \alpha_j(\Phi_t) [r_{j,t+1} - E(\tilde{r}_{j,t+1}|\Phi_t)],$$

which, from the “fair game” property of (5) has expectation,

$$E(\tilde{V}_{t+1}|\Phi_t) = \sum_{j=1}^n \alpha_j(\Phi_t) E(\tilde{z}_{j,t+1}|\Phi_t) = 0.$$

The expected return or “fair game” efficient markets model² has other important testable implications, but these are better saved for the later discussion of the empirical work. Now we turn to two special cases of the model, the submartingale and the random walk, that (as we shall see later) play an important role in the empirical literature.

2. Though we shall sometimes refer to the model summarized by (1) as the “fair game” model, keep in mind that the “fair game” properties of the model are *implications* of the assumptions that (i) the conditions of market equilibrium can be stated in terms of expected returns, and (ii) the information Φ_t is fully utilized by the market in forming equilibrium expected returns and thus current prices.

The role of “fair game” models in the theory of efficient markets was first recognized and studied rigorously by Mandelbrot [27] and Samuelson [38]. Their work will be discussed in more detail later.

B. *The Submartingale Model*

Suppose we assume in (1) that for all t and Φ_t

$$E(\tilde{p}_{j,t+1}|\Phi_t) \geq p_{jt}, \quad \text{or equivalently, } E(\tilde{r}_{j,t+1}|\Phi_t) \geq 0. \quad (6)$$

This is a statement that the price sequence $\{p_{jt}\}$ for security j follows a submartingale with respect to the information sequence $\{\Phi_t\}$, which is to say nothing more than that the expected value of next period's price, as projected on the basis of the information Φ_t , is equal to or greater than the current price. If (6) holds as an equality (so that expected returns and price changes are zero), then the price sequence follows a martingale.

A submartingale in prices has one important empirical implication. Consider the set of "one security and cash" mechanical trading rules by which we mean systems that concentrate on individual securities and that define the conditions under which the investor would hold a given security, sell it short, or simply hold cash at any time t . Then the assumption of (6) that expected returns conditional on Φ_t are non-negative directly implies that such trading rules based only on the information in Φ_t cannot have greater expected profits than a policy of always buying-and-holding the security during the future period in question. Tests of such rules will be an important part of the empirical evidence on the efficient markets model.³

C. *The Random Walk Model*

In the early treatments of the efficient markets model, the statement that the current price of a security "fully reflects" available information was assumed to imply that successive price changes (or more usually, successive one-period returns) are independent. In addition, it was usually assumed that successive changes (or returns) are identically distributed. Together the two hypotheses constitute the random walk model. Formally, the model says

$$f(r_{j,t+1}|\Phi_t) = f(r_{j,t+1}), \quad (7)$$

which is the usual statement that the conditional and marginal probability distributions of an independent random variable are identical. In addition, the density function f must be the same for all t .⁴

3. Note that the expected profitability of "one security and cash" trading systems vis-à-vis buy-and-hold is not ruled out by the general expected return or "fair game" efficient markets model. The latter rules out systems with expected profits in excess of equilibrium expected returns, but since in principle it allows equilibrium expected returns to be negative, holding cash (which always has zero actual and thus expected return) may have higher expected return than holding some security.

And negative equilibrium expected returns for some securities are quite possible. For example, in the Sharpe [40]-Lintner [24, 25] model (which is in turn a natural extension of the portfolio models of Markowitz [30] and Tobin [43]) the equilibrium expected return on a security depends on the extent to which the dispersion in the security's return distribution is related to dispersion in the returns on all other securities. A security whose returns on average move opposite to the general market is particularly valuable in reducing dispersion of portfolio returns, and so its equilibrium expected return may well be negative.

4. The terminology is loose. Prices will only follow a random walk if price changes are independent, identically distributed; and even then we should say "random walk with drift" since expected price changes can be non-zero. If one-period returns are independent, identically distributed, prices will not follow a random walk since the distribution of price changes will depend

Expression (7) of course says much more than the general expected return model summarized by (1). For example, if we restrict (1) by assuming that the expected return on security j is constant over time, then we have

$$E(\tilde{r}_{j,t+1}|\Phi_t) = E(\tilde{r}_{j,t+1}). \quad (8)$$

This says that the mean of the distribution of $r_{j,t+1}$ is independent of the information available at t , Φ_t , whereas the random walk model of (7) in addition says that the entire distribution is independent of Φ_t .⁵

We argue later that it is best to regard the random walk model as an extension of the general expected return or “fair game” efficient markets model in the sense of making a more detailed statement about the economic environment. The “fair game” model just says that the conditions of market equilibrium can be stated in terms of expected returns, and thus it says little about the details of the stochastic process generating returns. A random walk arises within the context of such a model when the environment is (fortuitously) such that the evolution of investor tastes and the process generating new information combine to produce equilibria in which return distributions repeat themselves through time.

Thus it is not surprising that empirical tests of the “random walk” model that are in fact tests of “fair game” properties are more strongly in support of the model than tests of the additional (and, from the viewpoint of expected return market efficiency, superfluous) pure independence assumption. (But it is perhaps equally surprising that, as we shall soon see, the evidence against the independence of returns over time is as weak as it is.)

D. *Market Conditions Consistent with Efficiency*

Before turning to the empirical work, however, a few words about the market conditions that might help or hinder efficient adjustment of prices to information are in order. First, it is easy to determine *sufficient* conditions for capital market efficiency. For example, consider a market in which (i) there are no transactions costs in trading securities, (ii) all available information is costlessly available to all market participants, and (iii) all agree on the implications of current information for the current price and distributions of future prices of each security. In such a market, the current price of a security obviously “fully reflects” all available information.

But a frictionless market in which all information is freely available and investors agree on its implications is, of course, not descriptive of markets met in practice. Fortunately, these conditions are sufficient for market efficiency, but not necessary. For example, as long as transactors take account of all

on the price level. But though rigorous terminology is usually desirable, our loose use of terms should not cause confusion; and our usage follows that of the efficient markets literature.

Note also that in the random walk literature, the information set Φ_t in (7) is usually assumed to include only the past return history, $r_{j,t}, r_{j,t-1}, \dots$

5. The random walk model does not say, however, that past information is of no value in *assessing* distributions of future returns. Indeed since return distributions are assumed to be stationary through time, past returns are the best source of such information. The random walk model does say, however, that the *sequence* (or the order) of the past returns is of no consequence in assessing distributions of future returns.

available information, even large transactions costs that inhibit the flow of transactions do not in themselves imply that when transactions do take place, prices will not “fully reflect” available information. Similarly (and speaking, as above, somewhat loosely), the market may be efficient if “sufficient numbers” of investors have ready access to available information. And disagreement among investors about the implications of given information does not in itself imply market inefficiency unless there are investors who can consistently make better evaluations of available information than are implicit in market prices.

But though transactions costs, information that is not freely available to all investors, and disagreement among investors about the implications of given information are not necessarily sources of market inefficiency, they are potential sources. And all three exist to some extent in real world markets. Measuring their effects on the process of price formation is, of course, the major goal of empirical work in this area.

III. THE EVIDENCE

All the empirical research on the theory of efficient markets has been concerned with whether prices “fully reflect” particular subsets of available information. Historically, the empirical work evolved more or less as follows. The initial studies were concerned with what we call *weak form* tests in which the information subset of interest is just past price (or return) histories. Most of the results here come from the random walk literature. When extensive tests seemed to support the efficiency hypothesis at this level, attention was turned to *semi-strong form* tests in which the concern is the speed of price adjustment to other obviously publicly available information (e.g., announcements of stock splits, annual reports, new security issues, etc.). Finally, *strong form* tests in which the concern is whether any investor or groups (e.g., managements of mutual funds) have monopolistic access to any information relevant for the formation of prices have recently appeared. We review the empirical research in more or less this historical sequence.

First, however, we should note that what we have called *the* efficient markets model in the discussions of earlier sections is the hypothesis that security prices at any point in time “fully reflect” *all* available information. Though we shall argue that the model stands up rather well to the data, it is obviously an extreme null hypothesis. And, like any other extreme null hypothesis, we do not expect it to be literally true. The categorization of the tests into weak, semi-strong, and strong form will serve the useful purpose of allowing us to pinpoint the level of information at which the hypothesis breaks down. And we shall contend that there is no important evidence against the hypothesis in the weak and semi-strong form tests (i.e., prices seem to efficiently adjust to obviously publicly available information), and only limited evidence against the hypothesis in the strong form tests (i.e., monopolistic access to information about prices does not seem to be a prevalent phenomenon in the investment community).

A. *Weak Form Tests of the Efficient Markets Model*

1. Random Walks and Fair Games: A Little Historical Background

As noted earlier, all of the empirical work on efficient markets can be considered within the context of the general expected return or “fair game” model, and much of the evidence bears directly on the special submartingale expected return model of (6). Indeed, in the early literature, discussions of the efficient markets model were phrased in terms of the even more special random walk model, though we shall argue that most of the early authors were in fact concerned with more general versions of the “fair game” model.

Some of the confusion in the early random walk writings is understandable. Research on security prices did not begin with the development of a theory of price formation which was then subjected to empirical tests. Rather, the impetus for the development of a theory came from the accumulation of evidence in the middle 1950’s and early 1960’s that the behavior of common stock and other speculative prices could be well approximated by a random walk. Faced with the evidence, economists felt compelled to offer some rationalization. What resulted was a theory of efficient markets stated in terms of random walks, but usually implying some more general “fair game” model.

It was not until the work of Samuelson [38] and Mandelbrot [27] in 1965 and 1966 that the role of “fair game” expected return models in the theory of efficient markets and the relationships between these models and the theory of random walks were rigorously studied.⁶ And these papers came somewhat after the major empirical work on random walks. In the earlier work, “theoretical” discussions, though usually intuitively appealing, were always lacking in rigor and often either vague or *ad hoc*. In short, until the Mandelbrot-Samuelson models appeared, there existed a large body of empirical results in search of a rigorous theory.

Thus, though his contributions were ignored for sixty years, the first statement and test of the random walk model was that of Bachelier [3] in 1900. But his “fundamental principle” for the behavior of prices was that speculation should be a “fair game”; in particular, the expected profits to the speculator should be zero. With the benefit of the modern theory of stochastic processes, we know now that the process implied by this fundamental principle is a martingale.

After Bachelier, research on the behavior of security prices lagged until the

6. Basing their analyses on futures contracts in commodity markets, Mandelbrot and Samuelson show that if the price of such a contract at time t is the expected value at t (given information Φ_t) of the spot price at the termination of the contract, then the futures price will follow a martingale with respect to the information sequence $\{\Phi_t\}$; that is, the expected price change from period to period will be zero, and the price changes will be a “fair game.” If the equilibrium expected return is not assumed to be zero, our more general “fair game” model, summarized by (1), is obtained.

But though the Mandelbrot-Samuelson approach certainly illuminates the process of price formation in commodity markets, we have seen that “fair game” expected return models can be derived in much simpler fashion. In particular, (1) is just a formalization of the assumptions that the conditions of market equilibrium can be stated in terms of expected returns and that the information Φ_t is used in forming market prices at t .

coming of the computer. In 1953 Kendall [21] examined the behavior of weekly changes in nineteen indices of British industrial share prices and in spot prices for cotton (New York) and wheat (Chicago). After extensive analysis of serial correlations, he suggests, in quite graphic terms:

The series looks like a wandering one, almost as if once a week the Demon of Chance drew a random number from a symmetrical population of fixed dispersion and added it to the current price to determine the next week's price [21, p. 13].

Kendall's conclusion had in fact been suggested earlier by Working [47], though his suggestion lacked the force provided by Kendall's empirical results. And the implications of the conclusion for stock market research and financial analysis were later underlined by Roberts [36].

But the suggestion by Kendall, Working, and Roberts that series of speculative prices may be well described by random walks was based on observation. None of these authors attempted to provide much economic rationale for the hypothesis, and, indeed, Kendall felt that economists would generally reject it. Osborne [33] suggested market conditions, similar to those assumed by Bachelier, that would lead to a random walk. But in his model, independence of successive price changes derives from the assumption that the decisions of investors in an individual security are independent from transaction to transaction—which is little in the way of an economic model.

Whenever economists (prior to Mandelbrot and Samuelson) tried to provide economic justification for the random walk, their arguments usually implied a “fair game.” For example, Alexander [8, p. 200] states:

If one were to start out with the assumption that a stock or commodity speculation is a “fair game” with equal expectation of gain or loss or, more accurately, with an expectation of zero gain, one would be well on the way to picturing the behavior of speculative prices as a random walk.

There is an awareness here that the “fair game” assumption is not sufficient to lead to a random walk, but Alexander never expands on the comment. Similarly, Cootner [8, p. 232] states:

If any substantial group of buyers thought prices were too low, their buying would force up the prices. The reverse would be true for sellers. Except for appreciation due to earnings retention, the conditional expectation of tomorrow's price, given today's price, is today's price.

In such a world, the only price changes that would occur are those that result from new information. Since there is no reason to expect that information to be non-random in appearance, the period-to-period price changes of a stock should be random movements, statistically independent of one another.

Though somewhat imprecise, the last sentence of the first paragraph seems to point to a “fair game” model rather than a random walk.⁷ In this light, the second paragraph can be viewed as an attempt to describe environmental conditions that would reduce a “fair game” to a random walk. But the specification imposed on the information generating process is insufficient for this purpose; one would, for example, also have to say something about investor

7. The appropriate conditioning statement would be “Given the sequence of historical prices.”

tastes. Finally, lest I be accused of criticizing others too severely for ambiguity, lack of rigor and incorrect conclusions,

By contrast, the stock market trader has a much more practical criterion for judging what constitutes important dependence in successive price changes. For his purposes the random walk model is valid as long as knowledge of the past behavior of the series of price changes cannot be used to increase expected gains. More specifically, the independence assumption is an adequate description of reality as long as the actual degree of dependence in the series of price changes is not sufficient to allow the past history of the series to be used to predict the future in a way which makes expected profits greater than they would be under a naive buy-and hold model [10, p 35].

We know now, of course, that this last condition hardly requires a random walk. It will in fact be met by the submartingale model of (6).

But one should not be too hard on the theoretical efforts of the early empirical random walk literature. The arguments were usually appealing; where they fell short was in awareness of developments in the theory of stochastic processes. Moreover, we shall now see that most of the empirical evidence in the random walk literature can easily be interpreted as tests of more general expected return or "fair game" models.⁸

2. Tests of Market Efficiency in the Random Walk Literature

As discussed earlier, "fair game" models imply the "impossibility" of various sorts of trading systems. Some of the random walk literature has been concerned with testing the profitability of such systems. More of the literature has, however, been concerned with tests of serial covariances of returns. We shall now show that, like a random walk, the serial covariances of a "fair game" are zero, so that these tests are also relevant for the expected return models.

If x_t is a "fair game," its unconditional expectation is zero and its serial covariance can be written in general form as:

$$E(\tilde{x}_{t+\tau} \tilde{x}_t) = \int_{x_t} x_t E(\tilde{x}_{t+\tau} | x_t) f(x_t) dx_t,$$

where f indicates a density function. But if x_t is a "fair game,"

$$E(\tilde{x}_{t+\tau} | x_t) = 0.^9$$

8. Our brief historical review is meant only to provide perspective, and it is, of course, somewhat incomplete. For example, we have ignored the important contributions to the early random walk literature in studies of warrants and other options by Sprenkle, Kruizenga, Boness, and others. Much of this early work on options is summarized in [8].

9. More generally, if the sequence $\{x_t\}$ is a fair game with respect to the information sequence $\{\Phi_t\}$, (i.e., $E(\tilde{x}_{t+1} | \Phi_t) = 0$ for all Φ_t); then x_t is a fair game with respect to any Φ'_t that is a subset of Φ_t (i.e., $E(\tilde{x}_{t+1} | \Phi'_t) = 0$ for all Φ'_t). To show this, let $\Phi_t = (\Phi'_t, \Phi''_t)$. Then, using Stieltjes integrals and the symbol F to denote cumulative distribution functions, the conditional expectation

$$E(\tilde{x}_{t+1} | \Phi'_t) = \int_{\Phi''_t} \int_{x_{t+1}} x_{t+1} dF(x_{t+1}, \Phi''_t | \Phi'_t) = \int_{\Phi''_t} \left[\int_{x_{t+1}} x_{t+1} dF(x_{t+1} | \Phi'_t, \Phi''_t) \right] dF(\Phi'_t | \Phi'_t).$$

From this it follows that for all lags, the serial covariances between lagged values of a "fair game" variable are zero. Thus, observations of a "fair game" variable are linearly independent.¹⁰

But the "fair game" model does not necessarily imply that the serial covariances of *one-period returns* are zero. In the weak form tests of this model the "fair game" variable is

$$z_{j,t} = r_{j,t} - E(\tilde{r}_{j,t} | r_{j,t-1}, r_{j,t-2}, \dots). \quad (\text{Cf. fn. 9}) \quad (9)$$

But the covariance between, for example, r_{jt} and $r_{j,t+1}$ is

$$E([\tilde{r}_{j,t+1} - E(\tilde{r}_{j,t+1})][\tilde{r}_{jt} - E(\tilde{r}_{jt})]) \\ = \int_{r_{jt}} [r_{jt} - E(\tilde{r}_{jt})][E(\tilde{r}_{j,t+1} | r_{jt}) - E(\tilde{r}_{j,t+1})] f(r_{jt}) dr_{jt},$$

and (9) does not imply that $E(\tilde{r}_{j,t+1} | r_{jt}) = E(\tilde{r}_{j,t+1})$: In the "fair game" efficient markets model, the deviation of the return for $t+1$ from its conditional expectation is a "fair game" variable, but the conditional expectation itself can depend on the return observed for t .¹¹

In the random walk literature, this problem is not recognized, since it is assumed that the expected return (and indeed the entire distribution of returns) is stationary through time. In practice, this implies estimating serial covariances by taking cross products of deviations of observed returns from the overall sample mean return. It is somewhat fortuitous, then, that this procedure, which represents a rather gross approximation from the viewpoint of the general expected return efficient markets model, does not seem to greatly affect the results of the covariance tests, at least for common stocks.¹²

But the integral in brackets is just $E(\tilde{x}_{t+1} | \Phi_t)$ which by the "fair game" assumption is 0, so that

$$E(x_{t+1} | \Phi'_t) = 0 \text{ for all } \Phi'_t \subset \Phi_t.$$

10. But though zero serial covariances are consistent with a "fair game," they do not imply such a process. A "fair game" also rules out many types of non linear dependence. Thus using arguments similar to those above, it can be shown that if x is a "fair game," $E(\tilde{x}_t \tilde{x}_{t+1} \dots \tilde{x}_{t+\tau}) = 0$ for all τ , which is not implied by $E(\tilde{x}_t \tilde{x}_{t+\tau}) = 0$ for all τ . For example, consider a three-period case where x must be either ± 1 . Suppose the process is $x_{t+2} = \text{sign}(x_t x_{t+1})$, i.e.,

$\frac{x_t}{+}$	$\frac{x_{t+1}}{+}$	$\rightarrow$	$\frac{x_{t+2}}{+}$
$+$	$+$	$\rightarrow$	$+$
$+$	$-$	$\rightarrow$	$-$
$-$	$+$	$\rightarrow$	$-$
$-$	$-$	$\rightarrow$	$+$

If probabilities are uniformly distributed across events,

$$E(\tilde{x}_{t+2} | x_{t+1}) = E(\tilde{x}_{t+2} | x_t) = E(\tilde{x}_{t+1} | x_t) = E(\tilde{x}_{t+2}) = E(\tilde{x}_{t+1}) = E(\tilde{x}_t) = 0,$$

so that all pairwise serial covariances are zero. But the process is not a "fair game," since $E(\tilde{x}_{t+2} | x_{t+1}, x_t) \neq 0$, and knowledge of (x_{t+1}, x_t) can be used as the basis of a simple "system" with positive expected profit.

11. For example, suppose the level of one-period returns follows a martingale so that

$$E(\tilde{r}_{j,t+1} | r_{jt}, r_{j,t-1}, \dots) = r_{jt}.$$

Then covariances between successive returns will be nonzero (though in this special case first differences of returns will be uncorrelated).

12. The reason is probably that for stocks, changes in equilibrium expected returns for the

TABLE 1 (from [10])
First-order Serial Correlation Coefficients for One-, Four-, Nine-, and Sixteen-Day
Changes in $\log_e$ Price

Stock	One	Differencing Interval (Days)			Sixteen
		Four	Nine		
Allied Chemical	.017	.029	— .091		— .118
Alcoa	.118*	.095	— .112		— .044
American Can	— .087*	— .124*	— .060		.031
A. T. & T.	— .039	— .010	— .009		— .003
American Tobacco	.111*	— .175*	.033		.007
Anaconda	.067*	— .068	— .125		.202
Bethlehem Steel	.013	— .122	— .148		.112
Chrysler	.012	.060	— .026		.040
Du Pont	.013	.069	— .043		— .055
Eastman Kodak	.025	— .006	— .053		— .023
General Electric	.011	.020	— .004		.000
General Foods	.061*	— .005	— .140		— .098
General Motors	— .004	— .128*	.009		— .028
Goodyear	— .123*	.001	— .037		.033
International Harvester	— .017	— .068	— .244*		.116
International Nickel	.096*	.038	.124		.041
International Paper	.046	.060	— .004		— .010
Johns Manville	.006	— .068	— .002		.002
Owens Illinois	— .021	— .006	.003		— .022
Procter & Gamble	.099*	— .006	.098		.076
Sears	.097*	— .070	— .113		.041
Standard Oil (Calif.)	.025	— .143*	— .046		.040
Standard Oil (N.J.)	.008	— .109	— .082		— .121
Swift & Co.	— .004	— .072	.118		— .197
Texaco	.094*	— .053	— .047		— .178
Union Carbide	.107*	.049	— .101		.124
United Aircraft	.014	— .190*	— .192*		— .040
U.S. Steel	.040	— .006	— .056		.236*
Westinghouse	— .027	— .097	— .137		.067
Woolworth	.028	— .033	— .112		.040

* Coefficient is twice its computed standard error.

For example, Table 1 (taken from [10]) shows the serial correlations between successive changes in the natural log of price for each of the thirty stocks of the Dow Jones Industrial Average, for time periods that vary slightly from stock to stock, but usually run from about the end of 1957 to September 26, 1962. The serial correlations of successive changes in $\log_e$ price are shown for differencing intervals of one, four, nine, and sixteen days.¹³

common differencing intervals of a day, a week, or a month, are trivial relative to other sources of variation in returns. Later, when we consider Roll's work [37], we shall see that this is not true for one week returns on U.S. Government Treasury Bills.

13. The use of changes in $\log_e$ price as the measure of return is common in the random walk literature. It can be justified in several ways. But for current purposes, it is sufficient to note that for price changes less than fifteen per cent, the change in $\log_e$ price is approximately the percentage price change or one-period return. And for differencing intervals shorter than one month, returns in excess of fifteen per cent are unusual. Thus [10] reports that for the data of Table 1, tests carried out on percentage or one-period returns yielded results essentially identical to the tests based on changes in $\log_e$ price.

The results in Table 1 are typical of those reported by others for tests based on serial covariances. (Cf. Kendall [21], Moore [31], Alexander [1], and the results of Granger and Morgenstern [17] and Godfrey, Granger and Morgenstern [16] obtained by means of spectral analysis.) Specifically, there is no evidence of substantial linear dependence between lagged price changes or returns. In absolute terms the measured serial correlations are always close to zero.

Looking hard, though, one can probably find evidence of statistically “significant” linear dependence in Table 1 (and again this is true of results reported by others). For the daily returns eleven of the serial correlations are more than twice their computed standard errors, and twenty-two out of thirty are positive. On the other hand, twenty-one and twenty-four of the coefficients for the four and nine day differences are negative. But with samples of the size underlying Table 1 ($N = 1200$ -1700 observations per stock on a daily basis) statistically “significant” deviations from zero covariance are not necessarily a basis for rejecting the efficient markets model. For the results in Table 1, the standard errors of the serial correlations were approximated as $(1/(N-1))^{1/2}$, which for the daily data implies that a correlation as small as .06 is more than twice its standard error. But a coefficient this size implies that a linear relationship with the lagged price change can be used to explain about .36% of the variation in the current price change, which is probably insignificant from an economic viewpoint. In particular, it is unlikely that the small absolute levels of serial correlation that are always observed can be used as the basis of substantially profitable trading systems.¹⁴

It is, of course, difficult to judge what degree of serial correlation would imply the existence of trading rules with substantial expected profits. (And indeed we shall soon have to be a little more precise about what is implied by “substantial” profits.) Moreover, zero serial covariances are consistent with a “fair game” model, but as noted earlier (fn. 10), there are types of nonlinear dependence that imply the existence of profitable trading systems, and yet do not imply nonzero serial covariances. Thus, for many reasons it is desirable to directly test the profitability of various trading rules.

The first major evidence on trading rules was Alexander’s [1, 2]. He tests a variety of systems, but the most thoroughly examined can be described as follows: If the price of a security moves up at least $y\%$, buy and hold the security until its price moves down at least $y\%$ from a subsequent high, at which time simultaneously sell and go short. The short position is maintained until the price rises at least $y\%$ above a subsequent low, at which time one covers the short position and buys. Moves less than $y\%$ in either direction are

14. Given the evidence of Kendall [21], Mandelbrot [28], Fama [10] and others that large price changes occur much more frequently than would be expected if the generating process were Gaussian, the expression $(1/(N-1))^{1/2}$ understates the sampling dispersion of the serial correlation coefficient, and thus leads to an overstatement of significance levels. In addition, the fact that sample serial correlations are predominantly of one sign or the other is not in itself evidence of linear dependence. If, as the work of King [23] and Blume [7] indicates, there is a market factor whose behavior affects the returns on all securities, the sample behavior of this market factor may lead to a predominance of signs of one type in the serial correlations for individual securities, even though the population serial correlations for both the market factor and the returns on individual securities are zero. For a more extensive analysis of these issues see [10].

ignored. Such a system is called a $y\%$ filter. It is obviously a “one security and cash” trading rule, so that the results it produces are relevant for the submartingale expected return model of (6).

After extensive tests using daily data on price indices from 1897 to 1959 and filters from one to fifty per cent, and after correcting some incorrect presumptions in the initial results of [1] (see fn. 25), in his final paper on the subject, Alexander concludes:

In fact, at this point I should advise any reader who is interested only in practical results, and who is not a floor trader and so must pay commissions, to turn to other sources on how to beat buy and hold. The rest of this article is devoted principally to a theoretical consideration of whether the observed results are consistent with a random walk hypothesis [8], p. 351).

Later in the paper Alexander concludes that there is some evidence in his results against the independence assumption of the random walk model. But market efficiency does not require a random walk, and from the viewpoint of the submartingale model of (6), the conclusion that the filters cannot beat buy-and-hold is support for the efficient markets hypothesis. Further support is provided by Fama and Blume [13] who compare the profitability of various filters to buy-and-hold for the individual stocks of the Dow-Jones Industrial Average. (The data are those underlying Table 1.)

But again, looking hard one can find evidence in the filter tests of both Alexander and Fama-Blume that is inconsistent with the submartingale efficient markets model, if that model is interpreted in a strict sense. In particular, the results for very small filters (1 per cent in Alexander's tests and .5, 1.0, and 1.5 per cent in the tests of Fama-Blume) indicate that it is possible to devise trading schemes based on very short-term (preferably intra-day but at most daily) price swings that will on average outperform buy-and-hold. The average profits on individual transactions from such schemes are miniscule, but they generate transactions so frequently that over longer periods and ignoring commissions they outperform buy-and-hold by a substantial margin. These results are evidence of persistence or positive dependence in very short-term price movements. And, interestingly, this is consistent with the evidence for slight positive linear dependence in successive daily price changes produced by the serial correlations.¹⁵

15. Though strictly speaking, such tests of pure independence are not directly relevant for expected return models, it is interesting that the conclusion that very short-term swings in prices persist slightly longer than would be expected under the martingale hypothesis is also supported by the results of non-parametric runs tests applied to the daily data of Table 1. (See [10], Tables 12-15.) For the daily price changes, the actual number of runs of price changes of the same sign is less than the expected number for 26 out of 30 stocks. Moreover, of the eight stocks for which the actual number of runs is more than two standard errors less than the expected number, five of the same stocks have positive daily, first order serial correlations in Table 1 that are more than twice their standard errors. But in both cases the statistical “significance” of the results is largely a reflection of the large sample sizes. Just as the serial correlations are small in absolute terms (the average is .026), the differences between the expected and actual number of runs on average are only three per cent of the total expected number.

On the other hand, it is also interesting that the runs tests do not support the suggestion of slight negative dependence in four and nine day changes that appeared in the serial correlations. In the runs tests such negative dependence would appear as a tendency for the actual number of runs to exceed the expected number. In fact, for the four and nine day price changes, for 17 and

But when one takes account of even the minimum trading costs that would be generated by small filters, their advantage over buy-and-hold disappears. For example, even a floor trader (i.e., a person who owns a seat) on the New York Stock Exchange must pay clearinghouse fees on his trades that amount to about .1 per cent per turnaround transaction (i.e., sales plus purchase). Fama-Blume show that because small filters produce such frequent trades, these minimum trading costs are sufficient to wipe out their advantage over buy-and-hold.

Thus the filter tests, like the serial correlations, produce empirically noticeable departures from the strict implications of the efficient markets model. But, in spite of any statistical significance they might have, from an economic viewpoint the departures are so small that it seems hardly justified to use them to declare the market inefficient.

3. Other Tests of Independence in the Random Walk Literature

It is probably best to regard the random walk model as a special case of the more general expected return model in the sense of making a more detailed specification of the economic environment. That is, the basic model of market equilibrium is the "fair game" expected return model, with a random walk arising when additional environmental conditions are such that distributions of one-period returns repeat themselves through time. From this viewpoint violations of the pure independence assumption of the random walk model are to be expected. But when judged relative to the benchmark provided by the random walk model, these violations can provide insights into the nature of the market environment.

For example, one departure from the pure independence assumption of the random walk model has been noted by Osborne [34], Fama ([10], Table 17 and Figure 8), and others. In particular, large daily price changes tend to be followed by large daily changes. The signs of the successor changes are apparently random, however, which indicates that the phenomenon represents a denial of the random walk model but not of the market efficiency hypothesis. Nevertheless, it is interesting to speculate why the phenomenon might arise. It may be that when important new information comes into the market it cannot always be immediately evaluated precisely. Thus, sometimes the initial price will overadjust to the information, and other times it will underadjust. But since the evidence indicates that the price changes on days following the initial large change are random in sign, the initial large change at least represents an unbiased adjustment to the ultimate price effects of the information, and this is sufficient for the expected return efficient markets model.

Niederhoffer and Osborne [32] document two departures from complete randomness in common stock price changes from transaction to transaction. First, their data indicate that reversals (pairs of consecutive price changes of opposite sign) are from two to three times as likely as continuations (pairs of consecutive price changes of the same sign). Second, a continuation is

18 of the 30 stocks in Table 1 the actual number of runs is less than the expected number. Indeed, runs tests in general show no consistent evidence of dependence for any differencing interval longer than a day, which seems especially pertinent in light of the comments in footnote 14.

slightly more frequent after a preceding continuation than after a reversal. That is, let $(+|++)$ indicate the occurrence of a positive price change, given two preceding positive changes. Then the events $(+|++)$ and $(-|--)$ are slightly more frequent than $(+|+-)$ or $(-|+-)$.¹⁶

Niederhoffer and Osborne offer explanations for these phenomena based on the market structure of the New York Stock Exchange (N.Y.S.E.). In particular, there are three major types of orders that an investor might place in a given stock: (a) buy limit (buy at a specified price or lower), (b) sell limit (sell at a specified price or higher), and (c) buy or sell at market (at the lowest selling or highest buying price of another investor). A book of unexecuted limit orders in a given stock is kept by the specialist in that stock on the floor of the exchange. Unexecuted sell limit orders are, of course, at higher prices than unexecuted buy limit orders. On both exchanges, the smallest non-zero price change allowed is $\frac{1}{8}$ point.

Suppose now that there is more than one unexecuted sell limit order at the lowest price of any such order. A transaction at this price (initiated by an order to buy at market¹⁷) can only be followed either by a transaction at the same price (if the next market order is to buy) or by a transaction at a lower price (if the next market order is to sell). Consecutive price increases can usually only occur when consecutive market orders to buy exhaust the sell limit orders at a given price.¹⁸ In short, the excessive tendency toward reversal for consecutive non-zero price changes could result from bunching of unexecuted buy and sell limit orders.

The tendency for the events $(+|++)$ and $(-|--)$ to occur slightly more frequently than $(+|+-)$ and $(-|+-)$ requires a more involved explanation which we shall not attempt to reproduce in full here. In brief, Niederhoffer and Osborne contend that the higher frequency of $(+|++)$ relative to $(+|+-)$ arises from a tendency for limit orders "to be concentrated at integers (26, 43), halves ($26\frac{1}{2}$, $43\frac{1}{2}$), quarters and odd eighths in descending order of preference."¹⁹ The frequency of the event $(+|++)$, which usually requires that sell limit orders be exhausted at at least two consecutively higher prices (the last of which is relatively more frequently at an odd eighth), more heavily reflects the absence of sell limit orders at odd eighths than the event $(+|+-)$, which usually implies that sell limit orders at only one price have been exhausted and so more or less reflects the average bunching of limit orders at all eighths.

But though Niederhoffer and Osborne present convincing evidence of sta-

16. On a transaction to transaction basis, positive and negative price changes are about equally likely. Thus, under the assumption that price changes are random, any pair of non-zero changes should be as likely as any other, and likewise for triplets of consecutive non-zero changes.

17. A buy limit order for a price equal to or greater than the lowest available sell limit price is effectively an order to buy at market, and is treated as such by the broker.

18. The exception is when there is a gap of more than $\frac{1}{8}$ between the highest unexecuted buy limit and the lowest unexecuted sell limit order, so that market orders (and new limit orders) can be crossed at intermediate prices.

19. Their empirical documentation for this claim is a few samples of specialists' books for selected days, plus the observation [34] that actual trading prices, at least for volatile high priced stocks, seem to be concentrated at integers, halves, quarters and odd eighths in descending order.

tistically significant departures from independence in price changes from transaction to transaction, and though their analysis of their findings presents interesting insights into the process of market making on the major exchanges, the types of dependence uncovered do not imply market inefficiency. The best documented source of dependence, the tendency toward excessive reversals in pairs of non-zero price changes, seems to be a direct result of the ability of investors to place limit orders as well as orders at market, and this negative dependence in itself does not imply the existence of profitable trading rules. Similarly, the apparent tendency for observed transactions (and, by implication, limit orders) to be concentrated at integers, halves, even eighths and odd eighths in descending order is an interesting fact about investor behavior, but in itself is not a basis on which to conclude that the market is inefficient.²⁰

The Niederhoffer-Osborne analysis of market making does, however, point clearly to the existence of market inefficiency, but with respect to strong form tests of the efficient markets model. In particular, the list of unexecuted buy and sell limit orders in the specialist's book is important information about the likely future behavior of prices, and this information is only available to the specialist. When the specialist is asked for a quote, he gives the prices and can give the quantities of the highest buy limit and lowest sell limit orders on his book, but he is prevented by law from divulging the book's full contents. The interested reader can easily imagine situations where the structure of limit orders in the book could be used as the basis of a profitable trading rule.²¹ But the record seems to speak for itself:

It should not be assumed that these transactions undertaken by the specialist, and in which he is involved as buyer or seller in 24 per cent of all market volume, are necessarily a burden to him. Typically, the specialist sells above his last purchase on 83 per cent of all his sales, and buys below his last sale on 81 per cent of all his purchases ([32], p. 908).

Thus it seems that the specialist has monopoly power over an important block of information, and, not unexpectedly, uses his monopoly to turn a profit. And this, of course, is evidence of market inefficiency in the strong form sense. The important economic question, of course, is whether the market making

20. Niederhoffer and Osborne offer little to refute this conclusion. For example ([32], p. 914):

Although the specific properties reported in this study have a significance from a statistical point of view, the reader may well ask whether or not they are helpful in a practical sense. Certain trading rules emerge as a result of our analysis. One is that limit and stop orders should be placed at odd eighths, preferably at $\frac{7}{8}$ for sell orders and at $\frac{1}{8}$ for buy orders. Another is to buy when a stock advances through a barrier and to sell when it sinks through a barrier.

The first "trading rule" tells the investor to resist his innate inclination to place orders at integers, but rather to place sell orders $\frac{7}{8}$ below an integer and buy orders $\frac{1}{8}$ above. Successful execution of the orders is then more likely, since the congestion of orders that occur at integers is avoided. But the cost of this success is apparent. The second "trading rule" seems no more promising, if indeed it can even be translated into a concrete prescription for action.

21. See, for example, ([32], p. 908). But it is unlikely that anyone but the specialist could earn substantial profits from knowledge of the structure of unexecuted limit orders on the book. The specialist makes trading profits by engaging in many transactions, each of which has a small average profit; but for any other trader, including those with seats on the exchange, these profits would be eaten up by commissions to the specialist.

function of the specialist could be fulfilled more economically by some non-monopolistic mechanism.²²

4. Distributional Evidence

At this date the weight of the empirical evidence is such that economists would generally agree that whatever dependence exists in series of historical returns cannot be used to make profitable predictions of the future. Indeed, for returns that cover periods of a day or longer, there is little in the evidence that would cause rejection of the stronger random walk model, at least as a good first approximation.

Rather, the last burning issue of the random walk literature has centered on the nature of the distribution of price changes (which, we should note immediately, is an important issue for the efficient markets hypothesis since the nature of the distribution affects both the types of statistical tools relevant for testing the hypothesis and the interpretation of any results obtained). A model implying normally distributed price changes was first proposed by Bachelier [3], who assumed that price changes from transaction to transaction are independent, identically distributed random variables with finite variances. If transactions are fairly uniformly spread across time, and if the number of transactions per day, week, or month is very large, then the Central Limit Theorem leads us to expect that these price changes will have normal or Gaussian distributions.

Osborne [33], Moore [31], and Kendall [21] all thought their empirical evidence supported the normality hypothesis, but all observed high tails (i.e., higher proportions of large observations) in their data distributions vis-à-vis what would be expected if the distributions were normal. Drawing on these findings and some empirical work of his own, Mandelbrot [28] then suggested that these departures from normality could be explained by a more general form of the Bachelier model. In particular, if one does not assume that distributions of price changes from transaction to transaction necessarily have finite variances, then the limiting distributions for price changes over longer differencing intervals could be any member of the stable class, which includes the normal as a special case. Non-normal stable distributions have higher tails than the normal, and so can account for this empirically observed feature of distributions of price changes. After extensive testing (involving the data from the stocks in Table 1), Fama [10] concludes that non-normal stable distributions are a better description of distributions of daily returns on common stocks than the normal. This conclusion is also supported by the empirical work of Blume [7] on common stocks, and it has been extended to U.S. Government Treasury Bills by Roll [37].

Economists have, however, been reluctant to accept these results,²³ primar-

22. With modern computers, it is hard to believe that a more competitive and economical system would not be feasible. It does not seem technologically impossible to replace the entire floor of the N.Y.S.E. with a computer, fed by many remote consoles, that kept all the books now kept by the specialists, that could easily make the entire book on any stock available to anybody (so that interested individuals could then compete to "make a market" in a stock) and that carried out transactions automatically.

23. Some have suggested that the long-tailed empirical distributions might result from processes

ily because of the wealth of statistical techniques available for dealing with normal variables and the relative paucity of such techniques for non-normal stable variables. But perhaps the biggest contribution of Mandelbrot's work has been to stimulate research on stable distributions and estimation procedures to be applied to stable variables. (See, for example, Wise [46], Fama and Roll [15], and Blattberg and Sargent [6], among others.) The advance of statistical sophistication (and the importance of examining distributional assumptions in testing the efficient markets model) is well illustrated in Roll [37], as compared, for example, with the early empirical work of Mandelbrot [28] and Fama [10].

5. "Fair Game" Models in the Treasury Bill Market

Roll's work is novel in other respects as well. Coming after the efficient markets models of Mandelbrot [27] and Samuelson [38], it is the first weak form empirical work that is consciously in the "fair game" rather than the random walk tradition.

More important, as we saw earlier, the "fair game" properties of the general expected return models apply to

$$z_{jt} = r_{jt} - E(\bar{r}_{jt} | \Phi_{t-1}). \quad (10)$$

For data on common stocks, tests of "fair game" (and random walk) properties seem to go well when the conditional expected return is estimated as the average return for the sample of data at hand. Apparently the variation in common stock returns about their expected values is so large relative to any changes in the expected values that the latter can safely be ignored. But, as Roll demonstrates, this result does not hold for Treasury Bills. Thus, to test the "fair game" model on Treasury Bills requires explicit economic theory for the evolution of expected returns through time.

Roll uses three existing theories of the term structure (the pure expectations hypothesis of Lutz [26] and two market segmentation hypotheses, one of which is the familiar "liquidity preference" hypothesis of Hicks [18] and Kessel [22]) for this purpose.²⁴ In his models r_{jt} is the rate observed from the term structure at period t for one week loans to commence at $t + j - 1$, and can be thought of as a "futures" rate. Thus $r_{j+1, t-1}$ is likewise the rate on

that are mixtures of normal distributions with different variances. Press [35], for example, suggests a Poisson mixture of normals in which the resulting distributions of price changes have long tails but finite variances. On the other hand, Mandelbrot and Taylor [29] show that other mixtures of normals can still lead to non-normal stable distributions of price changes for finite differencing intervals.

If, as Press' model would imply, distributions of price changes are long-tailed but have finite variances, then distributions of price changes over longer and longer differencing intervals should be progressively closer to the normal. No such convergence to normality was observed in [10] (though admittedly the techniques used were somewhat rough). Rather, except for origin and scale, the distributions for longer differencing intervals seem to have the same "high-tailed" characteristics as distributions for shorter differencing intervals, which is as would be expected if the distributions are non-normal stable.

24. As noted early in our discussions, all available tests of market efficiency are implicitly also tests of expected return models of market equilibrium. But Roll formulates explicitly the economic models underlying his estimates of expected returns, and emphasizes that he is simultaneously testing economic models of the term structure as well as market efficiency.

one week loans to commence at $t + j - 1$, but observed in this case at $t - 1$. Similarly, L_{jt} is the so-called “liquidity premium” in r_{jt} ; that is

$$r_{jt} = E(\tilde{r}_{0,t+j-1}|\Phi_t) + L_{jt}.$$

In words, the one-week “futures” rate for period $t + j - 1$ observed from the term structure at t is the expectation at t of the “spot” rate for $t + j - 1$ plus a “liquidity premium” (which could, however, be positive or negative).

In all three theories of the term structure considered by Roll, the conditional expectation required in (10) is of the form

$$E(\tilde{r}_{j,t}|\Phi_{t-1}) = r_{j+1,t-1} + E(\tilde{L}_{j,t}|\Phi_{t-1}) - L_{j+1,t-1}.$$

The three theories differ only in the values assigned to the “liquidity premiums.” For example, in the “liquidity preference” hypothesis, investors must always be paid a positive premium for bearing interest rate uncertainty, so that the L_{jt} are always positive. By contrast, in the “pure expectations” hypothesis, all liquidity premiums are assumed to be zero, so that

$$E(\tilde{r}_{j,t}|\Phi_{t-1}) = r_{j+1,t-1}.$$

After extensive testing, Roll concludes (i) that the two market segmentation hypotheses fit the data better than the pure expectations hypothesis, with perhaps a slight advantage for the “liquidity preference” hypothesis, and (ii) that as far as his tests are concerned, the market for Treasury Bills is efficient. Indeed, it is interesting that when the best fitting term structure model is used to estimate the conditional expected “futures” rate in (10), the resulting variable z_{jt} seems to be serially independent! It is also interesting that if he simply assumed that his data distributions were normal, Roll’s results would not be so strongly in support of the efficient markets model. In this case taking account of the observed high tails of the data distributions substantially affected the interpretation of the results.²⁵

6. Tests of a Multiple Security Expected Return Model

Though the weak form tests support the “fair game” efficient markets model, all of the evidence examined so far consists of what we might call “single security tests.” That is, the price or return histories of individual securities are examined for evidence of dependence that might be used as the basis of a trading system for *that* security. We have not discussed tests of whether securities are “appropriately priced” vis-à-vis one another.

But to judge whether differences between average returns are “appropriate” an economic theory of equilibrium expected returns is required. At the moment, the only fully developed theory is that of Sharpe [40] and Lintner [24,

25. The importance of distributional assumptions is also illustrated in Alexander’s work on trading rules. In his initial tests of filter systems [1], Alexander assumed that purchases could always be executed exactly (rather than at least) $y\%$ above lows and sales exactly $y\%$ below highs. Mandelbrot [28] pointed out, however, that though this assumption would do little harm with normally distributed price changes (since price series are then essentially continuous), with non-normal stable distributions it would introduce substantial positive bias into the filter profits (since with such distributions price series will show many discontinuities). In his later tests [2], Alexander does indeed find that taking account of the discontinuities (i.e., the presence of large price changes) in his data substantially lowers the profitability of the filters.

25] referred to earlier. In this model (which is a direct outgrowth of the mean-standard deviation portfolio models of investor equilibrium of Markowitz [30] and Tobin [43]), the expected return on security j from time t to $t + 1$ is

$$E(\tilde{r}_{j,t+1}|\Phi_t) = r_{f,t+1} + \left[\frac{E(\tilde{r}_{m,t+1}|\Phi_t) - r_{f,t+1}}{\sigma(\tilde{r}_{m,t+1}|\Phi_t)} \right] \frac{\text{cov}(\tilde{r}_{j,t+1}, \tilde{r}_{m,t+1}|\Phi_t)}{\sigma(\tilde{r}_{m,t+1}|\Phi_t)}, \quad (11)$$

where $r_{f,t+1}$ is the return from t to $t + 1$ on an asset that is riskless in money terms; $r_{m,t+1}$ is the return on the "market portfolio" m (a portfolio of all investment assets with each weighted in proportion to the total market value of all its outstanding units); $\sigma^2(\tilde{r}_{m,t+1}|\Phi_t)$ is the variance of the return on m ; $\text{cov}(\tilde{r}_{j,t+1}, \tilde{r}_{m,t+1}|\Phi_t)$ is the covariance between the returns on j and m ; and the appearance of Φ_t indicates that the various expected returns, variance and covariance, could in principle depend on Φ_t . Though Sharpe and Lintner derive (11) as a one-period model, the result is given a multiperiod justification and interpretation in [11]. The model has also been extended in (12) to the case where the one-period returns could have stable distributions with infinite variances.

In words, (11) says that the expected one-period return on a security is the one-period riskless rate of interest $r_{f,t+1}$ plus a "risk premium" that is proportional to $\text{cov}(\tilde{r}_{j,t+1}, \tilde{r}_{m,t+1}|\Phi_t)/\sigma(\tilde{r}_{m,t+1}|\Phi_t)$. In the Sharpe-Lintner model each investor holds some combination of the riskless asset and the market portfolio, so that, given a mean-standard deviation framework, the risk of an individual asset can be measured by its contribution to the standard deviation of the return on the market portfolio. This contribution is in fact $\text{cov}(\tilde{r}_{j,t+1}, \tilde{r}_{m,t+1}|\Phi_t)/\sigma(\tilde{r}_{m,t+1}|\Phi_t)$.²⁶ The factor

$$[E(\tilde{r}_{m,t+1}|\Phi_t) - r_{f,t+1}]/\sigma(\tilde{r}_{m,t+1}|\Phi_t),$$

which is the same for all securities, is then regarded as the market price of risk.

Published empirical tests of the Sharpe-Lintner model are not yet available, though much work is in progress. There is some published work, however, which, though not directed at the Sharpe-Lintner model, is at least consistent with some of its implications. The stated goal of this work has been to determine the extent to which the returns on a given security are related to the returns on other securities. It started (again) with Kendall's [21] finding that though common stock price changes do not seem to be serially correlated, there is a high degree of cross-correlation between the *simultaneous* returns of different securities. This line of attack was continued by King [23] who (using factor analysis of a sample of monthly returns on sixty N.Y.S.E. stocks for the period 1926-60) found that on average about 50% of the variance of an individual stock's returns could be accounted for by a "market factor" which affects the returns on all stocks, with "industry factors" accounting for at most an additional 10% of the variance.

26. That is,

$$\sum_j \text{cov}(\tilde{r}_{j,t+1}, \tilde{r}_{m,t+1}|\Phi_t)/\sigma(\tilde{r}_{m,t+1}|\Phi_t) = \sigma(\tilde{r}_{m,t+1}|\Phi_t).$$

For our purposes, however, the work of Fama, Fisher, Jensen, and Roll [14] (henceforth FFJR) and the more extensive work of Blume [7] on monthly return data is more relevant. They test the following “market model,” originally suggested by Markowitz [30]:

$$\tilde{r}_{j,t+1} = \alpha_j + \beta_j \tilde{r}_{M,t+1} + \tilde{u}_{j,t+1} \quad (12)$$

where $r_{j,t+1}$ is the rate of return on security j for month t , $r_{M,t+1}$ is the corresponding return on a market index M , α_j and β_j are parameters that can vary from security to security, and $u_{j,t+1}$ is a random disturbance. The tests of FFJR and subsequently those of Blume indicate that (12) is well specified as a linear regression model in that (i) the estimated parameters $\hat{\alpha}_j$ and $\hat{\beta}_j$ remain fairly constant over long periods of time (e.g., the entire post-World War II period in the case of Blume), (ii) $r_{M,t+1}$ and the estimated $\hat{u}_{j,t+1}$, are close to serially independent, and (iii) the $\hat{u}_{j,t+1}$ seem to be independent of $r_{M,t+1}$.

Thus the observed properties of the “market model” are consistent with the expected return efficient markets model, and, in addition, the “market model” tells us something about the process generating expected returns from security to security. In particular,

$$E(\tilde{r}_{j,t+1}) = \alpha_j + \beta_j E(\tilde{r}_{M,t+1}). \quad (13)$$

The question now is to what extent (13) is consistent with the Sharpe-Lintner expected return model summarized by (11). Rearranging (11) we obtain

$$E(\tilde{r}_{j,t+1}|\Phi_t) = \alpha_j(\Phi_t) + \beta_j(\Phi_t)E(\tilde{r}_{m,t+1}|\Phi_t), \quad (14)$$

where, noting that the riskless rate $r_{f,t+1}$ is itself part of the information set Φ_t , we have

$$\alpha_j(\Phi_t) = r_{f,t+1}[1 - \beta_j(\Phi_t)], \quad (15)$$

and

$$\beta_j(\Phi_t) = \frac{\text{cov}(\tilde{r}_{j,t+1}, \tilde{r}_{m,t+1}|\Phi_t)}{\sigma^2(\tilde{r}_{m,t+1}|\Phi_t)}. \quad (16)$$

With some simplifying assumptions, (14) can be reduced to (13). In particular, if the covariance and variance that determine $\beta_j(\Phi_t)$ in (16) are the same for all t and Φ_t , then $\beta_j(\Phi_t)$ in (16) corresponds to β_j in (12) and (13), and the least squares *estimate* of β_j in (12) is in fact just the ratio of the sample values of the covariance and variance in (16). If we also assume that $r_{f,t+1}$ is the same for all t , and that the behavior of the returns on the market portfolio m are closely approximated by the returns on some representative index M , we will have come a long way toward equating (13) and (11). Indeed, the only missing link is whether in the estimated parameters of (12)

$$\hat{\alpha}_j \cong r_f(1 - \hat{\beta}_j). \quad (17)$$

Neither FFJR nor Blume attack this question directly, though some of Blume’s evidence is at least promising. In particular, the magnitudes of the

estimated $\hat{\alpha}_j$ are roughly consistent with (17) in the sense that the estimates are always close to zero (as they should be with monthly return data).²⁷

In a sense, though, in establishing the apparent empirical validity of the “market model” of (12), both too much and too little have been shown *vis-à-vis* the Sharpe-Lintner expected return model of (11). We know that during the post-World War II period one-month interest rates on riskless assets (e.g., government bills with one month to maturity) have not been constant. Thus, if expected security returns were generated by a version of the “market model” that is fully consistent with the Sharpe-Lintner model, we would, according to (15), expect to observe some non-stationarity in the estimates of α_j . On a monthly basis, however, variation through time in one-period riskless interest rates is probably trivial relative to variation in other factors affecting monthly common stock returns, so that more powerful statistical methods would be necessary to study the effects of changes in the riskless rate.

In any case, since the work of FFJR and Blume on the “market model” was not concerned with relating this model to the Sharpe-Lintner model, we can only say that the results for the former are somewhat consistent with the implications of the latter. But the results for the “market model” are, after all, just a statistical description of the return generating process, and they are probably somewhat consistent with other models of equilibrium expected returns. Thus the only way to generate strong empirical conclusions about the Sharpe-Lintner model is to test it directly. On the other hand, any alternative model of equilibrium expected returns must be somewhat consistent with the “market model,” given the evidence in its support.

B. Tests of Martingale Models of the Semi-strong Form

In general, semi-strong form tests of efficient markets models are concerned with whether current prices “fully reflect” all obviously publicly available information. Each individual test, however, is concerned with the adjustment of security prices to one kind of information generating event (e.g., stock splits, announcements of financial reports by firms, new security issues, etc.). Thus each test only brings supporting evidence for the model, with the idea that by accumulating such evidence the validity of the model will be “established.”

In fact, however, though the available evidence is in support of the efficient markets model, it is limited to a few major types of information generating events. The initial major work is apparently the study of stock splits by Fama,

27. With least squares applied to monthly return data, the estimate of α_j in (12) is

$$\hat{\alpha}_j = \bar{r}_{j,t} - \hat{\beta}_j \bar{r}_{M,t}$$

where the bars indicate sample mean returns. But, in fact, Blume applies the market model to the wealth relatives $R_{jt} = 1 + r_{jt}$ and $R_{Mt} = 1 + r_{Mt}$. This yields precisely the same estimate of β_j as least squares applied to (12), but the intercept is now

$$\hat{\alpha}'_j = \bar{R}_{jt} - \hat{\beta}_j \bar{R}_{Mt} = 1 + \bar{r}_{jt} - \hat{\beta}_j (1 + \bar{r}_{Mt}) = 1 - \hat{\beta}_j + \hat{\alpha}_j.$$

Thus what Blume in fact finds is that for almost all securities, $\hat{\alpha}'_j + \hat{\beta}_j \cong 1$, which implies that $\hat{\alpha}_j$ is close to 0.

Fisher, Jensen, and Roll (FFJR) [14], and all the subsequent studies summarized here are adaptations and extensions of the techniques developed in FFJR. Thus, this paper will first be reviewed in some detail, and then the other studies will be considered.

1. Splits and the Adjustment of Stock Prices to New Information

Since the only apparent result of a stock split is to multiply the number of shares per shareholder without increasing claims to real assets, splits in themselves are not necessarily sources of new information. The presumption of FFJR is that splits may often be associated with the appearance of more fundamentally important information. The idea is to examine security returns around split dates to see first if there is any "unusual" behavior, and, if so, to what extent it can be accounted for by relationships between splits and other more fundamental variables.

The approach of FFJR to the problem relies heavily on the "market model" of (12). In this model if a stock split is associated with abnormal behavior, this would be reflected in the estimated regression residuals for the months surrounding the split. For a given split, define month 0 as the month in which the effective date of a split occurs, month 1 as the month immediately following the split month, month -1 as the month preceding, etc. Now define the average residual over all split securities for month m (where for each security m is measured relative to the split month) as

$$u_m = \sum_{j=1}^N \frac{\hat{u}_{jm}}{N},$$

where $\hat{u}_{jm}$ is the sample regression residual for security j in month m and N is the number of splits. Next, define the cumulative average residual U_m as

$$U_m = \sum_{k=-29}^m u_k.$$

The average residual u_m can be interpreted as the average deviation (in month m relative to split months) of the returns of split stocks from their normal relationships with the market. Similarly, U_m can be interpreted as the cumulative deviation (from month -29 to month m). Finally, define u_m^+ , u_m^- , U_m^+ , and U_m^- as the average and cumulative average residuals for splits followed by "increased" (+) and "decreased" (−) dividends. An "increase" is a case where the percentage change in dividends on the split share in the year after the split is greater than the percentage change for the N.Y.S.E. as a whole, while a "decrease" is a case of relative dividend decline.

The essence of the results of FFJR are then summarized in Figure 1, which shows the cumulative average residuals U_m , U_m^+ , and U_m^- for $-29 \leq m \leq 30$. The sample includes all 940 stock splits on the N.Y.S.E. from 1927-59, where the exchange was at least five new shares for four old, and where the security was listed for at least twelve months before and after the split.

For all three dividend categories the cumulative average residuals rise in

the 29 months prior to the split, and in fact the average residuals (not shown here) are uniformly positive. This cannot be attributed to the splitting process, since in only about ten per cent of the cases is the time between the announcement and effective dates of a split greater than four months. Rather, it seems that firms tend to split their shares during “abnormally” good times—that is, during periods when the prices of their shares have increased more than would

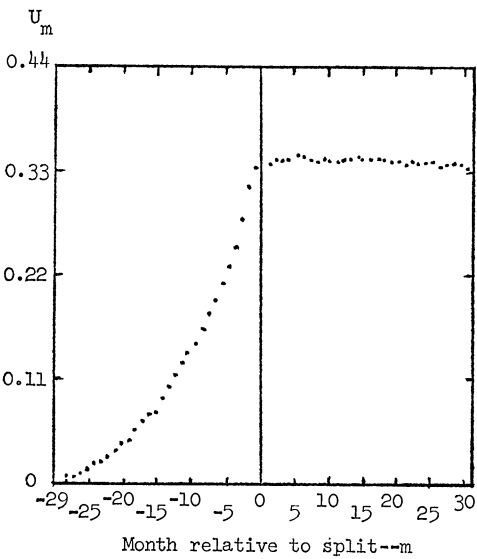


FIGURE 1a
 Cumulative average residuals—all splits.

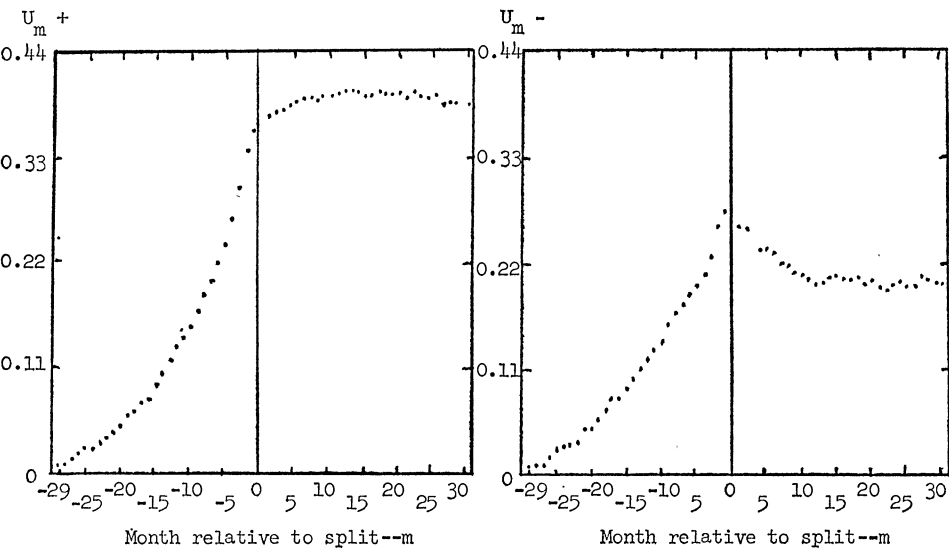


FIGURE 1b
 Cumulative average residuals for dividend “increases.”

FIGURE 1c
 Cumulative average residuals for dividend “decreases.”

be implied by their normal relationships with general market prices, which itself probably reflects a sharp improvement, relative to the market, in the earnings prospects of these firms sometime during the years immediately preceding a split.²⁸

After the split month there is almost no further movement in U_m , the cumulative average residual for all splits. This is striking, since 71.5 per cent (672 out of 940) of all splits experienced greater percentage dividend increases in the year after the split than the average for all securities on the N.Y.S.E. In light of this, FFJR suggest that when a split is announced the market interprets this (and correctly so) as a signal that the company's directors are probably confident that future earnings will be sufficient to maintain dividend payments at a higher level. Thus the large price increases in the months immediately preceding a split may be due to an alteration in expectations concerning the future earning potential of the firm, rather than to any intrinsic effects of the split itself.

If this hypothesis is correct, return behavior subsequent to splits should be substantially different for the cases where the dividend increase materializes than for the cases where it does not. FFJR argue that in fact the differences are in the directions that would be predicted. The fact that the cumulative average residuals for the "increased" dividends (Figure 1b) drift upward but only slightly in the year *after* the split is consistent with the hypothesis that when the split is *declared*, there is a price adjustment in anticipation of future dividend increases. But the behavior of the residuals for stock splits associated with "decreased" dividends offers even stronger evidence for the split hypothesis. The cumulative average residuals for these stocks (Figure 1c) rise in the few months before the split, but then fall dramatically in the few months after the split when the anticipated dividend increase is not forthcoming. When a year has passed after the split, the cumulative average residual has fallen to about where it was five months prior to the split, which is about the earliest time reliable information about a split is likely to reach the market. Thus by the time it becomes clear that the anticipated dividend increase is not forthcoming, the apparent effects of the split seem to have been wiped away, and the stock's returns have reverted to their normal relationship with market returns.

Finally, and most important, although the behavior of post-split returns will be very different depending on whether or not dividend "increases" occur, and in spite of the fact that a large majority of split securities do experience dividend "increases," when all splits are examined together (Figure 1a), subsequent to the split there is no net movement up or down in the cumulative

28. It is important to note, however, that as FFJR indicate, the persistent upward drift of the cumulative average residuals in the months preceding the split is not a phenomenon that could be used to increase expected trading profits. The reason is that the behavior of the average residuals is not representative of the behavior of the residuals for individual securities. In months prior to the split, successive sample residuals for individual securities seem to be independent. But in most cases, there are a few months in which the residuals are abnormally large and positive. The months of large residuals differ from security to security, however, and these differences in timing explain why the signs of the average residuals are uniformly positive for many months preceding the split.

average residuals. Thus, apparently the market makes unbiased forecasts of the implications of a split for future dividends, and these forecasts are fully reflected in the prices of the security by the end of the split month. After considerably more data analysis than can be summarized here, FFJR conclude that their results lend considerable support to the conclusion that the stock market is efficient, at least with respect to its ability to adjust to the information implicit in a split.

2. Other Studies of Public Announcements

Variants of the method of residual analysis developed in [14] have been used by others to study the effects of different kinds of public announcements, and all of these also support the efficient markets hypothesis.

Thus using data on 261 major firms for the period 1946-66, Ball and Brown [4] apply the method to study the effects of annual earnings announcements. They use the residuals from a time series regression of the annual earnings of a firm on the average earnings of all their firms to classify the firm's earnings for a given year as having "increased" or "decreased" relative to the market. Residuals from regressions of monthly common stock returns on an index of returns (i.e., the market model of (12)) are then used to compute cumulative average return residuals separately for the earnings that "increased," and those that "decreased." The cumulative average return residuals rise throughout the year in advance of the announcement for the earnings "increased" category, and fall for the earnings "decreased" category.²⁹ Ball and Brown [4, p. 175] conclude that in fact no more than about ten to fifteen percent of the information in the annual earnings announcement has not been anticipated by the month of the announcement.

On the macro level, Waud [45] has used the method of residual analysis to examine the effects of announcements of discount rate changes by Federal Reserve Banks. In this case the residuals are essentially just the deviations of the daily returns on the Standard and Poor's 500 Index from the average daily return. He finds evidence of a statistically significant "announcement effect" on stock returns for the first trading day following an announcement, but the magnitude of the adjustment is small, never exceeding .5%. More interesting from the viewpoint of the efficient markets hypothesis is his conclusion that, if anything, the market anticipates the announcements (or information is somehow leaked in advance). This conclusion is based on the non-random patterns of the signs of average return residuals on the days immediately preceding the announcement.

Further evidence in support of the efficient markets hypothesis is provided in the work of Scholes [39] on large secondary offerings of common stock (i.e., large underwritten sales of existing common stocks by individuals and institutions) and on new issues of stock. He finds that on average secondary issues are associated with a decline of between one and two per cent in the cumulative average residual returns for the corresponding common stocks. Since the magnitude of the price adjustment is unrelated to the size of the

29. But the comment of footnote 28 is again relevant here.

issue, Scholes concludes that the adjustment is not due to “selling pressure” (as is commonly believed), but rather results from negative information implicit in the fact that somebody is trying to sell a large block of a firm’s stock. Moreover, he presents evidence that the value of the information in a secondary depends to some extent on the vendor; somewhat as would be expected, by far the largest negative cumulative average residuals occur where the vendor is the corporation itself or one of its officers, with investment companies a distant second. But the identity of the vendor is not generally known at the time of the secondary, and corporate insiders need only report their transactions in their own company’s stock to the S.E.C. within six days after a sale. By this time the market on average has fully adjusted to the information in the secondary, as indicated by the fact that the average residuals behave randomly thereafter.

Note, however, that though this is evidence that prices adjust efficiently to public information, it is also evidence that corporate insiders at least sometimes have important information about their firm that is not yet publicly known. Thus Scholes’ evidence for secondary distributions provides support for the efficient markets model in the semi-strong form sense, but also some strong-form evidence against the model.

Though his results here are only preliminary, Scholes also reports on an application of the method of residual analysis to a sample of 696 new issues of common stock during the period 1926-66. As in the FFJR study of splits, the cumulative average residuals rise in the months preceding the new security offering (suggesting that new issues tend to come after favorable recent events)³⁰ but behave randomly in the months following the offering (indicating that whatever information is contained in the new issue is on average fully reflected in the price of the month of the offering).

In short, the available semi-strong form evidence on the effect of various sorts of public announcements on common stock returns is all consistent with the efficient markets model. The strong point of the evidence, however, is its consistency rather than its quantity; in fact, few different types of public information have been examined, though those treated are among the obviously most important. Moreover, as we shall now see, the amount of semi-strong form evidence is voluminous compared to the strong form tests that are available.

C. Strong Form Tests of the Efficient Markets Models

The strong form tests of the efficient markets model are concerned with whether all available information is fully reflected in prices in the sense that no individual has higher expected trading profits than others because he has monopolistic access to some information. We would not, of course, expect this model to be an exact description of reality, and indeed, the preceding discussions have already indicated the existence of contradictory evidence. In particular, Niederhoffer and Osborne [32] have pointed out that specialists on the N.Y.S.E. apparently use their monopolistic access to information concern-

30. Footnote 28 is again relevant here.

ing unfilled limit orders to generate monopoly profits, and Scholes' evidence [39] indicates that officers of corporations sometimes have monopolistic access to information about their firms.

Since we already have enough evidence to determine that the model is not strictly valid, we can now turn to other interesting questions. Specifically, how far down through the investment community do deviations from the model permeate? Does it pay for the average investor (or the average economist) to expend resources searching out little known information? Are such activities even generally profitable for various groups of market "professionals"? More generally, who are the people in the investment community that have access to "special information"?

Though this is a fascinating problem, only one group has been studied in any depth—the managements of open end mutual funds. Several studies are available (e.g., Sharpe [41, 42] and Treynor [44]), but the most thorough are Jensen's [19, 20], and our comments will be limited to his work. We shall first present the theoretical model underlying his tests, and then go on to his empirical results.

1. Theoretical Framework

In studying the performance of mutual funds the major goals are to determine (a) whether in general fund managers seem to have access to special information which allows them to generate "abnormal" expected returns, and (b) whether some funds are better at uncovering such special information than others. Since the criterion will simply be the ability of funds to produce higher returns than some norm with no attempt to determine what is responsible for the high returns, the "special information" that leads to high performance could be either keener insight into the implications of publicly available information than is implicit in market prices or monopolistic access to specific information. Thus the tests of the performance of the mutual fund industry are not strictly strong form tests of the efficient markets model.

The major theoretical (and practical) problem in using the mutual fund industry to test the efficient markets model is developing a "norm" against which performance can be judged. The norm must represent the results of an investment policy based on the assumption that prices fully reflect all available information. And if one believes that investors are generally risk averse and so on average must be compensated for any risks undertaken, then one has the problem of finding appropriate definitions of risk and evaluating each fund relative to a norm with its chosen level of risk.

Jensen uses the Sharpe [40]-Lintner [24, 25] model of equilibrium expected returns discussed above to derive a norm consistent with these goals. From (14)-(16), in this model the expected return on an asset or portfolio j from t to $t + 1$ is

$$E(\tilde{r}_{j,t+1}|\Phi_t) = r_{f,t+1} [1 - \beta_j(\Phi_t)] + E(\tilde{r}_{m,t+1}|\Phi_t)\beta_j(\Phi_t), \quad (18)$$

where the various symbols are as defined in Section III. A. 6. But (18) is an *ex ante* relationship, and to evaluate performance an *ex post* norm is needed.

One way the latter can be obtained is to substitute the realized return on the market portfolio for the expected return in (18) with the result³¹

$$E(\tilde{r}_{j,t+1}|\Phi_t, r_{m,t+1}) = r_{f,t+1} [1 - \beta_j(\Phi_t)] + r_{m,t+1}\beta_j(\Phi_t). \quad (19)$$

Geometrically, (19) says that within the context of the Sharpe-Lintner model, the expected return on j (given information Φ_t and the return $r_{m,t+1}$ on the market portfolio) is a linear function of its risk

$$\beta_j(\Phi_t) = \text{cov}(\tilde{r}_{j,t+1}, \tilde{r}_{m,t+1}|\Phi_t) / \sigma^2(\tilde{r}_{m,t+1}|\Phi_t),$$

as indicated in Figure 2. Assuming that the value of $\beta_j(\Phi_t)$ is somehow known, or can be reliably estimated, if j is a mutual fund, its *ex post* performance from t to $t + 1$ might now be evaluated by plotting its combination of realized return $r_{j,t+1}$ and risk in Figure 2. If (as for the point a) the combination falls above the expected return line (or, as it is more commonly called, the “market line”), it has done better than would be expected given its level of risk, while if (as for the point b) it falls below the line it has done worse.

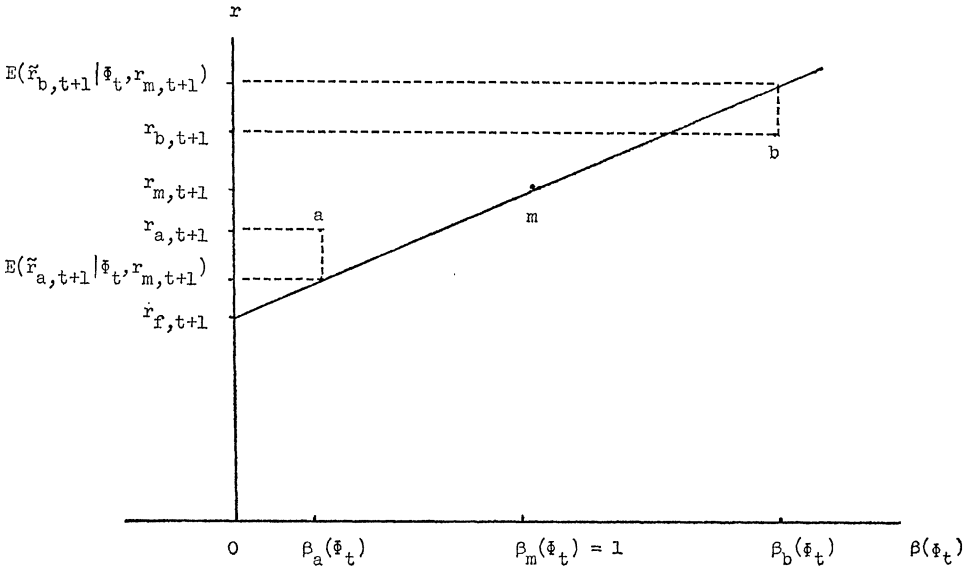


FIGURE 2
 Performance Evaluation Graph

Alternatively, the market line shows the combinations of return and risk provided by portfolios that are simple mixtures of the riskless asset and the market portfolio m . The returns and risks for such portfolios (call them c) are

$$r_{c,t+1} = \alpha r_{f,t+1} + (1 - \alpha) r_{m,t+1}$$

$$\beta_c(\Phi_t) = \frac{\text{cov}(\tilde{r}_{c,t+1}, \tilde{r}_{m,t+1}|\Phi_t)}{\sigma^2(\tilde{r}_{m,t+1}|\Phi_t)} = \frac{\text{cov}((1 - \alpha)\tilde{r}_{m,t+1}, \tilde{r}_{m,t+1}|\Phi_t)}{\sigma^2(\tilde{r}_{m,t+1}|\Phi_t)} = 1 - \alpha,$$

31. The assumption here is that the return $\tilde{r}_{j,t+1}$ is generated according to

$$\tilde{r}_{j,t+1} = r_{f,t+1}[1 - \beta_j(\Phi_t)] + r_{m,t+1}\beta_j(\Phi_t) + \tilde{u}_{j,t+1},$$

and

$$E(\tilde{u}_{j,t+1}|r_{m,t+1}) = 0 \text{ for all } r_{m,t+1}.$$

where α is the proportion of portfolio funds invested in the riskless asset. Thus, when $1 \geq \alpha \geq 0$ we obtain the combinations of return and risk along the market line from $r_{f,t+1}$ to m in Figure 2, while when $\alpha < 0$ (and under the assumption that investors can borrow at the same rate that they lend) we obtain the combinations of return and risk along the extension of the line through m . In this interpretation, the market line represents the results of a naive investment strategy, which the investor who thinks prices reflect all available information might follow. The performance of a mutual fund is then measured relative to this naive strategy.

2. Empirical Results

Jensen uses this risk-return framework to evaluate the performance of 115 mutual funds over the ten year period 1955-64. He argues at length for measuring return as the nominal ten year rate with continuous compounding (i.e., the natural log of the ratio of terminal wealth after ten years to initial wealth) and for using historical data on nominal one-year rates with continuous compounding to estimate risk. The Standard and Poor Index of 500 major common stocks is used as the proxy for the market portfolio.

The general question to be answered is whether mutual fund managements have any special insights or information which allows them to earn returns above the norm. But Jensen attacks the question on several levels. First, can the funds in general do well enough to compensate investors for loading charges, management fees, and other costs that might be avoided by simply choosing the combination of the riskless asset f and the market portfolio m with risk level comparable to that of the fund's actual portfolio? The answer seems to be an emphatic no. As far as net returns to investors are concerned, in 89 out of 115 cases, the fund's risk-return combination for the ten year period is below the market line for the period, and the average over all funds of the deviations of ten year returns from the market time is -14.6% . That is, on average the consumer's wealth after ten years of holding mutual funds is about fifteen per cent less than if he held the corresponding portfolios along the market line.

But the loading charge that an investor pays in buying into a fund is usually a pure salesman's commission that the fund itself never gets to invest. Thus one might ask whether, ignoring loading charges (i.e., assuming no such charges were paid by the investor), in general fund managements can earn returns sufficiently above the norm to cover all other expenses that are presumably more directly related to the management of the fund portfolios. Again, the answer seems to be no. Even when loading charges are ignored in computing returns, the risk-return combinations for 72 out of 115 funds are below the market line, and the average deviation of ten year returns from the market line is -8.9% .

Finally, as a somewhat stronger test of the efficient markets model, one would like to know if, ignoring all expenses, fund managements in general showed any ability to pick securities that outperformed the norm. Unfortunately, this question cannot be answered with precision for individual funds since, curiously, data on brokerage commissions are not published regularly.

But Jensen suggests the available evidence indicates that the answer to the question is again probably negative. Specifically, adding back all other published expenses of funds to their returns, the risk-return combinations for 58 out of 115 funds were below the market line, and the average deviation of ten year return from the line was -2.5% . But part of this result is due to the absence of a correction for brokerage commissions. Estimating these commissions from average portfolio turnover rates for all funds for the period 1953-58, and adding them back to returns for all funds increases the average deviation from the market line from -2.5% to $.09\%$, which still is not indicative of the existence of special information among mutual fund managers.

But though mutual fund managers in general do not seem to have access to information not already fully reflected in prices, perhaps there are individual funds that consistently do better than the norm, and so provide at least some strong form evidence against the efficient markets model. If there are such funds, however, they escape Jensen's search. For example, for individual funds, returns above the norm in one subperiod do not seem to be associated with performance above the norm in other subperiods. And regardless of how returns are measured (i.e., net or gross of loading charges and other expenses), the number of funds with large positive deviations of returns from the market line of Figure 2 is less than the number that would be expected by chance with 115 funds under the assumption that fund managements have no special talents in predicting returns.³²

Jensen argues that though his results apply to only one segment of the investment community, they are nevertheless striking evidence in favor of the efficient markets model:

Although these results certainly do not imply that the strong form of the martingale hypothesis holds for all investors and for all time, they provide strong evidence in support of that hypothesis. One must realize that these analysts are extremely well endowed. Moreover, they operate in the securities markets every day and have wide-ranging contacts and associations in both the business and financial communities. Thus, the fact that they are apparently unable to forecast returns accurately enough to recover their research and transactions costs is a striking piece of evidence in favor of the strong form of the martingale hypothesis—at least as far as the extensive subset of information available to these analysts is concerned [20, p. 170].

IV. SUMMARY AND CONCLUSIONS

The preceding (rather lengthy) analysis can be summarized as follows. In general terms, the theory of efficient markets is concerned with whether prices at any point in time “fully reflect” available information. The theory only has empirical content, however, within the context of a more specific model of

32. On the other hand, there is some suggestion in Scholes' [39] work on secondary issues that mutual funds may occasionally have access to “special information.” After corporate insiders, the next largest negative price changes occur when the secondary seller is an investment company (including mutual funds), though on average the price changes are much smaller (i.e., closer to 0) than when the seller is a corporate insider.

Moreover, Jensen's evidence itself, though not indicative of the existence of special information among mutual fund managers, is not sufficiently precise to conclude that such information never exists. This stronger conclusion would require exact data on unavoidable expenses (including brokerage commissions) of portfolio management incurred by funds.

market equilibrium, that is, a model that specifies the nature of market equilibrium when prices “fully reflect” available information. We have seen that all of the available empirical literature is implicitly or explicitly based on the assumption that the conditions of market equilibrium can be stated in terms of expected returns. This assumption is the basis of the expected return or “fair game” efficient markets models.

The empirical work itself can be divided into three categories depending on the nature of the information subset of interest. *Strong-form* tests are concerned with whether individual investors or groups have monopolistic access to any information relevant for price formation. One would not expect such an extreme model to be an exact description of the world, and it is probably best viewed as a benchmark against which the importance of deviations from market efficiency can be judged. In the less restrictive *semi-strong-form* tests the information subset of interest includes all obviously publicly available information, while in the *weak form* tests the information subset is just historical price or return sequences.

Weak form tests of the efficient market model are the most voluminous, and it seems fair to say that the results are strongly in support. Though statistically significant evidence for dependence in successive price changes or returns has been found, some of this is consistent with the “fair game” model and the rest does not appear to be sufficient to declare the market inefficient. Indeed, at least for price changes or returns covering a day or longer, there isn’t much evidence against the “fair game” model’s more ambitious offspring, the random walk.

Thus, there is consistent evidence of positive dependence in day-to-day price changes and returns on common stocks, and the dependence is of a form that can be used as the basis of marginally profitable trading rules. In Fama’s data [10] the dependence shows up as serial correlations that are consistently positive but also consistently close to zero, and as a slight tendency for observed numbers of runs of positive and negative price changes to be less than the numbers that would be expected from a purely random process. More important, the dependence also shows up in the filter tests of Alexander [1, 2] and those of Fama and Blume [13] as a tendency for very small filters to produce profits in excess of buy-and-hold. But any systems (like the filters) that attempt to turn short-term dependence into trading profits of necessity generate so many transactions that their expected profits would be absorbed by even the minimum commissions (security handling fees) that floor traders on major exchanges must pay. Thus, using a less than completely strict interpretation of market efficiency, this positive dependence does not seem of sufficient importance to warrant rejection of the efficient markets model.

Evidence in contradiction of the “fair game” efficient markets model for price changes or returns covering periods longer than a single day is more difficult to find. Cootner [9], and Moore [31] report preponderantly negative (but again small) serial correlations in weekly common stock returns, and this result appears also in the four day returns analyzed by Fama [10]. But it does not appear in runs tests of [10], where, if anything, there is some slight indication of positive dependence, but actually not much evidence of any

dependence at all. In any case, there is no indication that whatever dependence exists in weekly returns can be used as the basis of profitable trading rules.

Other existing evidence of dependence in returns provides interesting insights into the process of price formation in the stock market, but it is not relevant for testing the efficient markets model. For example, Fama [10] shows that large daily price changes tend to be followed by large changes, but of unpredictable sign. This suggests that important information cannot be completely evaluated immediately, but that the initial first day's adjustment of prices to the information is unbiased, which is sufficient for the martingale model. More interesting and important, however, is the Niederhoffer-Osborne [32] finding of a tendency toward excessive reversals in common stock price changes from transaction to transaction. They explain this as a logical result of the mechanism whereby orders to buy and sell at market are matched against existing limit orders on the books of the specialist. Given the way this tendency toward excessive reversals arises, however, there seems to be no way it can be used as the basis of a profitable trading rule. As they rightly claim, their results are a strong refutation of the theory of random walks, at least as applied to price changes from transaction to transaction, but they do not constitute refutation of the economically more relevant "fair game" efficient markets model.

Semi-strong form tests, in which prices are assumed to fully reflect all obviously publicly available information, have also supported the efficient markets hypothesis. Thus Fama, Fisher, Jensen, and Roll [14] find that the information in stock splits concerning the firm's future dividend payments is on average fully reflected in the price of a split share at the time of the split. Ball and Brown [4] and Scholes [39] come to similar conclusions with respect to the information contained in (i) annual earning announcements by firms and (ii) new issues and large block secondary issues of common stock. Though only a few different types of information generating events are represented here, they are among the more important, and the results are probably indicative of what can be expected in future studies.

As noted earlier, the strong-form efficient markets model, in which prices are assumed to fully reflect all available information, is probably best viewed as a benchmark against which deviations from market efficiency (interpreted in its strictest sense) can be judged. Two such deviations have in fact been observed. First, Niederhoffer and Osborne [32] point out that specialists on major security exchanges have monopolistic access to information on unexecuted limit orders and they use this information to generate trading profits. This raises the question of whether the "market making" function of the specialist (if indeed this is a meaningful economic function) could not as effectively be carried out by some other mechanism that did not imply monopolistic access to information. Second, Scholes [39] finds that, not unexpectedly, corporate insiders often have monopolistic access to information about their firms.

At the moment, however, corporate insiders and specialists are the only two groups whose monopolistic access to information has been documented. There is no evidence that deviations from the strong form of the efficient markets

model permeate down any further through the investment community. For the purposes of most investors the efficient markets model seems a good first (and second) approximation to reality.

In short, the evidence in support of the efficient markets model is extensive, and (somewhat uniquely in economics) contradictory evidence is sparse. Nevertheless, we certainly do not want to leave the impression that all issues are closed. The old saw, "much remains to be done," is relevant here as elsewhere. Indeed, as is often the case in successful scientific research, now that we know we've been in the past, we are able to pose and (hopefully) to answer an even more interesting set of questions for the future. In this case the most pressing field of future endeavor is the development and testing of models of market equilibrium under uncertainty. When the process generating equilibrium expected returns is better understood (and assuming that some expected return model turns out to be relevant), we will have a more substantial framework for more sophisticated intersecurity tests of market efficiency.

REFERENCES

1. Sidney S. Alexander. "Price Movements in Speculative Markets: Trends or Random Walks." *Industrial Management Review*, 2 (May 1961), 7-26. Also reprinted in [8], 199-218.
2. ———. "Price Movements in Speculative Markets: Trends or Random Walks. No. 2," in [8], 338-72.
3. Louis Bachelier. *Théorie de la Speculation* (Paris: Gauthier-Villars, 1900), and reprinted in English in [8], 17-78.
4. Ray Ball and Phillip Brown. "An Empirical Evaluation of Accounting Income Numbers." *Journal of Accounting Research*, 6 (Autumn, 1968), 159-78.
5. William Beaver. "The Information Content of Annual Earnings Announcements." *Empirical Research in Accounting: Selected Studies, 1968*, supplement to Vol. 7 of the *Journal of Accounting Research*, 67-92.
6. Robert Blattberg and Thomas Sargent. "Regression with Non-Gaussian Disturbances: Some Sampling Results," forthcoming in *Econometrica*.
7. Marshall Blume. "The Assessment of Portfolio Performance." Unpublished Ph.D. thesis, University of Chicago, 1968. A paper summarizing much of this work will appear in the April, 1970, *Journal of Business*.
8. Paul Cootner (ed.). *The Random Character of Stock Market Prices*. Cambridge: M.I.T., 1964.
9. ———. "Stock Prices: Random vs. Systematic Changes." *Industrial Management Review*, 3 (Spring 1962), 24-45. Also reprinted in [8], 231-52.
10. Eugene F. Fama. "The Behavior of Stock Market Prices." *Journal of Business*, 38 (January, 1965), 34-105.
11. ———. "Multiperiod Consumption-Investment Decisions." *American Economic Review*, (March, 1970).
12. ———. "Risk, Return and Equilibrium." Report No. 6831, University of Chicago, Center for Math. Studies in Business and Economics, June, 1968.
13. ——— and Marshall Blume. "Filter Rules and Stock Market Trading Profits." *Journal of Business*, 39 (Special Supplement, January, 1966), 226-41.
14. ———, Lawrence Fisher, Michael Jensen and Richard Roll. "The Adjustment of Stock Prices to New Information." *International Economic Review*, X (February, 1969), 1-21.
15. ——— and Richard Roll. "Some Properties of Symmetric Stable Distributions." *Journal of the American Statistical Association*, 63 (September, 1968), 817-36.
16. Michael D. Godfrey, C. W. J. Granger and O. Morgenstern. "The Random Walk Hypothesis of Stock Market Behavior." *Kyklos*, 17 (1964), 1-30.
17. C. W. J. Granger and O. Morgenstern. "Spectral Analysis of New York Stock Market Prices," *Kyklos*, 16 (1963), 1-27. Also reprinted in [8], 162-88.
18. John R. Hicks. *Value and Capital*. Oxford: The Clarendon Press, 1946.
19. Michael Jensen. "The Performance of Mutual Funds in the Period 1945-64," *Journal of Finance*, 23 (May, 1968), 389-416.

20. ———. "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios," *Journal of Business*, 42 (April, 1969), 167-247.
21. Maurice G. Kendall. "The Analysis of Economic Time-Series, Part I: Prices," *Journal of the Royal Statistical Society*, 96 (Part I, 1953), 11-25.
22. Ruben A. Kessel. "The Cyclical Behavior of the Term Structure of Interest Rates," National Bureau of Economic Research Occasional Paper No. 91. New York: Columbia University Press, 1965.
23. Benjamin F. King. "Market and Industry Factors in Stock Price Behavior," *Journal of Business*, 39 (Special Supplement January, 1966), 139-90.
24. John Lintner. "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20 (December, 1965), 587-615.
25. ———. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47 (February, 1965), 13-37.
26. Fredrich A. Lutz. "The Structure of Interest Rates," *Quarterly Journal of Economics*, 40 (1940-41).
27. Benoit Mandelbrot. "Forecasts of Future Prices, Unbiased Markets, and Martingale Models," *Journal of Business*, 39 (Special Supplement, January, 1966), 242-55.
28. ———. "The Variation of Certain Speculative Prices," *Journal of Business*, 36 (October, 1963), 394-419.
29. ——— and Howard M. Taylor. "On the Distribution of Stock Price Differences," *Operations Research*, 15 (November-December, 1967), 1057-62.
30. Harry Markowitz. *Portfolio Selection: Efficient Diversification of Investment*. New York: John Wiley & Sons, 1959.
31. Arnold Moore. "A Statistical Analysis of Common Stock Prices. Unpublished Ph.D. thesis, Graduate School of Business, University of Chicago, 1962.
32. Victor Niederhoffer and M. F. M. Osborne. "Market Making and Reversal on the Stock Exchange," *Journal of the American Statistical Association*, 61 (December, 1966), 897-916.
33. M. F. M. Osborne. "Brownian Motion in the Stock Market," *Operations Research*, 7 (March-April, 1959), 145-73. Also reprinted in [8], 100-28.
34. ———. "Periodic Structure in the Brownian Motion of Stock Prices," *Operations Research*, 10 (May-June, 1962), 345-79. Also reprinted in [8], 262-96.
35. S. James Press. "A compound Events Model for Security Prices," *Journal of Business*, 40 (July, 1968), 317-35.
36. Harry V. Roberts. "Stock Market 'Patterns' and Financial Analysis: Methodological Suggestions," *Journal of Finance*, 14 (March, 1959), 1-10.
37. Richard Roll. "The Efficient Market Model Applied to U.S. Treasury Bill Rates." Unpublished Ph.D. thesis, Graduate School of Business, University of Chicago, 1968.
38. Paul A. Samuelson. "Proof That Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review*, 6 (Spring, 1965), 41-9.
39. Myron Scholes. "A Test of the Competitive Hypothesis: The Market for New Issues and Secondary Offerings." Unpublished PH.D. thesis, Graduate School of Business, University of Chicago, 1969.
40. William F. Sharpe. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19 (September, 1964), 425-42.
41. ———. "Mutual Fund Performance," *Journal of Business*, 39 (Special Supplement, January, 1966), 119-38.
42. ———. "Risk Aversion in the Stock Market," *Journal of Finance*, 20 (September, 1965), 416-22.
43. James Tobin. "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies*, 25 (February, 1958), 65-85.
44. Jack L. Treynor. "How to Rate Management of Investment Funds," *Harvard Business Review*, 43 (January-February, 1965), 63-75.
45. Roger N. Waud. "Public Interpretation of Discount Rate Changes: Evidence on the 'Announcement Effect.'" forthcoming in *Econometrica*.
46. John Wise. "Linear Estimators for Linear Regression Systems Having Infinite Variances." Unpublished paper presented at the Berkeley-Stanford Mathematical Economics Seminar, October, 1963.
47. Holbrook Working. "A Random Difference Series for Use in the Analysis of Time Series," *Journal of the American Statistical Association*, 29 (March, 1934), 11-24.

The Efficient Market Hypothesis and Its Critics

Burton G. Malkiel

A generation ago, the efficient market hypothesis was widely accepted by academic financial economists; for example, see Eugene Fama's (1970) influential survey article, "Efficient Capital Markets." It was generally believed that securities markets were extremely efficient in reflecting information about individual stocks and about the stock market as a whole. The accepted view was that when information arises, the news spreads very quickly and is incorporated into the prices of securities without delay. Thus, neither technical analysis, which is the study of past stock prices in an attempt to predict future prices, nor even fundamental analysis, which is the analysis of financial information such as company earnings and asset values to help investors select "undervalued" stocks, would enable an investor to achieve returns greater than those that could be obtained by holding a randomly selected portfolio of individual stocks, at least not with comparable risk.

The efficient market hypothesis is associated with the idea of a "random walk," which is a term loosely used in the finance literature to characterize a price series where all subsequent price changes represent random departures from previous prices. The logic of the random walk idea is that if the flow of information is unimpeded and information is immediately reflected in stock prices, then tomorrow's price change will reflect only tomorrow's news and will be independent of the price changes today. But news is by definition unpredictable, and, thus, resulting price changes must be unpredictable and random. As a result, prices fully reflect all known information, and even uninformed investors buying a diversified portfolio at the tableau of prices given by the market will obtain a rate of return as generous as that achieved by the experts.

■ *Burton G. Malkiel is Chemical Bank Chairman's Professor of Economics, Princeton University, Princeton, New Jersey. His e-mail address is <bmalkiel@princeton.edu>.*

The way I put it in my book, *A Random Walk Down Wall Street*, first published in 1973, a blindfolded chimpanzee throwing darts at the *Wall Street Journal* could select a portfolio that would do as well as the experts. Of course, the advice was not literally to throw darts, but instead to throw a towel over the stock pages—that is, to buy a broad-based index fund that bought and held all the stocks in the market and that charged very low expenses.

By the start of the twenty-first century, the intellectual dominance of the efficient market hypothesis had become far less universal. Many financial economists and statisticians began to believe that stock prices are at least partially predictable. A new breed of economists emphasized psychological and behavioral elements of stock-price determination, and they came to believe that future stock prices are somewhat predictable on the basis of past stock price patterns as well as certain “fundamental” valuation metrics. Moreover, many of these economists were even making the far more controversial claim that these predictable patterns enable investors to earn excess risk adjusted rates of return.

This paper examines the attacks on the efficient market hypothesis and the belief that stock prices are partially predictable. While I make no attempt to present a complete survey of the purported regularities or anomalies in the stock market, I will describe the major statistical findings as well as their behavioral underpinnings, where relevant, and also examine the relationship between predictability and efficiency. I will also describe the major arguments of those who believe that markets are often irrational by analyzing the “crash of 1987,” the Internet “bubble” of the *fin de siècle* and other specific irrationalities often mentioned by critics of efficiency. I conclude that our stock markets are far more efficient and far less predictable than some recent academic papers would have us believe. Moreover, the evidence is overwhelming that whatever anomalous behavior of stock prices may exist, it does not create a portfolio trading opportunity that enables investors to earn extraordinary risk adjusted returns.

At the outset, it is important to make clear what I mean by the term “efficiency.” I will use as a definition of efficient financial markets that such markets do not allow investors to earn above-average returns without accepting above-average risks. A well-known story tells of a finance professor and a student who come across a \$100 bill lying on the ground. As the student stops to pick it up, the professor says, “Don’t bother—if it were really a \$100 bill, it wouldn’t be there.” The story well illustrates what financial economists usually mean when they say markets are efficient. Markets can be efficient in this sense even if they sometimes make errors in valuation, as was certainly true during the 1999–early 2000 Internet “bubble.” Markets can be efficient even if many market participants are quite irrational. Markets can be efficient even if stock prices exhibit greater volatility than can apparently be explained by fundamentals such as earnings and dividends. Many of us economists who believe in efficiency do so because we view markets as amazingly successful devices for reflecting new information rapidly and, for the most part, accurately. Above all, we believe that financial markets are efficient because they don’t allow investors to earn above-average risk adjusted returns. In short, we

believe that \$100 bills are not lying around for the taking, either by the professional or the amateur investor.

What I do not argue is that the market pricing is always perfect. After the fact, we know that markets have made egregious mistakes, as I think occurred during the recent Internet “bubble.” Nor do I deny that psychological factors influence securities prices. But I am convinced that Benjamin Graham (1965) was correct in suggesting that while the stock market in the short run may be a voting mechanism, in the long run it is a weighing mechanism. True value will win out in the end. Before the fact, there is no way in which investors can reliably exploit any anomalies or patterns that might exist. I am skeptical that any of the “predictable patterns” that have been documented in the literature were ever sufficiently robust so as to have created profitable investment opportunities, and after they have been discovered and publicized, they will certainly not allow investors to earn excess returns.

A Nonrandom Walk Down Wall Street

In this section, I review some of the patterns of possible predictability suggested by studies of the behavior of past stock prices.

Short-Term Momentum, Including Underreaction to New Information

The original empirical work supporting the notion of randomness in stock prices looked at measures of short-run serial correlations between successive stock price changes. In general, this work supported the view that the stock market has no memory—that is, the way a stock price behaved in the past is not useful in divining how it will behave in the future; for example, see the survey of articles contained in Cootner (1964). More recent work by Lo and MacKinlay (1999) finds that short-run serial correlations are not zero and that the existence of “too many” successive moves in the same direction enable them to reject the hypothesis that stock prices behave as true random walks. There does seem to be some momentum in short-run stock prices. Moreover, Lo, Mamaysky and Wang (2000) also find, through the use of sophisticated nonparametric statistical techniques that can recognize patterns, some of the stock price signals used by “technical analysts,” such as “head and shoulders” formations and “double bottoms,” may actually have some modest predictive power.

Economists and psychologists in the field of behavioral finance find such short-run momentum to be consistent with psychological feedback mechanisms. Individuals see a stock price rising and are drawn into the market in a kind of “bandwagon effect.” For example, Shiller (2000) describes the rise in the U.S. stock market during the late 1990s as the result of psychological contagion leading to irrational exuberance. The behavioralists offered another explanation for patterns of short-run momentum—a tendency for investors to underreact to new information. If the full impact of an important news announcement is only grasped over a period of time, stock prices will exhibit the positive serial correlation found by

investigators. As behavioral finance became more prominent as a branch of the study of financial markets, momentum, as opposed to randomness, seemed reasonable to many investigators.

However, several factors should prevent us from interpreting the empirical results reported above as an indication that markets are inefficient. First, while the stock market may not be a mathematically perfect random walk, it is important to distinguish statistical significance from economic significance. The statistical dependencies giving rise to momentum are extremely small and are not likely to permit investors to realize excess returns. Anyone who pays transactions costs is unlikely to fashion a trading strategy based on the kinds of momentum found in these studies that will beat a buy-and-hold strategy. Indeed, Odean (1999) suggests that momentum investors do not realize excess returns. Quite the opposite—a sample of such investors suggests that such traders did far worse than buy-and-hold investors even during a period where there was clear statistical evidence of positive momentum. This is because of the large transactions costs involved in attempting to exploit whatever momentum exists. Similarly, Lesmond, Schill and Zhou (2001) find that standard “relative strength” strategies are not profitable because of the trading costs involved in their execution.

Second, while behavioral hypotheses about bandwagon effects and underreaction to new information may sound plausible enough, the evidence that such effects occur systematically in the stock market is often rather thin. For example, Eugene Fama (1998) surveys the considerable body of empirical work on “event studies” that seeks to determine if stock prices respond efficiently to information. The “events” include such announcements as earnings surprises, stock splits, dividend actions, mergers, new exchange listings and initial public offerings. Fama finds that apparent underreaction to information is about as common as overreaction, and postevent continuation of abnormal returns is as frequent as postevent reversals. He also shows that many of the return “anomalies” arise only in the context of some very particular model and that the results tend to disappear when exposed to different models for expected “normal” returns, different methods to adjust for risk and when different statistical approaches are used to measure them. For example, a study that gives equal weight to postannouncement returns of many stocks can produce different results from a study that weights the stocks according to their value. Certainly, whatever momentum displayed by stock prices does not appear to offer investors a dependable way to earn abnormal returns.

The key factor is whether any patterns of serial correlation are consistent over time. Momentum strategies, which refer to buying stocks that display positive serial correlation and/or positive relative strength, appeared to produce positive relative returns during some periods of the late 1990s, but highly negative relative returns during 2000. It is far from clear that any stock price patterns are useful for investors in fashioning an investment strategy that will dependably earn excess returns.

Many predictable patterns seem to disappear after they are published in the finance literature. Schwert (2001) points out two possible explanations for such a pattern. One explanation may be that researchers are always sifting through

mountains of financial data. Their normal tendency is to focus on results that challenge perceived wisdom, and every now and again, a combination of a certain sample and a certain technique will produce a statistically significant result that seems to challenge the efficient markets hypothesis. Alternatively, perhaps practitioners learn quickly about any true predictable pattern and exploit it to the extent that it becomes no longer profitable. My own view is that such apparent patterns were never sufficiently large or stable to guarantee consistently superior investment results, and certainly, such patterns will never be useful for investors after they have received considerable publicity. The so-called “January effect,” for example, in which stock prices rose in early January, seems to have disappeared soon after it was discovered.

Long-Run Return Reversals

In the short-run, when stock returns are measured over periods of days or weeks, the usual argument against market efficiency is that some positive serial correlation exists. But many studies have shown evidence of negative serial correlation—that is, return reversals—over longer holding periods. For example, Fama and French (1988) found that 25 to 40 percent of the variation in long holding period returns can be predicted in terms of a negative correlation with past returns. Similarly, Poterba and Summers (1988) found substantial mean reversion in stock market returns at longer horizons.

Some studies have attributed this forecastability to the tendency of stock market prices to “overreact.” DeBondt and Thaler (1985), for example, argue that investors are subject to waves of optimism and pessimism that cause prices to deviate systematically from their fundamental values and later to exhibit mean reversion. They suggest that such overreaction to past events is consistent with the behavioral decision theory of Kahneman and Tversky (1979), where investors are systematically overconfident in their ability to forecast either future stock prices or future corporate earnings. These findings give some support to investment techniques that rest on a “contrarian” strategy, that is, buying the stocks, or groups of stocks, that have been out of favor for long periods of time and avoiding those stocks that have had large run-ups over the last several years.

There is indeed considerable support for long-run negative serial correlation in stock returns. However, the finding of mean reversion is not uniform across studies and is quite a bit weaker in some periods than it is for other periods. Indeed, the strongest empirical results come from periods including the Great Depression—which may be a time with patterns that do not generalize well. Moreover, such return reversals for the market as a whole may be quite consistent with the efficient functioning of the market since they could result, in part, from the volatility of interest rates and the tendency of interest rates to be mean reverting. Since stock returns must rise or fall to be competitive with bond returns, there is a tendency when interest rates go up for prices of both bond and stocks to go down, and as interest rates go down for prices of bonds and stocks to go up. If interest rates revert to the mean over time, this pattern will tend to generate return

reversals, or mean reversion, in a way that is quite consistent with the efficient functioning of markets.

Moreover, it may not be possible to profit from the tendency for individual stocks to exhibit return reversals. Fluck, Malkiel and Quandt (1997) simulated a strategy of buying stocks over a 13-year period during the 1980s and early 1990s that had particularly poor returns over the past three to five years. They found that stocks with very low returns over the past three to five years had higher returns in the next period and that stocks with very high returns over the past three to five years had lower returns in the next period. Thus, they confirmed the very strong *statistical* evidence of return reversals. However, they also found that returns in the next period were similar for both groups, so they could not confirm that a contrarian approach would yield higher-than-average returns. There was a statistically strong pattern of return reversal, but not one that implied an inefficiency in the market that would enable investors to make excess returns.

Seasonal and Day-of-the-Week Patterns

A number of researchers have found that January has been a very unusual month for stock market returns. Returns from an equally weighted stock index have tended to be unusually high during the first two weeks of the year. The return premium has been particularly evident for stocks with relatively small total capitalizations (Keim, 1983). Haugen and Lakonishok (1988) documented the high January returns in a book titled *The Incredible January Effect*. There also appear to be a number of day-of-the-week effects. For example, French (1980) documents significantly higher Monday returns. There appear to be significant differences in average daily returns in countries other than the United States (Hawawini and Keim, 1995). There also appear to be some patterns in returns around the turn of the month (Lakonishok and Smidt, 1988), as well as around holidays (Ariel, 1990).

The general problem with these predictable patterns or anomalies, however, is that they are *not* dependable from period to period. Wall Street traders now joke that the “January effect” is more likely to occur on the previous Thanksgiving. Moreover, these nonrandom effects (even if they were dependable) are very small relative to the transactions costs involved in trying to exploit them. They do not appear to offer arbitrage opportunities that would enable investors to make excess risk adjusted returns.

Predictable Patterns Based on Valuation Parameters

Considerable empirical research has been conducted to determine if future stock returns can be predicted on the basis of initial valuation parameters. It is claimed that valuation ratios, such as the price-earnings multiple or the dividend yield of the stock market as a whole, have considerable predictive power. This section examines the body of work based on time series analyses.

Predicting Future Returns from Initial Dividend Yields

Formal statistical tests of the ability of dividend yields (that is, the ratio of dividend to stock price) to forecast future returns have been conducted by Fama and French (1988) and Campbell and Shiller (1988). Depending on the forecast horizon involved, as much as 40 percent of the variance of future returns for the stock market as a whole can be predicted on the basis of the initial dividend yield of the market index.

An interesting way of presenting the results is shown in the top panel of Exhibit 1. The exhibit was produced by measuring the dividend yield of the broad U.S. stock market Standard & Poor's 500 Stock Index each quarter since 1926 and then calculating the market's subsequent ten-year total return through the year 2001. The observations were then divided into deciles depending upon the level of the initial dividend yield. In general, the exhibit shows that investors have earned a higher rate of return from the stock market when they purchased a market basket of equities with an initial dividend yield that was relatively high and relatively low future rates of return when stocks were purchased at low dividend yields.

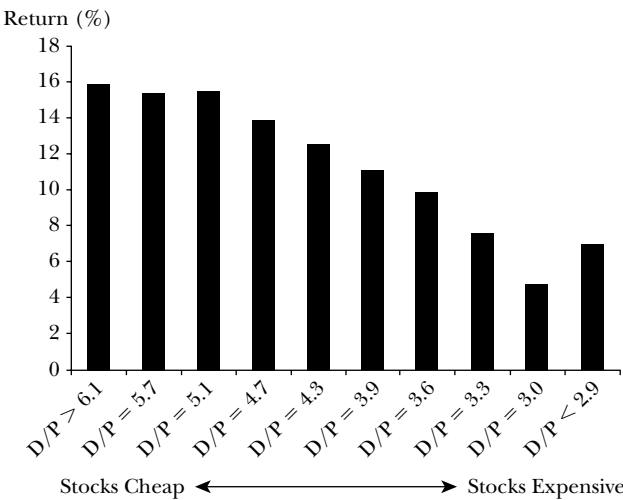
These findings are not necessarily inconsistent with efficiency. Dividend yields of stocks tend to be high when interest rates are high, and they tend to be low when interest rates are low. Consequently, the ability of initial yields to predict returns may simply reflect the adjustment of the stock market to general economic conditions. Moreover, the use of dividend yields to predict future returns has been ineffective since the mid-1980s. Dividend yields have been at the 3 percent level or below continuously since the mid-1980s, indicating very low forecasted returns. In fact, for all ten-year periods from 1985 through 1992 that ended June 30, 2002, realized annual equity returns from the market index have averaged over 15 percent. One possible explanation is that the dividend behavior of U.S. corporations may have changed over time (Bagwell and Shoven, 1989; Fama and French, 2001). Companies in the twenty-first century may be more likely to institute a share repurchase program rather than increase their dividends. Thus, dividend yield may not be as meaningful as in the past as a useful predictor of future equity returns.

Finally, it is worth noting that this phenomenon does *not* work consistently with individual stocks (Fluck, Malkiel and Quandt, 1997). Investors who simply purchase a portfolio of individual stocks with the highest dividend yields in the market will *not* earn a particularly high rate of return. One popular implementation of such a "high dividend" strategy in the United States is the "Dogs of the Dow Strategy," which involves buying the ten stocks in the Dow Jones Industrial Average with the highest dividend yields. For some past periods, this strategy handily outpaced the overall average, and so several "Dogs of the Dow" mutual funds were brought to market and aggressively sold to individual investors. However, such funds generally underperformed the market averages during the 1995–1999 period.

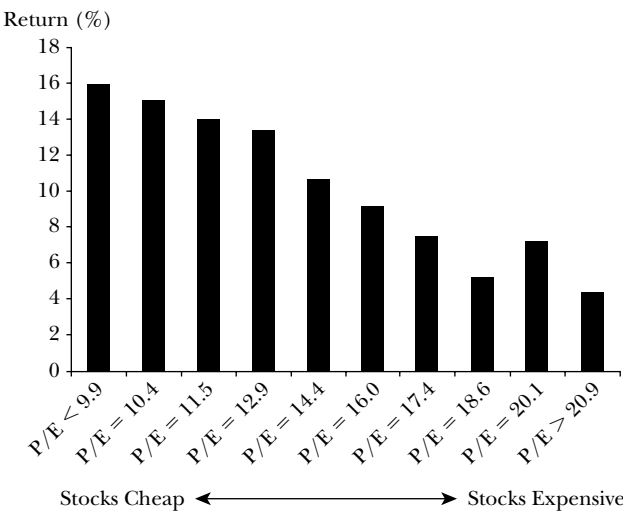
Predicting Market Returns from Initial Price-Earnings Multiples

The same kind of predictability for the market as a whole, as was demonstrated for dividends, has been shown for price-earnings ratios. The data are shown in the

Exhibit 1
The Future 10-Year Rates of Return When Stocks are Purchased at Alternative Initial Dividend Yields (D/P)



The Future 10-Year Rates of Return When Stocks are Purchased at Alternative Initial Price-to-Earnings (P/E) Multiples



Source: The Leuthold Group

bottom half of Exhibit 1. The exhibit presents a decile analysis similar to that described for dividend yields above. Investors have tended to earn larger long-horizon returns when purchasing the market basket of stocks at relatively low price-earnings multiples. Campbell and Shiller (1998) report that initial P/E ratios

explained as much as 40 percent of the variance of future returns. They conclude that equity returns have been predictable in the past to a considerable extent.

Consider, however, the recent experience of investors who have attempted to undertake investment strategies based either on the level of the price-earnings multiple or the dividend yield to predict future long-horizon returns. Price-earnings multiples for the Standard & Poor's 500 stock index rose into the low 20s on June 30, 1987 (suggesting very low long-horizon returns). Dividend yields fell below 3 percent. Price-earnings multiples rose into the low 20s. The average annual total return from the index over the next 10 years was an extraordinarily generous 16.7 percent. Dividend yields, again, fell to 3 percent in June 1992. Price-earnings multiples rose to the mid-20s. The subsequent return through June 2002 was 11.4 percent. The yield of the index fluctuated between 2 and 3 percent from 1993 through 1995 and earnings multiples remained in the mid-20s, yet long-horizon returns through June 30, 2002, fluctuated between 11 and 12 percent. Even from early December 1996, the date of Campbell and Shiller's presentation to the Federal Reserve suggesting near zero returns for the S&P500, the index provided almost a 7 percent annual return through mid-2002. Such results suggest to me a very cautious assessment of the extent to which stock market returns are predictable on this basis.

Other Predictable Time Series Patterns

Studies have found some amount of predictability of stock returns based on various financial statistics. For example, Fama and Schwert (1977) found that short-term interest rates were related to future stock returns. Campbell (1987) found that term structure of interest rates spreads contained useful information for forecasting stock returns. Keim and Stambaugh (1986) found that risk spreads between high-yield corporate bonds and short rates had some predictive power. Again, even if some predictability exists, it may reflect time varying risk premiums and required rates of return for stock investors rather than an inefficiency. Moreover, it is far from clear that any of these results can be used to generate profitable trading strategies. Whether such historical statistical relations give investors reliable and useful guides to appropriate asset allocation is far from clear.

Cross-Sectional Predictable Patterns Based on Firm Characteristics and Valuation Parameters

A large number of patterns that are claimed to be predictable are based on firm characteristics and different valuation parameters.

The Size Effect

One of the strongest effects investigators have found is the tendency over long periods of time for smaller-company stocks to generate larger returns than those of

large-company stocks. Since 1926, small-company stocks in the United States have produced annual rates of return over 1 percentage point larger than the returns from large stocks (Keim, 1983). Fama and French (1993) examined data from 1963 to 1990 and divided all stocks into deciles according to their size as measured by total capitalization. Decile one contained the smallest 10 percent of all stocks, while decile ten contained the largest stocks. The results, plotted in Exhibit 2, show a clear tendency for the deciles made up of portfolios of smaller stocks to generate higher average monthly returns than deciles made up of larger stocks.

The crucial issue here is the extent to which the higher returns of small companies represent a predictable pattern that will allow investors to generate excess risk-adjusted returns. According to the capital asset pricing model, the correct measure of risk for a stock is its “beta”—that is, the extent to which the return of the stock varies with the return for the market as a whole. If the beta measure of systematic risk from the capital asset pricing model is accepted as the correct risk measurement statistic, the size effect can be interpreted as indicating an anomaly and a market inefficiency, because using this measure portfolios consisting of smaller stocks have excess risk-adjusted returns. Fama and French (1993) point out, however, that the average relationship between beta and return during the 1963–1990 period was flat—not upward sloping as the capital asset pricing model predicts. Moreover, if stocks are divided up by beta deciles, ten portfolios constructed by size display the same kind of positive relationship shown in Exhibit 2. On the other hand, within size deciles, the relationship between beta and return continues to be flat. Fama and French suggest that size may be a far better proxy for risk than beta, and therefore that their findings should not be interpreted as indicating that markets are inefficient.

The dependability of the size phenomenon is also open to question. From the mid-1980s through the decade of the 1990s, there has been no gain from holding smaller stocks. Indeed, in most world markets, larger capitalization stocks produced larger rates of return. It may be that the growing institutionalization of the market led portfolio managers to prefer larger companies with more liquidity to smaller companies where it would be difficult to liquidate significant blocks of stock. Finally, it is also possible that some studies of the small-firm effect have been affected by survivorship bias. Today’s computerized databases of companies include only small firms that have survived, not the ones that later went bankrupt. Thus, a researcher who examined the ten-year performance of today’s small companies would be measuring the performance of those companies that survived—not the ones that failed.

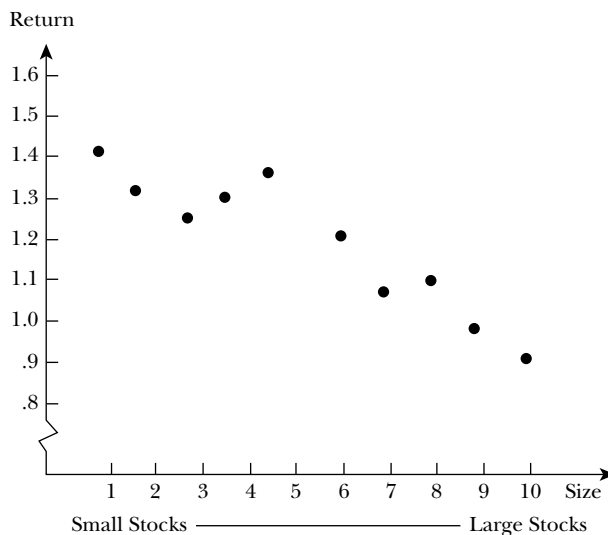
Value Stocks

Several studies suggest that “value” stocks have higher returns than so-called “growth” stocks. The most common two methods of identifying value stocks have been price-earnings ratios and price-to-book-value ratios.

Stocks with low price-earnings multiples (often called “value” stocks) appear to provide higher rates of return than stocks with high price-to-earnings ratios, as first

Exhibit 2

Average Monthly Returns for Portfolios Formed on the Basis of Size: 1963–1990



Source: Fama and French (1992).

shown by Nicholson (1960) and later confirmed by Ball (1978) and Basu (1983). This finding is consistent with the views of behavioralists that investors tend to be overconfident of their ability to project high earnings growth and thus overpay for “growth” stocks (for example, Kahneman and Riepe, 1998). The finding is also consistent with the views of Graham and Dodd (1934), first expounded in their classic book on security analysis and later championed by the legendary U.S. investor Warren Buffett. Similar results have been shown for price/cash flow multiples, where cash flow is defined as earnings plus depreciation and amortization (Hawawini and Keim, 1995).

The ratio of stock price to book value, defined as the value of a firm’s assets minus its liabilities divided by the number of shares outstanding, has also been found to be a useful predictor of future returns. Low price-to-book is considered to be another hallmark of so-called “value” in equity securities and is also consistent with the view of behavioralists that investors tend to overpay for “growth” stocks that subsequently fail to live up to expectations. Fama and French (1993) concluded that size and price-to-book-value together provide considerable explanatory power for future returns, and once they are accounted for, little additional influence can be attributed to P/E multiples. Fama and French (1997) also conclude that the P/BV effect is important in many world stock markets other than the United States.

Such results raise questions about the efficiency of the market if one accepts the capital asset pricing model, as Lakonishok, Shleifer and Vishny (1994) point out. But these findings do not necessarily imply inefficiency. They may simply indicate failure of the capital asset pricing model to capture all the dimensions of

risk. For example, Fama and French (1993) suggest that the price-to-book-value ratio may reflect another risk factor that is priced into the market and not captured by the capital asset pricing model. Companies in some degree of financial distress, for example, are likely to sell at low prices relative to book values. Fama and French (1993) argue that a three-factor asset-pricing model (including price-to-book-value and size as measures of risk) is the appropriate benchmark against which anomalies should be measured.

We also need to keep in mind that the results of published studies—even those done over decades—may still be time-dependent and ask whether the return patterns of academic studies can actually be generated with real money. Exhibit 3 presents average actual returns generated by mutual funds classified by either their “growth” or “value” objectives. “Value” funds are so classified if they buy stocks with price-to-earnings and price-to-book-value multiples that are below the averages for the whole stock market. Over a period running back to the 1930s, it does not appear that investors could actually have realized higher rates of return from mutual funds specializing in “value” stocks. Indeed, the exhibit suggests that the Fama-French (1993) period from the early 1960s through 1990 may have been a unique period in which value stocks rather consistently produced higher rates of return.

Schwert (2001) points out that the investment firm of Dimensional Fund Advisors actually began a mutual fund that selected value stocks quantitatively according to the Fama and French (1993) criteria. The abnormal return of such a portfolio (adjusting for beta, the capital asset pricing model measure of risk) was a negative 0.2 percent per month over the 1993–1998 period. The absence during that period of an excess return to the “value” stocks is consistent with the results from “actively managed” value mutual funds shown in Exhibit 3.

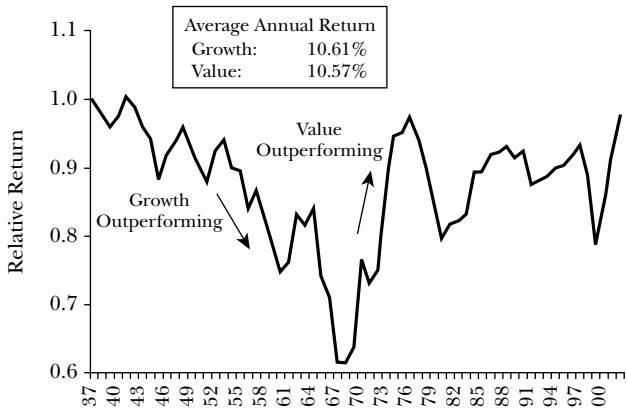
The Equity Risk Premium Puzzle

Another puzzle that is often used to suggest that markets are less than fully rational is the existence of a very large historical equity risk premium that seems inconsistent with the actual riskiness of common stocks as can be measured statistically. For example, using the Ibbotson data on stock returns from 1926 through 2001, common stocks have produced rates of return of approximately 10.5 percent, while high-grade bonds have returned only about 5.5 percent. I believe that this finding is the result of a combination of perceived equity risk being considerably higher during the early years of the period and of average equity returns being much higher than had been forecast by investors.

It is easy to say 50 to 75 years later that common stocks were underpriced during the 1930s and 1940s. But remember that the annual average of almost 6 percent growth in corporate earnings and dividends that has occurred since 1926 was hardly a foregone conclusion during a period of severe depression and world war. Indeed, the U.S. stock market is almost unique in that it is one of the few world markets that remained in continuous operation during the entire period and the measured risk premium results, in part, from survivorship bias. One must be very

Exhibit 3

Reversion to the Mean: Relative Performance of “Value” vs. “Growth” Mutual Funds, 1937–June 2002



Source: Lipper Analytic Services and Bogle Research Institute Valley Forge, Pennsylvania.
Note: The exhibit shows the cumulative value of one dollar invested in the average “value” fund divided by the same statistic calculated for the average “growth” fund.

careful to distinguish between expected risk premiums and such premiums measured after the fact. Fama and French (2002) argue that the high average realized returns result in part from large *unexpected* capital gains. Economists such as Shiller have suggested that during the early 2000s, the expected equity risk premium was, if anything, irrationally too low.

Summarizing the “Anomalies” and Predictable Patterns

The preceding sections have pointed out many “anomalies” and statistically significant predictable patterns in the stock returns that have been uncovered in the literature. However, these patterns are not robust and dependable in different sample periods, and some of the patterns based on fundamental valuation measures of individual stocks may simply reflect better proxies for measuring risk.

Moreover, many of these patterns, even if they did exist, could self-destruct in the future, as many of them have already done. Indeed, this is the logical reason why one should be cautious not to overemphasize these anomalies and predictable patterns. Suppose, for example, one of the anomalies or predictable patterns appears to be robust. Suppose there is a truly dependable and exploitable January effect, that the stock market—especially stocks of small companies—will generate extraordinary returns during the first five days of January. What will investors do? They will buy on the last day of December and sell on January 5. But then investors find that the market rallied on the last day of December, and so they will need to begin to buy on the next-to-last day of December; and because there is so much “profit taking” on January 5, investors will have to sell on January 4 to take advantage of this effect. Thus, to beat the gun, investors will have to be buying

earlier and earlier in December and selling earlier and earlier in January so that eventually the pattern will self-destruct. Any truly repetitive and exploitable pattern that can be discovered in the stock market and can be arbitrated away will self-destruct. Indeed, the January effect became undependable after it received considerable publicity.

Similarly, suppose there is a general tendency for stock prices to underreact to certain new events, leading to abnormal returns to investors who exploit the lack of full immediate adjustment (DeBondt and Thaler, 1995; Campbell, Lo and MacKinlay, 1977). “Quantitative” investment managers will then develop trading strategies to exploit the pattern. Indeed, the more potentially profitable a discoverable pattern is, the less likely it is to survive.

Many of the predictable patterns that have been discovered may simply be the result of data mining. The ease of experimenting with financial databanks of almost every conceivable dimension makes it quite likely that investigators will find some seemingly significant but wholly spurious correlation between financial variables or among financial and nonfinancial data sets. Given enough time and massaging of data series, it is possible to tease almost any pattern out of most data sets. Moreover, the published literature is likely to be biased in favor of reporting such results. Significant effects are likely to be published in professional journals while negative results, or boring confirmations of previous findings, are relegated to the file drawer or discarded. Data-mining problems are unique to nonexperimental sciences, such as economics, which rely on statistical analysis for their insights and cannot test hypotheses by running repeated controlled experiments.

An exchange at a symposium about a decade ago between Robert Shiller, an economist who is sympathetic to the argument that stock prices are partially predictable and skeptical about market efficiency, and Richard Roll, an academic financial economist who also is a portfolio manager, is quite revealing (Roll and Shiller, 1992). After Shiller stressed the importance of inefficiencies in the pricing of stocks, Roll responded as follows:

I have personally tried to invest money, my client’s money and my own, in every single anomaly and predictive device that academics have dreamed up. . . . I have attempted to exploit the so-called year-end anomalies and a whole variety of strategies supposedly documented by academic research. *And I have yet to make a nickel on any of these supposed market inefficiencies . . .* a true market *inefficiency* ought to be an exploitable opportunity. If there’s nothing investors can exploit in a systematic way, time in and time out, then it’s very hard to say that information is not being properly incorporated into stock prices.

Seemingly Irrefutable Cases of Inefficiency

Critics of efficiency argue that there are several instances of recent market history where market prices could not plausibly have been set by rational investors

and that psychological considerations must have played the dominant role. It is alleged, for example, that the stock market lost about one-third of its value from early to mid-October 1987 with essentially no change in the general economic environment. How could market prices be efficient both at the start of October and during the middle of the month? Similarly, it is widely believed that the pricing of Internet stocks in early 2000 could only be explained by the behavior of irrational investors. Do such events make a belief in efficient markets untenable?

The Market Crash of October 1987

Can the October 1987 market crash be explained by rational considerations, or does such a rapid and significant change in market valuations prove the dominance of psychological rather than logical factors in understanding the stock market? Behaviorists would say that the one-third drop in market prices, which occurred early in October 1987, can only be explained by relying on psychological considerations, since the basic elements of the valuation equation did not change rapidly over that period. It is, of course, impossible to rule out the existence of behavioral or psychological influences on stock market pricing. But logical considerations can explain a sharp change in market valuations such as occurred during the first weeks of October 1987.

A number of factors could rationally have changed investors' views about the proper value of the stock market in October 1987. For one thing, yields on long-term Treasury bonds increased from about 9 percent to almost 10.5 percent in the two months prior to mid-October. Moreover, a number of events may rationally have increased risk perceptions during the first two weeks of October. Early in the month, Congress threatened to impose a "merger tax" that would have made merger activity prohibitively expensive and could well have ended the merger boom. The risk that merger activity might be curtailed increased risks throughout the stock market by weakening the discipline over corporate management that potential takeovers provide. Also, in early October 1987, then Secretary of the Treasury James Baker had threatened to encourage a further fall in the exchange value of the dollar, increasing risks for foreign investors and frightening domestic investors as well. While it is impossible to correlate each day's movement in stock prices to specific news events, it is not unreasonable to ascribe the sharp decline in mid-October to the cumulative effect of a number of unfavorable "fundamental" events. As Merton Miller (1991) has written, "... on October 19, some weeks of external events, minor in themselves . . . cumulatively signaled a possible change in what had been up to then a very favorable political and economic climate for equities . . . and . . . many investors simultaneously came to believe they were holding too large a share of their wealth in risky equities."

Share prices can be highly sensitive as a result of rational responses to small changes in interest rates and risk perceptions. Suppose stocks are priced as the present value of the expected future stream of dividends. For a long-term holder of stocks, this rational principle of valuation translates to a formula:

$$r = D/P + g,$$

where r is the rate of return, D/P is the (expected) dividend yield, and g is the long-term growth rate. For present purposes, consider r to be the required rate of return for the market as a whole. Suppose initially that the “riskless” rate of interest on government bonds is 9 percent and that the required additional risk premium for equity investors is 2 percentage points. In this case, r will be 11 percent ($0.09 + 0.02 = 0.11$). If a typical stock’s expected growth rate, g , is 7 percent and if the dividend is \$4 per share, we can solve for the appropriate price of the stock index (P), obtaining

$$0.11 = \frac{\$4}{P} + 0.07$$

$$P = \$100.$$

Now assume that yields on government bonds rise from 9 to 10.5 percent, with no increase in expected inflation, and that risk perceptions increase so that stock-market investors now demand a premium of 2.5 percentage points instead of the 2 points in the previous example. The appropriate rate of return or discount rate for stocks, r , rises then from 11 percent to 13 percent ($0.105 + 0.025$), and the price of our stock index falls from \$100 to \$66.67:

$$0.13 = \frac{\$4}{P} + 0.07$$

$$P = \$66.67.$$

The price must fall to raise the dividend yield from 4 to 6 percent so as to raise the total return by the required 2 percentage points. Clearly, no irrationality is required for share prices to suffer quite dramatic declines with the sorts of changes in interest rates and risk perceptions that occurred in October 1987. Of course, even a very small decline in anticipated growth would have magnified these declines in warranted share valuations.

This is not to say that psychological factors were irrelevant in explaining the sharp drop in prices during October 1987—they undoubtedly played a role. But it would be a mistake to dismiss the significant change in the external environment, which can provide an entirely rational explanation for a significant decline in the appropriate values for common stocks.

The Internet Bubble of the Late 1990s

Another stock market event often cited by behavioralists as clear evidence of the irrationality of markets is the Internet “bubble” of the late 1990s. Surely, the

remarkable market values assigned to Internet and related high-tech companies seem inconsistent with rational valuation. I have some sympathy with behavioralists in this instance, and in reviewing Robert Shiller's (2000) *Irrational Exuberance*, I agreed that it was in the high-tech sector of the market that his thesis could be supported. But even here, when we know after the fact that major errors were made, there were certainly no arbitrage opportunities available to rational investors before the "bubble" popped.

Equity valuations rest on uncertain future forecasts. Even if all market participants rationally price common stocks as the present value of all future cash flows expected, it is still possible for excesses to develop. We know now, with the benefit of hindsight, that outlandish and unsupportable claims were being made regarding the growth of the Internet (and the related telecommunications structure needed to support it). We know now that projections for the rates and duration of growth of these "new economy" companies were unsustainable. But remember, sharp-penciled professional investors argued that the valuations of high-tech companies were proper. Many of Wall Street's most respected security analysts, including those independent of investment banking firms, were recommending Internet stocks to the firm's institutional and individual clients as being fairly valued. Professional pension fund and mutual fund managers overweighted their portfolios with high-tech stocks.

While it is now clear in retrospect that such professionals were egregiously wrong, there was certainly no obvious arbitrage opportunity available. One could disagree with the projected growth rates of security analysts. But who could be sure, with the use of the Internet for a time doubling every several months, that the extraordinary growth rates that could justify stock valuations were impossible? After all, even Alan Greenspan was singing the praises of the new economy. Nothing is ever as clear in prospect as it is in retrospect. The extent of the "bubble" was only clear in retrospect.

Not only is it almost impossible to judge with confidence what the proper fundamental value is for any security, but potential arbitrageurs face additional risks. Shleifer (2000) has argued that "noise trader risk"—the risk from traders who are attempting to buy into rising markets and sell into declining markets—limits the extent to which one should expect arbitrage to bring prices quickly back to rational values even in the presence of an apparent bubble. Professional arbitrageurs will be loath to sell short a stock they believe is trading at two times its "fundamental" value when it is always possible that some greater fools may be willing to pay three times the stock's value. Arbitrageurs are quite likely to have short horizons, since even temporary losses may induce their clients to withdraw their money.

While there were no profitable and predictable arbitrage opportunities available during the Internet "bubble," and while stock prices eventually did adjust to levels that more reasonably reflected the likely present value of their cash flows, an argument can be maintained that the asset prices did remain "incorrect" for a period of time. The result was that too much new capital flowed to Internet and

related telecommunications companies. Thus, the stock market may well have temporarily failed in its role as an efficient allocator of equity capital. Fortunately, “bubble” periods are the exception rather than the rule, and acceptance of such occasional mistakes is the necessary price of a flexible market system that usually does a very effective job of allocating capital to its most productive uses.

Other Illustrations of Irrational Pricing

Are there not some illustrations of irrational pricing that can be clearly ascertained as they arise, not simply after a “bubble” has burst? My favorite illustration concerns the spinoff of Palm Pilot from its parent 3Com Corporation during the height of the Internet boom in early 2000. Initially, only 5 percent of the Palm Pilot shares were distributed to the public; the other 95 percent remained on 3Com’s balance sheet. As Palm Pilot began trading, enthusiasm for the shares was so great that the 95 percent of its shares still owned by 3Com had a market value considerably more than the entire market capitalization of 3Com, implying that all the rest of its business had a negative value. Other illustrations involve ticker symbol confusion. Rasches (2001) finds clear evidence of comovement of stocks with similar ticker symbols; for example, the stock of MCI Corporation (ticker symbol MCIC) moves in tandem with an unrelated closed-end bond investment fund Mass Mutual Corporate Investors (ticker symbol MCI). In a charming article entitled “A Rose.com by Any Other Name,” Cooper, Dimitrov and Rau (2001) found positive stock price reactions during 1998 and 1999 on corporate name changes when “dot com” was added to the corporate title. Finally, it has been argued that closed-end funds sell at irrational discounts from their net asset values (for example, Shleifer, 2000).

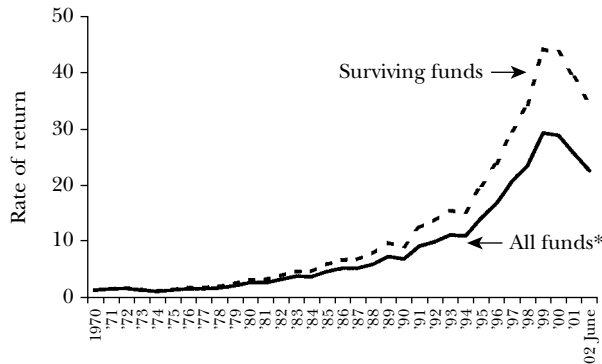
But none of these illustrations should shake our faith that exploitable arbitrage opportunities should not exist in an efficient market. The apparent arbitrage in the Palm Pilot case (sell Palm Pilot short and buy 3Com) could not be undertaken because not enough Palm stock was outstanding to make borrowing the stock possible to effectuate a short sale. The “anomaly” disappeared once 3Com spun off more of Palm stock. Moreover, the potential profits from name or ticker symbol confusion are extremely small relative to the transactions costs that would be required to exploit them. Finally, the “closed-end fund puzzle” is not really a puzzle today. Discounts have narrowed from historical averages for funds with assets traded in liquid markets, and researchers such as Ross (forthcoming) have suggested that they can largely be explained by fund management fees. Perhaps the more important puzzle today is why so many investors buy high-expense, actively managed mutual funds instead of low-cost index funds.

The Performance of Professional Investors

For me, the most direct and most convincing tests of market efficiency are direct tests of the ability of professional fund managers to outperform the market

Exhibit 4

The Records of Surviving Funds Overstates the Success of Active Management



* Includes those liquidated or merged out of existence.

Source: Lipper Analytic Services, as compiled by The Vanguard Group.

as a whole. Surely, if market prices were determined by irrational investors and systematically deviated from rational estimates of the present value of corporations, and if it were easy to spot predictable patterns in security returns or anomalous security prices, then professional fund managers should be able to beat the market. Direct tests of the actual performance of professionals, who often are compensated with strong incentives to outperform the market, should represent the most compelling evidence of market efficiency.

A remarkably large body of evidence suggests that professional investment managers are not able to outperform index funds that buy and hold the broad stock market portfolio. The first study of mutual fund performance was undertaken by Jensen (1968). He found that active mutual fund managers were unable to add value and, in fact, tended to underperform the market by approximately the amount of their added expenses. I repeated Jensen's study with data from a subsequent period and confirmed the earlier results (Malkiel, 1995). Moreover, I found that the degree of "survivorship bias" in the data was substantial; that is, poorly performing funds tend to be merged into other funds in the mutual fund's family complex, thus burying the records of many of the underperformers. Exhibit 4 updates the study I performed through mid-2002, showing that the return for surviving funds is quite a bit better than the actual return for all funds, including funds liquidated or merged out of existence. Survivorship bias makes the interpretation of long-run mutual fund data sets very difficult. But even using data sets with some degree of survivorship bias, one cannot sustain the argument that professional investors can beat the market.

Exhibit 5 presents the percentage of actively managed mutual funds that have been outperformed by the Standard & Poor's 500 and the Wilshire stock indexes. Throughout the past decade, about three-quarters of actively managed funds have

Exhibit 5
Percentage of Large Capitalization Equity Funds Outperformed by Index Ending 6/30/2002

	1 year	3 years	5 years	10 years
S&P 500 vs. Large Cap Equity Funds	63%	56%	70%	79%
Wilshire 5000 vs. Large Cap Equity Funds	72%	64%	69%	74%

Source: Lipper Analytic Services.
 Note: All large capitalization mutual funds in existence are covered with the exception of “sector” funds and funds investing in foreign securities.

Exhibit 6
Median Total Returns Ending 12/31/2001

	10 years	15 years	20 years
Large Cap Equity Funds	10.98%	11.95%	13.42%
S&P 500 Index	12.94%	13.74%	15.24%

Source: Lipper Analytic Services, Wilshire Associates, Standard & Poor’s and The Vanguard Group.

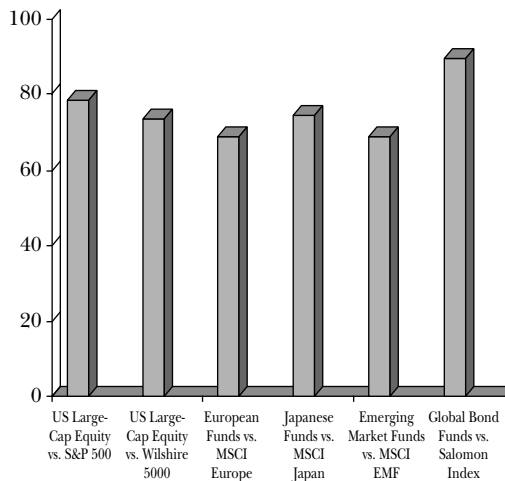
failed to beat the index. Similar results obtain for earlier decades. Exhibit 6 shows that the median large-capitalization professionally managed equity fund has underperformed the S&P 500 index by almost 2 percentage points over the past 10-, 15- and 20-year periods. Exhibit 7 shows similar results in different markets and against different benchmarks.

Managed funds are regularly outperformed by broad index funds, with equivalent risk. Moreover, those funds that produce excess returns in one period are not likely to do so in the next. There is no dependable persistence in performance. During the 1970s, the top 20 mutual funds enjoyed almost double the performance of the index. During the 1980s, those same funds underperformed the index. The best performing funds of the 1980s similarly underperformed during the 1990s. A more dramatic example of the lack of persistence in performance is shown in Exhibit 8. These mutual funds during 1998 and 1999 enjoyed three times the performance of the index. During 2000 and 2001, they did three times worse than the index. Over the long run, the results are even more devastating to active managers. One can count on the fingers of one hand the number of professional portfolio managers who have managed to beat the market by any significant amount. Exhibit 9 shows the distribution of return. Of the original 355 funds, only five of them outperformed the market by 2 percent per year or more.

The record of professionals does not suggest that sufficient predictability exists in the stock market or that there are recognizable and exploitable irrationalities sufficient to produce excess returns.

Exhibit 7

Percentage of Various Actively Managed Funds Outperformed by Benchmark Index 10 Years to 12/31/01



Source: Lipper Analytic Services and Micropal.

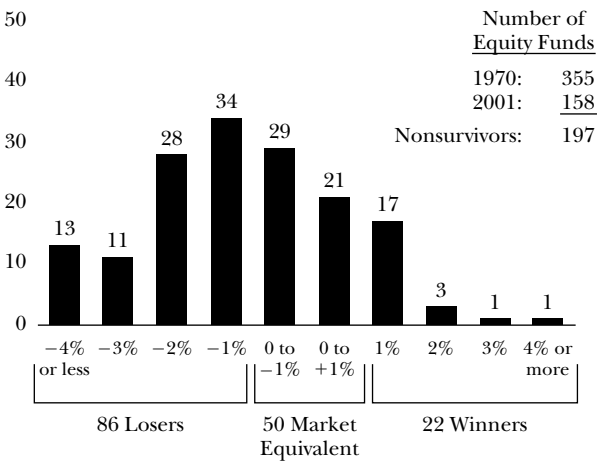
Exhibit 8

Getting Burned by Hot Funds

Fund Name	1998–1999		2000–2001	
	Rank	Average Annual Return	Rank	Average Annual Return
Van Wagoner:Emrg Growth	1	105.52	1106	−43.54
Rydex:OTC Fund;Inv	2	93.43	1103	−36.31
TCW Galileo:AGr Eq;Instl	3	92.78	1098	−34.00
RS Inv:Emerg Growth	4	90.19	1055	−26.17
PBHG:Large Cap 20	5	84.56	1078	−29.03
Janus Olympus Fund	6	77.24	1061	−27.03
Van Kampen Aggr Gro;A	7	76.70	1067	−28.04
Janus Mercury	8	76.31	1057	−26.35
PBHG:Sel Equity	9	76.21	1097	−33.19
WM:Growth;A	10	74.77	1046	−25.82
Berger New Generatn;Inv	11	73.31	1107	−45.96
Janus Enterprise	12	72.28	1101	−35.40
Janus Venture	13	72.22	1091	−30.89
Fidelity Aggr Growth	14	70.56	1105	−38.02
Janus Twenty	15	69.09	1090	−30.83
Amer Cent:New Oppty.	16	67.64	1033	−24.11
Morg Stan Sm Cap Gro;B	17	66.59	1102	−35.96
Van Kampen Emerg Gro;A	18	65.67	1021	−22.70
TCW Galileo:SC Gro;Instl.	19	64.87	1099	−34.77
BlackRock:MdCp Gro;Instl.	20	64.44	1009	−22.18
Average Fund Return		76.72		−31.52
S&P 500 Return		24.75		−10.50

Source: Lipper Analytic Services and Bogle Research Institute, Valley Forge, PA.

Exhibit 9
 The Odds of Success: Returns of Surviving Mutual Funds vs. S&P 500, 1970–2001



Source: Data from Lipper Analytic Services and The Vanguard Group.

Conclusion

As long as stock markets exist, the collective judgment of investors will sometimes make mistakes. Undoubtedly, some market participants are demonstrably less than rational. As a result, pricing irregularities and even predictable patterns in stock returns can appear over time and even persist for short periods. Moreover, the market cannot be perfectly efficient, or there would be no incentive for professionals to uncover the information that gets so quickly reflected in market prices, a point stressed by Grossman and Stiglitz (1980). Undoubtedly, with the passage of time and with the increasing sophistication of our databases and empirical techniques, we will document further apparent departures from efficiency and further patterns in the development of stock returns.

But I suspect that the end result will not be an abandonment of the belief of many in the profession that the stock market is remarkably efficient in its utilization of information. Periods such as 1999 where “bubbles” seem to have existed, at least in certain sectors of the market, are fortunately the exception rather than the rule. Moreover, whatever patterns or irrationalities in the pricing of individual stocks that have been discovered in a search of historical experience are unlikely to persist and will not provide investors with a method to obtain extraordinary returns. If any \$100 bills are lying around the stock exchanges of the world, they will not be there for long.

■ *I wish to thank J. Bradford De Long, Timothy Taylor and Michael Waldman for their extremely helpful observations. While they may not agree with all of the conclusions in this paper, they have strengthened my arguments in important ways.*

References

- Ariel, Robert A.** 1990. "High Stock Returns Before Holidays: Existence and Evidence on Possible Causes." *Journal of Finance*. December, 45:5, pp. 1611–626.
- Bagwell, Laurie Simon and John B. Shoven.** 1989. "Cash Distributions to Shareholders." *Journal of Economic Perspectives*. Summer, 3:3, pp. 129–40.
- Ball, Ray.** 1978. "Anomalies in Relationships Between Securities' Yields and Yield-Surrogates." *Journal of Financial Economics*. July, 6, pp. 103–26.
- Basu, Sanjoy.** 1983. "The Relationship Between Earnings' Yield, Market Value and the Returns for NYSE Common Stocks: Further Evidence." *Journal of Financial Economics*. June, 12:1, pp. 129–56.
- Campbell, John Y.** 1987. "Stock Returns and the Term Structure." *Journal of Financial Economics*. June, 18, pp. 373–400.
- Campbell, John Y. and Robert J. Shiller.** 1988. "Stock Prices, Earnings, and Expected Dividends." *Journal of Finance*. 43:3, pp. 661–76.
- Campbell, John Y. and Robert J. Shiller.** 1998. "Valuation Ratios and the Long-Run Stock Market Outlook." *Journal of Portfolio Management*. Winter, 24, pp. 11–26.
- Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay.** 1997. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.
- Cooper, Michael, Orlin Dimitrov and P.R. Rau.** 2001. "A Rose.com by Any Other Name." *Journal of Finance*. June 15, 56, pp. 2371–388.
- Cootner, Paul, ed.** 1964. *The Random Character of Stock Market Prices*. Cambridge, Mass.: MIT Press.
- DeBondt, Werner F. M. and Richard Thaler.** 1985. "Does the Stock Market Overreact?" *Journal of Finance*. July, 40, pp. 793–805.
- DeBondt, Werner F. M. and Richard Thaler.** 1995. "Financial Decision-Making in Markets and Firms: A Behavioral Perspective," in *Handbook in OR & MS, Volume 9*. R. Jarrow et al., eds. New York: Elsevier Science, Chapter 13.
- Fama, Eugene.** 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*. 25:2, pp. 383–417.
- Fama, Eugene.** 1998. "Market Efficiency, Long-Term Returns, and Behavioral Finance." *Journal of Financial Economics*. 49:3, pp. 283–306.
- Fama, Eugene and Kenneth French.** 1988. "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy*. 96:2, pp. 246–73.
- Fama, Eugene and Kenneth French.** 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance*. June, 47, pp. 427–65.
- Fama, Eugene and Kenneth French.** 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Finance*. 33:1, pp. 3–56.
- Fama, Eugene and Kenneth French.** 1997. "Value vs. Growth: The International Evidence." *Journal of Finance*. December, 53, pp. 1975–999.
- Fama, Eugene and Kenneth French.** 2001. "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay." *Journal of Financial Economics*. April, 60:1, pp. 3–43.
- Fama, Eugene and Kenneth French.** 2002. "The Equity Premium." *Journal of Finance*. 62:2, pp. 637–59.
- Fama, Eugene and G. William Schwert.** 1977. "Asset Returns and Inflation." *Journal of Financial Economics*. November, 5:2, pp. 55–69.
- Fluck, Zsuzsanna, Burton Malkiel and Richard Quandt.** 1997. "The Predictability of Stock Returns: A Cross-Sectional Simulation." *Review of Economics and Statistics*. 79:2, pp. 176–83.
- French, Kenneth.** 1980. "Stock Returns and the Weekend Effect." *Journal of Financial Economics*. March, 8, pp. 55–69.
- Graham, Benjamin and David L. Dodd.** 1934. *Security Analysis: Principles and Techniques*. New York: McGraw Hill.
- Graham, Benjamin and David L. Dodd.** 1965. *The Intelligent Investor*. New York: Harper & Row.
- Grossman, Sanford J. and Joseph E. Stiglitz.** 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review*. 70:3, pp. 393–408.
- Haugen, Robert A. and Josef Lakonishok.** 1988. *The Incredible January Effect*. Homewood: Dow Jones-Irwin.
- Hawawini, Gabriel and Donald B. Keim.** 1995. "On the Predictability of Common Stock Returns: Worldwide Evidence," in *Handbooks in Operations Research & Management Science, Volume 9*. R. Jarrow et al., eds. Amsterdam: Elsevier Science B. V., pp. 497–544.
- Jensen, Michael.** 1968. "The Performance of Mutual Funds in the Period 1945–64." *Journal of Finance*. May, 23, pp. 389–416.
- Kahneman, Daniel and Mark W. Riepe.** 1998. "Aspects of Investor Psychology." *Journal of Portfolio Management*. Summer, 24:4, pp. 52–65.
- Kahneman, Daniel and Amos Tversky.** 1973. "On the Psychology of Prediction." *Psychological Review*. 80, pp. 237–51.
- Kahneman, Daniel and Amos Tversky.** 1979.

"Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*. 47:2, pp. 263–91.

Keim, Donald B. 1983. "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence." *Journal of Financial Economics*. June, 12, pp. 13–25.

Keim, Donald B. and Robert T. Stambaugh. 1986. "Predicting Returns in Stock and Bond Markets." *Journal of Financial Economics*. December, 17, pp. 357–90.

Lakonishok, Josef and S. Smidt. 1988. "Are Seasonal Anomalies Real? A Ninety-Year Perspective." *Review of Financial Studies*. Winter, 1:4, pp. 403–25.

Lakonishok, Josef, Andrei Shleifer and Robert Vishny. 1994. "Contrarian Investment, Extrapolation, and Risk." *Journal of Finance*. December, 49, pp. 1541–578.

Lesmond, David, Michael Schill and Chunsheng Zhou. 2001. "The Illusory Nature of Momentum Profits." Unpublished manuscript, Tulane University.

Lo, Andrew W. and A. Craig MacKinlay. 1999. *A Non-Random Walk Down Wall Street*. Princeton: Princeton University Press.

Lo, Andrew W., Harry Mamaysky and Jiang Wang. 2000. "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation." *Journal of Finance*. August, 55:4, pp. 1705–765.

Malkiel, Burton G. 1973. *A Random Walk Down Wall Street*. New York: W. W. Norton & Co.

Malkiel, Burton G. 1995. "Returns from Investing in Equity Mutual Funds 1971 to 1991." *Journal of Finance*. June, 50:2, pp. 549–72.

Malkiel, Burton G. 2000. "Review of Robert J. Shiller's *Irrational Exuberance*." *Wall Street Journal*. April 4.

Miller, Merton. 1991. *Financial Innovations and Market Volatility*. Cambridge: Blackwell.

Nicholson, S. F. 1960. "Price-Earnings Ratios."

Financial Analysts Journal. July/August, 16, pp. 43–50.

Odean, Terrance. 1999. "Do Investors Trade Too Much?" *American Economic Review*. December, 89:5, pp. 1279–298.

Poterba, James and Lawrence Summers. 1988. "Mean Reversion in Stock Returns: Evidence and Implications." *Journal of Financial Economics*. 22:1, pp. 27–59.

Rasches, Michael. 2001. "Massively Confused Investors Making Conspicuously Ignorant Choices (MCI-MCIC)." *Journal of Finance*. 56:5, pp. 1911–927.

Roll, Richard and Robert J. Shiller. 1992. "Comments: Symposium on Volatility in U.S. and Japanese Stock Markets." *Journal of Applied Corporate Finance*. 5:1, pp. 25–29.

Ross, Stephen. Forthcoming. *Princeton Lectures in Finance 2001*. Princeton: Princeton University Press.

Samuelson, Paul. 1965. "Proof that Properly Anticipated Prices Fluctuate Randomly." *Industrial Management Review*. Spring, 6, pp. 41–49.

Schwert, G. William. 2001. "Anomalies and Market Efficiency," in *Handbook of the Economics of Finance*. G. Constantinides et al., eds. Amsterdam: North Holland, Chapter 17.

Shiller, Robert J. 1981. "Do Stock Prices Move so Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*. 71:3, pp. 421–36.

Shiller, Robert J. 1996. "Price-Earnings Ratios as Forecasts of Returns: The Stock Market Outlook in 1996." Unpublished manuscript, Yale University.

Shiller, Robert J. 2000. *Irrational Exuberance*. Princeton: Princeton University Press.

Shleifer, Andrei. 2000. *Inefficient Markets: An Introduction to Behavioral Finance*. New York: Oxford University Press.

Northwestern Utilities Ltd. v. Edmonton (City), 1929 CanLII 39 (SCC), [1929] SCR 186

Source:	Supreme Court of Canada
Date:	1929-02-05
Other citations:	[1929] 2 DLR 4 — [1929] SCJ No 3 (QL) — [1929] ACS no 3
Citation:	Northwestern Utilities Ltd. v. Edmonton (City), 1929 CanLII 39 (SCC), [1929] SCR 186, < https://canlii.ca/t/2f2nx >, retrieved on 2026-06-05

Supreme Court of Canada Northwestern Utilities Ltd. v. Edmonton (City), [1929] S.C.R. 186 Date: 1929-02-05

Northwestern Utilities, Limited Appellant;

and

The City of Edmonton and Board of Public Utility Commissioners of Alberta Respondents

The City of Edmonton Appellant;

and

Northwestern Utilities, Limited, and Board of Public Utility Commissioners of Alberta Respondents

1928: October 24; 1929: February 5.

Present:—Anglin C.J.C. and Mignault, Rinfret, Lamont and Smith JJ.

ON APPEAL FROM THE APPELLATE DIVISION OF THE SUPREME COURT OF ALBERTA

Public utilities—Public Utilities Act, Alta.—Hearings and investigations by Board of Public Utility Commissioners—Powers of Board—Obtaining of evidence—Absence of evidence—Order of Board fixing rates for gas supply in municipality by franchise holder—Return on investment—Inclusion in "rate base" of discount on sale of bonds—Appeal from Board's order—"Question of law."

[Page 186]

The Board of Public Utility Commissioners of Alberta made an order in 1922 fixing rates chargeable for gas proposed to be supplied in the city of Edmonton by the predecessor of the appellant company. The Board fixed the rates on the basis of an allowance of 10% as a fair return on the investment in the enterprise, and in determining the "rate base" (the amount to be considered as invested in the enterprise) it included as a capital expenditure a sum which was the discount on the sale of the company's bonds. The rates were to continue in force for three years from the date on which gas was first

[Page 187]

supplied. In 1926 the appellant company applied for continuation of the rates. On this application the city objected to such a high rate of return and to the inclusion in the rate base of the item for bond discount. The Board continued said item in the rate base, but reduced the return to 9% "in view of the elements which go to make up the rate base, and in view of the altered conditions of the money market." The parties appealed (by leave) to the Appellate Division, Alta., and then to this Court, the company against the reduction of the rate of return, and the city against the inclusion of the bond discount item in the rate base. The company contended that no evidence was adduced before the Board of "altered conditions of the money market," and that, without hearing evidence upon the point and giving the company opportunity to establish that the conditions of the money market had remained unaltered since 1922, the Board acted without jurisdiction in making the reduction. Under s. 47 of *The Public Utilities Act, 1923*, Alta., c. 53, as amended 1927, c. 39, an appeal lies from the Board upon a question "of jurisdiction" or "of law," upon leave obtained.

Held 1. The company's last mentioned contention involved a "question of law," and therefore it had a right to appeal.

2. The city's appeal failed; the question raised thereon was not one of jurisdiction or law.

3. The company's appeal failed. The Board had power to reduce the rate of return, notwithstanding that at the hearing before it no witnesses testified as to altered conditions of the money market. The company's contention that to alter the rate of return would be unfair to its shareholders who had invested in the enterprise after the order fixing the rates in 1922, was not a matter open for consideration upon the appeal, as it did not involve a question of jurisdiction or law.

Per Rinfret and Lamont JJ.: A consideration of ss. 21 (4) (5), 25, 43, and 44 of the said Act, the purposes of the Act, and the extent of the powers vested in the Board, leads to the conclusion that the intention of the legislature was to leave it largely to the Board's discretion to say in what manner it should obtain the information required for the proper exercise of its functions; it was not to be bound by the technical rules of legal evidence, but was to be governed by such rules as, in its discretion, it thought fit to adopt. An inference that it had not the proper evidence before it as to the altered conditions of the money market could not be drawn from the fact that no oral testimony in respect thereof was given at the hearing. The company had notice that a reduction was sought and that the city was attacking the methods and principles adopted in fixing the rate of return in 1922. This put the whole question of a fair return at large and informed the company that it would have to establish to the Board's satisfaction every element and condition necessary to justify a continuation of the 10% rate; and there was nothing in the record to justify the conclusion that the company had not the opportunity of making proof at the hearing as to the conditions of the money market.

Per Smith J.: The Board has power to reduce the rate of return without evidence; the question of a fair rate of return is largely one of opinion, hardly capable of being reduced to certainty by evidence, and appears to be one of the things entrusted by the statute to the judgment of the Board.

[Page 188]

APPEALS by Northwestern Utilities, Limited, and the City of Edmonton, respectively, from the dismissal by the Appellate Division of the Supreme Court of Alberta of their respective appeals from the award of the Board of Public Utility Commissioners for the Province of Alberta fixing rates to be paid by consumers of natural gas, for the supply of which within the city of Edmonton the said company, Northwestern Utilities, Limited, has a franchise.

The company applied to the Board for an order continuing the rates which had been fixed for a certain period by an order of the Board made in 1922. The Board made an award fixing the rates, from which each party appealed to the Appellate Division. Under s. 47 of *The Public Utilities Act* of Alberta, 1923, c. 53, as amended 1927, c. 39, an appeal lies from the Board to the Appellate Division "upon a question of jurisdiction or upon a question of law," if leave to appeal is obtained as therein provided. Such leave to appeal was obtained, it being reserved to each party to move before the Appellate Division to set aside the order granting leave to the other party, on the ground that the matters as to which leave to appeal was given did not involve any question of law or jurisdiction.

The company's objection to the Board's award was that it fixed the rates on the basis of an allowance of only 9%, instead of 10% which was allowed under the order made in 1922, as the "rate of return" on the investment in the enterprise. The Board in its award said:—

In view of the elements which go to make up the rate base, and in view of the altered conditions of the money market, the Board believes it is justified in reducing the rate of return that the company shall be allowed, to nine per cent., and the Board's estimates are on that basis.

The company contended that there was before the Board no evidence of any "altered conditions of the money market," that the "elements which go to make up the rate base" were the same as in 1922, and afforded no reason for changing the rate of return, that to reduce the rate of return would be unfair to its shareholders, who had invested in the enterprise after the order fixing the rates in 1922, that the money was invested and the plant constructed on the strength of the principles laid down in the 1922 award, and that it was clearly understood that the principles then adopted would govern all future revisions.

[Page 189]

The city's objection to the award was that, in determining the "rate base" (the amount to be considered as invested in the enterprise) it included (as it had done in the 1922 award) as a capital expenditure a sum which was the discount on the sale of the company's bonds.

The Appellate Division dismissed both appeals (no written reasons being given). Subsequently it made separate orders giving each party leave to appeal to the Supreme Court of Canada. On an application by both parties in the Supreme Court of Canada, the appeals were consolidated.

By the judgment of this Court both appeals were dismissed with costs.

E. Lafleur K.C. and H. R. Milner K.C. for Northwestern Utilities, Limited.
O. M. Biggar K.C. for the City of Edmonton.

The judgment of Anglin C.J.C. and Mignault J., was delivered by

ANGLIN C.J.C.—While, with my brother Smith, I incline to the view that the appellant company may have some reason to complain of unfairness in the judgment of the Board of Public Utility Commissioners reducing the rate of return from 10% to 9%, I agree with the conclusion reached by my brother Lamont and concurred in by my brother Smith that it is not open to us to entertain the appeal of the company on that ground. It does not seem to raise either a question of law or jurisdiction within the purview of the statute on which the right of appeal rests. I would dismiss the appeal.

The judgment of Rinfret and Lamont JJ. was delivered by

LAMONT J.—These are separate but consolidated appeals by the Northwestern Utilities, Limited (hereinafter called the Company) and the City of Edmonton, respectively, from the dismissal by the Appellate Division of the Supreme Court of Alberta of their respective appeals against the award made by the Board of Public Utility Commissioners on an application by the company for an

[Page 190]

order fixing the price to be paid by the consumers of natural gas within the city. Subsequent to the dismissal of the appeals, the Appellate Division made separate orders giving each party leave to appeal to this Court. By a further order the appeals were consolidated.

The company is the successor of the Northern Alberta Natural Gas Development Company, which held a franchise from the city for the supply of natural gas to the inhabitants thereof.

Disputes having arisen between the Development Company and the city, and an action having been commenced, the parties, on August 28, 1922, agreed to a settlement of their difficulties. One of the terms of the settlement was that the prices or rates to be paid by the inhabitants of the city should be fixed by the Board of Public Utility Commissioners. An application was accordingly made to the Board, the parties were heard, and, on November 27, 1922, an order was made fixing the rates to be paid. These rates were to continue in force for three years from the date on which gas was first supplied to consumers.

In order to fix just and reasonable rates, which it was the duty of the Board to fix, the Board had to consider certain elements which must always be taken into account in fixing

a rate which is fair and reasonable to the consumer and to the company. One of these is the rate base, by which is meant the amount which the Board considers the owner of the utility has invested in the enterprise and on which he is entitled to a fair return. Another is the percentage to be allowed as a fair return.

In the award of 1922, which came into operation in the fall of 1923, the Board included in the rate base as a capital expenditure the sum of \$283,900 (10% of the cost of plant) as, "an allowance for the promotion and financing" of the company, and the sum of \$650,000 which was the discount on the sale of the Development Company's bonds. It also determined that 10% was a fair return on the investment. The rates thus fixed by the Board, with certain alterations made with the consent of all parties, continued in force for three years. In October, 1926, the appellant company, which had succeeded to the rights of the Development Company, applied to the Board for an order continuing the rates for such period as the Board might see fit. In its

[Page 191]

reply to the application the city submitted (par. 23) that the order of November, 1922, should in certain respects be disregarded. One of these was the following:—

(e) Rate of Return. It is submitted that the methods and principles adopted in the fixing of the rate of return are erroneous and that the rate of return allowed is too high.

The city also protested against including in the rate base the item for the promotion and financing of the company and the item for bond discount.

In its answer to the city's reply the company alleged (par. 10) that at the hearing in 1922 the city was fully and adequately represented, that it had submitted evidence, that upon the award being delivered it raised no objection to any part thereof, and, therefore, was now estopped from contending that the principles then laid down were wrong in principle or in fact.

In its award the Board continued both the above mentioned sums in the rate base, but reduced the rate of return to the company from 10% to 9%. The reason assigned by the Board for this reduction is as follows:—

In view of the elements which go to make up the rate base, and in view of the altered conditions of the money market, the Board believes it is justified in reducing the rate of return that the Company shall be allowed, to nine per cent., (and the Board's estimates are on that basis.

From the award the parties appealed, first to the Appellate Division of the Supreme Court of Alberta, and now to this Court. The company appealed against the reduction of the rate of return on its capital expenditure to 9%. Referring to the reasons given by the Board for making the reduction the company in its factum says:—

1. The city adduced no evidence as to "altered conditions of the money market" and
2. "The elements which go to make up the rate base" in 1927 are the same as in 1922.

The city appealed against the inclusion in the rate base of the item of the bond discount above mentioned.

The *Public Utilities Act* allows an appeal from the Board only upon a question of jurisdiction, or upon a question of law, and even then only when leave to appeal has first been obtained from a judge of the Appellate Division.

As against the company's appeal the city raises the preliminary objection that no question either of jurisdiction or law is involved therein. In my opinion the objection cannot be sustained. The substance of the company's

[Page 192]

appeal is that the Board in making a reduction in the rate of return did so for two reasons, one of which was the "altered conditions of the money market," and that of this no evidence was adduced before the Board. The company contends that, without hearing evidence upon the point, and without giving it an opportunity to establish that the conditions of the money market had remained unaltered since 1922, the Board was without jurisdiction to make the reduction. This contention was not stated in this form in the order granting leave to appeal to the Appellate Division, but the fixing of the rate of return at 9% only, was there set out as an error of the Board in respect of which leave to appeal was granted.

Whether or not the Board can properly base an order (in part at least) on the existence of a state of fact of which no evidence was adduced before it at the hearing and as to which the party affected has not had any opportunity of being heard is, in my opinion, a question of law which depends for its answer upon the construction to be placed upon the *Public Utilities Act*.

I am, therefore, of opinion that the company had a right to appeal.

The question involved in this appeal is: Had the Board jurisdiction to find as a fact how the conditions of the money market had altered between November, 1922, and July, 1927, without any witness testifying at the hearing that an alteration had taken place.

As the Board was determining what would be a fair return on the capital invested by the company in the enterprise, and as it reduced the return from 10% to 9%, it can, I think, be taken that by "the altered conditions of the money market" the Board meant that the returns for money invested in securities in which moneys were ordinarily invested had decreased during the period in question. In other words, that the rate of interest

obtainable for moneys furnished for investment was, generally speaking, lower by a certain percentage in 1927 than it was in 1922. That, in my opinion, is all that is involved in the finding.

The duty of the Board was to fix fair and reasonable rates; rates which, under the circumstances, would be fair to the consumer on the one hand, and which, on the other

[Page 193]

hand, would secure to the company a fair return for the capital invested. By a fair return is meant that the company will be allowed as large a return on the capital invested in its enterprise (which will be net to the company) as it would receive if it were investing the same amount in other securities possessing an attractiveness, stability and certainty equal to that of the company's enterprise. In fixing this net return the Board should take into consideration the rate of interest which the company is obliged to pay upon its bonds as a result of having to sell them at a time when the rate of interest payable thereon exceeded that payable on bonds issued at the time of the hearing. To properly fix a fair return the Board must necessarily be informed of the rate of return which money would yield in other fields of investment. Having gone into the matter fully in 1922, and having fixed 10% as a fair return under the conditions then existing, all the Board needed to know, in order to fix a proper return in 1927, was whether or not the conditions of the money market had altered, and, if so, in what direction, and to what extent.

For the city it was argued that, as one of the statutory powers of the Board was to deal with the financial affairs of local authorities (s. 20 (d)), and as this included the power to authorize the issue of new debentures by these authorities and to determine the rate of interest to be paid thereon and also the power to order a variation of the rate of interest payable upon any debt of the local authority (s. 103), the Board must necessarily be familiar with the rate of interest prevailing from time to time and therefore did not require to have witnesses called to furnish it with information which in the regular performance of its duty it was obliged to possess. In view of the powers and duties of the Board under the Act there is, in my opinion, considerable to be said for the city's contention. It is not necessary, however, to determine this question, for in the statute itself I find sufficient to justify the conclusion that the intention of the Legislature was to leave it largely to the discretion of the Board to say in what manner it should obtain the information required for the proper exercise of its functions.

[Page 194]

The material provisions of the Act on this point are as follows:—

21. (4) The Board may in its discretion accept and act upon evidence by affidavit or written affirmation or by the report of any officer or engineer appointed by it or obtained in such other manner as it may decide.

(5) All hearings and investigations before the Board shall be governed by rules adopted by the Board, and in the conduct thereof the Board shall not be bound by the technical rules of legal evidence.

Section 25 provides that upon a complaint being made to the Board that any proprietor of a public utility has unlawfully done or unlawfully failed to do something relating to a matter over which the Board has jurisdiction, the Board shall "after hearing such evidence as it may think fit to require" make such order as it thinks fit under the circumstances. Section 43 provides that the Board may "appoint or direct any person to make an inquiry and report upon any application * * * before the Board." And by section 44 the Board may "review, rescind, change, alter or vary any decision or order made by it." A perusal of these statutory provisions and a consideration of the purposes of the Act and the extent of the powers vested in the Board leads me to the conclusion that the Legislature intended to create a Board which in the exercise of its functions should not be bound by the technical rules of legal evidence but which would be governed by such rules as, in its discretion, it thought fit to adopt (s. 21 (5)). We have not been made acquainted with the rules, if any, adopted by the Board to govern its investigations. Nor do we know what information it possessed as to the altered conditions of the money market; but, as it had authority to act on evidence "obtained in such manner as it may decide" (s. 21 (4)), an inference that it had not the proper evidence before it cannot be drawn from the fact that no oral testimony in respect thereof was given at the hearing. If, in this case, the Board had asked its secretary to inquire from the various financial institutions in Edmonton if there had been any alteration in the conditions of the money market between 1922 and 1927, and the secretary had reported that there had been a certain decrease in the returns from invested capital, would it have been necessary to call witnesses to verify the report? In my opinion it would not. Nor would it have been necessary to afford to either party an opportunity to controvert before the

[Page 195]

Board the information so obtained. Then would it have been necessary to mention in the award that the fact that such altered conditions had been established to the satisfaction of the Board by a report of its secretary? I can find nothing in the Act requiring mention to be made of the evidence or of the manner of obtaining it.

Reference was made to s. 86, which provides that no order involving any outlay, loss or depreciation to the proprietor of any public utility or to any municipality or person shall be made without due notice and full opportunity to all parties concerned to make proof to be heard at a public sitting of the Board, except in the case of urgency. A reduction in the rate of return to the company would, in my opinion, come within this section. The Board was, therefore, without jurisdiction to make the reduction unless the company had notice that a reduction was sought and had an opportunity of proving that under the circumstances existing at the time of the hearing the existing rate of return was

fair and reasonable. That the company had notice that the city was demanding a reduction is beyond question (par. 23 (e)). It had more. It had notice that the city was attacking the methods and principles adopted in fixing the rate of return in 1922. This, in my opinion, put the whole question of a fair return at large and informed the company that it would have to establish to the satisfaction of the Board every element and condition necessary to justify a continuation of the 10% rate. The company does not say that it was refused an opportunity of putting in evidence as to the conditions of the money market. Nowhere does it deny that it could have put in evidence had it so desired. What it does say is that the city did not adduce evidence on the point and that no witnesses were called to testify before the Board in regard thereto. There is nothing before us to justify an inference that the company was not at liberty to call witnesses as to the conditions of the money market had it so desired. Moreover, in the order which the company obtained giving it leave to appeal it did not even suggest that it had no opportunity of submitting evidence as to the existing market conditions. The ground upon which the company relied to meet the city's demand for a reduction, as set out in the answer which it filed, was that as the city had ac-

[Page 196]

cepted the award when it was delivered and had raised no objection thereto, it was now precluded from seeking to set aside the principles upon which the rate of return was based. In its factum it went further and contended that, even if there was no estoppel, the principles then adopted should now be adhered to because it was on the strength of their having been adopted that the shareholders of the company invested their money in the enterprise. This contention cannot be made effective. In the first place, it involves neither a question of jurisdiction nor of law. In the second place, it is the duty of the Board to fix rates which, in its opinion, will be fair and reasonable at the time the order is made and for the period for which they are fixed. If any wrong principle or erroneous view has been adopted it is the duty of the Board at the next revision to correct the error. The argument that it would be unfair to the shareholders now to alter the rate of return is not a matter open for consideration on appeal. Moreover, when these shareholders invested their money they knew that the rates fixed were to be in force for three years only and that it would be the duty of the Board on the next revision to fix rates which at that time would be fair and reasonable under the circumstances then existing.

Our attention was also called to s. 47 (1a) as indicating an intention that evidence must be taken on all material points. That subsection reads as follows:—

(1a) On the hearing of any appeal referred to in subsection 1 of this section no evidence other than the evidence which was submitted to the Board upon the making of the order appealed from shall be admitted, and the Court shall proceed either to confirm or vacate the order appealed from, and in the latter event shall refer the matter back to the Board for further consideration and redetermination.

In my opinion this subsection means no more than that no new evidence is to be admitted on appeal.

The appeal of the company should therefore be dismissed with costs.

The appeal of the city should likewise be dismissed with costs. The items which should be included in the rate base cannot, in my opinion, be considered a question of jurisdiction or of law.

SMITH J.—The City of Edmonton had made an agreement with the Northern Alberta Natural Gas Development Company, by which the company obtained a fran-

[Page 197]

chise to supply natural gas to the city, and agreed to construct the necessary works. The company failed to construct the works, and the city sued for damages for breach of contract. The actions were settled by an agreement dated 22nd August, 1922, under which the determination of the rates to be charged by the company for gas was referred to the Board of Public Utility Commissioners, and the company was, within six months after the fixing of the rates, to deposit \$50,000 with the city, which was to be forfeited to the city as liquidated damages in case the company did not complete the construction of the works as agreed.

A rate hearing was held by the Board after this settlement, at which the company and the city were represented, and the Board made an award, setting out a rate basis and fixing prices for gas on this basis.

The difficulty about proceeding with the works had been the procuring of capital on the basis of prices provided in the original agreement and amendments made. The whole object of fixing a rate base and prices in advance of construction was to facilitate financing by the company. It would necessarily be on the basis of the award that investors would buy bonds and stock of the company. The company had the option of proceeding with the works or abandoning them and forfeiting the \$50,000, after seeing the award. In July following the making of the award, the company assigned its franchise and property to the appellant, the Northwestern Utilities, Limited, which, by sale of its bonds and stock, raised the necessary capital, constructed the works, and put them in operation. The rate to be charged for gas was fixed by the award for three years, and at the end of this period the company applied to the Board for continuation of the rates fixed by the award. The rate base fixed by the Board in the award of 1922 contained many items, such as total investment, operating cost, depletion reserve, reserve for repayment of cost of plant, total necessary revenue, amounts of gas to be sold, and the rate of return on capital to be allowed. It is evident that, with the exception of the last of these items, the amounts fixed must have been estimates, liable to be varied by actual results.

[Page 198]

The rate of return to be allowed on capital was fixed in the award at 10%, not based on the ordinary rate of money on the market at the time or on an estimated future rate, but on consideration of the rate that would induce investors to risk their capital in an extremely hazardous and doubtful venture. At the hearing before the Board in 1922, the company had asked a 12% rate of return on capital, and the city had conceded 10%, which the Board fixed, though it stated that under the circumstances a return of more than 10% would not seem to be unjust. The reason set out for not fixing this higher rate was that it might so restrict the market that the higher rate would not compensate for the restriction of the market, and would therefore not be to the advantage of the company. It is, however, stated that in case of future revision, it may be found desirable, under certain circumstances, to increase this rate.

On the revision at the end of three years, this rate was not increased, but was reduced from 10% to 9%, at the instance of the city, and this reduction constitutes the ground of appeal.

In the reasons given by the Board in fixing the new rates, it is pointed out that, where rates have been fixed in advance of construction and financing, the Board is not precluded from subsequently making changes that may appear from subsequent reconsideration to be necessary, and it is then stated that

those investing in such a case must depend on the fairness of the Board in seeing that the Company is allowed a fair and reasonable return upon its investment, but the Board may, and indeed it should, take into consideration the circumstances under which such investment was made.

In discussing these circumstances in reference to a request by the city for elimination from the rate base of the 1922 award of the item for bond discount, the Board says:

There is, moreover, an additional factor to be considered in the present case and that is, that in 1922 the inclusion of the allowance for bond discount was practically agreed to by the city in its case and the item was not questioned by the city until at the recent hearing. It is only fair to assume that the fact of the inclusion of the bond discount in the rate base formed part of the inducement for the making of the investment. Under the circumstances, therefore, the Board does not feel justified in adopting the City's contention in this regard.

This lays down a principle with which one heartily agrees, and which applies exactly to the city's application for reduction of the rate of return on capital fixed in the award

[Page 199]

of 1922 at 10%. The Board fixed this rate with the assent of the city, and this rate, coupled with the suggestion by the Board that it might be increased, "formed part of the inducement for the making of the investment."

The altered condition of the money market, given as a reason for the reduction of the rate to 9%, seems to me to have no bearing on the matter. The representation to the investor in 1922 was, for the risk you take in placing your capital in a hazardous undertaking, you will be allowed as a basis in fixing rates to be charged for gas a return of 10%. What the regular money market might be three years later could have nothing to do with the decision to invest. The whole question was, viewing the risk, and the chances, as matters then stood, was the chance of 10% on the money worth the risk of a bad investment, with the possibility of the loss of all or part of the capital?

The Board then, in my opinion, laid down a proper principle, and applied it in other instances, but failed to apply it to this item, as to which I think it was particularly applicable. The question is, can this Court set aside the finding of the Board as to this item on the appeal? I agree with my brother Lamont that, whether or not under the Act the Board was entitled to reduce the rate to 9% without evidence, because of a change in money market conditions, is a question of law, and that there is therefore a right of appeal, and it is with some regret that I feel bound to agree with him that the Board had jurisdiction to make the change in rate without evidence, and without giving the company an opportunity to offer evidence. The question of a fair rate of return on a risky investment is largely a matter of opinion, and is hardly capable of being reduced to certainty by evidence, and appears to be one of the things entrusted by the statute to the judgment of the Board.

I am not entirely in accord with the observations of my brother Lamont in reference to the sending out of someone to gather evidence of the state of the money market and acting on that party's report without the knowledge of the company. The objection in such a case would not be the failure to set out in the award the fact of such evidence and its nature, but the failure to disclose it to the company with an opportunity to answer it. If it were a case where, evi-

[Page 200]

dence toeing necessary, it had been taken in the manner suggested, or otherwise, and a finding based on it without disclosure of it to the company and an opportunity to answer it, I would regard such a proceeding as contrary to elementary principles of justice, and as affording, under the statute, a ground for setting the award as to this item aside and referring it back for reconsideration. It does not, however, appear that any evidence was taken, and as stated, I have concluded that there was power to make the change without evidence.

I therefore concur with my brother Lamont in the disposal of this appeal.

Appeals dismissed with costs.

© Copyright 2026, vLex. All Rights Reserved.

Copy for use in the context of the business of the vLex customer only. Otherwise, distribution or reproduction is not permitted

TransCanada Pipelines Ltd. v. National Energy Board et al., (2004) 319 N.R. 171 (FCA)

Jurisdiction:	Federal Jurisdiction (Canada)
Court:	Court of Appeal (Canada)
Judge:	Rothstein, Noël and Sharlow, JJ.A.
Citation:	(2004), 319 N.R. 171 (FCA), 2004 FCA 149
Date:	16 February 2004

vLex Document Id: VLEX-681540865

Link: <https://app.vlex.com/vid/transcan-pipelines-v-nat-681540865>

TransCan. Pipelines v. Nat. Energy Bd. (2004), 319 N.R. 171 (FCA)

MLB headnote and full text

Temp. Cite: [2004] N.R. TBEd. AP.040

TransCanada Pipelines Limited (appellant) v. The National Energy Board,
Canadian Association of Petroleum Producers, Centra Gas Manitoba Inc.,
Coral Energy Canada Inc., Industrial Gas Users Association, Mirant
Canada Energy Marketing, Ltd., and Ontario Minister of Energy
(respondents)

(A-327-03; 2004 FCA 149)

**Indexed As: TransCanada Pipelines Ltd. v. National Energy Board
et al.**

Federal Court of Appeal
Rothstein, Noël and Sharlow, JJ.A.
February 16, 2004.

Summary:

TransCanada Pipelines obtained leave to appeal an order of the National Energy Board. The first issue was whether the Board erred in taking customer or consumer interests into account in determining the rate of return on capital it would allow the appellant's Canadian Mainline natural gas transmission system to earn. The second issue was whether the Board erred by fettering its discretion by refusing to depart from an automatic adjustment mechanism it had used to establish the Mainline's rate of return on equity.

The Federal Court of Appeal dismissed the appeal. TransCanada failed to show that the Board took customer or consumer interests into account. The Board did not fetter its discretion.

Administrative Law - Topic 2088

Natural justice - Constitution of board or tribunal - Bias - Apprehension of - Tolls charged by TransCanada Pipelines for transporting natural gas on its pipeline were regulated by the National Energy Board on a cost of service basis - The costs of service included the pipeline's cost of capital - The rate of return on equity was one component of cost of capital - The rate of return on equity had to be estimated by the Board - The Board established a mechanism to automatically adjust the rate of return on equity (the 1995 decision) - In 2001, TransCanada applied for a review of the 1995 decision - The Board rejected TransCanada's proposed new methodology for determining cost of capital and determined that it should continue to use the 1995 automatic adjustment mechanism - TransCanada argued that the Board fettered its discretion - The Federal Court of Appeal held that an affected party wishing to change the process had the onus to demonstrate that its proposal was preferable to the one that was the subject of a Board order - That was not an improper onus - It did not reflect a fettering of discretion by the Board - It did not give rise to any apprehension of bias by the Board - See paragraphs 44 to 57.

Administrative Law - Topic 2145

Natural justice - Administrative decisions or findings - Failure to consider evidence - Tolls charged by TransCanada Pipelines for transporting natural gas on its pipeline were regulated by the National Energy Board on a cost of service basis - The costs of service included the pipeline's cost of capital - The rate of return on equity was one component of cost of capital - The rate of return on equity had to be estimated by the Board - The Federal Court of Appeal stated that in cost of capital proceedings, the Board could, based on the evidence and its own judgment, choose a methodology for determining cost of capital and estimate the cost of capital for a forthcoming year - The Board's estimate might not reflect the precise estimates of the parties or the witnesses - As long as the Board's estimate was within the range of estimates put forward and the Board demonstrated that it considered those estimates, it could not be said to have ignored evidence - Even if the Board's estimate was outside that range, if the Board showed that it considered the evidence and provided adequate reasons for its opinion, the Board would not be found to have ignored evidence - See paragraph 58.

Administrative Law - Topic 8264

Administrative powers - Discretionary powers - Fettering of discretion - [See **Administrative Law - Topic 2088**].

Mines and Minerals - Topic 8024

Oil and gas - Regulation - National Energy Board - Hearings - [See **Administrative Law - Topic 2088**].

Mines and Minerals - Topic 8548

Oil and gas - Pipelines - Construction and operation of - Rates and charges - Tolls charged by TransCanada Pipelines for transporting natural gas on its pipeline were regulated by the National Energy Board on a cost of service basis - The costs of service included the pipeline's cost of capital - The rate of return on equity was one component of cost of capital - The rate of return on equity had to be estimated by the Board - The Federal Court of Appeal agreed that the impact on customers or consumers could not be a factor in determining the rate of return on equity - However, any resulting increase in tolls might be a relevant factor for the Board to consider in determining the way in

which a utility should recover its costs - Therefore, increases could be phased in over a period of time, provided that there was no economic loss to the utility in the process -See paragraphs 26 to 43.

Mines and Minerals - Topic 8548

Oil and gas - Pipelines - Construction and operation of - Rates and charges - [See **Administrative Law - Topic 2088** and **Administrative Law - Topic 2145**].

Cases Noticed:

British Columbia Hydro & Power Authority v. Westcoast Transmission Co. et al., [1981] 2 F.C. 646; 36 N.R. 33 (F.C.A.), refd to. [para. 30].

Northwestern Utilities Ltd. v. Edmonton (City), [1929] S.C.R. 186, refd to. [para. 33].

Hemlock Valley Electrical Services Ltd. v. Utilities Commission (B.C.) et al. (1992), 12 B.C.A.C. 1; 23 W.A.C. 1; 66 B.C.L.R.(2d) 1 (C.A.), refd to. [para. 43].

Counsel:

Alan J. Lenczner, Risa M. Kirshblum and Wendy Moreland, for the appellant;

Margery Fowke, for the respondent, National Energy Board;

John J. Marshall, Q.C., and Don Davies, for the respondent, Canadian Association of Petroleum Producers;

Alan Mark, for the respondent, Coral Energy Inc.;

Peter C.P. Thompson, Q.C., and Vincent J. DeRose, for the respondent, Industrial Gas Users Association;

Keith F. Miller, for the respondent, Mirant Energy Marketing Canada Inc.;

John Turcun and Sara Blake, for the respondent, Ontario Minister of Energy.

Solicitors of Record:

Lenczner Slaght Royce, Toronto, Ontario, for the appellant;

McLeod Dixon LLP, Calgary, Alberta, for the respondent, Canadian Association of Petroleum Producers;

Burnet Duckworth & Palmer LLP, Calgary, Alberta, for the respondent, Mirant Energy Marketing Canada Inc.;

Ontario Ministry of the Attorney General, Toronto, Ontario, for the respondent, Ontario Minister of Energy;

National Energy Board, Calgary, Alberta, for the respondent, National Energy Board;

Ogilvy Renault, Toronto, Ontario, for the respondent, Coral Energy Inc.;

Manitoba Hydro Law Department, Winnipeg, Manitoba, for the respondent, Centra Gas Manitoba Inc.;

Borden Ladner Gervais LLP, Ottawa, Ontario, for the respondent, Industrial Gas Users Association.

This appeal was heard at Toronto, Ontario, on February 16, 2004, by Rothstein, Noël and Sharlow, J.J.A., of the Federal Court of Appeal. Rothstein, J.A., delivered the following decision for the court on April 5, 2004.

Introduction

[1] Rothstein, J.A. : This is an appeal from a February 2003 decision of the National Energy Board (RH-R-1-2002), pursuant to leave granted by this Court under section 22 of the National Energy Board Act , R.S.C. 1985, c. -7.

[2] There are two issues in the appeal. The first is whether the National Energy Board ("Board") erred in taking customer or consumer interests into account in determining the rate of return on capital it would allow the appellant's Canadian Mainline natural gas transmission system ("the Mainline") to earn. The second is whether the Board erred by fettering its discretion by refusing to depart from an automatic adjustment mechanism it had used to establish the Mainline's rate of return on equity.

[3] In order to understand the issues under appeal, it is first necessary to provide some background and the procedural history leading to the February 2003 decision.

Background

[4] The National Energy Board regulates interprovincial natural gas transmission pipelines. The Mainline is considered a Group 1 pipeline by the Board. Group 1 pipelines are major pipelines which are audited by the Board on a regular basis and whose operating results are continuously monitored by the Board.

[5] The tolls charged for transporting natural gas on the Mainline are regulated by the Board on a cost of service basis. That means that for a future period, referred to as a "test" year, the Board, based on the evidence before it, estimates the costs to be incurred by the Mainline. The tolls which the Board allows the Mainline to charge its customers are designed to generate sufficient revenue to recover these approved costs while at the same time fairly allocating charges to users in relation to the costs and benefits of dif-

ferent services. Included in the cost of service, and indeed, the largest single component of the Mainline's costs, is the Mainline's cost of capital.

[6] The cost of capital to a utility is equivalent to the aggregate return on investment investors require in order to keep their capital invested in the utility and to invest new capital in the utility. That return will be made in the form of interest on debt and dividends and capital appreciation on equity. Usually, that return is expressed as the rate of return investors require on their debt or equity investments.

[7] The rate of return on debt is not usually controversial. It normally consists of the weighted average interest rate for the test year on the utility's outstanding long-term debt. On the other hand, the rate of return on equity is often the subject of controversy and of much debate by expert witnesses.

[8] Unlike debt, where the interest rate payable is directly observable, the rate of return on equity cannot be accurately determined in advance. There are various methods experts use to estimate the rate of return on equity required by investors. The one adopted by the Board is an Equity Risk Premium methodology whereby the Board estimates a risk-free rate based on government bond rates and adds a risk premium to account for the risk associated with equity investment in a "benchmark" pipeline.

[9] Once the separate rates of return on debt and equity are established, they are consolidated into a composite rate of return on capital, based on the relative amounts of debt and equity in the utility's capital structure. In order to account for varying levels of risk between pipelines, the Board constructs for each pipeline a capital structure, i.e. the relative portions of debt and equity capital needed to finance its prudently acquired assets plus its working capital, on the basis of expert evidence. The greater the risk attributed to each pipeline, the greater the required equity component of its capital structure. That is because bond investors, who are more risk averse than equity investors, will not lend funds to an enterprise unless there is sufficient equity capital invested in the enterprise to give them confidence that they will be able to recover their investment from the assets of the enterprise in the event of default.

[10] For example, if the required rate of return on debt is 5%, the required rate of return on equity is 10% and the utility's capital structure, as determined by the Board, consists of 60% debt and 40% equity, the composite rate of return on capital would be $5\% \times 0.60 + 10\% \times 0.40 = 7\%$.

[11] The composite rate of return on capital is then multiplied by a rate base which consists of the Board's determination, according to its accounting regulations, of the net book value of the utility's prudently acquired assets plus its working capital. Multiplying the rate of return required by investors by this rate base gives the total dollar amount of return required by investors. The product is equivalent to the utility's estimated cost of capital for the test year. That cost is added to all other costs to get the utility's total cost of service. The total is then allocated amongst the utility's customers.

[12] Even though cost of capital may be more difficult to estimate than some other costs, it is a real cost that the utility must be able to recover through its revenues. If the Board does not permit the utility to recover its cost of capital, the utility will be unable to raise new capital or engage in refinancing as it will be unable to offer investors the same rate of return as other investments of similar risk. As well, existing shareholders will insist that retained earnings not be reinvested in the utility.

[13] In the long run, unless a regulated enterprise is allowed to earn its cost of capital, both debt and equity, it will be unable to expand its operations or even maintain existing ones. Eventually, it will go out of business. This will harm not only its shareholders, but also the customers it will no longer be able to service. The impact on customers and ultimately consumers will be even more significant where there is insufficient competition in the market to provide adequate alternative service.

Procedural History

[14] In 1994, the Board conducted a public hearing into the cost of capital of certain Group 1 pipelines including the Mainline. The purpose of the hearing was to fix the cost of capital for those pipelines for the period commencing January 1, 1995, and to establish, if possible, an automatic mechanism to adjust the rate of return on equity in the future in order to avoid the expense of litigating annual or biennial changes to the rate of return on equity.

[15] As a result of that proceeding, the Board issued reasons for decision (RH-2-94) in March 1995 fixing the Mainline's return on equity for the 1995 test year at 12.25% based on a deemed capital structure of 70% debt and 30% equity. The Board's deemed capital structure did not provide for any explicit preferred share capital. Therefore, all references to equity refer to common equity.

[16] The Board also established an adjustment mechanism by which the rate of return on equity would be adjusted on January 1 in 1996 and each subsequent calendar year. This mechanism was based upon the Equity Risk Premium methodology whereby:

1. a risk free (Government of Canada) bond yield forecast would be forecasted for the forthcoming year;
2. this bond yield forecast would be deducted from the bond yield forecast of the immediately preceding year;
3. this difference would be multiplied by a factor of 0.75 to determine the adjustment to the rate of return on equity;
4. the product derived in step 3 would be added to or deducted from the rate of return on equity determined by the Board for the preceding year;
5. the sum resulting from step 4 would be rounded to the nearest 25 basis points (1/100th of a percent).

[17] The Mainline's rate of return on equity was adjusted according to this formula in 1996 and subsequent years, although in 1997, the Board abandoned the rounding adjustment, i.e. step 5 above.

[18] By 2001, the appellant had concluded that application of the formula was understating its required rate of return on capital. Therefore, the appellant applied, pursuant to subsection 21(1) of the National Energy Board Act, for "review and variance of the [1995 decision] to allow for the determination of a fair return for TransCanada for the years 2001 and 2002." Subsection 21(1) provides:

"21(1) Subject to subsection (2), the Board may review, vary or rescind any decision or order made by it or rehear any application before deciding it."

"21(1) Sous réserve du paragraphe (2), l'Office peut réviser, annuler ou modifier ses ordonnances ou décisions, ou procéder à une nouvelle audition avant de statuer sur une demande."

[19] The appellant submitted that the Board should approve a new methodology for determining the Mainline's cost of capital - the After-Tax Weighted-Average Cost of Capital (ATWACC) methodology. Alternatively, if the ATWACC methodology was not accepted, the appellant submitted that the required rate of return on equity for the Mainline should be 12.5% for 2001 and 2002 and that based on its risk, the deemed equity component of the Mainline's capital structure should be increased to 40%.

[20] As a result of the appellant's submissions, the Board conducted a hearing in February, March and April 2002. The issues at the hearing were:

"1. Is the Rate of Return on Common Equity (ROE) formula, established by the Board in its RH-2-94 Decision, still appropriate for determining TransCanada's ROE?

"2. Is the After Tax Weighted-Average Cost of Capital (ATWACC) methodology an appropriate regulatory approach to determining cost of capital?

"3. In the event the Board decides to adopt the ATWACC methodology, what is the appropriate ATWACC for TransCanada?

"4. In the event the Board declines to adopt the ATWACC methodology and it is determined that the ROE formula is no longer suitable:

a) What would be an appropriate methodology for determining return on capital and capital structure for TransCanada?

b) In applying the above-determined methodology, what would be an appropriate return on capital and capital structure for TransCanada?

"5. What is the appropriate effective date for changes to TransCanada's cost of capital? (RH-4-2001 at 4)."

[21] By reasons for decision (RH-4-2001) dated June 2002, the Board:

1. rejected the appellant's ATWACC proposal;
2. determined that the rate of return on equity for the Mainline should continue to be based on the adjustment formula established in its 1995 decision; and
3. increased the deemed equity component of the Mainline's capital structure from 30% to 33% to account for increased business risk.

[22] By application to the Board dated September 16, 2002, the appellant applied for a review and variance of the 2002 decision. This application was also made pursuant to subsection 21(1).

[23] Section 44 of the National Energy Board Rules of Practice and Procedure, 1995, SOR/95-208 sets out the requirements for a review application. Subsection 44(2) provides:

"44(2) An application for review or rehearing shall contain

.....

(b) the grounds that the applicant considers sufficient, in the case of an application for review, to raise a doubt as to the correctness of the decision or order ... including

(i) any error of law or of jurisdiction,"

"(2) La demande de révision ou de nouvelle audition contient les éléments suivants:

.....

b) les motifs que le demandeur juge suffisants pour mettre en doute le bien-fondé de la décision ou de l'ordonnance, s'il s'agit d'une demande de révision, ... notamment:

(i) une erreur de droit ou de compétence,"

[24] In its decision on the review & variance application (RH-R-1-2002), dated February 2003, the Board found that the appellant had not raised a doubt as to the correctness of its 2002 decision and dismissed the application for review and variance.

[25] The appellant was granted leave to appeal the Board's 2003 decision to this Court.

Analysis

1. Standard of Review and Approach to the Decision Being Appealed

[26] In view of my conclusion that the appeal should be dismissed, it is not necessary to conduct an extensive standard of review analysis. Even on the most intrusive standard of review (correctness), it has not been demonstrated that the Board erred in law.

[27] There is also a question of the extent to which the Court should consider the Board's 2002 decision, which itself was not appealed. Normally, the Court is to restrict itself to a consideration of the decision under appeal. However, when the question is whether the Board erred or came to an unreasonable or patently unreasonable result in finding in its 2003 decision that the appellant had not raised a doubt as to the correctness of the prior 2002 decision, it is necessary to have regard, at least to some extent, to that prior decision. Rather than becoming bogged down into the intricacies of the scope of the Court's review, I am satisfied, even on a unrestricted consideration of both the 2002 and 2003 decisions, that the Board made no error of law in either case.

2. Did the Board err in considering customer or consumer interests in determining the Mainline's rate of return on capital?

[28] As a preliminary point, the appellant drew a distinction between its customers and the ultimate consumers. For purposes of this decision, such a distinction is immaterial. The appellant's position is that the Mainline's return on capital should be determined solely from the perspective of the Mainline, without considering other interests, whether they be direct customers or ultimate consumers.

a) The Board is not required to adopt any specific methodology in determining tolls .

[29] The National Energy Board Act contains no provisions or directions which require the Board to determine a pipeline's rate of return on capital. The Act only requires that "all tolls be just and reasonable." Subsections 60(1) and section 62 provide:

"60(1) A company shall not charge any tolls except tolls that are

(a) specified in a tariff that has been filed with the Board and is in effect; or

(b) approved by an order of the Board.

"62. All tolls shall be just and reasonable, and shall always, under substantially similar circumstances and conditions with respect to all traffic of the same description carried over the same route, be charged equally to all persons at the same rate."

"60(1) Les seuls droits qu'une compagnie peut imposer sont ceux qui sont:

- a) soit spécifiés dans un tarif produit auprès de l'Office et en vigueur;
- b) soit approuvés par ordonnance de l'Office.

"62. Tous les droits doivent être justes et raisonnables et, dans des circonstances et conditions essentiellement similaires, être exigés de tous, au même taux, pour tous les transports de même nature sur le même parcours."

[30] The authority of the Board to determine just and reasonable tolls is not limited by any statutory directions. The broad authority of the Board was well articulated by Thurlow, C.J., in *British Columbia Hydro & Power Authority v. Westcoast Transmission Co. et al.*, [1981] 2 F.C. 646; 36 N.R. 33, at 655-56 (F.C.A.):

"There are no like provisions in part IV of the National Energy Board Act . Under it, tolls are to be just and reasonable and may be charged only as specified in a tariff that has been filed with the Board and is in effect. The Board is given authority in the broadest of terms to make orders with respect to all matters relating to them. Plainly, the Board has authority to make orders designed to ensure that the tolls to be charged by a pipeline company will be just and reasonable. But its power in that respect is not trammelled or fettered by statutory rules or directions as to how that function is to be carried out or how the purpose is to be achieved. In particular, there are no statutory directions that, in considering whether tolls that a pipeline company propose to charge are just and reasonable, the Board must adopt any particular accounting approach or device or that it must do so by determining cost of service and a rate base and fixing a fair return thereon."

[31] The Board has adopted a cost of service method for determining the Mainline's tolls. Before this Court, counsel for a number of the respondents suggested different methodologies for determining just and reasonable tolls that would be open to the Board, such as:

- 1. tolls based on agreements between pipelines and shippers;
- 2. tolls based on charges of other pipelines;
- 3. use of base year tolls adjusted for inflation;
- 4. tolls based on mechanisms to encourage utilities towards greater efficiency.

As no particular methodology is required by the National Energy Board Act , the Board could have adopted a different methodology for determining just and reasonable tolls for the Mainline.

b) Having adopted a cost of service methodology, the costs determined by the Board must be just and reasonable to both the Mainline and its users .

[32] In the case of the Mainline, the Board has adopted a cost of service methodology whereby the Mainline is to be compensated through tolls for its prudently incurred costs, including its cost of capital, and in particular, its cost of equity capital. Once it did so, it had to faithfully determine the Mainline's costs based on the evidence and its own sound judgment.

[33] Cost of equity for a future year cannot be directly measured and therefore must be based on estimates. The Board must choose an estimate that allows the Mainline to earn what has been termed a "fair return." In *Northwestern Utilities Ltd. v. Edmonton (City)* , [1929] S.C.R. 186, at 192-93, the Supreme Court defined a fair return in the following terms:

"The duty of the Board was to fix fair and reasonable rates; rates which, under the circumstances, would be fair to the consumer on the one hand, and which, on the other hand, would secure to the company a fair return for the capital invested. By a fair return is meant that the company will be allowed as large a return on the capital invested in its enterprise (which will be net to the company) as it would receive if it were investing the same amount in other securities possessing an attractiveness, stability and certainty equal to that of the company's enterprise."

Tolls which reflect a fair return on capital will be just and reasonable to both the Mainline and its users.

[34] To put the matter another way, when the cost of service methodology is used to determine just and reasonable tolls, if the Board does not permit the Mainline to recover its costs because it has understated the Mainline's cost of equity capital, the Mainline will be unable to earn a fair return on equity. The tolls will therefore not be just and reasonable from the Mainline's point of view. On the other hand, the tolls

must also be just and reasonable from the point of view of the Mainline's customers and the ultimate consumers who rely on service from the Mainline. Therefore, customers and consumers have an interest in ensuring that the Mainline's costs are not overstated. As respondents' counsel pointed out, there are numerous costing issues that may be subject to challenge. Questions may arise about, among other things, the allocation of costs between the Mainline and other divisions of the appellant; whether costs have been, or are being, prudently incurred; and whether the Mainline's compensation plans are reasonable. And, specific to this appeal, customers and consumers have an interest in ensuring that the Mainline's cost of equity is not overstated.

c) The Board did not improperly consider the impact on customers or consumers of increasing tolls to reflect the appellant's costs .

[35] In oral argument, the appellant conceded that it does not object to its customers having input into the Board's cost determinations and in particular, its cost of capital determination, provided the issues in dispute are restricted to the costs of the Mainline. However, the appellant does object to the Board taking the impact of tolls on customers and consumers into account in determining the Mainline's cost of equity capital. The appellant says that the required rate of return on equity must be determined solely on the basis of the Mainline's cost of equity capital. The impact of any resulting toll increases on customers or consumers is an irrelevant consideration in that determination. The appellant does concede that when the final tolls are being fixed, the impact on the customers and consumers may be relevant, but insists that it is irrelevant when determining the required return on equity.

[36] I think that this argument is sound and in keeping with the decision of the Supreme Court in *Northwestern Utilities* . The cost of equity capital does not change because allowing the Mainline to recover it would cause an increase in tolls. Under the Board's Equity Risk Premium methodology, the cost of equity capital is driven by the Board's estimate of the risk-free interest rate and the degree of risk investors perceive in the "benchmark" pipeline. The higher the risk, the higher their required rate of return. The degree of risk specific to the Mainline is accounted for by adjustments to its deemed capital structure. Accordingly, the cost to the Mainline of providing that rate of return on the equity component of its deemed capital structure is unaffected by the impact of tolls on customers or consumers.

[37] The appellant has not demonstrated that the Board took the impact on customers or consumers into account in making its determination of the Mainline's required rate of return on equity.

[38] It is true that in its 2002 decision, the Board did state:

"In respect of the appropriate balance of customer and investor interests, the Board notes that customer interest in rate of return matters relates most directly to the impact the approved return will have on tolls. The Board is of the view that the impact of the rate of return on tolls is a relevant factor in the determination of a fair return." (RH-4-2001 at 12).

[39] The appellant says it cannot tell if the Board took the impact on customers or consumers into account in making its determination of the Mainline's required rate of return on equity. There is certainly no indication in its 2002 reasons that the Board adjusted its estimate of the required rate of return on equity based upon the impact it would have on tolls. In fact, the Board simply applied the automatic adjustment formula adopted in its 1995 decision. That formula does not take into account the impact of tolls on customers or consumers.

[40] It is also true that, in relation to an adjustment the Board made in the Mainline's deemed capital structure in its 2002 decision, the Board did state:

"In light of the above, the Board is of the view that it would be appropriate to increase the Mainline's deemed common equity ratio from 30% to 33%. The Board notes that this increase will raise the Mainline's annual cost of service and tolls by approximately 2%. The Board has determined that the toll increase is warranted by the prospective business risk facing the Mainline and that it will not impose an undue burden on shippers." (RH-4-2001 at 59).

[41] As I understand the Board's reasons, in view of the Mainline's increased business risk, the equity component of its deemed capital structure was increased from 30% to 33%. Because the required rate of return on equity was greater than the required rate of return on debt, this increased the overall estimate of the Mainline's required rate of return on capital, resulting in a 2% increase in tolls.

[42] While the Board observed that the increase would not be an undue burden on shippers, there is no suggestion that the increase in the equity component of the Mainline's deemed capital structure was in any way suppressed by considerations of its impact on customers or consumers. Nor, as I have said, is there any indication that the Board determined a required rate of return on equity for the Mainline and then adjusted it downward based on the impact it would have on tolls. In the absence of some indication in the Board's reasons, there is no basis for such an assumption.

d) The Board may adopt temporary measures to ameliorate "rate shock" so long as the utility eventually recovers its costs .

[43] I would add one further point. While I agree with the appellant that the impact on customers or consumers cannot be a factor in the determination of the cost of equity capital, any resulting increase in tolls may be a relevant factor for the Board to consider in determining the way in which a utility should recover its costs. It may be that an increase is so significant that it would lead to "rate shock" if implemented all at once and therefore should be phased in over time. It is quite proper for the Board to take such considerations into account, provided that there is, over a reasonable period of time, no economic loss to the utility in the process. In other words, the phased in tolls would have to compensate the utility for deferring recovery of its cost of capital. In the end, where a cost of service method is used, the utility must recover its costs over a reasonable period of time, regardless of any impact those costs may have on customers or consumers (see *Hemlock Valley Electrical Services Ltd. v. Utilities Commission (B.C.) et al.* (1992), 12 B.C.A.C. 1; 23 W.A.C. 1; 66 B.C.L.R.(2d) 1 (C.A.), at 20-21 (C.A.)). In this case, however, there is no suggestion that the Board sought to phase in or otherwise understate the Mainline's cost of capital.

3. Did the Board fetter its discretion?

a) Appellant's arguments

[44] The appellant's second alleged error of law is that the Board fettered its discretion. The appellant submits that the Board placed an inappropriate onus on the appellant to demonstrate that the cost of equity adjustment formula established by the Board in its 1995 decision, but not expressed in the National Energy Board Act or in any judicial authority, was to govern unless the appellant could persuade the Board otherwise.

[45] In its factum, the appellant states that the high onus of reversal placed on it by the Board caused the Board to act "inconsistently with its obligations of impartiality as an administrative tribunal." Some of the respondents characterised this as an allegation of bias against the Board.

[46] In oral argument, the appellant added that the Board wrongly discarded evidence of both the appellant and the respondents because the Board was not open to reviewing the adjustment formula.

b) The intended duration of the automatic adjustment mechanism .

[47] In its 1995 decision, the Board was expressly addressing "what simplified procedure should be implemented to effect an annual adjustment to the rate of return applicable to pipelines between cost of capital proceedings" (RH-2-94 at 1). The Board explained its reasons for seeking an automatic adjustment mechanism in the following words:

"In setting this matter down for hearing, it was the Board's intention to put in place means of improving the efficacy of the toll setting process for the year 1995 and beyond. The Board expressed the desire to avoid annual hearings on the cost of capital and was of the view that some automatic mechanism to adjust the return on common equity could be the most appropriate way to ensure that this return continued to be fair to all parties, while avoiding the expense of litigating annual or biennial changes in the rate of return. The Board therefore included as an issue in the RH-2-94 proceeding, the design and implementation of a predetermined adjustment mechanism to the rate of return on the common equity component. The Board's objective in this regard was to conduct detailed examinations of the pipelines' cost of capital only when significant changes had occurred in financial markets, business circumstances, or in general economic conditions." (RH-2-94 at 1-2)

[48] After an extensive hearing in which it considered the submissions of pipelines, shippers, governments and others, the Board established the automatic adjustment mechanism whereby the cost of equity capital would be determined. As to how long the automatic adjustment mechanism would remain in place, the Board stated:

"The Board is not setting a limit on the life of the mechanism and it does not expect to reassess the rate of return on common equity in a formal hearing for at least three years. The Board has confidence that the adjustment mechanism adopted will provide an appropriate balance between the interests of pipeline company shareholders and those of shippers." (RH-2-94 at 32).

[49] In its 1995 decision, the Board also established a deemed capital structure for the Group 1 pipelines. As discussed above, the Mainline was deemed to have a capital structure made up of 70% debt and 30% equity. The Board expressed the view that its capital structure determination would endure for an extended period of years, but that the Board would be prepared to consider a re-assessment of capital structure if requested by a pipeline, its shippers or another interested party:

"The Board also expects that the capital structure set in this hearing for each of the pipelines will endure for an extended period of years. The Board will be prepared to consider a reassessment of capital structures, likely on an individual basis, in the event of a significant change in business risk, in corporate structure or in corporate financial fundamentals. The Board does not favour routine reassessments of capital structure. For these reasons, the Board has not set out a specific date or any criteria for capital structure re-evaluation. Any reassessment of capital structure, for reasons such as those expressed above, must be at the request of the pipeline itself, its shippers or some other interested party. It would then be for the Board to assess the merits of such a request." (RH-2-94 at 32).

[50] The Board's Order TG/TO-1-95, which implemented the 1995 decision, set the Mainline's deemed capital structure and required that the Mainline's cost of equity capital for 1996 and subsequent years be determined through the application of the adjustment formula. The Order contained no time limit and therefore continues in force until reviewed or varied by the Board.

c) The appellant did bear the burden of showing that the automatic adjustment mechanism should no longer apply .

[51] The Board applied its automatic adjustment mechanism annually until 2001 when the appellant brought its fair rate of return application, seeking a review and variance of the 1995 decision and the adoption of a new means of determining its cost of capital.

[52] The appellant's position seems to be that when it brought its fair rate of return application in 2001, the Board was required to disregard entirely the automatic adjustment mechanism and start fresh - with a clean slate as it were - to determine the appropriate method by which to estimate the Mainline's cost of capital.

[53] However, the adjustment formula was part of an order that continued to bind the appellant. Subsection 23(1) of the National Energy Board Act provides:

"23(1) Except as provided in this Act, every decision or order of the Board is final and conclusive."

"23(1) Sauf exceptions prévues à la présente loi, les décisions ou ordonnances de l'Office sont définitives et sans appel."

Section 22 allows for appeals to the Federal Court of Appeal while subsection 21(1) allows the Board to review, vary and rescind its decisions and orders. Neither the Board's 1995 decision nor the order implementing it were appealed. The adjustment formula therefore continued to apply until the appellant demonstrated to the Board that it should be replaced.

[54] The hearing conducted by the Board on the appellant's fair return application was extensive. Written evidence was filed and the oral hearing proceeded for more than a month. The Board's 2002 decision was 64 pages long. The Board considered the appellant's ATWACC proposal and its alternative increased rate of return on equity proposal, reviewed the evidence of the witnesses and ultimately concluded that utilization of the automatic adjustment formula continued to yield a rate of return on equity that the Board considered to be appropriate for the Mainline.

[55] However, the Board did, to some extent, accept the appellant's argument that the Mainline's business risk had increased. In order to take account of the increased risk, the Board increased the equity component of the Mainline's deemed capital structure from 30% to 33% so that the capital structure would be 33% equity and 67% debt.

[56] I can detect no fettering of discretion or the placing of an improper onus on the appellant in the Board's reasons. In its 1995 decision, the Board stated that its automatic adjustment formula was to reflect a simplified procedure to determine annual adjustments to pipeline rates of return on common equity. It was therefore to continue indefinitely. When an affected party wishes to change the process, it has the onus to demonstrate that its proposal is preferable to the one which is the subject of a binding Board order. That is not an improper onus. Nor does it reflect a fettering of discretion by the Board. Most importantly, it does not give rise to any apprehension of impartiality or bias on the part of the Board.

[57] In reviewing the 2002 decision, the Review and Variance Panel found in its 2003 decision that the onus was on the appellant to demonstrate that the automatic adjustment formula was no longer appropriate and that the appellant had failed to do so:

"The Fair Return Application was, among other things, an application for review of the RH-2-94 Decision and related orders, pursuant to subsection 21(1) of the Act . The onus was on TransCanada to prove to the Board in RH-4-2001 that the RH-2-94 Formula was no longer appropriate for determining the Mainline's

return on equity. Neither the intervenors nor the Board had the onus in the RH-4-2001 proceeding to justify the continued use of the Formula. The Formula was appropriate unless and until TransCanada persuaded the Board otherwise.

"TransCanada failed to meet the burden and accordingly, the RH-2-94 Formula continued to apply. The Board was not required in the RH-4-2001 Decision to justify that the Formula was appropriate; that determination was made in the RH-2-94 proceeding." (RH-R-1-2002 at 24).

I find no error on the part of the Board in that analysis or conclusion.

d) The Board did not disregard or ignore evidence .

[58] As to the appellant's argument that the Board disregarded evidence, I agree that the Board did not adopt the evidence of any particular witness for or against the appellant. But that does not mean that the evidence was discarded or ignored. In cost of capital proceedings, the Board is entitled, on the basis of the evidence before it and the use of its own judgment, to choose a methodology for determining cost of capital and to estimate the cost of capital for a forthcoming year. Very often, the Board's estimate will not reflect the precise estimates of one side or the other or of one witness or another. Having regard to all the evidence, the Board will determine its own estimate. As long as that estimate is within the range of estimates put forward in the evidence and the Board demonstrates that it considered the estimates put forward, the Board cannot be said to have ignored evidence. Indeed, even if the Board's estimate is outside that range, if the Board shows that it considered the evidence submitted and provides adequate reasons for its opinion, the Board will not be found to have ignored evidence.

[59] In this case, the estimates in the evidence of the required rate of return on equity ranged from 8.28% to 12.50%. The Board's reasons indicate that it considered the estimates put forward. Using its automatic adjustment formula, the Board calculated that the required rate of return on equity for the Mainline would be 9.61% in 2001 and 9.53% in 2002. I cannot see that the Board disregarded or ignored evidence in deciding to continue to utilize the automatic adjustment formula to determine the required rate of return on equity for the Mainline.

Conclusion

[60] I would dismiss this appeal with costs.

Appeal dismissed.

Editor: David Weir/clh

[End of document]

Ontario Energy Board *Appellant*

Commission de l'énergie de l'Ontario
Appelante

v.

c.

**Ontario Power Generation Inc.,
Power Workers' Union, Canadian Union
of Public Employees, Local 1000 and
Society of Energy Professionals** *Respondents*

**Ontario Power Generation Inc.,
Syndicat des travailleurs et travailleuses
du secteur énergétique, Syndicat canadien
de la fonction publique, section locale 1000 et
Society of Energy Professionals** *Intimés*

and

et

Ontario Education Services Corporation
Intervener

**Corporation des services en éducation
de l'Ontario** *Intervenante*

**INDEXED AS: ONTARIO (ENERGY BOARD) v.
ONTARIO POWER GENERATION INC.**

**RÉPERTORIÉ : ONTARIO (COMMISSION DE
L'ÉNERGIE) c. ONTARIO POWER GENERATION INC.**

2015 SCC 44

2015 CSC 44

File No.: 35506.

N° du greffe : 35506.

2014: December 3; 2015: September 25.

2014 : 3 décembre; 2015 : 25 septembre.

Present: McLachlin C.J. and Abella, Rothstein,
Cromwell, Moldaver, Karakatsanis and Gascon JJ.

Présents : La juge en chef McLachlin et les juges Abella,
Rothstein, Cromwell, Moldaver, Karakatsanis et Gascon.

ON APPEAL FROM THE COURT OF APPEAL FOR
ONTARIO

EN APPEL DE LA COUR D'APPEL DE L'ONTARIO

Public utilities — Electricity — Rate-setting decision by utilities regulator — Utility seeking to recover incurred or committed compensation costs in utility rates set by Ontario Energy Board — Whether Board bound to apply particular prudence test in evaluating utility costs — Whether Board's decision to disallow \$145 million in labour compensation costs related to utility's nuclear operations reasonable — Ontario Energy Board Act, 1998, S.O. 1998, c. 15, Sch. B, s. 78.1(5), (6).

Services publics — Électricité — Décision d'un organisme de réglementation des services publics relative-ment à l'établissement de tarifs — Demande d'un service public en vue d'obtenir le recouvrement de dépenses de rémunération faites ou convenues grâce aux tarifs établis par la Commission de l'énergie de l'Ontario — La Com-mission avait-elle l'obligation d'employer une méthode particulière axée sur la prudence pour apprécier les dépenses du service public? — Le refus de la Commis-sion d'approuver 145 millions de dollars au titre des dé-penses de rémunération liées aux installations nucléaires du service public était-il raisonnable? — Loi de 1998 sur la Commission de l'énergie de l'Ontario, L.O. 1998, c. 15, ann. B, art. 78.1(5), (6).

Administrative law — Boards and tribunals — Ap-peals — Standing — Whether Ontario Energy Board acted improperly in pursuing appeal and in arguing in favour of reasonableness of its own decision — Whether

Droit administratif — Organismes et tribunaux ad-ministratifs — Appels — Qualité pour agir — La Com-mission de l'énergie de l'Ontario a-t-elle agi de manière inappropriée en se pourvoyant en appel et en faisant valoir

Board attempted to use appeal to “bootstrap” its original decision by making additional arguments on appeal.

In Ontario, utility rates are regulated through a process by which a utility seeks approval from the Ontario Energy Board for costs the utility has incurred or expects to incur in a specified period of time. Where the Board approves of the costs, they are incorporated into utility rates such that the utility receives payment amounts to cover the approved expenditures. The Board disallowed certain payment amounts applied for by Ontario Power Generation (“OPG”) as part of its rate application covering the 2011-2012 operating period. Specifically, the Board disallowed \$145 million in labour compensation costs related to OPG’s nuclear operations on the grounds that OPG’s labour costs were out of step with those of comparable entities in the regulated power generation industry. A majority of the Ontario Divisional Court dismissed OPG’s appeal and upheld the decision of the Board. The Court of Appeal set aside the decisions of the Divisional Court and the Board and remitted the matter to the Board for redetermination in accordance with its reasons.

The crux of OPG’s argument here is that the Board is legally required to compensate OPG for all of its prudently committed or incurred costs. OPG asserts that prudence in this context has a particular methodological meaning that requires the Board to assess the reasonableness of OPG’s decision to incur or commit to costs at the time the decisions to incur or commit to the costs were made and that OPG ought to benefit from a presumption of prudence. The Board on the other hand argues that a particular prudence test methodology is not compelled by law, and that in any case the costs disallowed here were not committed nuclear compensation costs, but are better characterized as forecast costs.

OPG also raises concerns regarding the Board’s role in acting as a party on appeal from its own decision, arguing that the Board’s aggressive and adversarial defence of its decision was improper, and the Board attempted to use the appeal to bootstrap its original decision by making additional arguments on appeal. The Board argues that the structure of utilities regulation in Ontario makes it necessary and important for it to argue the merits of its decision on appeal.

le caractère raisonnable de sa propre décision? — A-t-elle tenté de se servir de l’appel pour « s’auto-justifier » en formulant de nouveaux arguments à l’appui de sa décision initiale?

En Ontario, la tarification d’un service public est réglementée, de sorte que ce dernier doit obtenir de la Commission de l’énergie de l’Ontario l’approbation des dépenses qu’il a faites ou qu’il prévoit faire pendant une période donnée. Lorsque cette approbation est obtenue, les tarifs sont rajustés de manière que le service public touche des paiements qui correspondent à ses dépenses. La Commission a refusé certains paiements sollicités par Ontario Power Generation (« OPG ») dans sa décision sur la demande d’établissement des tarifs pour la période 2011-2012. Elle a en fait refusé à OPG le recouvrement de 145 millions de dollars au titre des dépenses de rémunération liées aux installations nucléaires du service public au motif que ces dépenses étaient en rupture avec celles d’organismes comparables dans le secteur réglementé de la production d’énergie. Les juges majoritaires de la Cour divisionnaire de l’Ontario ont rejeté l’appel d’OPG et confirmé la décision de la Commission. La Cour d’appel a annulé les décisions de la Cour divisionnaire et de la Commission, puis renvoyé le dossier à la Commission afin qu’elle rende une nouvelle décision conforme à ses motifs.

La thèse d’OPG en l’espèce veut essentiellement que la Commission soit légalement tenue de l’indemniser de la totalité des dépenses faites ou convenues avec prudence. OPG prétend que, dans ce contexte, la prudence se définit selon une méthode particulière qui exige de la Commission qu’elle détermine si, au moment où elles ont été prises, les décisions de faire les dépenses ou de convenir des dépenses étaient raisonnables. Elle soutient en outre qu’une présomption de prudence doit s’appliquer à son bénéfice. La Commission prétend pour sa part que la loi ne l’oblige pas à employer quelque méthode fondée sur le principe de la prudence et que, de toute manière, les dépenses de rémunération des employés du secteur nucléaire refusées en l’espèce n’étaient pas des dépenses convenues, mais bien des dépenses prévues.

OPG exprime en outre des préoccupations sur la participation de la Commission à l’appel de sa propre décision et fait valoir que la manière agressive et conflictuelle dont la Commission a défendu sa décision initiale n’était pas justifiée et que l’organisme a tenté de se servir de l’appel pour s’auto-justifier en formulant de nouveaux arguments à l’appui de sa décision initiale. La Commission soutient que la manière dont les services publics sont réglementés en Ontario fait en sorte qu’il est nécessaire et important qu’elle défende la justesse de ses décisions portées en appel.

Held (Abella J. dissenting): The appeal should be allowed. The decision of the Court of Appeal is set aside and the decision of the Board is reinstated.

Per McLachlin C.J. and Rothstein, Cromwell, Moldaver, Karakatsanis and Gascon JJ.: The first issue is the appropriateness of the Board's participation in the appeal. The concerns with regard to tribunal participation on appeal from the tribunal's own decision should not be read to establish a categorical ban. A discretionary approach provides the best means of ensuring that the principles of finality and impartiality are respected without sacrificing the ability of reviewing courts to hear useful and important information and analysis. Because of their expertise and familiarity with the relevant administrative scheme, tribunals may in many cases be well positioned to help the reviewing court reach a just outcome. Further, some cases may arise in which there is simply no other party to stand in opposition to the party challenging the tribunal decision. In a situation where no other well-informed party stands opposed, the presence of a tribunal as an adversarial party may help the court ensure it has heard the best of both sides of a dispute. The following factors are relevant in informing the court's exercise of its discretion: statutory provisions addressing the structure, processes and role of the particular tribunal and the mandate of the tribunal, that is, whether the function of the tribunal is to adjudicate individual conflicts between parties or whether it serves a policy-making, regulatory or investigative role, or acts on behalf of the public interest. The importance of fairness, real and perceived, weighs more heavily against tribunal standing where the tribunal served an adjudicatory function in the proceeding. Tribunal standing is a matter to be determined by the court conducting the first-instance review in accordance with the principled exercise of that court's discretion. In exercising its discretion, the court is required to balance the need for fully informed adjudication against the importance of maintaining tribunal impartiality.

Consideration of these factors in the context of this case leads to the conclusion that it was not improper for the Board to participate in arguing in favour of the reasonableness of its decision on appeal. The Board was the only respondent in the initial review of its decision. It had no alternative but to step in if the decision was to be defended on the merits. Also, the Board was exercising

Arrêt (la juge Abella est dissidente) : Le pourvoi est accueilli. La décision de la Cour d'appel est annulée et celle de la Commission est rétablie.

La juge en chef McLachlin et les juges Rothstein, Cromwell, Moldaver, Karakatsanis et Gascon : Se pose en premier lieu la question du caractère approprié de la participation de la Commission au pourvoi. Les préoccupations relatives à la participation d'un tribunal administratif à l'appel de sa propre décision ne sauraient fonder l'interdiction absolue d'une telle participation. La démarche discrétionnaire offre le meilleur moyen d'assurer le caractère définitif de la décision et l'impartialité du décideur sans que la cour de révision ne soit alors privée de données et d'analyses à la fois utiles et importantes. Vu ses compétences spécialisées et sa connaissance approfondie du régime administratif en cause, le tribunal administratif peut, dans bien des cas, être bien placé pour aider la cour de révision à rendre une juste décision. Qui plus est, dans certains cas, il n'y a tout simplement personne pour s'opposer à la partie qui conteste la décision du tribunal administratif. Lorsqu'aucune autre partie bien au fait des enjeux ne fait valoir le point de vue opposé, la participation du tribunal administratif à titre de partie adverse peut contribuer à faire en sorte que la cour statue après avoir entendu les arguments les plus convaincants de chacune des deux parties au litige. Les considérations suivantes permettent de délimiter l'exercice du pouvoir discrétionnaire de la cour de révision : les dispositions législatives portant sur la structure, le fonctionnement et la mission du tribunal en cause et le mandat du tribunal, à savoir si sa fonction consiste soit à trancher des différends individuels opposant plusieurs parties, soit à élaborer des politiques, à réglementer ou à enquêter, ou à défendre l'intérêt public. L'importance de l'équité, réelle et perçue, milite davantage contre la reconnaissance de la qualité pour agir du tribunal administratif qui a exercé une fonction juridictionnelle dans l'instance. Il appartient à la cour de première instance chargée du contrôle judiciaire de décider de la qualité pour agir d'un tribunal administratif en exerçant son pouvoir discrétionnaire de manière raisonnée. Dans l'exercice de son pouvoir discrétionnaire, la cour doit établir un équilibre entre la nécessité d'une décision bien éclairée et l'importance d'assurer l'impartialité du tribunal administratif.

L'application de ces principes à la situation considérée en l'espèce mène à la conclusion qu'il n'était pas inapproprié que la Commission participe à l'appel pour défendre le caractère raisonnable de sa décision. La Commission était la seule partie intimée lors du contrôle judiciaire initial de sa décision. Elle n'avait d'autre choix que de prendre part à l'instance pour que sa décision

a regulatory role by setting just and reasonable payment amounts to a utility. In this case, the Board's participation in the instant appeal was not improper.

The issue of tribunal “bootstrapping” is closely related to the question of when it is proper for a tribunal to act as a party on appeal or judicial review of its decision. The standing issue concerns the types of argument a tribunal may make, while the bootstrapping issue concerns the content of those arguments. A tribunal engages in bootstrapping where it seeks to supplement what would otherwise be a deficient decision with new arguments on appeal. A tribunal may not defend its decision on a ground that it did not rely on in the decision under review. The principle of finality dictates that once a tribunal has decided the issues before it and provided reasons for its decision, absent a power to vary its decision or rehear the matter, it cannot use judicial review as a chance to amend, vary, qualify or supplement its reasons. While a permissive stance towards new arguments by tribunals on appeal serves the interests of justice insofar as it ensures that a reviewing court is presented with the strongest arguments in favour of both sides, to permit bootstrapping may undermine the importance of reasoned, well-written original decisions. In this case, the Board did not impermissibly step beyond the bounds of its original decision in its arguments before the Court. The arguments raised by the Board on appeal do not amount to impermissible bootstrapping.

The merits issue concerns whether the appropriate methodology was followed by the Board in its disallowance of \$145 million in labour compensation costs sought by OPG. The just-and-reasonable approach to recovery of the cost of services provided by a utility captures the essential balance at the heart of utilities regulation: to encourage investment in a robust utility infrastructure and to protect consumer interests, utilities must be allowed, over the long run, to earn their cost of capital, no more, no less. In order to ensure the balance between utilities' and consumers' interests is struck, just and reasonable rates must be those that ensure consumers are paying what the Board expects it to cost to efficiently provide the services they receive, taking account of both operating and capital costs. In that way, consumers may be assured that, overall, they are paying no more than what is necessary for the service they receive, and

soit défendue au fond. Aussi, la Commission a exercé sa fonction de réglementation en établissant les paiements justes et raisonnables auxquels un service public avait droit. Sa participation au pourvoi n'avait rien d'inapproprié en l'espèce.

La question de l'« autojustification » est étroitement liée à celle de savoir à quelles conditions le tribunal administratif est en droit d'agir comme partie à l'appel ou au contrôle judiciaire de sa décision. Statuer sur la qualité pour agir d'un tribunal c'est décider de ce qu'il peut faire valoir, alors que l'autojustification touche à la teneur des prétentions. Un tribunal s'autojustifie lorsqu'il cherche, par la présentation de nouveaux arguments en appel, à étoffer une décision qui, sinon, serait lacunaire. Un tribunal ne peut défendre sa décision en invoquant un motif qui n'a pas été soulevé dans la décision faisant l'objet du contrôle. Le caractère définitif de la décision veut que, dès lors qu'il a tranché les questions dont il était saisi et qu'il a motivé sa décision, à moins qu'il ne soit investi du pouvoir de modifier sa décision ou d'entendre à nouveau l'affaire, un tribunal ne puisse profiter d'un contrôle judiciaire pour modifier, changer, nuancer ou compléter ses motifs. Même s'il est dans l'intérêt de la justice de permettre au tribunal de présenter de nouveaux arguments en appel, la cour de révision étant alors saisie des arguments les plus convaincants à l'appui de chacune des thèses, autoriser l'autojustification risque de compromettre l'importance de décisions bien étayées et bien rédigées au départ. Dans la présente affaire, la Commission n'a pas indûment outrepassé les limites de sa décision initiale lorsqu'elle a présenté ses arguments devant la Cour. Les arguments qu'elle a invoqués en appel n'équivalent pas à une autojustification inadmissible.

La question de fond est celle de savoir si la Commission a employé une méthode appropriée pour refuser à OPG le recouvrement de 145 millions de dollars au titre des dépenses de rémunération. L'approche fondée sur le caractère juste et raisonnable des dépenses qu'un service public peut recouvrer rend compte de l'équilibre essentiel recherché dans la réglementation des services publics : pour encourager l'investissement dans une infrastructure robuste et protéger l'intérêt des consommateurs, un service public doit pouvoir, à long terme, toucher l'équivalent du coût du capital, ni plus, ni moins. Lorsqu'il s'agit d'assurer l'équilibre entre les intérêts du service public et ceux du consommateur, la tarification juste et raisonnable est celle qui fait en sorte que le consommateur paie ce que la Commission prévoit qu'il en coûtera pour la prestation efficace du service, compte tenu à la fois des dépenses d'exploitation et des coûts en

utilities may be assured of an opportunity to earn a fair return for providing those services.

The *Ontario Energy Board Act, 1998* does not prescribe the methodology the Board must use to weigh utility and consumer interests when deciding what constitutes just and reasonable payment amounts to the utility. However, the *Ontario Energy Board Act, 1998* places the burden on the applicant utility to establish that payment amounts approved by the Board are just and reasonable. It would thus seem inconsistent with the statutory scheme to presume that utility decisions to incur costs were prudent. The Board has broad discretion to determine the methods it may use to examine costs — but it cannot shift the burden of proof contrary to the statutory scheme.

The issue is whether the Board was bound to use a no-hindsight, presumption of prudence test to determine whether labour compensation costs were just and reasonable. The prudent investment test, or prudence review, is a valid and widely accepted tool that regulators may use when assessing whether payments to a utility would be just and reasonable. However, there is no support in the statutory scheme for the notion that the Board should be required as a matter of law, under the *Ontario Energy Board Act, 1998* to apply the prudence test such that the mere decision not to apply it when considering committed costs would render its decision on payment amounts unreasonable. Where a statute requires only that the regulator set “just and reasonable” payments, as the *Ontario Energy Board Act, 1998* does in Ontario, the regulator may make use of a variety of analytical tools in assessing the justness and reasonableness of a utility’s proposed payment amounts. This is particularly so where, as here, the regulator has been given express discretion over the methodology to be used in setting payment amounts.

Where the regulator has discretion over its methodological approach, understanding whether the costs at issue are “forecast” or “committed” may be helpful in reviewing the reasonableness of a regulator’s choice of methodology. Here, the labour compensation costs which led to the \$145 million disallowance are best understood

capital. Ainsi, le consommateur a l’assurance que, globalement, il ne paie pas plus que ce qui est nécessaire pour obtenir le service, et le service public a l’assurance de pouvoir toucher une juste contrepartie pour la prestation du service.

La *Loi de 1998 sur la Commission de l’énergie de l’Ontario* ne prescrit pas la méthode que doit utiliser la Commission pour sopeser les intérêts respectifs du service public et du consommateur lorsqu’elle décide ce qui constitue des paiements justes et raisonnables. Suivant cette loi, il incombe cependant au service public requérant d’établir que les paiements qu’il demande à la Commission d’approuver sont justes et raisonnables. Il semble donc contraire au régime législatif de présumer que la décision du service public de faire les dépenses était prudente. La Commission jouit d’un grand pouvoir discrétionnaire qui lui permet d’arrêter la méthode à employer dans l’examen des dépenses, mais elle ne peut tout simplement pas inverser le fardeau de la preuve qu’établit le régime législatif.

La question à trancher est celle de savoir si la Commission était tenue à l’application d’un critère excluant le recul et présumant la prudence pour décider si les dépenses de rémunération du personnel étaient justes et raisonnables. Le critère de l’investissement prudent — ou contrôle de la prudence — offre aux organismes de réglementation un moyen valable et largement reconnu d’apprécier le caractère juste et raisonnable des paiements sollicités par un service public. Toutefois, aucun élément du régime législatif n’appuie l’idée que la Commission devrait être tenue en droit, suivant la *Loi de 1998 sur la Commission de l’énergie de l’Ontario*, d’appliquer le critère de la prudence de sorte que la seule décision de ne pas l’appliquer pour apprécier des dépenses convenues rendrait déraisonnable sa décision sur les paiements. Lorsqu’un texte législatif — telle la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* en Ontario — exige seulement qu’il fixe des paiements « justes et raisonnables », l’organisme de réglementation peut avoir recours à divers moyens d’analyse pour apprécier le caractère juste et raisonnable des paiements sollicités par le service public. Cela est particulièrement vrai lorsque, comme en l’espèce, l’organisme de réglementation se voit accorder expressément un pouvoir discrétionnaire quant à la méthode à appliquer pour fixer les paiements.

Lorsque l’organisme de réglementation possède un pouvoir discrétionnaire quant à la méthode à employer, la qualification des dépenses — « prévues » ou « convenues » — peut constituer une étape importante pour statuer sur le caractère raisonnable de la méthode retenue. Dans la présente affaire, il convient mieux de voir dans

as partly committed costs and partly costs subject to management discretion. They are partly committed because they resulted from collective agreements entered into between OPG and two of its unions, and partly subject to management discretion because OPG retained some flexibility to manage total staffing levels in light of, among other things, projected attrition of the workforce. It is not reasonable to treat these costs as entirely forecast. However, the Board was not bound to apply a particular prudence test in evaluating these costs. It is not necessarily unreasonable, in light of the particular regulatory structure established by the *Ontario Energy Board Act, 1998*, for the Board to evaluate committed costs using a method other than a no-hindsight prudence review. Applying a presumption of prudence would have conflicted with the burden of proof in the *Ontario Energy Board Act, 1998* and would therefore not have been reasonable. The question of whether it was reasonable to assess a particular cost using hindsight should turn instead on the circumstances of that cost.

In this case, the nature of the disputed costs and the environment in which they arose provide a sufficient basis to find that the Board did not act unreasonably in not applying the prudent investment test in determining whether it would be just and reasonable to compensate OPG for these costs and disallowing them. Since the costs at issue are operating costs, there is little danger that a disallowance of these costs will have a chilling effect on OPG's willingness to incur operating costs in the future, because costs of the type disallowed here are an inescapable element of operating a utility. Further, the costs at issue arise in the context of an ongoing repeat-player relationship between OPG and its employees. Such a context supports the reasonableness of a regulator's decision to weigh all evidence it finds relevant in striking a just and reasonable balance between the utility and consumers, rather than confining itself to a no-hindsight approach. There is no dispute that collective agreements are "immutable" between employees and the utility. However, if the legislature had intended for costs under collective agreements to also be inevitably imposed on consumers, it would not have seen fit to grant the Board oversight of utility compensation costs. The Board's decision in no way purports to force OPG to break its contractual commitments to unionized employees. It was not unreasonable

les dépenses de rémunération dont le recouvrement a été refusé à raison de 145 millions de dollars en partie des dépenses convenues et en partie des dépenses relevant du pouvoir discrétionnaire de la direction. Elles sont en partie convenues parce qu'elles résultent de conventions collectives intervenues entre OPG et deux de ses syndicats, et elles relèvent en partie de la discrétion de la direction parce qu'OPG conservait une certaine marge de manœuvre dans la gestion des niveaux de dotation globale compte tenu, entre autres, de l'attrition projetée de l'effectif. Il est déraisonnable de considérer qu'il s'agit en totalité de dépenses prévues. Cependant, la Commission n'était pas tenue d'appliquer un principe de prudence donné pour apprécier ces dépenses. Il n'est pas nécessairement déraisonnable, à la lumière du cadre réglementaire établi par la *Loi de 1998 sur la Commission de l'énergie de l'Ontario*, que la Commission se prononce sur les dépenses convenues en employant une autre méthode que l'application d'un critère de prudence qui exclut le recul. Présumer la prudence aurait été incompatible avec le fardeau de la preuve que prévoit la *Loi de 1998 sur la Commission de l'énergie de l'Ontario* et, de ce fait, déraisonnable. Qu'il soit raisonnable ou non d'apprécier certaines dépenses avec recul devrait plutôt dépendre des circonstances de la décision dont s'originent ces dépenses.

Dans la présente affaire, la nature des dépenses litigieuses et le contexte dans lequel elles ont vu le jour permettent de conclure que la Commission n'a pas agi de manière déraisonnable en n'appliquant pas le critère de l'investissement prudent pour décider s'il était juste et raisonnable d'indemniser OPG de ces dépenses et en refusant le recouvrement de celles-ci. Puisque les dépenses en cause sont des dépenses d'exploitation, il est peu probable que le refus essuyé dissuade OPG de faire de telles dépenses à l'avenir, car les dépenses de la nature de celles dont le recouvrement a été refusé sont inhérentes à l'exploitation d'un service public. Aussi, les dépenses en cause découlent d'une relation continue entre OPG et ses employés. Pareil contexte milite en faveur du caractère raisonnable de la décision de l'organisme de réglementation de soupeser toute preuve qu'il juge pertinente aux fins d'établir un équilibre juste et raisonnable entre le service public et les consommateurs, au lieu de s'en tenir à une approche excluant le recul. Nul ne conteste que les conventions collectives intervenues entre le service public et ses employés sont « immuables ». Toutefois, si le législateur avait voulu que les dépenses qui en sont issues se répercutent inévitablement sur les consommateurs, il n'aurait pas jugé opportun d'investir la Commission du pouvoir de surveiller les dépenses de

for the Board to adopt a mixed approach that did not rely on quantifying the exact share of compensation costs that fell into the forecast and committed categories. Such an approach represents an exercise of the Board's methodological discretion in addressing a challenging issue where these costs did not fit easily into one category or the other.

The Board's disallowance may have adversely impacted OPG's ability to earn its cost of capital in the short run. Nevertheless, the disallowance was intended to send a clear signal that OPG must take responsibility for improving its performance. Such a signal may, in the short run, provide the necessary impetus for OPG to bring its compensation costs in line with what, in the Board's opinion, consumers should justly expect to pay for an efficiently provided service. Sending such a signal is consistent with the Board's market proxy role and its objectives under s. 1 of the *Ontario Energy Board Act, 1998*.

Per Abella J. (dissenting): The Board's decision was unreasonable because the Board failed to apply the methodology set out for itself for evaluating just and reasonable payment amounts. It both ignored the legally binding nature of the collective agreements between Ontario Power Generation and the unions and failed to distinguish between committed compensation costs and those that were reducible.

The Board stated in its reasons that it would use two kinds of review in order to determine just and reasonable payment amounts. As to "forecast costs", that is, those over which a utility retains discretion and can still be reduced or avoided, the Board explained that it would review such costs using a wide range of evidence, and that the onus would be on the utility to demonstrate that its forecast costs were reasonable. A different approach, however, would be applied to those costs the company could not "take action to reduce". These costs, sometimes called "committed costs", represent binding commitments that leave a utility with no discretion about whether to make the payment. The Board explained that it would evaluate these costs using a "prudence review". The application of a prudence review does not shield

rémunération d'un service public. La Commission n'entend aucunement, par sa décision, contraindre OPG à se soustraire à ses engagements contractuels envers ses employés syndiqués. Il n'était pas déraisonnable que la Commission opte pour une démarche hybride qui ne se fonde pas sur la répartition exacte des dépenses de rémunération entre celles qui sont prévues et celles qui sont convenues. Pareille démarche correspond à un exercice du pouvoir discrétionnaire de la Commission sur le plan méthodologique lorsqu'elle est appelée à se prononcer sur une question épineuse et que les dépenses en cause ne sont pas aisément assimilables à l'une ou l'autre de ces catégories.

Le refus de la Commission a pu nuire à la possibilité qu'OPG obtienne à court terme l'équivalent de son coût du capital. Toutefois, il visait à signifier clairement à OPG qu'il lui incombe d'accroître sa performance. L'envoi d'un tel message peut, à court terme, donner à OPG l'impulsion nécessaire pour rapprocher ses dépenses de rémunération de ce que, selon la Commission, les consommateurs devraient à bon droit s'attendre à payer pour la prestation efficace du service. L'envoi d'un tel message est conforme au rôle de substitut du marché de la Commission et à ses objectifs selon l'article premier de la *Loi de 1998 sur la Commission de l'énergie de l'Ontario*.

La juge Abella (dissidente) : La Commission a rendu une décision déraisonnable en ce qu'elle n'a pas appliqué la méthode qu'elle avait elle-même établie pour déterminer le montant de paiements justes et raisonnables. Elle a à la fois méconnu le caractère contraignant en droit des conventions collectives liant Ontario Power Generation et les syndicats et omis de distinguer les dépenses de rémunération convenues de celles qui étaient réductibles.

Dans ses motifs, la Commission a dit recourir à deux examens pour arrêter le montant de paiements justes et raisonnables. En ce qui concerne les « dépenses prévues », soit celles à l'égard desquelles le service public conserve un pouvoir discrétionnaire et qu'il peut toujours réduire ou éviter, la Commission a expliqué qu'elle examinait ces dépenses au regard d'une vaste gamme d'éléments de preuve et qu'il incombait au service public d'en démontrer le caractère raisonnable. Cependant, une démarche différente était suivie pour les dépenses à l'égard desquelles la société ne pouvait « prendre de mesures de réduction ». Ces dépenses, parfois appelées « dépenses convenues », résultent d'obligations contractuelles qui excluent tout pouvoir discrétionnaire permettant au service public de ne pas les acquitter. La Commission a expliqué

these costs from scrutiny, but it does include a presumption that the costs were prudently incurred.

Rather than apply the methodology it set out for itself, however, the Board assessed *all* compensation costs in Ontario Power Generation's collective agreements as adjustable forecast costs, without determining whether any of them were costs for which there is no opportunity for the company to take action to reduce. The Board's failure to separately assess the compensation costs committed as a result of the collective agreements from other compensation costs, ignored not only its own methodological template, but labour law as well.

The compensation costs for approximately 90 per cent of Ontario Power Generation's regulated workforce were established through legally binding collective agreements which obligated the utility to pay fixed levels of compensation, regulated staffing levels, and provided unionized employees with employment security. The obligations contained in these collective agreements were immutable and legally binding commitments. The agreements therefore did not just leave the utility with limited flexibility regarding overall compensation or staffing levels, they made it *illegal* for the utility to alter the compensation and staffing levels of 90 per cent of its regulated workforce in a manner that was inconsistent with its commitments under the agreements.

The Board, however, applying the methodology it said it would use for the utility's forecast costs, put the onus on Ontario Power Generation to prove the reasonableness of all its compensation costs and concluded that it had failed to provide compelling evidence or documentation or analysis to justify compensation levels. Had the Board used the approach it said it would use for costs the company had no opportunity to reduce, it would have used an after-the-fact prudence review, with a rebuttable presumption that the utility's expenditures were reasonable.

It may well be that Ontario Power Generation has the ability to manage some staffing levels through attrition or other mechanisms that did not breach the utility's commitments under its collective agreements, and that these costs may therefore properly be characterized as forecast

qu'elle appréciait ces dépenses en se livrant à un « contrôle de la prudence ». L'application du principe de la prudence ne soustrait pas ces dépenses à tout examen, mais elle présume que les dépenses ont été faites de manière prudente.

Toutefois, au lieu d'appliquer la méthode qu'elle avait elle-même établie, la Commission a considéré *toutes* les dépenses de rémunération issues des conventions collectives d'Ontario Power Generation comme des dépenses prévues ajustables sans se demander s'il s'agissait en partie de dépenses pour lesquelles la société ne pouvait prendre de mesures de réduction. Par son omission d'apprécier les dépenses de rémunération issues des conventions collectives séparément des autres dépenses de rémunération, la Commission a méconnu à la fois son propre cadre méthodologique et le droit du travail.

Les dépenses de rémunération visant environ 90 p. 100 de l'effectif obligatoire d'Ontario Power Generation étaient établies par des conventions collectives contraignantes en droit qui imposaient des barèmes de rémunération fixes, qui déterminaient les niveaux de dotation et qui garantissaient la sécurité d'emploi des employés syndiqués. Les obligations contractées dans ces conventions collectives constituaient des engagements immuables ayant force obligatoire. Ces conventions ne laissaient pas seulement au service public peu de marge de manœuvre quant aux barèmes de rémunération et aux niveaux de dotation dans leur ensemble, elles rendaient *illégal* la modification par le service public — d'une manière incompatible avec les engagements qu'il y prenait — des barèmes de rémunération et des niveaux de dotation quant à 90 p. 100 de son effectif obligatoire.

Or, en appliquant la méthode qu'elle avait dit qu'elle utiliserait à l'égard des dépenses prévues du service public, la Commission a en fait obligé Ontario Power Generation à prouver le caractère raisonnable de toutes ses dépenses de rémunération et a conclu que l'entreprise n'avait présenté ni preuve convaincante, ni documents ou analyses qui justifiaient les barèmes de rémunération. Si elle avait eu recours à l'approche qu'elle avait dit qu'elle utiliserait pour les dépenses à l'égard desquelles la société ne pouvait « prendre de mesures de réduction », la Commission aurait contrôlé la prudence des dépenses après coup et appliqué la présomption réfutable selon laquelle elles étaient raisonnables.

Il se peut fort bien qu'Ontario Power Generation puisse modifier certains niveaux de dotation par voie d'attrition ou grâce à d'autres mécanismes qui ne vont pas à l'encontre de ses obligations suivant les conventions collectives. Il se peut fort bien aussi que les dépenses puissent

costs. But no factual findings were made by the Board about the extent of any such flexibility. There is in fact no evidence in the record, nor any evidence cited in the Board's decision, setting out what proportion of Ontario Power Generation's compensation costs were fixed and what proportion remained subject to the utility's discretion. Given that collective agreements are legally binding, it was unreasonable for the Board to assume that Ontario Power Generation could reduce the costs fixed by these contracts in the absence of any evidence to that effect.

Selecting a test which is more likely to confirm the Board's assumption that collectively-bargained costs are excessive, misconceives the point of the exercise, namely, to determine whether those costs were in fact excessive. Blaming collective bargaining for what are *assumed* to be excessive costs, imposes the appearance of an ideologically-driven conclusion on what is intended to be a principled methodology based on a distinction between committed and forecast costs, not between costs which are collectively bargained and those which are not. While the Board has wide discretion to fix payment amounts that are just and reasonable and, subject to certain limitations, to establish the methodology used to determine such amounts, once the Board establishes a methodology, it is, at the very least, required to faithfully apply it.

Absent methodological clarity and predictability, Ontario Power Generation would be unable to know how to determine what expenditures and investments to make and how to present them to the Board for review. Wandering sporadically from approach to approach, or failing to apply the methodology it declares itself to be following, creates uncertainty and leads, inevitably, to needlessly wasting public time and resources in constantly having to anticipate and respond to moving regulatory targets. Whether or not one can fault the Board for failing to use a particular methodology, what the Board can unquestionably be analytically faulted for, is evaluating all compensation costs fixed by collective agreements as being amenable to adjustment. Treating these compensation costs as reducible was unreasonable.

donc être assimilées à juste titre à des dépenses prévues. La Commission n'a toutefois tiré aucune conclusion de fait sur l'étendue d'une telle marge de manœuvre. En fait, aucun élément du dossier ou de la preuve invoquée par la Commission n'indique dans quelle proportion les dépenses de rémunération d'Ontario Power Generation étaient fixes et dans quelle proportion elles demeuraient assujetties au pouvoir discrétionnaire du service public. Comme les conventions collectives sont contraignantes en droit, il était déraisonnable que la Commission présume qu'Ontario Power Generation pouvait réduire les dépenses déterminées par ces contrats en l'absence de toute preuve en ce sens.

En choisissant un critère éminemment susceptible de confirmer l'hypothèse de la Commission selon laquelle les dépenses issues de négociations collectives sont excessives, on se méprend sur l'objectif de la démarche, qui est de déterminer si ces dépenses étaient bel et bien excessives. Imputer à la négociation collective ce que l'on *suppose* constituer des dépenses excessives revient à substituer ce qui a l'apparence d'une conclusion idéologique à ce qui est censé résulter d'une méthode d'analyse raisonnée qui distingue entre les dépenses convenues et les dépenses prévues, non entre les dépenses issues de négociations collectives et celles qui ne le sont pas. Même si la Commission jouit d'un vaste pouvoir discrétionnaire lui permettant de déterminer les paiements qui sont justes et raisonnables et, à l'intérieur de certaines limites, de définir la méthode utilisée pour établir le montant de ces paiements, dès lors qu'elle a établi une telle méthode, elle doit à tout le moins l'appliquer avec constance.

En l'absence de clarté et de prévisibilité quant à la méthode à appliquer, Ontario Power Generation ne peut savoir comment déterminer les dépenses et les investissements à faire et de quelle manière les soumettre à l'examen de la Commission. Passer sporadiquement d'une approche à une autre ou ne pas appliquer la méthode que l'on prétend appliquer crée de l'incertitude et mène inévitablement au gaspillage inutile du temps et des ressources publics en ce qu'il faut constamment anticiper un objectif réglementaire fluctuant et s'y ajuster. On peut reprocher ou non à la Commission de ne pas avoir appliqué une certaine méthode, mais on peut assurément lui reprocher, sur le plan analytique, d'avoir considéré toutes les dépenses de rémunération déterminées par des conventions collectives comme des dépenses ajustables. Voir dans ces dépenses des dépenses réductibles est à mon sens déraisonnable.

The appeal should accordingly be dismissed, the Board's decision set aside, and the matter remitted to the Board for reconsideration.

Cases Cited

By Rothstein J.

Considered: *Enbridge Gas Distribution Inc. v. Ontario Energy Board* (2006), 210 O.A.C. 4; *Northwestern Utilities Ltd. v. City of Edmonton*, [1979] 1 S.C.R. 684; **referred to:** *Toronto Hydro-Electric System Ltd. v. Ontario (Energy Board)*, 2010 ONCA 284, 99 O.R. (3d) 481; *Northwestern Utilities Ltd. v. City of Edmonton*, [1929] S.C.R. 186; *TransCanada Pipelines Ltd. v. National Energy Board*, 2004 FCA 149, 319 N.R. 171; *Ontario Power Generation Inc. (Re)*, EB-2007-0905, November 3, 2008 (online: <http://www.ontarioenergyboard.ca/>); *CAIMAW v. Paccar of Canada Ltd.*, [1989] 2 S.C.R. 983; *B.C.G.E.U. v. Indust. Rel. Council* (1988), 26 B.C.L.R. (2d) 145; *McLean v. British Columbia (Securities Commission)*, 2013 SCC 67, [2013] 3 S.C.R. 895; *Ellis-Don Ltd. v. Ontario (Labour Relations Board)*, 2001 SCC 4, [2001] 1 S.C.R. 221; *Tremblay v. Quebec (Commission des affaires sociales)*, [1992] 1 S.C.R. 952; *Ontario (Children's Lawyer) v. Ontario (Information and Privacy Commissioner)* (2005), 75 O.R. (3d) 309; *Canada (Attorney General) v. Quadrini*, 2010 FCA 246, [2012] 2 F.C.R. 3; *Leon's Furniture Ltd. v. Information and Privacy Commissioner (Alta.)*, 2011 ABCA 94, 502 A.R. 110; *Henthorne v. British Columbia Ferry Services Inc.*, 2011 BCCA 476, 344 D.L.R. (4th) 292; *United Brotherhood of Carpenters and Joiners of America, Local 1386 v. Bransen Construction Ltd.*, 2002 NBCA 27, 249 N.B.R. (2d) 93; *Chandler v. Alberta Association of Architects*, [1989] 2 S.C.R. 848; *Dunsmuir v. New Brunswick*, 2008 SCC 9, [2008] 1 S.C.R. 190; *Alberta (Information and Privacy Commissioner) v. Alberta Teachers' Association*, 2011 SCC 61, [2011] 3 S.C.R. 654; *Tervita Corp. v. Canada (Commissioner of Competition)*, 2015 SCC 3, [2015] 1 S.C.R. 161; *Bell Canada v. Bell Aliant Regional Communications*, 2009 SCC 40, [2009] 2 S.C.R. 764; *Re General Increase in Freight Rates* (1954), 76 C.R.T.C. 12; *ATCO Gas and Pipelines Ltd. v. Alberta (Energy and Utilities Board)*, 2006 SCC 4, [2006] 1 S.C.R. 140; *State of Missouri ex rel. Southwestern Bell Telephone Co. v. Public Service Commission of Missouri*, 262 U.S. 276 (1923); *Duquesne Light Co. v. Barasch*, 488 U.S. 299 (1989); *U.S. West Communications, Inc. v. Public Service Commission of Utah*, 901 P.2d 270 (1995); *British Columbia Electric Railway Co. v. Public Utilities Commission of British Columbia*, [1960] S.C.R.

Je serais donc d'avis de rejeter le pourvoi, d'annuler la décision de la Commission et de renvoyer l'affaire à la Commission pour réexamen.

Jurisprudence

Citée par le juge Rothstein

Arrêts examinés : *Enbridge Gas Distribution Inc. c. Ontario Energy Board* (2006), 210 O.A.C. 4; *Northwestern Utilities Ltd. c. Ville d'Edmonton*, [1979] 1 R.C.S. 684; **arrêts mentionnés :** *Toronto Hydro-Electric System Ltd. c. Ontario (Energy Board)*, 2010 ONCA 284, 99 O.R. (3d) 481; *Northwestern Utilities Ltd. c. City of Edmonton*, [1929] R.C.S. 186; *TransCanada Pipelines Ltd. c. Office national de l'Énergie*, 2004 CAF 149; *Ontario Power Generation Inc. (Re)*, EB-2007-0905, November 3, 2008 (en ligne : <http://www.ontarioenergyboard.ca/>); *CAIMAW c. Paccar of Canada Ltd.*, [1989] 2 R.C.S. 983; *B.C.G.E.U. c. Indust. Rel. Council* (1988), 26 B.C.L.R. (2d) 145; *McLean c. Colombie-Britannique (Securities Commission)*, 2013 CSC 67, [2013] 3 R.C.S. 895; *Ellis-Don Ltd. c. Ontario (Commission des relations de travail)*, 2001 CSC 4, [2001] 1 R.C.S. 221; *Tremblay c. Québec (Commission des affaires sociales)*, [1992] 1 R.C.S. 952; *Ontario (Children's Lawyer) c. Ontario (Information and Privacy Commissioner)* (2005), 75 O.R. (3d) 309; *Canada (Procureur général) c. Quadrini*, 2010 CAF 246, [2012] 2 R.C.F. 3; *Leon's Furniture Ltd. c. Information and Privacy Commissioner (Alta.)*, 2011 ABCA 94, 502 A.R. 110; *Henthorne c. British Columbia Ferry Services Inc.*, 2011 BCCA 476, 344 D.L.R. (4th) 292; *United Brotherhood of Carpenters and Joiners of America, Local 1386 c. Bransen Construction Ltd.*, 2002 NBCA 27, 249 R.N.-B. (2^e) 93; *Chandler c. Alberta Association of Architects*, [1989] 2 R.C.S. 848; *Dunsmuir c. Nouveau-Brunswick*, 2008 CSC 9, [2008] 1 R.C.S. 190; *Alberta (Information and Privacy Commissioner) c. Alberta Teachers' Association*, 2011 CSC 61, [2011] 3 R.C.S. 654; *Tervita Corp. c. Canada (Commissaire de la concurrence)*, 2015 CSC 3, [2015] 1 R.C.S. 161; *Bell Canada c. Bell Aliant Communications régionales*, 2009 CSC 40, [2009] 2 R.C.S. 764; *Re General Increase in Freight Rates* (1954), 76 C.R.T.C. 12; *ATCO Gas and Pipelines Ltd. c. Alberta (Energy and Utilities Board)*, 2006 CSC 4, [2006] 1 R.C.S. 140; *State of Missouri ex rel. Southwestern Bell Telephone Co. c. Public Service Commission of Missouri*, 262 U.S. 276 (1923); *Duquesne Light Co. c. Barasch*, 488 U.S. 299 (1989); *U.S. West Communications, Inc. c. Public Service Commission of Utah*, 901 P.2d 270 (1995); *British Columbia Electric Railway Co. c. Public Utilities Commission of British Columbia*, [1960]

837; *Nova Scotia Power Inc., Re*, 2005 NSUARB 27; *Nova Scotia Power Inc. (Re)*, 2012 NSUARB 227.

R.C.S. 837; *Nova Scotia Power Inc., Re*, 2005 NSUARB 27; *Nova Scotia Power Inc. (Re)*, 2012 NSUARB 227.

By Abella J. (dissenting)

Citée par la juge Abella (dissidente)

Verizon Communications Inc. v. Federal Communications Commission, 535 U.S. 467 (2002); *Northwestern Utilities Ltd. v. City of Edmonton*, [1929] S.C.R. 186; *State of Missouri ex rel. Southwestern Bell Telephone Co. v. Public Service Commission of Missouri*, 262 U.S. 276 (1923); *Enersource Hydro Mississauga Inc. (Re)*, 2012 LNONOEB 373 (QL); *Enbridge Gas Distribution Inc. (Re)*, 2002 LNONOEB 4 (QL); *Enbridge Gas Distribution Inc. v. Ontario Energy Board* (2006), 210 O.A.C. 4; *Ontario Power Generation v. Society of Energy Professionals*, [2011] O.L.A.A. No. 117 (QL); *TransCanada Pipelines Ltd. v. National Energy Board*, 2004 FCA 149, 319 N.R. 171.

Verizon Communications Inc. c. Federal Communications Commission, 535 U.S. 467 (2002); *Northwestern Utilities Ltd. c. City of Edmonton*, [1929] R.C.S. 186; *State of Missouri ex rel. Southwestern Bell Telephone Co. c. Public Service Commission of Missouri*, 262 U.S. 276 (1923); *Enersource Hydro Mississauga Inc. (Re)*, 2012 LNONOEB 373 (QL); *Enbridge Gas Distribution Inc. (Re)*, 2002 LNONOEB 4 (QL); *Enbridge Gas Distribution Inc. c. Ontario Energy Board* (2006), 210 O.A.C. 4; *Ontario Power Generation c. Society of Energy Professionals*, [2011] O.L.A.A. No. 117 (QL); *TransCanada Pipelines Ltd. c. Office national de l'Énergie*, 2004 CAF 149.

Statutes and Regulations Cited

Labour Relations Act, 1995, S.O. 1995, c. 1, Sch. A, ss. 56, 69.
Nuclear Safety and Control Act, S.C. 1997, c. 9.
Ontario Energy Board Act, 1998, S.O. 1998, c. 15, Sch. B, ss. 1, 33(3), 78.1.
Payments Under Section 78.1 of the Act, O. Reg. 53/05, ss. 3, 6.
Public Utilities Act, R.S.B.C. 1948, c. 277 [rep. 1973, c. 29, s. 187], s. 16(1)(b).

Lois et règlements cités

Loi de 1995 sur les relations de travail, L.O. 1995, c. 1, ann. A, art. 56, 69.
Loi de 1998 sur la Commission de l'énergie de l'Ontario, L.O. 1998, c. 15, ann. B, art. 1, 33(3), 78.1.
Loi sur la sûreté et la réglementation nucléaires, L.C. 1997, c. 9.
Payments Under Section 78.1 of the Act, O. Reg. 53/05, art. 3, 6.
Public Utilities Act, R.S.B.C. 1948, c. 277 [abr. 1973, c. 29, art. 187], art. 16(1)(b).

Authors Cited

Burns, Robert E., et al. *The Prudent Investment Test in the 1980s*, report NRRI-84-16. Columbus, Ohio: National Regulatory Research Institute, April 1985.
 Chaykowski, Richard P. *An Assessment of the Industrial Relations Context and Outcomes at OPG*, file No. EB-2013-0321, exhibit F4-03-01, attachment 1, September 2013 (online: http://www.opg.com/about/regulatory-affairs/Documents/2014-2015/F4-03-01_Attachment%201.pdf).
 Clark, Ron W., Scott A. Stoll and Fred D. Cass. *Ontario Energy Law: Electricity*. Markham, Ont.: LexisNexis, 2012.
 Falzon, Frank A. V. "Tribunal Standing on Judicial Review" (2008), 21 *C.J.A.L.P.* 21.
 Jacobs, Laverne A., and Thomas S. Kuttner. "Discovering What Tribunals Do: Tribunal Standing Before the Courts" (2002), 81 *Can. Bar Rev.* 616.

Doctrine et autres documents cités

Burns, Robert E., et al. *The Prudent Investment Test in the 1980s*, report NRRI-84-16, Columbus (Ohio), National Regulatory Research Institute, April 1985.
 Chaykowski, Richard P. *An Assessment of the Industrial Relations Context and Outcomes at OPG*, file No. EB-2013-0321, exhibit F4-03-01, attachment 1, September 2013 (en ligne : http://www.opg.com/about/regulatory-affairs/Documents/2014-2015/F4-03-01_Attachment%201.pdf).
 Clark, Ron W., Scott A. Stoll and Fred D. Cass. *Ontario Energy Law : Electricity*, Markham (Ont.), LexisNexis, 2012.
 Falzon, Frank A. V. « Tribunal Standing on Judicial Review » (2008), 21 *R.C.D.A.P.* 21.
 Jacobs, Laverne A., and Thomas S. Kuttner. « Discovering What Tribunals Do : Tribunal Standing Before the Courts » (2002), 81 *R. du B. can.* 616.

Kahn, Jonathan. “Keep *Hope* Alive: Updating the Prudent Investment Standard for Allocating Nuclear Plant Cancellation Costs” (2010), 22 *Fordham Envtl. L. Rev.* 43.

Mullan, David. “Administrative Law and Energy Regulation”, in Gordon Kaiser and Bob Heggie, eds., *Energy Law and Policy*. Toronto: Carswell, 2011, 35.

Ontario. Office of the Auditor General of Ontario. *2011 Annual Report*. Toronto: Queen’s Printer, 2011.

Reid, Laurie, and John Todd. “New Developments in Rate Design for Electricity Distributors”, in Gordon Kaiser and Bob Heggie, eds., *Energy Law and Policy*. Toronto: Carswell, 2011, 519.

Simple, Noel. “The Case for Tribunal Standing in Canada” (2007), 20 *C.J.A.L.P.* 305.

APPEAL from a judgment of the Ontario Court of Appeal (Rosenberg, Goudge and Blair JJ.A.), 2013 ONCA 359, 116 O.R. (3d) 793, 365 D.L.R. (4th) 247, 307 O.A.C. 109, [2013] O.J. No. 3917 (QL), 2013 CarswellOnt 9792 (WL Can.), setting aside a decision of the Divisional Court (Aitken, Swinton and Hoy JJ.), 2012 ONSC 729, 109 O.R. (3d) 576, 347 D.L.R. (4th) 355, [2012] O.J. No. 862 (QL), 2012 CarswellOnt 2710 (WL Can.), and setting aside a decision of the Ontario Energy Board, EB-2010-0008, March 10, 2011 (online: <http://www.ontarioenergyboard.ca/>), 2011 LNONOEB 57 (QL), 2011 CarswellOnt 3723 (WL Can.). Appeal allowed, Abella J. dissenting.

Glenn Zacher, Patrick Duffy and James Wilson, for the appellant.

John B. Laskin, Crawford Smith, Myriam Seers and Carlton Mathias, for the respondent Ontario Power Generation Inc.

Richard P. Stephenson and Emily Lawrence, for the respondent the Power Workers’ Union, Canadian Union of Public Employees, Local 1000.

Paul J. J. Cavalluzzo and Amanda Darrach, for the respondent the Society of Energy Professionals.

Mark Rubenstein, for the intervener.

Kahn, Jonathan. « Keep *Hope* Alive : Updating the Prudent Investment Standard for Allocating Nuclear Plant Cancellation Costs » (2010), 22 *Fordham Envtl. L. Rev.* 43.

Mullan, David. « Administrative Law and Energy Regulation », in Gordon Kaiser and Bob Heggie, eds., *Energy Law and Policy*, Toronto, Carswell, 2011, 35.

Ontario. Bureau du vérificateur général de l’Ontario. *Rapport annuel 2011*, Toronto, Imprimeur de la Reine, 2011.

Reid, Laurie, and John Todd. « New Developments in Rate Design for Electricity Distributors », in Gordon Kaiser and Bob Heggie, eds., *Energy Law and Policy*, Toronto, Carswell, 2011, 519.

Simple, Noel. « The Case for Tribunal Standing in Canada » (2007), 20 *R.C.D.A.P.* 305.

POURVOI contre un arrêt de la Cour d’appel de l’Ontario (les juges Rosenberg, Goudge et Blair), 2013 ONCA 359, 116 O.R. (3d) 793, 365 D.L.R. (4th) 247, 307 O.A.C. 109, [2013] O.J. No. 3917 (QL), 2013 CarswellOnt 9792 (WL Can.), qui a infirmé une décision de la Cour divisionnaire (les juges Aitken, Swinton et Hoy), 2012 ONSC 729, 109 O.R. (3d) 576, 347 D.L.R. (4th) 355, [2012] O.J. No. 862 (QL), 2012 CarswellOnt 2710 (WL Can.), et une décision de la Commission de l’énergie de l’Ontario, EB-2010-0008, 10 mars 2011 (en ligne : <http://www.ontarioenergyboard.ca/>), 2011 LNONOEB 57 (QL), 2011 CarswellOnt 3723 (WL Can.). Pourvoi accueilli, la juge Abella est dissidente.

Glenn Zacher, Patrick Duffy et James Wilson, pour l’appelante.

John B. Laskin, Crawford Smith, Myriam Seers et Carlton Mathias, pour l’intimée Ontario Power Generation Inc.

Richard P. Stephenson et Emily Lawrence, pour l’intimé le Syndicat des travailleurs et travailleuses du secteur énergétique, Syndicat canadien de la fonction publique, section locale 1000.

Paul J. J. Cavalluzzo et Amanda Darrach, pour l’intimée Society of Energy Professionals.

Mark Rubenstein, pour l’intervenante.

The judgment of McLachlin C.J. and Rothstein, Cromwell, Moldaver, Karakatsanis and Gascon JJ. was delivered by

Version française du jugement de la juge en chef McLachlin et des juges Rothstein, Cromwell, Moldaver, Karakatsanis et Gascon rendu par

[1] ROTHSTEIN J. — In Ontario, utility rates are regulated through a process by which a utility seeks approval from the Ontario Energy Board (“Board”) for costs the utility has incurred or expects to incur in a specified period of time. Where the Board approves of costs, they are incorporated into utility rates such that the utility receives payment amounts to cover the approved expenditures. This case concerns the decision of the Board to disallow certain payment amounts applied for by Ontario Power Generation Inc. (“OPG”) as part of its rate application covering the 2011-2012 operating period. Specifically, the Board disallowed \$145 million in labour compensation costs related to OPG’s nuclear operations on the grounds that OPG’s labour costs were out of step with those of comparable entities in the regulated power generation industry.

[1] LE JUGE ROTHSTEIN — En Ontario, la tarification d’un service public est réglementée, de sorte que ce dernier doit obtenir de la Commission de l’énergie de l’Ontario (« Commission ») l’approbation des dépenses qu’il a faites ou qu’il prévoit faire pendant une période donnée. Lorsque cette approbation est obtenue, les tarifs sont rajustés de manière que l’entreprise touche des paiements qui correspondent à ses dépenses. Le présent pourvoi vise la décision de la Commission de refuser certains paiements à Ontario Power Generation Inc. (« OPG ») par suite de sa demande d’approbation de tarifs pour la période 2011-2012. Plus particulièrement, la Commission a refusé d’approuver des dépenses de 145 millions de dollars au titre de la rémunération du personnel affecté aux installations nucléaires au motif que le coût de la main-d’œuvre d’OPG était en rupture avec celui d’organismes comparables dans le secteur réglementé de la production d’énergie.

[2] OPG appealed the Board’s decision to the Ontario Divisional Court. A majority of the court dismissed the appeal and upheld the decision of the Board. OPG then appealed that decision to the Ontario Court of Appeal, which set aside the decisions of the Divisional Court and the Board and remitted the matter to the Board for redetermination in accordance with its reasons. The Board now appeals to this Court.

[2] OPG en a appelé devant la Cour divisionnaire de l’Ontario, dont les juges majoritaires ont rejeté l’appel et confirmé la décision de la Commission. OPG s’est alors adressée à la Cour d’appel de l’Ontario, qui a annulé les décisions de la Cour divisionnaire et de la Commission, puis renvoyé le dossier à la Commission afin qu’elle rende une nouvelle décision conforme à ses motifs. La Commission interjette aujourd’hui appel devant notre Cour.

[3] OPG asserts that the Board’s decision to disallow these labour compensation costs was unreasonable. The crux of OPG’s argument is that the Board is legally required to compensate OPG for all of its prudently committed or incurred costs. OPG asserts that prudence in this context has a particular methodological meaning that requires the Board to assess the reasonableness of OPG’s decisions to incur or commit to costs at the time the decisions to incur or commit to the costs were made and that OPG ought to benefit from a presumption of prudence. Because the Board did not employ this prudence

[3] OPG soutient que le refus de la Commission d’approuver ces dépenses de rémunération de ses employés est déraisonnable. Sa thèse veut essentiellement que la Commission soit légalement tenue de l’indemniser de la totalité des dépenses faites ou convenues avec prudence. OPG prétend que, dans ce contexte, la prudence se définit selon une méthode particulière qui exige de la Commission qu’elle détermine si, au moment où elles ont été prises, les décisions de faire les dépenses ou de convenir des dépenses étaient raisonnables. Elle soutient en outre qu’une présomption de prudence

methodology, OPG argues that its decision was unreasonable.

[4] The Board argues that a particular “prudence test” methodology is not compelled by law, and that in any case the costs disallowed here were not “committed” nuclear compensation costs, but are better characterized as forecast costs.

[5] OPG also raises concerns regarding the Board’s role in acting as a party on appeal from its own decision. OPG argues that in this case, the Board’s aggressive and adversarial defence of its original decision was improper, and that the Board attempted to use the appeal to “bootstrap” its original decision by making additional arguments on appeal.

[6] The Board asserts that the scope of its authority to argue on appeal was settled when it was granted full party rights in connection with the granting of leave by this Court. Alternatively, the Board argues that the structure of utilities regulation in Ontario makes it necessary and important for it to argue the merits of its decisions on appeal.

[7] In my opinion, the labour compensation costs which led to the \$145 million disallowance are best understood as partly committed costs and partly costs subject to management discretion. They are partly committed because they resulted from collective agreements entered into between OPG and two of its unions, and partly subject to management discretion because OPG retained some flexibility to manage total staffing levels in light of, among other things, projected attrition of the workforce. It is not reasonable to treat these costs as entirely forecast. However, I do not agree with OPG that the Board was bound to apply a particular prudence test in evaluating these costs. The *Ontario Energy Board Act, 1998*, S.O. 1998, c. 15, Sch. B, and associated regulations give the Board broad latitude to determine the methodology it uses in assessing utility

doit s’appliquer à son bénéfice. La Commission n’ayant pas eu recours à pareille méthode pour se prononcer sur la prudence d’OPG, sa décision serait déraisonnable.

[4] La Commission rétorque que la loi ne l’oblige pas à employer quelque méthode pour appliquer le « principe de la prudence » et que, de toute manière, les dépenses de rémunération des employés du secteur nucléaire refusées en l’espèce n’étaient pas des dépenses « convenues », mais bien des dépenses prévues.

[5] OPG déplore par ailleurs que la Commission soit partie à l’appel de sa propre décision. Selon elle, la manière agressive et conflictuelle dont la Commission a défendu sa décision initiale n’était pas justifiée, et la Commission tente de se servir de l’appel pour « s’auto-justifier » en formulant de nouveaux arguments à l’appui de sa décision initiale.

[6] La Commission fait valoir que la Cour a circonscrit la faculté qu’elle avait de plaider en appel lorsqu’elle lui a reconnu tous les droits d’une partie au moment d’autoriser le pourvoi. Subsidiairement, elle soutient que la manière dont les services publics sont réglementés en Ontario fait en sorte qu’il est nécessaire et important qu’elle défende la justesse de ses décisions portées en appel.

[7] Il convient mieux, à mon sens, de voir dans les dépenses de rémunération qui ont été refusées à raison de 145 millions de dollars en partie des dépenses convenues et en partie des dépenses relevant du pouvoir discrétionnaire de la direction. Elles sont en partie convenues parce qu’elles résultent de conventions collectives intervenues entre OPG et deux syndicats, et elles relèvent en partie de la discrétion de la direction parce qu’OPG conserve une certaine marge de manœuvre dans la gestion des niveaux de dotation globale compte tenu, entre autres, de l’attrition projetée de l’effectif. Il est déraisonnable de considérer qu’il s’agit en totalité de dépenses prévues. Je ne crois cependant pas, malgré ce qu’affirme OPG, que la Commission était tenue d’appliquer un principe de prudence donné pour apprécier les dépenses. La *Loi de 1998 sur la*

costs, subject to the Board’s ultimate duty to ensure that payment amounts it orders be just and reasonable to both the utility and consumers.

[8] In this case, the nature of the disputed costs and the environment in which they arose provide a sufficient basis to find that the Board did not act unreasonably in disallowing the costs.

[9] Regarding the Board’s role on appeal, I do not find that the Board acted improperly in arguing the merits of this case, nor do I find that the arguments raised on appeal amount to impermissible “bootstrapping”.

[10] Accordingly, I would allow the appeal, set aside the decision of the Court of Appeal, and reinstate the decision of the Board.

I. Regulatory Framework

[11] The *Ontario Energy Board Act, 1998* establishes the Board as a regulatory body with authority to oversee, among other things, electricity generation in the province of Ontario. Section 1 sets out the objectives of the Board in regulating electricity, which include:

1. (1) . . .

1. To protect the interests of consumers with respect to prices and the adequacy, reliability and quality of electricity service.
2. To promote economic efficiency and cost effectiveness in the generation, transmission, distribution, sale and demand management of electricity and to facilitate the maintenance of a financially viable electricity industry.

Accordingly, the Board must ensure that it regulates with an eye to balancing both consumer interests and the efficiency and financial viability of the

Commission de l’énergie de l’Ontario, L.O. 1998, c. 15, ann. B, et ses règlements connexes accordent à la Commission une grande latitude dans le choix d’une méthode pour apprécier les dépenses d’un service public, sous réserve de l’obligation de faire en sorte que, au final, les paiements qu’elle ordonne soient justes et raisonnables vis-à-vis à la fois du service public et du consommateur.

[8] Dans la présente affaire, la nature des dépenses litigieuses et le contexte dans lequel elles ont vu le jour permettent de conclure que la Commission n’a pas agi de manière déraisonnable en refusant de les approuver.

[9] En ce qui concerne la participation de la Commission au pourvoi, je ne crois pas qu’il soit inapproprié qu’elle défende la justesse de sa décision, ni que les arguments qu’elle invoque en appel équivalent à une « autojustification » inadmissible.

[10] Je suis donc d’avis d’accueillir le pourvoi, d’annuler la décision de la Cour d’appel et de rétablir la décision de la Commission.

I. Cadre réglementaire

[11] La *Loi de 1998 sur la Commission de l’énergie de l’Ontario* fait de la Commission un organisme de réglementation investi du pouvoir de surveiller, entre autres choses, la production d’électricité en Ontario. Son article premier énonce les objectifs de la Commission dans la réglementation de l’électricité, dont les suivants :

1. (1) . . .

1. Protéger les intérêts des consommateurs en ce qui concerne les prix, ainsi que la suffisance, la fiabilité et la qualité du service d’électricité.
2. Promouvoir l’efficacité économique et la rentabilité dans les domaines de la production, du transport, de la distribution et de la vente d’électricité ainsi que de la gestion de la demande d’électricité et faciliter le maintien d’une industrie de l’électricité financièrement viable.

La Commission doit donc s’acquitter de sa fonction de réglementation dans le souci d’établir un équilibre entre l’intérêt du consommateur, d’une part,

electricity industry. The Board’s role has also been described as that of a “market proxy”: 2012 ONSC 729, 109 O.R. (3d) 576, at para. 54; 2013 ONCA 359, 116 O.R. (3d) 793, at para. 38. In this sense, the Board’s role is to emulate as best as possible the forces to which a utility would be subject in a competitive landscape: *Toronto Hydro-Electric System Ltd. v. Ontario (Energy Board)*, 2010 ONCA 284, 99 O.R. (3d) 481, at para. 48.

[12] One of the Board’s most powerful tools to achieve its objectives is its authority to fix the amount of payments utilities receive in exchange for the provision of service. Section 78.1(5) of the *Ontario Energy Board Act, 1998* provides in relevant part:

(5) The Board may fix such other payment amounts as it finds to be just and reasonable,

(a) on an application for an order under this section, if the Board is not satisfied that the amount applied for is just and reasonable; . . .

[13] Section 78.1(6) provides: “. . . the burden of proof is on the applicant in an application made under this section”.

[14] As I read these provisions, the utility applies for payment amounts for a future period (called the “test period”). The Board will accept the payment amounts applied for unless the Board is not satisfied that the amounts are just and reasonable. Where the Board is not satisfied, s. 78.1(5) empowers it to fix other payment amounts which it finds to be just and reasonable.

[15] This Court has had the occasion to consider the meaning of similar statutory language in *Northwestern Utilities Ltd. v. City of Edmonton*, [1929] S.C.R. 186. In that case, the Court held that “fair and reasonable” rates were those “which, under the circumstances, would be fair to the consumer on

et l’efficacité et la viabilité financière du secteur de l’électricité, d’autre part. On lui attribue aussi un rôle de « substitut du marché » (2012 ONSC 729, 109 O.R. (3d) 576, par. 54; 2013 ONCA 359, 116 O.R. (3d) 793, par. 38). Sa fonction consiste alors à reproduire au mieux les forces auxquelles serait soumis un service public dans un contexte concurrentiel (*Toronto Hydro-Electric System Ltd. c. Ontario (Energy Board)*, 2010 ONCA 284, 99 O.R. (3d) 481, par. 48).

[12] L’un des leviers les plus puissants dont dispose la Commission pour atteindre ses objectifs réside dans son pouvoir de fixer le montant des paiements que touche l’entreprise pour la prestation du service. Voici l’extrait pertinent du par. 78.1(5) de la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* :

(5) La Commission peut fixer les autres paiements qu’elle estime justes et raisonnables :

a) dans le cadre d’une requête en vue d’obtenir une ordonnance prévue au présent article, si elle n’est pas convaincue que le montant du paiement qui fait l’objet de la requête est juste et raisonnable; . . .

[13] Le paragraphe 78.1(6) dispose pour sa part : « . . . le fardeau de la preuve incombe au requérant dans une requête présentée en vertu du présent article ».

[14] Suivant mon interprétation de ces dispositions, le service public demande des paiements pour une période à venir (appelée « période de référence »). La Commission fait droit à la demande, sauf lorsqu’elle n’est pas convaincue que les paiements demandés sont justes et raisonnables. Lorsqu’elle n’en est pas convaincue, le par. 78.1(5) lui permet de déterminer les paiements qui lui paraissent justes et raisonnables.

[15] Dans l’arrêt *Northwestern Utilities Ltd. c. City of Edmonton*, [1929] R.C.S. 186, la Cour a eu l’occasion de se prononcer sur le sens d’un libellé législatif semblable. Elle a alors statué que la tarification « juste et raisonnable » était celle [TRADUCTION] « qui, dans les circonstances, était juste pour le

the one hand, and which, on the other hand, would secure to the company a fair return for the capital invested” (pp. 192-93).

[16] This means that the utility must, over the long run, be given the opportunity to recover, through the rates it is permitted to charge, its operating and capital costs (“capital costs” in this sense refers to all costs associated with the utility’s invested capital). This case is concerned primarily with operating costs. If recovery of operating costs is not permitted, the utility will not earn its cost of capital, which represents the amount investors require by way of a return on their investment in order to justify an investment in the utility. The required return is one that is equivalent to what they could earn from an investment of comparable risk. Over the long run, unless a regulated utility is allowed to earn its cost of capital, further investment will be discouraged and it will be unable to expand its operations or even maintain existing ones. This will harm not only its shareholders, but also its customers: *TransCanada Pipelines Ltd. v. National Energy Board*, 2004 FCA 149, 319 N.R. 171.

[17] This of course does not mean that the Board must accept every cost that is submitted by the utility, nor does it mean that the rate of return to equity investors is guaranteed. In the short run, return on equity may vary, for example if electricity consumption by the utility’s customers is higher or lower than predicted. Similarly, a disallowance of any operating costs to which the utility has committed itself will negatively impact the return to equity investors. I do not intend to enter into a detailed analysis of how the cost of equity capital should be treated by utility regulators, but merely to observe that any disallowance of costs to which a utility has committed itself has an effect on equity investor returns. This effect must be carefully considered in light of the long-run necessity that utilities be able to attract investors and retain earnings in order to survive and operate efficiently and effectively, in accordance with the statutory objectives of the Board in regulating electricity in Ontario.

consommateur, d’une part, et qui permettait à l’entreprise d’obtenir un juste rendement sur les capitaux investis, d’autre part » (p. 192-193).

[16] Dès lors, le service public doit pouvoir à long terme recouvrer, grâce à la tarification approuvée, ses dépenses d’exploitation et ses coûts en capital, ces derniers s’entendant alors de tous les coûts liés aux capitaux investis par le service public. Le pourvoi vise principalement les dépenses d’exploitation. Si leur recouvrement n’est pas autorisé, le service public n’obtient pas l’équivalent du coût du capital, soit le rendement exigé par les investisseurs pour investir dans le service public. Le rendement exigé équivaut à celui qu’ils pourraient réaliser sur un investissement comportant un risque comparable. À long terme, à moins que le service public réglementé ne puisse obtenir l’équivalent du coût du capital, les nouveaux investissements seront découragés et l’entreprise ne pourra accroître ses activités, ni même les poursuivre. Ce sont non seulement ses actionnaires, mais aussi ses clients, qui en souffriront (*TransCanada Pipelines Ltd. c. Office national de l’Énergie*, 2004 CAF 149).

[17] Évidemment, la Commission n’est pas tenue pour autant d’accepter toute dépense avancée par le service public, et le rendement obtenu par les actionnaires n’est pas non plus garanti. À court terme, ce rendement peut fluctuer, notamment lorsque la consommation d’électricité est supérieure ou inférieure à celle prévue. De même, le refus d’approuver des dépenses d’exploitation dont le service public a convenu aura un effet défavorable sur le rendement des actions. Je n’entends pas me livrer à une analyse détaillée de la manière dont le coût du capital-actions devrait être considéré par les organismes qui réglementent les services publics, mais seulement faire observer que tout refus d’approuver une dépense dont un service public a convenu a un effet sur le rendement des actions. Cet effet justifie une grande attention au vu de la nécessité qu’un service public attire les investissements à long terme et réinvestisse ses bénéfices afin de survivre et de fonctionner de manière efficace et rentable, conformément aux objectifs légaux de la Commission applicables à la réglementation de l’électricité en Ontario.

[18] As noted above, the burden is on the utility to satisfy the Board that the payment amounts it applies for are just and reasonable. If it fails to do so, the Board may disallow the portion of the application that it finds is not for amounts that are just and reasonable.

[19] Where applied-for operating costs are disallowed, the utility, if it is able to do so, may forego the expenditure of such costs. Where the expenditure cannot be foregone, the shareholders of the utility will have to absorb the reduction in the form of receiving less than their anticipated rate of return on their investment, i.e. the utility's cost of equity capital. In such circumstances it will be the management of the utility that will be responsible in the future for bringing its costs into line with what the Board considers just and reasonable.

[20] In order to ensure that the balance between utilities' and consumers' interests is struck, just and reasonable rates must be those that ensure consumers are paying what the Board expects it to cost to efficiently provide the services they receive, taking account of both operating and capital costs. In that way, consumers may be assured that, overall, they are paying no more than what is necessary for the service they receive, and utilities may be assured of an opportunity to earn a fair return for providing those services.

II. Facts

[21] OPG is Ontario's largest energy generator, and is subject to rate regulation by the Board. OPG came into being in 1999 as one of the successor corporations to Ontario Hydro. It operates Board-regulated nuclear and hydroelectric facilities that generate approximately half of Ontario's electricity. Its sole shareholder is the Province of Ontario.

[22] It employs approximately 10,000 people in connection with its regulated facilities, 95 percent of whom work in its nuclear business. Approximately 90 percent of its employees in its regulated

[18] Rappelons qu'il incombe au service public de convaincre la Commission du caractère juste et raisonnable des paiements qu'il sollicite. S'il n'y parvient pas, la Commission peut rejeter la demande en partie à raison du montant qui, selon elle, n'est pas juste et raisonnable.

[19] En cas de refus d'approbation, le service public peut renoncer, si cela lui est possible, aux dépenses d'exploitation en cause. S'il ne peut y renoncer, ses actionnaires absorbent le déficit en touchant un rendement inférieur à celui prévu, c'est-à-dire le coût du capital-actions pour le service public. Il appartient dès lors à la direction de ce dernier de faire en sorte que ses dépenses correspondent à celles que la Commission tient pour justes et raisonnables.

[20] Lorsqu'il s'agit d'assurer l'équilibre entre les intérêts du service public et ceux du consommateur, la tarification juste et raisonnable est celle qui fait en sorte que le consommateur paie ce que la Commission prévoit qu'il en coûtera pour la prestation efficace du service, compte tenu à la fois des dépenses d'exploitation et des coûts en capital. Ainsi, le consommateur a l'assurance que, globalement, il ne paie pas plus que ce qui est nécessaire pour obtenir le service, et le service public a l'assurance de pouvoir toucher une juste contrepartie pour la prestation du service.

II. Faits

[21] OPG est le plus grand producteur d'énergie de l'Ontario, et sa tarification est réglementée par la Commission. Elle a vu le jour en 1999 et fait partie des entreprises qui ont succédé à Ontario Hydro. Elle exploite des installations nucléaires et hydroélectriques soumises à la réglementation de la Commission qui produisent environ la moitié de l'électricité consommée dans la province. Son unique actionnaire est la province d'Ontario.

[22] Son effectif se compose d'environ 10 000 personnes pour ses activités réglementées, dont 95 p. 100 travaillent dans le secteur nucléaire. Environ 90 p.100 des employés affectés à ses activités

businesses are unionized, with approximately two thirds of unionized employees represented by the Power Workers' Union, Canadian Union of Public Employees, Local 1000 ("PWU"), and one third represented by the Society of Energy Professionals ("Society").

[23] Since early in its existence as an independent utility, OPG has been aware of the importance of improving its corporate performance. As part of a general effort to improve its business, OPG undertook efforts to benchmark its nuclear performance against comparable power plants around the world. In a memorandum of agreement ("MOA") with the Province of Ontario dated August 17, 2005, OPG committed to the following:

OPG will seek continuous improvement in its nuclear generation business and internal services. OPG will benchmark its performance in these areas against CANDU nuclear plants worldwide as well as against the top quartile of private and publicly-owned nuclear electricity generators in North America. OPG's top operational priority will be to improve the operation of its existing nuclear fleet.

(A.R., vol. III, at p. 215)

[24] As part of OPG's first-ever rate application with the Board in 2007, for a test period covering the years 2008 and 2009, OPG sought approval for a \$6.4 billion "revenue requirement"; this term refers to "the total revenue that is required by the company to pay all of its allowable expenses and also to recover all costs associated with its invested capital": L. Reid and J. Todd, "New Developments in Rate Design for Electricity Distributors", in G. Kaiser and B. Heggie, eds., *Energy Law and Policy* (2011), 519, at p. 521. This constituted an increase of \$1 billion over the revenue requirement that it had sought and was granted under the regulatory scheme in place prior to the Board's assumption of regulatory authority over OPG: EB-2007-0905, Decision with Reasons, November 3, 2008 ("Board 2008-2009 Decision") (online), at pp. 5-6.

réglementées sont syndiqués, dont approximativement les deux tiers sont représentés par le Syndicat des travailleurs et travailleuses du secteur énergétique, Syndicat canadien de la fonction publique, section locale 1000 (« STTSE »), et le tiers par Society of Energy Professionals (« Society »).

[23] Dès ses débuts en tant que service public indépendant, OPG a eu conscience de l'importance d'accroître sa performance d'entreprise. Dans le cadre de mesures générales prises à cette fin, elle a entrepris de comparer le rendement de son secteur nucléaire à celui de centrales comparables dans le monde. Dans un protocole d'accord intervenu avec la province d'Ontario le 17 août 2005, OPG a pris l'engagement suivant :

[TRADUCTION] OPG visera l'amélioration constante de son secteur nucléaire et de ses services internes. Elle comparera sa performance dans ces domaines à celle de l'exploitation des réacteurs CANDU à travers le monde ainsi qu'à celle des producteurs privés et publics d'électricité d'origine nucléaire appartenant au quartile supérieur en Amérique du Nord. Sa priorité première sera d'améliorer l'exploitation de son parc nucléaire actuel.

(d.a., vol. III, p. 215)

[24] Dans la toute première demande qu'elle a présentée à la Commission en 2007 pour la période de référence 2008-2009, OPG a sollicité l'approbation de « recettes nécessaires » se chiffrant à 6,4 milliards de dollars; ce poste correspond [TRADUCTION] « aux recettes dont l'entreprise a besoin au total pour le paiement de toutes ses dépenses susceptibles d'approbation et, également, pour recouvrer tous les coûts liés aux capitaux investis » (L. Reid et J. Todd, « New Developments in Rate Design for Electricity Distributors », dans G. Kaiser et B. Heggie, dir., *Energy Law and Policy* (2011), 519, p. 521). Il s'agissait d'une majoration d'un milliard de dollars par rapport à ce qu'OPG avait demandé et obtenu en application du régime de réglementation en vigueur avant que la Commission ne soit investie de son pouvoir de réglementation vis-à-vis d'elle (EB-2007-0905, décision motivée, 3 novembre 2008 (« décision 2008-2009 de la Commission ») (en ligne), p. 5-6).

[25] The Board found that OPG was not meeting the nuclear performance expectations of its sole shareholder and that it had done little to conduct benchmarking of its performance against that of its peers, despite its commitment to do so dating back to 2005. Indeed, the only evidence of benchmarking that OPG submitted as part of its rate application was a 2006 report from Navigant Consulting, Inc. (“Navigant Report”), which found that OPG was overstaffed by 12 percent in comparison to its peers. The Board found that OPG had not acted on the recommendations of the Navigant Report and had not commissioned subsequent benchmarking studies to assess its performance (Board 2008-2009 Decision, at pp. 27 and 30). The Board also found that operating costs at OPG’s Pickering nuclear facilities were “far above industry averages” (p. 29). The Board thus disallowed \$35 million of OPG’s proposed revenue requirement and directed OPG to prepare benchmarking studies for use in future applications (p. 31).

[26] In explaining the importance of benchmarking, the Board stated: “The reason why the MOA emphasized benchmarking was because such studies can and do shine a light on inefficiencies and lack of productivity improvement” (Board 2008-2009 Decision, at p. 30).

[27] On May 5, 2010, shortly before OPG was set to file its second rate application, which is the subject of this appeal, the Ontario Minister of Energy and Infrastructure wrote to the President and CEO of OPG to ensure that OPG would demonstrate in its upcoming rate application “concerted efforts to identify cost saving opportunities and focus [its] forthcoming rate application on those items that are essential to the safe and reliable operation of [its] existing assets and projects already under development” (A.R., vol. IV, at p. 38).

[25] La Commission a estimé qu’OPG ne satisfaisait pas aux attentes de son unique actionnaire quant à la performance de son secteur nucléaire et qu’elle avait peu fait pour comparer sa performance à celle de ses pairs, alors qu’elle s’y était engagée dès 2005. De fait, la seule preuve d’une démarche en ce sens présentée par OPG dans le cadre de sa demande d’approbation de tarifs était un rapport établi par Navigant Consulting Inc. en 2006 (« rapport Navigant ») et selon lequel l’effectif d’OPG dépassait de 12 p. 100 celui de ses pairs. La Commission a conclu qu’OPG n’avait pas donné suite aux recommandations du rapport Navigant, ni commandé d’études comparatives ultérieures pour évaluer sa performance (décision 2008-2009 de la Commission, p. 27 et 30). Elle a aussi jugé les coûts d’exploitation d’OPG aux installations nucléaires de Pickering [TRADUCTION] « bien supérieurs à la moyenne du secteur » (p. 29). Elle a donc refusé d’approuver 35 millions de dollars au chapitre des recettes nécessaires et enjoint à OPG de réaliser des études comparatives pour étayer ses demandes ultérieures (p. 31).

[26] Pour expliquer l’importance de la comparaison, la Commission dit ce qui suit : [TRADUCTION] « La raison pour laquelle le protocole d’accord insiste sur la conduite d’une étude comparative est qu’une telle étude peut faire et fait ressortir toute inefficacité ou absence d’accroissement de la productivité » (décision 2008-2009 de la Commission, p. 30).

[27] Le 5 mai 2010, peu avant qu’OPG ne dépose sa deuxième demande d’approbation de tarifs — qui est l’objet du pourvoi —, le ministre de l’Énergie et de l’Infrastructure de l’Ontario a écrit au président-directeur général du service public afin que ce dernier fasse état, dans sa demande, [TRADUCTION] « d’efforts concertés pour trouver des moyens de réaliser des économies et mette l’accent sur les postes de dépense qui sont essentiels à l’exploitation sûre et fiable de ses actifs existants et de ses installations projetées déjà en cours de réalisation » (d.a., vol. IV, p. 38).

[28] On May 26, 2010, OPG filed its payment amounts application for the 2011-2012 test period. As part of its evidence before the Board, OPG submitted two reports by ScottMadden Inc., a general management consulting firm specializing in benchmarking and business planning for nuclear facilities. The Phase 1 report compared OPG's nuclear operational and financial performance against that of external peers using industry performance metrics. The Phase 2 final report discussed performance improvement targets with the intent of improving OPG's nuclear business. OPG collaborated with ScottMadden on the Phase 1 and 2 reports, which were released on July 2, 2009 and September 11, 2009, respectively.

[29] OPG's rate application pertained to a test period beginning on January 1, 2011 and ending on December 31, 2012. OPG sought approval of a \$6.9 billion revenue requirement, which represented an increase of 6.2 percent over OPG's then-current revenue based on the preceding year's approved utility rates. Of the \$6.9 billion revenue requirement sought by OPG, \$2.8 billion pertained to compensation costs, of which approximately \$2.4 billion concerned OPG's nuclear business.

[30] A substantial portion of OPG's wage and compensation expenses was fixed by OPG's collective agreements with the unions, PWU and the Society. At the time of its application, OPG was party to a collective agreement with PWU, effective from April 2009 through March 2012, while its collective agreement with the Society expired on December 31, 2010. These collective agreements provided annual wage increases between 2 percent and 3 percent. OPG forecast an additional 1 percent increase for step progressions and promotions of unionized staff. Following the Board's hearing in this case, an interest arbitrator ordered a new collective agreement between OPG and the Society, effective February 3, 2011. This collective agreement provided wage increases that varied between 1 percent and 3 percent.

[28] Le 26 mai 2010, OPG a déposé sa demande de paiements pour la période de référence 2011-2012. Elle a présenté à l'appui deux rapports de ScottMadden Inc., un cabinet-conseil en gestion générale spécialisé dans la comparaison et la planification opérationnelle d'installations nucléaires. Le rapport de la phase 1 compare la performance opérationnelle et financière d'OPG à celle d'autres entreprises à partir de mesures de la performance dans le secteur d'activité. Le rapport final de la phase 2 porte sur les objectifs d'accroissement de la performance dans l'optique d'une amélioration de l'exploitation du secteur nucléaire. OPG a collaboré avec ScottMadden pour l'établissement des rapports des phases 1 et 2, qui ont respectivement été publiés les 2 juillet et 11 septembre 2009.

[29] La demande visait la période allant du 1^{er} janvier 2011 au 31 décembre 2012. OPG y demandait l'approbation de recettes nécessaires de 6,9 milliards de dollars, soit une augmentation de 6,2 p. 100 par rapport aux recettes d'alors compte tenu des tarifs approuvés pour la période précédente. Des 6,9 milliards de dollars sollicités au titre des recettes nécessaires, 2,8 milliards auraient été affectés à la rémunération, dont environ 2,4 milliards dans le secteur nucléaire.

[30] Une grande partie des dépenses d'OPG au chapitre des salaires et de la rémunération était déterminée par des conventions collectives intervenues avec les syndicats (STTSE et Society). Lors du dépôt de la demande, OPG était liée par une convention collective conclue avec le STTSE en vigueur d'avril 2009 à mars 2012, alors que la convention collective qui la liait à Society avait expiré le 31 décembre 2010. Ces conventions collectives prévoyaient des augmentations annuelles de salaires se situant entre 2 et 3 p. 100, auxquelles s'ajoutait 1 p. 100 pour les changements d'échelon et l'avancement. Après l'audition de la demande par la Commission dans la présente affaire, un arbitre a ordonné l'application d'une nouvelle convention collective liant OPG et Society à compter du 3 février 2011. La convention collective prévoyait des augmentations de salaires de 1 à 3 p. 100.

III. Judicial History

A. *Ontario Energy Board: 2011 LNNOEB 57 (QL)* (“Board Decision”)

[31] In its decision concerning OPG’s rate application for the 2011-2012 test period, the Board stated that it enjoyed broad discretion pursuant to Ontario Regulation 53/05 (*Payments Under Section 78.1 of the Act*) and s. 78.1 of the *Ontario Energy Board Act, 1998* to “adopt the mechanisms it judges appropriate in setting just and reasonable rates” (para. 73). The Board recognized that different tests could apply depending on whether its analysis concerned the recovery of forecast costs or an after-the-fact review of costs already incurred. In this rate application, it was appropriate to take into consideration all evidence that the Board deemed relevant to assess the reasonableness of OPG’s revenue requirement.

[32] The Board rejected OPG’s proposed revenue requirement of \$6.9 billion, reducing it by \$145 million over the test period “to send a clear signal that OPG must take responsibility for improving its performance” (para. 350). Key to its disallowance was the Board’s finding that OPG was overstaffed and that its compensation levels were excessive.

[33] Regarding the number of staff, the Board pointed out that a benchmarking study commissioned by OPG itself, the ScottMadden Phase 2 final report, suggested that certain staff positions could be reduced or eliminated altogether. The Board suggested that OPG could review its organizational structure and reassign or eliminate positions in the coming years, as 20 percent to 25 percent of its staff were set to retire between 2010 and 2014 and it was possible to make greater use of external contractors. Regarding compensation, the Board found that OPG had not submitted compelling evidence justifying the benchmarking of its salaries of non-management employees to the 75th percentile of a survey of

III. Historique judiciaire

A. *Commission de l’énergie de l’Ontario (2011 LNNOEB 57 (QL))* (« décision de la Commission »)

[31] Dans sa décision relative à la demande d’approbation de tarifs d’OPG pour la période de référence 2011-2012, la Commission dit que le règlement 53/05 de l’Ontario (*Payments Under Section 78.1 of the Act*) (« règlement 53/05 ») et l’art. 78.1 de la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* lui confèrent un vaste pouvoir discrétionnaire quant [TRADUCTION] « au choix d’une méthode indiquée pour fixer des tarifs justes et raisonnables » (para. 73). Elle reconnaît que différents principes peuvent s’appliquer selon qu’il s’agit du recouvrement de dépenses prévues ou de l’examen après coup de dépenses déjà faites. Pour statuer sur la demande dont elle était saisie, il convenait de tenir compte de tout élément de preuve que la Commission jugeait pertinent pour apprécier le caractère raisonnable des recettes nécessaires d’OPG.

[32] La Commission refuse d’approuver les 6,9 milliards de dollars demandés par OPG au titre des recettes nécessaires, les réduisant de 145 millions de dollars pour la période référence [TRADUCTION] « afin de signifier clairement à OPG qu’il lui incombe d’accroître sa performance » (par. 350). Cette décision défavorable tient surtout à l’opinion de la Commission selon laquelle OPG compte trop d’employés et ses niveaux de rémunération sont excessifs.

[33] Au sujet de la taille de l’effectif, la Commission relève que, selon une étude comparative qu’OPG a elle-même commandée (le rapport final de la phase 2 de ScottMadden), la dotation de certains postes peut être réduite, voire supprimée. Elle recommande à OPG de revoir sa structure organisationnelle et de réaffecter du personnel ou de supprimer des postes au cours des années suivantes. Vingt à vingt-cinq pour cent du personnel d’OPG devait en effet partir à la retraite entre 2010 et 2014 et il était possible de recourir davantage à la sous-traitance. Au chapitre de la rémunération, elle estime qu’OPG n’a pas présenté d’éléments convaincants pour justifier que les salaires de son personnel opérationnel

industry salaries conducted by Towers Perrin. Instead, the Board considered the proper benchmark to be the 50th percentile, the same percentile against which OPG benchmarks management compensation. In determining the appropriate disallowance, the Board acknowledged that OPG may not have been able to achieve the full \$145 million in savings for the test period through the reduction of compensation levels alone because of its collective agreements with the unions.

B. *Ontario Superior Court of Justice, Divisional Court: 2012 ONSC 729, 109 O.R. (3d) 576*

[34] OPG appealed the Board Decision on the basis that it was unreasonable and that the reasons provided were inadequate. OPG argued that the Board should have conducted a prudent investment test — that is, it should have restricted its review of compensation costs to a consideration of whether the collective agreements that prescribed the compensation costs were prudent at the time they were entered into. OPG also argued that the Board should have presumed that the costs were prudent.

[35] The panel of three Divisional Court judges was split. Justice Hoy (as she then was), for the majority, found the Board Decision reasonable because management had the ability to reduce total compensation costs in the future within the framework of the collective agreement. Applying a strict prudent investment test would not permit the Board to fulfill its statutory objective of promoting cost effectiveness in the generation of electricity. It was particularly important for the Board to exercise its authority to set just and reasonable rates given the “double monopoly” dynamic at play:

The collective agreements were concluded between a regulated monopoly, which passes costs on to consumers, not a competitive enterprise, and two unions which account for approximately 90 per cent of the employees and amount to a near, second monopoly, based on terms

se situent au 75^e percentile des salaires versés dans le secteur selon une étude de Towers Perrin. Selon la Commission, ils devraient se situer au 50^e percentile, soit le même que pour le personnel de direction. Pour décider de la réduction qui s’impose, elle reconnaît qu’OPG pourrait ne pas être en mesure, pendant la période de référence, de réaliser des économies de 145 millions de dollars par la réduction de sa seule masse salariale à cause des conventions collectives en vigueur.

B. *Cour supérieure de Justice de l’Ontario, Cour divisionnaire (2012 ONSC 729, 109 O.R. (3d) 576)*

[34] OPG a fait appel de la décision au motif que celle-ci était déraisonnable et mal motivée. Elle a soutenu que la Commission aurait dû appliquer le principe de l’investissement prudent, c’est-à-dire que, dans son examen des dépenses de rémunération, elle aurait dû seulement s’interroger sur la prudence de conclure, à l’époque, les conventions collectives qui commandaient ces dépenses. Elle a ajouté que la Commission aurait dû présumer que les dépenses étaient prudentes.

[35] La décision de la formation de trois juges de la Cour divisionnaire est partagée. Au nom des juges majoritaires, la juge Hoy (aujourd’hui Juge en chef adjointe de l’Ontario) conclut que la décision de la Commission est raisonnable, car il était possible à la direction d’OPG de réduire ultérieurement ses dépenses globales de rémunération dans le respect des conventions collectives. L’application stricte du principe de l’investissement prudent n’aurait pas permis à la Commission d’atteindre son objectif, d’origine législative, de favoriser la rentabilité de la production d’électricité. Vu la présence de « deux monopoles », il importait particulièrement que la Commission exerce son pouvoir de fixer des tarifs justes et raisonnables :

[TRADUCTION] Les conventions collectives sont intervenues entre un monopole réglementé qui refile ses coûts au consommateur et qui n’est pas soumis à la concurrence, et deux syndicats qui représentent environ 90 p. 100 des salariés et qui constituent presque un second monopole

inherited from Ontario Hydro and in face of the reality that running a nuclear operation without the employees would be extremely difficult. [para. 54]

[36] Justice Aitken dissented, finding that,

to the extent that [nuclear compensation] costs were pre-determined, in the sense that they were locked in as a result of collective agreements entered prior to the date of the application and the test period, OPG only had to prove their prudence or reasonableness based on the circumstances that were known or that reasonably could have been anticipated at the time the decision to enter those collective agreements was made. [para. 83]

She would have held that the Board’s failure to undertake a separate and explicit prudence review for the committed portion of nuclear compensation costs, coupled with its consideration of hindsight factors in assessing the reasonableness of these costs, rendered the Board Decision unreasonable.

C. Ontario Court of Appeal: 2013 ONCA 359, 116 O.R. (3d) 793

[37] The Ontario Court of Appeal reversed the Divisional Court’s decision and remitted the case to the Board. The court drew a distinction between forecast costs and committed costs, with committed costs being those that the utility “is committed to pay in [the test period]” and that “cannot be managed or reduced by the utility in that time frame, usually because of contractual obligations” (para. 29). Although costs may not require actual payment until the future, as in this case, costs that have been “contractually incurred to be paid over the time frame are nonetheless committed even though they have not yet been paid” (para. 29). When reviewing such costs, the court held that the Board must undertake a prudence review as described in *Enbridge Gas Distribution Inc. v. Ontario Energy Board* (2006), 210 O.A.C. 4 (paras. 15-16). By failing to follow this jurisprudence and by requiring that OPG “manage costs that, by law, it cannot manage”, the Board acted unreasonably (para. 37).

étant donné les conditions héritées d’Ontario Hydro et le fait qu’il serait extrêmement difficile d’exploiter des installations nucléaires sans les salariés. [par. 54]

[36] Dissidente, la juge Aitken opine que,

[TRADUCTION] dans la mesure où les coûts [de rémunération des employés du secteur nucléaire] étaient déterminés à l’avance, c’est-à-dire qu’ils étaient arrêtés par des conventions collectives conclues avant la demande et la période de référence, OPG devait seulement prouver la prudence ou le caractère raisonnable de la décision de conclure ces conventions au vu des circonstances connues ou qui auraient pu raisonnablement être prévues au moment de prendre la décision. [par. 83]

Elle aurait statué que l’omission de la Commission d’appliquer séparément et expressément le principe de la prudence à la partie des dépenses de rémunération du secteur nucléaire dont elle avait convenu, jumelée à son appréciation avec le recul du caractère raisonnable de ces dépenses, a rendu la décision de la Commission déraisonnable.

C. Cour d’appel de l’Ontario (2013 ONCA 359, 116 O.R. (3d) 793)

[37] La Cour d’appel de l’Ontario infirme le jugement de la Cour divisionnaire et renvoie le dossier à la Commission. Elle établit une distinction entre les dépenses prévues et les dépenses convenues, ces dernières correspondant à celles que le service public [TRADUCTION] « a convenu d’acquitter pendant [la période de référence] » et qu’il « ne peut modifier ou réduire pendant cette période, généralement à cause d’obligations contractuelles » (par. 29). Même si les dépenses n’ont pas à être acquittées dans l’immédiat, comme en l’espèce, celles qui, « par contrat, doivent être acquittées pendant la période de référence constituent néanmoins des dépenses convenues, même si elles n’ont pas encore été acquittées » (par. 29). La Cour d’appel statue que la Commission doit, dans son examen de ces dépenses, appliquer le principe de la prudence énoncé dans *Enbridge Gas Distribution Inc. c. Ontario Energy Board* (2006), 210 O.A.C. 4 (par. 15-16). En ne respectant pas ce précédent et en obligeant OPG à « modifier des dépenses qu’elle ne peut juridiquement modifier », la Commission a agi déraisonnablement (par. 37).

IV. Issues

[38] The Board raises two issues on appeal:

1. What is the appropriate standard of review?
2. Was the Board’s decision to disallow \$145 million of OPG’s revenue requirement reasonable?

[39] Before this Court, OPG has argued that the Board stepped beyond the appropriate role of a tribunal in an appeal from its own decision, which raises the following additional issue:

3. Did the Board act impermissibly in pursuing its appeal in this case?

V. Analysis

[40] It is logical to begin by considering the appropriateness of the Board’s participation in the appeal. I will next consider the appropriate standard of review, and then the merits issue of whether the Board’s decision in this case was reasonable.

A. *The Appropriate Role of the Board in This Appeal*

(1) Tribunal Standing

[41] In *Northwestern Utilities Ltd. v. City of Edmonton*, [1979] 1 S.C.R. 684 (“*Northwestern Utilities*”), per Estey J., this Court first discussed how an administrative decision-maker’s participation in the appeal or review of its own decisions may give rise to concerns over tribunal impartiality. Estey J. noted that “active and even aggressive participation can have no other effect than to discredit the impartiality of an administrative tribunal either in the case where the matter is referred back to it, or in future proceedings involving similar interests and

IV. Questions en litige

[38] La Commission soulève deux questions dans le cadre du pourvoi :

1. Quelle est la norme de contrôle applicable?
2. Sa décision de retrancher 145 millions de dollars des recettes nécessaires d’OPG est-elle raisonnable?

[39] Devant notre Cour, OPG fait valoir que la Commission outrepassa le rôle qui sied à un tribunal administratif dans le cadre d’un appel de sa propre décision, ce qui soulève la question supplémentaire suivante :

3. La Commission a-t-elle agi de manière inacceptable en se pourvoyant en tant que partie à l’appel en l’espèce?

V. Analyse

[40] Il convient en toute logique d’examiner d’abord le caractère approprié de la participation de la Commission au pourvoi. J’examinerai ensuite la norme de contrôle applicable, puis la question de fond de savoir si la décision de la Commission est raisonnable.

A. *Le rôle qui sied à la Commission dans le cadre du pourvoi*

(1) La qualité pour agir d’un tribunal administratif

[41] Dans *Northwestern Utilities Ltd. c. Ville d’Edmonton*, [1979] 1 R.C.S. 684 (« *Northwestern Utilities* »), sous la plume du juge Estey, notre Cour se demande pour la première fois en quoi la participation d’un décideur administratif à l’appel ou au contrôle de sa propre décision peut soulever des doutes sur son impartialité. Pour reprendre les propos du juge Estey, « [u]ne participation aussi active ne peut que jeter le discrédit sur l’impartialité d’un tribunal administratif lorsque l’affaire lui est renvoyée ou lorsqu’il est saisi d’autres procédures

issues or the same parties” (p. 709). He further observed that tribunals already receive an opportunity to make their views clear in their original decisions: “. . . it abuses one’s notion of propriety to countenance its participation as a full-fledged litigant in this Court” (p. 709).

[42] The Court in *Northwestern Utilities* ultimately held that the Alberta Public Utilities Board — which, like the Ontario Energy Board, had a statutory right to be heard on judicial appeal (see *Ontario Energy Board Act, 1998*, s. 33(3)) — was limited in the scope of the submissions it could make. Specifically, Estey J. observed that

[i]t has been the policy in this Court to limit the role of an administrative tribunal whose decision is at issue before the Court, even where the right to appear is given by statute, to an explanatory role with reference to the record before the Board and to the making of representations relating to jurisdiction. [p. 709]

[43] This Court further considered the issue of agency standing in *CAIMAW v. Paccar of Canada Ltd.*, [1989] 2 S.C.R. 983, which involved judicial review of a British Columbia Labour Relations Board decision. Though a majority of the judges hearing the case did not endorse a particular approach to the issue, La Forest J., Dickson C.J. concurring, accepted that a tribunal had standing to explain the record and advance its view of the appropriate standard of review and, additionally, to argue that its decision was reasonable.

[44] This finding was supported by the need to make sure the Court’s decision on review of the tribunal’s decision was fully informed. La Forest J. cited *B.C.G.E.U. v. Indust. Rel. Council* (1988), 26 B.C.L.R. (2d) 145 (C.A.), at p. 153, for the proposition that the tribunal is the party best equipped to draw the Court’s attention to

concernant des intérêts et des questions semblables ou impliquant les mêmes parties » (p. 709). Il ajoute que le tribunal administratif avait déjà le loisir de s’expliquer clairement dans sa décision initiale et « [qu’il] enfreint de façon inacceptable la réserve dont [il doit] faire preuve lorsqu’[il] particip[e] aux procédures comme partie à part entière » (p. 709).

[42] Dans *Northwestern Utilities*, notre Cour statue finalement que la portée des observations que pouvait présenter l’Alberta Public Utilities Board — qui, à l’instar de la Commission de l’énergie de l’Ontario, jouissait légalement du droit d’être entendue en appel devant une cour de justice (voir la *Loi de 1998 sur la Commission de l’énergie de l’Ontario*, par. 33(3)) — était limitée. Le juge Estey fait remarquer ce qui suit :

Cette Cour, à cet égard, a toujours voulu limiter le rôle du tribunal administratif dont la décision est contestée à la présentation d’explications sur le dossier dont il était saisi et d’observations sur la question de sa compétence, même lorsque la loi lui confère le droit de comparaître. [p. 709]

[43] Dans *CAIMAW c. Paccar of Canada Ltd.*, [1989] 2 R.C.S. 983, qui porte sur le contrôle judiciaire d’une décision de la commission des relations de travail de la Colombie-Britannique, notre Cour approfondit la question de la qualité pour agir d’un organisme administratif. Même si les juges majoritaires qui ont entendu le pourvoi n’adoptent pas d’approche particulière pour se prononcer, le juge La Forest, avec l’appui du juge en chef Dickson, reconnaît qu’un tribunal administratif a qualité non seulement pour expliquer le dossier et faire valoir son point de vue sur la norme de contrôle applicable, mais aussi pour soutenir que sa décision est raisonnable.

[44] Cette conclusion repose sur la nécessité de faire en sorte que la cour de révision rende un jugement parfaitement éclairé sur la décision du tribunal administratif. Le juge La Forest invoque l’arrêt *B.C.G.E.U. c. Indust. Rel. Council* (1988), 26 B.C.L.R. (2d) 145 (C.A.), p. 153, pour avancer que le tribunal administratif est le mieux placé pour attirer l’attention de la cour

those considerations, rooted in the specialized jurisdiction or expertise of the tribunal, which may render reasonable what would otherwise appear unreasonable to someone not versed in the intricacies of the specialized area.

(*Paccar*, at p. 1016)

La Forest J. found, however, that the tribunal could not go so far as to argue that its decision was correct (p. 1017). Though La Forest J. did not command a majority, L'Heureux-Dubé J. also commented on tribunal standing in her dissent, and agreed with the substance of La Forest J.'s analysis (p. 1026).

[45] Trial and appellate courts have struggled to reconcile this Court's statements in *Northwestern Utilities* and *Paccar*. Indeed, while this Court has never expressly overturned *Northwestern Utilities*, on some occasions, it has permitted tribunals to participate as full parties without comment: see, e.g., *McLean v. British Columbia (Securities Commission)*, 2013 SCC 67, [2013] 3 S.C.R. 895; *Ellis-Don Ltd. v. Ontario (Labour Relations Board)*, 2001 SCC 4, [2001] 1 S.C.R. 221; *Tremblay v. Quebec (Commission des affaires sociales)*, [1992] 1 S.C.R. 952; see also *Ontario (Children's Lawyer) v. Ontario (Information and Privacy Commissioner)* (2005), 75 O.R. (3d) 309 (C.A.) ("*Goodis*"), at para. 24.

[46] A number of appellate decisions have grappled with this issue and "for the most part now display a more relaxed attitude in allowing tribunals to participate in judicial review proceedings or statutory appeals in which their decisions were subject to attack": D. Mullan, "Administrative Law and Energy Regulation", in G. Kaiser and B. Heggie, 35, at p. 51. A review of three appellate decisions suffices to establish the rationale behind this shift.

[47] In *Goodis*, the Children's Lawyer urged the court to refuse or limit the standing of the Information and Privacy Commissioner, whose decision

sur les considérations, enracinées dans la compétence ou les connaissances spécialisées du tribunal, qui peuvent rendre raisonnable ce qui autrement paraîtrait déraisonnable à quelqu'un qui n'est pas versé dans les complexités de ce domaine spécialisé.

(*Paccar*, p. 1016)

Toutefois, le juge La Forest conclut que le tribunal administratif ne peut aller jusqu'à défendre le bien-fondé de sa décision (p. 1017). Sa thèse ne convainc pas une majorité de ses collègues, mais la juge L'Heureux-Dubé, dissidente, qui se prononce elle aussi sur la qualité pour agir du tribunal administratif, souscrit à son analyse sur le fond (p. 1026).

[45] Juridictions de première instance et d'appel ont tenté tant bien que mal de concilier les opinions exprimées par les juges de la Cour dans les arrêts *Northwestern Utilities* et *Paccar*. De fait, même si notre Cour n'est jamais expressément revenue sur *Northwestern Utilities*, elle a parfois autorisé un tribunal administratif à participer à l'instance à titre de partie à part entière sans expliquer sa décision (voir p. ex. *McLean c. Colombie-Britannique (Securities Commission)*, 2013 CSC 67, [2013] 3 R.C.S. 895; *Ellis-Don Ltd. c. Ontario (Commission des relations de travail)*, 2001 CSC 4, [2001] 1 R.C.S. 221; *Tremblay c. Québec (Commission des affaires sociales)*, [1992] 1 R.C.S. 952; voir également *Ontario (Children's Lawyer) c. Ontario (Information and Privacy Commissioner)* (2005), 75 O.R. (3d) 309 (C.A.) (« *Goodis* »), par. 24).

[46] Dans un certain nombre de décisions, les cours d'appel se sont attaquées à la question et, [TRADUCTION] « pour la plupart, elles sont désormais plus enclines à autoriser un tribunal administratif à participer au contrôle judiciaire ou à l'appel, prévu par la loi, de sa propre décision » (D. Mullan, « Administrative Law and Energy Regulation », dans G. Kaiser et B. Heggie, 35, p. 51). Le survol de trois arrêts de juridictions d'appel suffit à établir la raison d'être de ce revirement.

[47] Dans *Goodis*, le Bureau de l'avocate des enfants demandait à la cour de ne pas reconnaître ou de restreindre la qualité pour agir du Commissaire

was under review. The Ontario Court of Appeal declined to apply any formal, fixed rule that would limit the tribunal to certain categories of submissions and instead adopted a contextual, discretionary approach: *Goodis*, at paras. 32-34. The court found no principled basis for the categorical approach, and observed that such an approach may lead to undesirable consequences:

For example, a categorical rule denying standing if the attack asserts a denial of natural justice could deprive the court of vital submissions if the attack is based on alleged deficiencies in the structure or operation of the tribunal, since these are submissions that the tribunal is uniquely placed to make. Similarly, a rule that would permit a tribunal standing to defend its decision against the standard of reasonableness but not against one of correctness, would allow unnecessary and prevent useful argument. Because the best argument that a decision is reasonable may be that it is correct, a rule based on this distinction seems tenuously founded at best as Robertson J.A. said in *United Brotherhood of Carpenters and Joiners of America, Local 1386 v. Bransen Construction Ltd.*, [2002] N.B.J. No. 114, 249 N.B.R. (2d) 93 (C.A.), at para. 32.

(*Goodis*, at para. 34)

[48] The court held that *Northwestern Utilities* and *Paccar* should be read as the source of “fundamental considerations” that should guide the court’s exercise of discretion in the context of the case: *Goodis*, at para. 35. The two most important considerations, drawn from those cases, were the “importance of having a fully informed adjudication of the issues before the court” (para. 37), and “the importance of maintaining tribunal impartiality”: para. 38. The court should limit tribunal participation if it will undermine future confidence in its objectivity. The court identified a list of factors, discussed further below, that may aid in determining whether and

à l’information et à la protection de la vie privée dont la décision faisait l’objet d’une demande de contrôle. La Cour d’appel de l’Ontario a refusé de se montrer formaliste et d’appliquer une règle fixe qui aurait obligé le tribunal administratif à s’en tenir à des observations d’un certain type et elle a adopté plutôt une approche contextuelle et discrétionnaire (*Goodis*, par. 32-34). Elle a conclu que l’approche catégorique n’avait pas de fondement rationnel et a fait remarquer qu’une telle approche pouvait avoir des conséquences fâcheuses :

[TRADUCTION] Par exemple, la règle catégorique qui refuse au tribunal administratif la qualité pour agir lorsque la contestation allègue le déni de justice naturelle peut priver la cour d’observations capitales lorsque la contestation se fonde des défaillances alléguées de la structure ou du fonctionnement du tribunal administratif, car ce sont des sujets sur lesquels ce dernier est particulièrement bien placé pour formuler des observations. De même, la règle qui reconnaît à un tribunal administratif la qualité pour défendre sa décision au regard du critère de la raisonabilité, mais non du critère de la décision correcte, permet le débat inutile et empêche le débat utile. Parce que le meilleur moyen d’établir la raisonabilité d’une décision peut être de démontrer qu’elle est correcte, une règle fondée sur cette distinction semble au mieux ténue, comme l’affirme le juge Robertson dans *Fraternité unie des charpentiers et menuisiers d’Amérique, section locale 1386 c. Bransen Construction Ltd.*, [2002] A.N.-B. n° 114, 249 R.N.-B. (2^e) 93 (C.A.), par. 32.

(*Goodis*, par. 34)

[48] La Cour d’appel statue qu’il faut voir dans les arrêts *Northwestern Utilities* et *Paccar* la source de [TRADUCTION] « considérations fondamentales » qui doivent guider l’exercice de son pouvoir discrétionnaire eu égard au contexte de l’affaire (*Goodis*, par. 35). Les deux considérations les plus importantes, selon ces arrêts, sont « la nécessité de faire en sorte que la cour rende une décision parfaitement éclairée sur les questions en litige » (par. 37) et « celle d’assurer l’impartialité du tribunal administratif » (par. 38). La cour doit limiter la participation du tribunal administratif lorsque cette participation est de nature à miner la confiance

to what extent the tribunal should be permitted to make submissions: paras. 36-38.

[49] In *Canada (Attorney General) v. Quadrini*, 2010 FCA 246, [2012] 2 F.C.R. 3, Stratas J.A. identified two common law restrictions that, in his view, restricted the scope of a tribunal's participation on appeal from its own decision: finality and impartiality. Finality, the principle whereby a tribunal may not speak on a matter again once it has decided upon it and provided reasons for its decision, is discussed in greater detail below, as it is more directly related to concerns surrounding "bootstrapping" rather than agency standing itself.

[50] The principle of impartiality is implicated by tribunal argument on appeal, because decisions may in some cases be remitted to the tribunal for further consideration. Stratas J.A. found that "[s]ubmissions by the tribunal in a judicial review proceeding that descend too far, too intensely, or too aggressively into the merits of the matter before the tribunal may disable the tribunal from conducting an impartial redetermination of the merits later": *Quadrini*, at para. 16. However, he ultimately found that these principles did not mandate "hard and fast rules", and endorsed the discretionary approach set out by the Ontario Court of Appeal in *Goodis*: *Quadrini*, at paras. 19-20.

[51] A third example of recent judicial consideration of this issue may be found in *Leon's Furniture Ltd. v. Information and Privacy Commissioner (Alta.)*, 2011 ABCA 94, 502 A.R. 110. In this case, Leon's Furniture challenged the Commissioner's standing to make submissions on the merits of the appeal (para. 16). The Alberta Court of Appeal, too, adopted the position that the law should respond to the fundamental concerns raised in *Northwestern*

ultérieure des citoyens dans son objectivité. La Cour d'appel énumère les considérations — sur lesquelles je reviendrai — qui jouent dans la décision d'autoriser ou non le tribunal administratif à présenter des observations et dans la détermination de la mesure dans laquelle il lui est permis de le faire, le cas échéant (par. 36-38).

[49] Dans *Canada (Procureur général) c. Quadrini*, 2010 CAF 246, [2012] 2 R.C.F. 3, le juge Stratas relève deux considérations qui, en common law, limitent selon lui la participation éventuelle d'un tribunal administratif à l'appel de sa propre décision : le caractère définitif et l'impartialité. Le principe du caractère définitif veut qu'un tribunal ne puisse se prononcer de nouveau dans une affaire une fois qu'il a rendu sa décision, motifs à l'appui. J'y reviendrai plus en détail, car j'estime que ce principe se rapporte plus directement à l'« autojustification » de sa décision par le tribunal administratif qu'à sa qualité pour agir comme telle.

[50] Le principe de l'impartialité entre en jeu lorsque le tribunal administratif défend une thèse en appel car, dans certains cas, sa décision peut lui être renvoyée pour réexamen. Le juge Stratas conclut que « [l]es observations que le tribunal administratif présente dans une instance en contrôle judiciaire et qui plongent trop loin, trop intensément ou trop énergiquement dans le bien-fondé de l'affaire soumise au tribunal administratif risquent d'empêcher celui-ci de procéder par la suite à un réexamen impartial du bien-fondé de l'affaire » (*Quadrini*, par. 16). Il conclut toutefois au final que les principes applicables n'imposaient pas de « règles absolues », et il souscrit à l'approche discrétionnaire de la Cour d'appel de l'Ontario dans *Goodis* (*Quadrini*, par. 19-20).

[51] L'arrêt *Leon's Furniture Ltd. c. Information and Privacy Commissioner (Alta.)*, 2011 ABCA 94, 502 A.R. 110, constitue un troisième exemple récent où une cour de justice est appelée à se pencher sur le sujet. Leon's Furniture a contesté la qualité du commissaire intimé de plaider sur le fond en appel (par. 16). La Cour d'appel de l'Alberta estime elle aussi que le droit applicable doit donner suite aux considérations fondamentales soulevées dans

Utilities but should nonetheless approach the question of tribunal standing with discretion, to be exercised in view of relevant contextual considerations: paras. 28-29.

[52] The considerations set forth by this Court in *Northwestern Utilities* reflect fundamental concerns with regard to tribunal participation on appeal from the tribunal's own decision. However, these concerns should not be read to establish a categorical ban on tribunal participation on appeal. A discretionary approach, as discussed by the courts in *Goodis*, *Leon's Furniture*, and *Quadrini*, provides the best means of ensuring that the principles of finality and impartiality are respected without sacrificing the ability of reviewing courts to hear useful and important information and analysis: see N. Semple, "The Case for Tribunal Standing in Canada" (2007), 20 *C.J.A.L.P.* 305; L. A. Jacobs and T. S. Kuttner, "Discovering What Tribunals Do: Tribunal Standing Before the Courts" (2002), 81 *Can. Bar Rev.* 616; F. A. V. Falzon, "Tribunal Standing on Judicial Review" (2008), 21 *C.J.A.L.P.* 21.

[53] Several considerations argue in favour of a discretionary approach. Notably, because of their expertise and familiarity with the relevant administrative scheme, tribunals may in many cases be well positioned to help the reviewing court reach a just outcome. For example, a tribunal may be able to explain how one interpretation of a statutory provision might impact other provisions within the regulatory scheme, or the factual and legal realities of the specialized field in which they work. Submissions of this type may be harder for other parties to present.

[54] Some cases may arise in which there is simply no other party to stand in opposition to the party challenging the tribunal decision. Our judicial review processes are designed to function best when both sides of a dispute are argued vigorously before the reviewing court. In a situation where no other well-informed party stands opposed, the presence of a tribunal as an adversarial party may help the court

l'arrêt *Northwestern Utilities*, mais que la question de la qualité pour agir d'un tribunal administratif relève néanmoins d'un pouvoir discrétionnaire qu'il faut exercer eu égard aux éléments contextuels applicables (par. 28-29).

[52] Les considérations énoncées par notre Cour dans *Northwestern Utilities* témoignent de préoccupations fondamentales quant à la participation d'un tribunal administratif à l'appel de sa propre décision. Or, ces préoccupations ne sauraient fonder l'interdiction absolue d'une telle participation. La démarche discrétionnaire préconisée dans *Goodis*, *Leon's Furniture* et *Quadrini* offre le meilleur moyen d'assurer le caractère définitif de la décision et l'impartialité du décideur sans que la cour de révision ne soit alors privée de données et d'analyses à la fois utiles et importantes (voir N. Semple, « The Case for Tribunal Standing in Canada » (2007), 20 *R.C.D.A.P.* 305; L. A. Jacobs et T. S. Kuttner, « Discovering What Tribunals Do : Tribunal Standing Before the Courts » (2002), 81 *R. du B. can.* 616; F. A. V. Falzon, « Tribunal Standing on Judicial Review » (2008), 21 *R.C.D.A.P.* 21).

[53] Plusieurs considérations militent en faveur d'une démarche discrétionnaire. En particulier, vu ses compétences spécialisées et sa connaissance approfondie du régime administratif en cause, le tribunal administratif peut, dans bien des cas, être bien placé pour aider la cour de révision à rendre une juste décision. Par exemple, il peut être en mesure d'expliquer en quoi une certaine interprétation de la disposition législative en cause peut avoir une incidence sur d'autres dispositions du régime de réglementation ou sur les réalités factuelles et juridiques de son domaine de spécialisation. Il pourrait être plus difficile d'obtenir de tels éléments d'information d'autres parties.

[54] Dans certains cas, il n'y a tout simplement personne pour s'opposer à la partie qui conteste la décision du tribunal administratif. Le contrôle judiciaire se révèle optimal lorsque les deux facettes du litige sont vigoureusement défendues devant la cour de révision. Lorsqu'aucune autre partie bien au fait des enjeux ne fait valoir le point de vue opposé, la participation du tribunal administratif à titre de

ensure it has heard the best of both sides of a dispute.

[55] Canadian tribunals occupy many different roles in the various contexts in which they operate. This variation means that concerns regarding tribunal partiality may be more or less salient depending on the case at issue and the tribunal’s structure and statutory mandate. As such, statutory provisions addressing the structure, processes and role of the particular tribunal are key aspects of the analysis.

[56] The mandate of the Board, and similarly situated regulatory tribunals, sets them apart from those tribunals whose function it is to adjudicate individual conflicts between two or more parties. For tribunals tasked with this latter responsibility, “the importance of fairness, real and perceived, weighs more heavily” against tribunal standing: *Henthorne v. British Columbia Ferry Services Inc.*, 2011 BCCA 476, 344 D.L.R. (4th) 292, at para. 42.

[57] I am thus of the opinion that tribunal standing is a matter to be determined by the court conducting the first-instance review in accordance with the principled exercise of that court’s discretion. In exercising its discretion, the court is required to balance the need for fully informed adjudication against the importance of maintaining tribunal impartiality.

[58] In this case, as an initial matter, the *Ontario Energy Board Act, 1998* expressly provides that “[t]he Board is entitled to be heard by counsel upon the argument of an appeal” to the Divisional Court: s. 33(3). This provision neither expressly grants the Board standing to argue the merits of the decision on appeal, nor does it expressly limit the Board to jurisdictional or standard-of-review arguments as was the case for the relevant statutory provision in *Quadrini*: see para. 2.

partie adverse peut contribuer à faire en sorte que la cour statue après avoir entendu les arguments les plus convaincants de chacune des deux parties au litige.

[55] Les tribunaux administratifs canadiens tiennent nombre de rôles différents dans les contextes variés où ils évoluent, de sorte que la crainte d’une partialité de leur part peut être plus ou moins grande selon l’affaire en cause, ainsi que la structure du tribunal et son mandat légal. Dès lors, les dispositions législatives portant sur la structure, le fonctionnement et la mission d’un tribunal en particulier sont cruciales aux fins de l’analyse.

[56] Le mandat de la Commission, comme celui des tribunaux administratifs qui lui sont apparentés, la différencie des tribunaux administratifs appelés à trancher des différends individuels opposant plusieurs parties. Dans le cas de ces derniers, [TRADUCTION] « l’importance de l’équité, réelle et perçue, milite davantage » contre la reconnaissance de leur qualité pour agir (*Henthorne c. British Columbia Ferry Services Inc.*, 2011 BCCA 476, 344 D.L.R. (4th) 292, par. 42).

[57] Par conséquent, je suis d’avis qu’il appartient à la cour de première instance chargée du contrôle judiciaire de décider de la qualité pour agir d’un tribunal administratif en exerçant son pouvoir discrétionnaire de manière raisonnée. Dans l’exercice de son pouvoir discrétionnaire, la cour doit établir un équilibre entre la nécessité d’une décision bien éclairée et l’importance d’assurer l’impartialité du tribunal administratif.

[58] Dans la présente affaire, le par. 33(3) de la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* prévoit à titre préliminaire que « [l]a Commission a le droit d’être représentée par un avocat lors de l’audition de l’appel » devant la Cour divisionnaire. Cette disposition ne confère pas expressément à la Commission une qualité pour agir qui permet de faire valoir le bien-fondé de sa décision en appel, ni ne limite expressément la thèse qu’elle peut défendre à la présentation d’arguments relatifs à la compétence ou à la norme de contrôle comme le fait la disposition en cause dans l’affaire *Quadrini* (voir par. 2).

[59] In accordance with the foregoing discussion of tribunal standing, where the statute does not clearly resolve the issue, the reviewing court must rely on its discretion to define the tribunal's role on appeal. While not exhaustive, I would find the following factors, identified by the courts and academic commentators cited above, are relevant in informing the court's exercise of this discretion:

- (1) If an appeal or review were to be otherwise unopposed, a reviewing court may benefit by exercising its discretion to grant tribunal standing.
- (2) If there are other parties available to oppose an appeal or review, and those parties have the necessary knowledge and expertise to fully make and respond to arguments on appeal or review, tribunal standing may be less important in ensuring just outcomes.
- (3) Whether the tribunal adjudicates individual conflicts between two adversarial parties, or whether it instead serves a policy-making, regulatory or investigative role, or acts on behalf of the public interest, bears on the degree to which impartiality concerns are raised. Such concerns may weigh more heavily where the tribunal served an adjudicatory function in the proceeding that is the subject of the appeal, while a proceeding in which the tribunal adopts a more regulatory role may not raise such concerns.

[60] Consideration of these factors in the context of this case leads me to conclude that it was not improper for the Board to participate in arguing in favour of the reasonableness of its decision on appeal. First, the Board was the only respondent in the initial review of its decision. Thus, it had no alternative but to step in if the decision was to be defended on the merits. Unlike some other provinces, Ontario has no designated utility consumer advocate, which left the Board — tasked by statute

[59] Au vu de cette analyse de la qualité pour agir d'un tribunal administratif, lorsque le texte législatif applicable n'est pas clair sur ce point, la cour de révision s'en remet à son pouvoir discrétionnaire pour délimiter les attributs du tribunal administratif en appel. Voici quelles sont, entre autres, les considérations — relevées par les juridictions et les auteurs précités — qui délimitent l'exercice de ce pouvoir discrétionnaire :

- (1) lorsque, autrement, l'appel ou la demande de contrôle serait non contesté, il peut être avantageux que la cour de révision exerce le pouvoir discrétionnaire qui lui permet de reconnaître la qualité pour agir du tribunal administratif;
- (2) lorsque d'autres parties sont susceptibles de contester l'appel ou la demande de contrôle et qu'elles ont les connaissances et les compétences spécialisées nécessaires pour bien avancer une thèse ou la réfuter, la qualité pour agir du tribunal administratif peut revêtir une importance moindre pour l'obtention d'une issue juste;
- (3) le fait que la fonction du tribunal administratif consiste soit à trancher des différends individuels opposant deux parties, soit à élaborer des politiques, à réglementer ou enquêter ou à défendre l'intérêt public influe sur la mesure dans laquelle l'impartialité soulève des craintes ou non. Ces craintes peuvent jouer davantage lorsque le tribunal a exercé une fonction juridictionnelle dans l'instance visée par l'appel, et moins lorsque son rôle s'est révélé d'ordre réglementaire.

[60] Au vu de ces considérations, je conclus qu'il n'était pas inapproprié que la Commission participe à l'appel pour défendre le caractère raisonnable de sa décision. Premièrement, la Commission était la seule partie intimée lors du contrôle judiciaire initial de sa décision. Elle n'avait donc d'autre choix que de prendre part à l'instance pour que sa décision soit défendue au fond. Contrairement à d'autres provinces, l'Ontario n'a nommé aucun défenseur des droits des clients des services publics,

with acting to safeguard the public interest — with few alternatives but to participate as a party.

[61] Second, the Board is tasked with regulating the activities of utilities, including those in the electricity market. Its regulatory mandate is broad. Among its many roles: it licenses market participants, approves the development of new transmission and distribution facilities, and authorizes rates to be charged to consumers. In this case, the Board was exercising a regulatory role by setting just and reasonable payment amounts to a utility. This is unlike situations in which a tribunal may adjudicate disputes between two parties, in which case the interests of impartiality may weigh more heavily against full party standing.

[62] The nature of utilities regulation further argues in favour of full party status for the Board here, as concerns about the appearance of partiality are muted in this context. As noted by Doherty J.A., “[l]ike all regulated bodies, I am sure Enbridge wins some and loses some before the [Board]. I am confident that Enbridge fully understands the role of the regulator and appreciates that each application is decided on its own merits by the [Board]”: *Enbridge*, at para. 28. Accordingly, I do not find that the Board’s participation in the instant appeal was improper. It remains to consider whether the content of the Board’s arguments was appropriate.

(2) Bootstrapping

[63] The issue of tribunal “bootstrapping” is closely related to the question of when it is proper for a tribunal to act as a party on appeal or judicial review of its decision. The standing issue concerns what types of argument a tribunal may make, i.e.

si bien que la Commission — qui est légalement garante de l’intérêt public — n’avait pas vraiment d’autre avenue que celle de se constituer partie à l’instance.

[61] Deuxièmement, la Commission a pour mandat de réglementer les activités de services publics, y compris ceux qui appartiennent au domaine de l’électricité. Son mandat de réglementation est large. Au nombre de ses nombreuses fonctions, mentionnons l’octroi de permis aux participants du marché, l’approbation de nouvelles installations de transport et de distribution et l’autorisation des tarifs exigés des consommateurs. Dans la présente affaire, la Commission a exercé sa fonction de réglementation en établissant les paiements justes et raisonnables auxquels un service public avait droit. Il s’agit d’une situation différente de celle où le tribunal administratif est habilité à trancher un différend entre deux parties, le souci d’impartialité pouvant alors militer davantage contre la qualité d’agir comme partie à part entière.

[62] L’objet de la réglementation est un autre élément qui milite en faveur de la pleine reconnaissance de la qualité pour agir de la Commission, puisque la crainte d’apparence de partialité est faible en l’espèce. Pour reprendre les propos du juge Doherty dans *Enbridge*, par. 28, [TRADUCTION] « [à] l’instar de tout organisme réglementé, je suis certain que [la Commission] donne parfois raison à Enbridge et lui donne parfois tort. J’ose croire qu’Enbridge comprend parfaitement le rôle de l’organisme de réglementation et sait que [la Commission] statue sur chaque demande en fonction des faits qui lui sont propres ». Je conclus donc que la participation de la Commission au pourvoi n’a rien d’inapproprié. Reste à savoir si les arguments de la Commission sont appropriés.

(2) L’autojustification

[63] La question de l’« autojustification » est étroitement liée à celle de savoir à quelles conditions le tribunal administratif (ci-après le « tribunal ») est en droit d’agir comme partie à l’appel ou au contrôle judiciaire de sa décision. Statuer sur la

jurisdictional or merits arguments, while the bootstrapping issue concerns the content of those arguments.

[64] As the term has been understood by the courts who have considered it in the context of tribunal standing, a tribunal engages in bootstrapping where it seeks to supplement what would otherwise be a deficient decision with new arguments on appeal: see, e.g., *United Brotherhood of Carpenters and Joiners of America, Local 1386 v. Bransen Construction Ltd.*, 2002 NBCA 27, 249 N.B.R. (2d) 93. Put differently, it has been stated that a tribunal may not “defen[d] its decision on a ground that it did not rely on in the decision under review”: *Goodis*, at para. 42.

[65] The principle of finality dictates that once a tribunal has decided the issues before it and provided reasons for its decision, “absent a power to vary its decision or rehear the matter, it has spoken finally on the matter and its job is done”: *Quadrini*, at para. 16, citing *Chandler v. Alberta Association of Architects*, [1989] 2 S.C.R. 848. Under this principle, the court found that tribunals could not use judicial review as a chance to “amend, vary, qualify or supplement its reasons”: *Quadrini*, at para. 16. In *Leon’s Furniture*, Slatter J.A. reasoned that a tribunal could “offer interpretations of its reasons or conclusion, [but] cannot attempt to reconfigure those reasons, add arguments not previously given, or make submissions about matters of fact not already engaged by the record”: para. 29.

[66] By contrast, in *Goodis*, Goudge J.A. found on behalf of a unanimous court that while the Commissioner had relied on an argument not expressly set out in her original decision, this argument was available for the Commissioner to make on appeal. Though he recognized that “[t]he importance of reasoned decision making may be undermined if, when attacked in court, a tribunal can simply offer different, better, or even contrary reasons to

qualité pour agir d’un tribunal c’est décider de ce qu’il peut faire valoir (p. ex. des prétentions relatives à sa compétence ou à la justesse de sa décision), alors que l’« autojustification » touche à la teneur des prétentions.

[64] Suivant le sens attribué à cette notion par les cours de justice qui l’ont examinée dans le contexte de la qualité pour agir, un tribunal « s’autojustifie » lorsqu’il cherche, par la présentation de nouveaux arguments en appel, à étoffer une décision qui, sinon, serait lacunaire (voir p. ex. *United Brotherhood of Carpenters and Joiners of America, Local 1386 c. Bransen Construction Ltd.*, 2002 NBCA 27, 249 R.N.-B. (2^e) 93). Autrement dit, un tribunal ne pourrait [TRADUCTION] « défendre sa décision en invoquant un motif qui n’a pas été soulevé dans la décision faisant l’objet du contrôle » (*Goodis*, par. 42).

[65] Le caractère définitif de la décision veut que, dès lors qu’il a tranché les questions dont il était saisi et qu’il a motivé sa décision, le tribunal ait statué définitivement et que son travail soit terminé, « à moins qu’il ne soit investi du pouvoir de modifier sa décision ou d’entendre à nouveau l’affaire » (*Quadrini*, par. 16, citant *Chandler c. Alberta Association of Architects*, [1989] 2 R.C.S. 848). Partant, la cour a conclu qu’un tribunal ne peut profiter d’un contrôle judiciaire pour « modifier, changer, nuancer ou compléter ses motifs » (*Quadrini*, par. 16). Dans l’arrêt *Leon’s Furniture*, le juge Slatter affirme qu’un tribunal peut [TRADUCTION] « offrir différentes interprétations de ses motifs ou de sa conclusion, [mais] non tenter de remanier ses motifs, invoquer de nouveaux arguments ou se prononcer sur des questions de fait que ne soulève pas déjà le dossier » (par. 29).

[66] En revanche, le juge Goudge conclut, dans l’arrêt *Goodis*, avec l’accord de tous ses collègues, que même si la commissaire invoque un argument qui ne figure pas expressément dans sa décision initiale, elle peut le soulever en appel. Il reconnaît que [TRADUCTION] « [l’]importance de décisions bien étayées pourrait être compromise si un tribunal pouvait simplement offrir, à l’appui de sa décision attaquée devant une cour de justice,

support its decision” (para. 42), Goudge J.A. ultimately found that the Commissioner was permitted to raise a new argument on judicial review. The new argument presented was “not inconsistent with the reason offered in the decision. Indeed it could be said to be implicit in it”: para. 55. “It was therefore proper for the Commissioner to be permitted to raise this argument before the Divisional Court and equally proper for the court to decide on that basis”: para. 58.

[67] There is merit in both positions on the issue of bootstrapping. On the one hand, a permissive stance toward new arguments by tribunals on appeal serves the interests of justice insofar as it ensures that a reviewing court is presented with the strongest arguments in favour of both sides: *Semple*, at p. 315. This remains true even if those arguments were not included in the tribunal’s original reasons. On the other hand, to permit bootstrapping may undermine the importance of reasoned, well-written original decisions. There is also the possibility that a tribunal, surprising the parties with new arguments in an appeal or judicial review after its initial decision, may lead the parties to see the process as unfair. This may be particularly true where a tribunal is tasked with adjudicating matters between two private litigants, as the introduction of new arguments by the tribunal on appeal may give the appearance that it is “ganging up” on one party. As discussed, however, it may be less appropriate in general for a tribunal sitting in this type of role to participate as a party on appeal.

[68] I am not persuaded that the introduction of arguments by a tribunal on appeal that interpret or were implicit but not expressly articulated in its original decision offends the principle of finality. Similarly, it does not offend finality to permit a tribunal to explain its established policies and practices to the reviewing court, even if those were not described in the reasons under review. Tribunals need not repeat explanations of such practices in every decision merely to guard against charges of bootstrapping should they be called upon to explain them on appeal or review. A tribunal may also

des motifs différents, plus convaincants, voire opposés » (par. 42), mais il conclut finalement que la commissaire peut présenter un nouvel argument dans le cadre d’un contrôle judiciaire. Le nouvel argument n’est toutefois « pas incompatible avec les motifs formulés dans la décision, car on peut en effet affirmer qu’il en fait implicitement partie » (par. 55). « La commissaire pouvait donc soulever l’argument devant la Cour divisionnaire, et celle-ci pouvait en tenir compte pour se prononcer » (par. 58).

[67] Les deux thèses avancées sur l’autojustification se défendent. D’une part, il est dans l’intérêt de la justice de permettre au tribunal de présenter de nouveaux arguments en appel, car la cour de révision est alors saisie des arguments les plus convaincants à l’appui de chacune des thèses (*Semple*, p. 315). Cela demeure vrai même si ces arguments ne figurent pas dans la décision initiale. D’autre part, autoriser l’autojustification risque de compromettre l’importance de décisions bien étayées et bien rédigées au départ. Permettre au tribunal de présenter de nouveaux arguments en appel ou dans le cadre du contrôle judiciaire de sa décision initiale peut aussi amener les parties à conclure que le processus n’est pas équitable. Il peut surtout en être ainsi lorsque le tribunal est appelé à trancher des différends opposant deux personnes privées, puisque la présentation de nouveaux arguments en appel peut donner l’impression que le tribunal « se ligue » contre l’une des parties. Or, je le rappelle, il ne convient généralement pas que le tribunal doté d’un tel mandat participe en tant que partie à l’appel.

[68] Je ne suis pas convaincu que la formulation en appel de nouveaux arguments qui interprètent la décision initiale ou qui l’étaient implicitement, mais non expressément, va à l’encontre du principe du caractère définitif. De même, il n’est pas contraire à ce principe de permettre au tribunal d’expliquer à la cour de révision quelles sont ses politiques et pratiques établies, même lorsque les motifs contestés n’en font pas mention. Le tribunal n’a pas à les expliquer systématiquement dans chaque décision à la seule fin de se prémunir contre une allégation d’autojustification advenant qu’il

respond to arguments raised by a counterparty. A tribunal raising arguments of these types on review of its decision does so in order to uphold the initial decision; it is not reopening the case and issuing a new or modified decision. The result of the original decision remains the same even if a tribunal seeks to uphold that effect by providing an interpretation of it or on grounds implicit in the original decision.

[69] I am not, however, of the opinion that tribunals should have the unfettered ability to raise entirely new arguments on judicial review. To do so may raise concerns about the appearance of unfairness and the need for tribunal decisions to be well reasoned in the first instance. I would find that the proper balancing of these interests against the reviewing courts' interests in hearing the strongest possible arguments in favour of each side of a dispute is struck when tribunals do retain the ability to offer interpretations of their reasons or conclusions and to make arguments implicit within their original reasons: see *Leon's Furniture*, at para. 29; *Goodis*, at para. 55.

[70] In this case, I do not find that the Board impermissibly stepped beyond the bounds of its original decision in its arguments before this Court. In its reply factum, the Board pointed out — correctly, in my view — that its submissions before this Court simply highlight what is apparent on the face of the record, or respond to arguments raised by the respondents.

[71] I would, however, urge the Board, and tribunal parties in general, to be cognizant of the tone they adopt on review of their decisions. As Goudge J.A. noted in *Goodis*:

... if an administrative tribunal seeks to make submissions on a judicial review of its decision, it [should] pay careful attention to the tone with which it does so. Although this is not a discrete basis upon which its standing might be limited, there is no doubt that the tone

soit appelé à les préciser en appel ou en contrôle judiciaire. Il peut aussi répondre aux arguments de la partie adverse dans le cadre du contrôle judiciaire de sa décision car il le fait dans le but de faire confirmer sa décision initiale, non de rouvrir le dossier et de rendre une nouvelle décision ou de modifier la décision initiale. L'effet de la décision initiale demeure inchangé même lorsque le tribunal demande sa confirmation en offrant une interprétation de cette décision ou en invoquant des motifs qui la sous-tendent implicitement.

[69] Cependant, je ne crois pas qu'un tribunal devrait avoir la possibilité inconditionnelle de présenter une thèse entièrement nouvelle dans le cadre d'un contrôle judiciaire, car lui reconnaître cette faculté pourrait l'exposer à des allégations d'iniquité et nuire au prononcé de décisions bien motivées au départ. Je suis d'avis qu'il y a un juste équilibre entre ces considérations et celles voulant que la cour de révision entende les arguments les plus convaincants de chacune des parties lorsqu'il est permis au tribunal d'offrir différentes interprétations de ses motifs ou de ses conclusions ou de présenter des arguments qui sous-tendent implicitement ses motifs initiaux (voir *Leon's Furniture*, par. 29; *Goodis*, par. 55).

[70] Je ne crois pas que, dans la présente affaire, la Commission a indûment outrepassé les limites de sa décision initiale lorsqu'elle a présenté ses arguments devant notre Cour. Dans son mémoire en réplique, la Commission signale — à juste titre, selon moi — que ses observations mettent simplement en évidence ce qui ressort du dossier ou répondent aux arguments des intimées.

[71] J'exhorte toutefois la Commission et, de façon générale, tout tribunal qui se constitue partie à une instance à se soucier du ton qu'il adopte lors du contrôle judiciaire de sa décision. Comme le fait remarquer le juge Goudge dans l'arrêt *Goodis*,

[TRADUCTION] le tribunal administratif qui veut faire valoir son point de vue lors du contrôle judiciaire de sa décision [doit] porte[r] une attention particulière au ton qu'il adopte. Bien qu'il ne s'agisse pas d'un motif précis pour lequel sa qualité pourrait être restreinte, il ne

of the proposed submissions provides the background for the determination of that issue. A tribunal that seeks to resist a judicial review application will be of assistance to the court to the degree its submissions are characterized by the helpful elucidation of the issues, informed by its specialized position, rather than by the aggressive partisanship of an adversary. [para. 61]

[72] In this case, the Board generally acted in such a way as to present helpful argument in an adversarial but respectful manner. However, I would sound a note of caution about the Board’s assertion that the imposition of the prudent investment test “would in all likelihood not change the result” if the decision were remitted for reconsideration (A.F., at para. 99). This type of statement may, if carried too far, raise concerns about the principle of impartiality such that a court would be justified in exercising its discretion to limit tribunal standing so as to safeguard this principle.

B. *Standard of Review*

[73] The parties do not dispute that reasonableness is the appropriate standard of review for the Board’s actions in applying its expertise to set rates and approve payment amounts under the *Ontario Energy Board Act, 1998*. I agree. In addition, to the extent that the resolution of this appeal turns on the interpretation of the *Ontario Energy Board Act, 1998*, the Board’s home statute, a standard of reasonableness presumptively applies: *Dunsmuir v. New Brunswick*, 2008 SCC 9, [2008] 1 S.C.R. 190, at para. 54; *Alberta (Information and Privacy Commissioner) v. Alberta Teachers’ Association*, 2011 SCC 61, [2011] 3 S.C.R. 654, at para. 30; *Tervita Corp. v. Canada (Commissioner of Competition)*, 2015 SCC 3, [2015] 1 S.C.R. 161, at para. 35. Nothing in this case suggests the presumption should be rebutted.

[74] This appeal involves two distinct uses of the term “reasonable”. One concerns the standard of review: on appeal, this Court is charged with evaluating the “justification, transparency and intelligibility” of the Board’s reasoning, and “whether the

fait aucun doute que le ton des observations proposées offre une toile de fond à cet égard. Le tribunal qui désire contester une demande de contrôle judiciaire sera utile à la cour dans la mesure où ses observations permettront d’éclaircir les questions et où elles seront fondées sur ses connaissances spécialisées, au lieu d’être empreintes d’un parti pris agressif contre la partie adverse. [par. 61]

[72] En l’espèce, la Commission a généralement présenté des arguments utiles dans le cadre d’un débat contradictoire, mais respectueux. Une mise en garde s’impose toutefois selon moi en ce qui concerne l’affirmation de la Commission selon laquelle l’application du critère de l’investissement prudent [TRADUCTION] « ne changerait vraisemblablement pas l’issue de l’affaire » si la décision lui était renvoyée pour réexamen (m.a., par. 99). Une telle affirmation peut, si elle est poussée trop loin, faire douter de l’impartialité du tribunal au point où une cour de justice serait justifiée d’exercer son pouvoir discrétionnaire et de limiter la qualité pour agir du tribunal de manière à préserver son impartialité.

B. *Norme de contrôle*

[73] Les parties conviennent que la norme de contrôle qui s’applique aux actes de la Commission lorsqu’elle fait appel à son expertise pour fixer les tarifs et approuver des paiements sur le fondement de la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* est celle de la décision raisonnable. Je suis d’accord. En outre, dans la mesure où l’issue du pourvoi repose sur l’interprétation de cette loi — la loi constitutive de la Commission —, l’application de la norme de la décision raisonnable doit être présumée (*Dunsmuir c. Nouveau-Brunswick*, 2008 CSC 9, [2008] 1 R.C.S. 190, par. 54; *Alberta (Information and Privacy Commissioner) c. Alberta Teachers’ Association*, 2011 CSC 61, [2011] 3 R.C.S. 654, par. 30; *Tervita Corp. c. Canada (Commissaire de la concurrence)*, 2015 CSC 3, [2015] 1 R.C.S. 161, par. 35). Rien ne donne à penser en l’espèce que la présomption soit réfutée.

[74] Le pourvoi fait intervenir deux notions distinctes de ce qui est « raisonnable ». L’une est liée à la norme de contrôle : en appel, la Cour doit apprécier la « justification [. . .], [. . .] la transparence et [. . .] l’intelligibilité » du raisonnement de la

decision falls within a range of possible, acceptable outcomes which are defensible in respect of the facts and law” (*Dunsmuir*, at para. 47). The other is statutory: the Board’s rate-setting powers are to be used to ensure that, in its view, a just and reasonable balance is struck between utility and consumer interests. These reasons will attempt to keep the two uses of the term distinct.

C. Choice of Methodology Under the Ontario Energy Board Act, 1998

[75] The question of whether the Board’s decision to disallow recovery of certain costs was reasonable turns on how that decision relates to the Board’s statutory and regulatory powers to approve payments to utilities and to have these payments reflected in the rates paid by consumers. The Board’s general rate- and payment-setting powers are described above under the “Regulatory Framework” heading.

[76] The just-and-reasonable approach to recovery of the cost of services provided by a utility captures the essential balance at the heart of utilities regulation: to encourage investment in a robust utility infrastructure and to protect consumer interests, utilities must be allowed, over the long run, to earn their cost of capital, no more, no less.

[77] The *Ontario Energy Board Act, 1998* does not, however, either in s. 78.1 or elsewhere, prescribe the methodology the Board must use to weigh utility and consumer interests when deciding what constitutes just and reasonable payment amounts to the utility. Indeed, s. 6(1) of O. Reg. 53/05 expressly permits the Board, subject to certain exceptions set out in s. 6(2), to “establish the form, methodology, assumptions and calculations used in making an order that determines payment amounts for the purpose of section 78.1 of the Act”.

Commission et se demander si la décision appartient « aux issues possibles acceptables pouvant se justifier au regard des faits et du droit » (*Dunsmuir*, par. 47). L’autre est d’origine législative : la Commission doit utiliser son pouvoir de fixation des tarifs de manière à établir un équilibre qu’elle considère juste et raisonnable entre les intérêts du service public et ceux des consommateurs. Je m’efforce ci-après de respecter cette distinction.

C. Choix de la méthode suivant la Loi de 1998 sur la Commission de l’énergie de l’Ontario

[75] La question de savoir si le refus de la Commission d’approuver le recouvrement de certaines dépenses est raisonnable ou non dépend du lien de ce refus avec les pouvoirs légaux et réglementaires de la Commission d’approuver des paiements au service public et de répercuter ces paiements sur les tarifs exigés des consommateurs. Les pouvoirs généraux de la Commission en matière de fixation des tarifs et des paiements sont énoncés précédemment à la rubrique « Cadre réglementaire ».

[76] L’approche fondée sur le caractère juste et raisonnable des dépenses qu’un service public peut recouvrer rend compte de l’équilibre essentiel recherché dans la réglementation des services publics : pour encourager l’investissement dans une infrastructure robuste et protéger l’intérêt des consommateurs, un service public doit pouvoir, à long terme, toucher l’équivalent du coût du capital, ni plus, ni moins.

[77] Or, la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* ne prévoit ni à l’art. 78.1 ni à quelque autre article la méthode que doit utiliser la Commission pour sopeser les intérêts respectifs du service public et des consommateurs lorsqu’elle décide ce qui constitue des paiements justes et raisonnables. Certes, sous réserve de certaines exceptions prévues au par. 6(2), le par. 6(1) du règlement 53/05 permet expressément à la Commission de [TRADUCTION] « définir la forme, la méthode, les hypothèses et les calculs utilisés pour rendre une ordonnance qui établit le montant du paiement aux fins de l’article 78.1 de la Loi ».

[78] As a contrasting example, para. 4.1 of s. 6(2) of O. Reg. 53/05 establishes a specific methodology for use when the Board reviews “costs incurred and firm financial commitments made in the course of planning and preparation for the development of proposed new nuclear generation facilities”. When reviewing such costs, the Board must be satisfied that “the costs were prudently incurred” and that “the financial commitments were prudently made”: para. 4.1 of s. 6(2). The provision thus establishes a specific context in which the Board’s analysis is focused on the prudence of the decision to incur or commit to certain costs. The absence of such language in the more general s. 6(1) provides further reason to read the regulation as providing broad methodological discretion to the Board in making orders for payment amounts where the specific provisions of s. 6(2) do not apply.

[79] Regarding whether a presumption of prudence must be applied to OPG’s decisions to incur costs, neither the *Ontario Energy Board Act, 1998* nor O. Reg. 53/05 expressly establishes such a presumption. Indeed, the *Ontario Energy Board Act, 1998* places the burden on the applicant utility to establish that payment amounts approved by the Board are just and reasonable: s. 78.1(6) and (7). It would thus seem inconsistent with the statutory scheme to presume that utility decisions to incur costs were prudent.

[80] Justice Abella concludes that the Board’s review of OPG’s costs should have consisted of “an after-the-fact prudence review, with a rebuttable presumption that the utility’s expenditures were reasonable”: para. 150. Such an approach is contrary to the statutory scheme. While the Board has considerable methodological discretion, it does not have the freedom to displace the burden of proof established by s. 78.1(6) of the *Ontario Energy Board Act, 1998*: “. . . the burden of proof is on the applicant in an application made under this section”. Of course, this does not imply that the applicant must systematically prove that every single cost is just

[78] En revanche, la disposition 4.1 du par. 6(2) du règlement 53/05 prescrit le recours à une méthode particulière lorsque la Commission examine [TRADUCTION] « les dépenses faites et les engagements financiers fermes pris dans le cadre de la planification et de la préparation relatives à la réalisation d’installations nucléaires projetées ». La Commission doit être convaincue que « les dépenses ont été faites de manière prudente » et que « les engagements financiers ont été pris de manière prudente » (la disposition 4.1 du par. 6(2)). La disposition établit donc un cadre précis où l’analyse de la Commission est axée sur la prudence de la décision de faire certaines dépenses ou de convenir de certaines dépenses. L’absence d’un libellé en ce sens dans la disposition générale qu’est le par. 6(1) constitue un autre motif de considérer que le règlement confère à la Commission un large pouvoir discrétionnaire quant à la méthode à employer pour ordonner un paiement lorsque les dispositions particulières du par. 6(2) ne s’appliquent pas.

[79] Pour ce qui concerne la question de savoir si la présomption de prudence doit s’appliquer aux décisions d’OPG de faire des dépenses, ni la *Loi de 1998 sur la Commission de l’énergie de l’Ontario*, ni le règlement 53/05 n’établissent expressément une telle présomption. D’ailleurs, suivant cette loi, il incombe au service public requérant d’établir que les paiements qu’il demande à la Commission d’approuver sont justes et raisonnables (par. 78.1(6) et (7)). Il semble donc contraire au régime législatif de présumer que la décision de faire des dépenses est prudente.

[80] La juge Abella conclut que l’examen des dépenses d’OPG par la Commission aurait dû consister à « contrôl[er] la prudence des dépenses après coup et [à] appliqu[er] la présomption réfutable selon laquelle elles étaient raisonnables » (par. 150). Or, une telle approche est contraire au régime législatif. La Commission jouit certes d’une grande marge de manœuvre quant au choix d’une méthode, mais elle n’a pas la faculté d’inverser le fardeau de la preuve établi au par. 78.1(6) de la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* : « . . . le fardeau de la preuve incombe au requérant dans une requête présentée en vertu du

and reasonable. The Board has broad discretion to determine the methods it may use to examine costs — it just cannot shift the burden of proof contrary to the statutory scheme.

[81] In judicially reviewing a decision of the Board to allow or disallow payments to a utility, the court’s role is to assess whether the Board reasonably determined that a certain payment amount was “just and reasonable” for both the utility and the consumers. Such an approach is consistent with this Court’s rate-setting jurisprudence in other regulatory domains in which the regulator is given methodological discretion, where it has been observed that “[t]he obligation to act is a question of law, but the choice of the method to be adopted is a question of discretion with which, under the statute, no Court of law may interfere”: *Bell Canada v. Bell Aliant Regional Communications*, 2009 SCC 40, [2009] 2 S.C.R. 764, at para. 40 (concerning telecommunication rate-setting), quoting *Re General Increase in Freight Rates* (1954), 76 C.R.T.C. 12 (S.C.C.), at p. 13 (concerning railway freight rates). Of course, today this statement must be understood to permit intervention by a court where the exercise of discretion rendered a decision unreasonable. Accordingly, it remains to determine whether the Board’s analytical approach to disallowing the costs at issue in this case rendered the Board’s decision unreasonable under the “just and reasonable” standard.

D. Characterization of Costs at Issue

[82] Forecast costs are costs which the utility has not yet paid, and over which the utility still retains discretion as to whether the disbursement will be

présent article ». Il ne s’ensuit pas, bien sûr, que le requérant doit systématiquement prouver le caractère juste et raisonnable de chacune de ses dépenses, individuellement. La Commission jouit d’un grand pouvoir discrétionnaire qui lui permet d’arrêter les méthodes à employer dans l’examen des dépenses, mais elle ne peut tout simplement pas inverser le fardeau de la preuve qu’établit le régime législatif.

[81] La cour de justice appelée à contrôler la décision de la Commission d’approuver ou non des paiements à un service public doit se demander si la conclusion de la Commission selon laquelle un paiement d’un certain montant est « juste et raisonnable » tant pour le service public que pour le consommateur est raisonnable ou non. Cette approche concorde avec les décisions de notre Cour sur l’établissement de tarifs dans d’autres secteurs réglementés où l’organisme de réglementation dispose d’un pouvoir discrétionnaire qui lui permet de recourir à une méthode ou à une autre. Dans ces décisions, la Cour signale que « [l]’obligation d’agir est une question de droit, mais le choix de la méthode est une question relevant de l’exercice du pouvoir discrétionnaire et à l’égard de laquelle, selon le texte de loi, aucun tribunal judiciaire ne peut intervenir » (*Bell Canada c. Bell Aliant Communications régionales*, 2009 CSC 40, [2009] 2 R.C.S. 764, par. 40 (tarification des télécommunications), citant *Re General Increase in Freight Rates* (1954), 76 C.R.T.C. 12 (C.S.C.), p. 13 (tarification du transport ferroviaire des marchandises)). Certes, de nos jours, il faut voir dans ces propos la reconnaissance du pouvoir d’une cour de justice d’intervenir lorsqu’elle estime que l’exercice du pouvoir discrétionnaire a débouché sur une décision déraisonnable. Reste donc à décider si la méthode d’analyse retenue par la Commission pour refuser d’approuver les dépenses en l’espèce a rendu sa décision déraisonnable selon la norme du paiement « juste et raisonnable ».

D. Qualification des dépenses en cause

[82] Les dépenses prévues sont celles que le service public n’a pas encore acquittées et qu’un pouvoir discrétionnaire lui permet de renoncer à faire.

made. A disallowance of such costs presents a utility with a choice: it may change its plans and avoid the disallowed costs, or it may incur the costs regardless of the disallowance with the knowledge that the costs will ultimately be borne by the utility's shareholders rather than its ratepayers. By contrast, committed costs are those for which, if a regulatory board disallows recovery of the costs in approved payments, the utility and its shareholders will have no choice but to bear the burden of those costs themselves. This result may occur because the utility has already spent the funds, or because the utility entered into a binding commitment or was subject to other legal obligations that leave it with no discretion as to whether to make the payment in the future.

[83] There is disagreement between the parties as to how the costs disallowed by the Board in this matter should be characterized. The Board asserts that compensation costs for the test period are forecast insofar as they have not yet been disbursed, while OPG asserts that the costs should be characterized as committed, because OPG is under a contractual obligation to pay those amounts when they become due. This disagreement is important because a “no hind-sight” prudence review, which is discussed in detail below, has developed in the context of “committed” costs. Indeed, it makes no sense to apply such a test where a utility still retains discretion over whether the costs will ultimately be incurred; the decision to commit the utility to such costs has not yet been made. Accordingly, where the regulator has discretion over its methodological approach, understanding whether the costs at issue are “forecast” or “committed” may be helpful in reviewing the reasonableness of a regulator's choice of methodology.

[84] In this case, at least some of the compensation costs that the Board found to be excessive were driven by collective agreements to which OPG had committed before the application at issue, and which established compensation costs that were, in aggregate, above the 75th percentile for comparable positions at other utilities. The collective agreements left OPG with limited flexibility

Lorsque leur approbation est refusée, le service public peut soit modifier ses plans et renoncer aux dépenses, soit les faire malgré le refus étant entendu qu'elles seront assumées par les actionnaires plutôt que par les consommateurs. À l'opposé, les dépenses convenues sont celles que ses actionnaires et lui n'auront d'autre choix que d'assumer si l'organisme de réglementation refuse de permettre leur recouvrement et d'approuver les paiements sollicités. Cela peut advenir lorsque le service public a déjà déboursé la somme en cause ou qu'il a pris un engagement contraignant ou était assujéti à d'autres obligations qui écartent tout pouvoir discrétionnaire lui permettant de ne pas acquitter la somme ultérieurement.

[83] Les parties ne s'entendent pas sur la qualification des dépenses que la Commission a refusé d'approuver. Selon cette dernière, les dépenses de rémunération pour la période de référence sont des dépenses prévues dans la mesure où elles n'ont pas encore été acquittées. OPG soutient plutôt qu'il s'agit de dépenses convenues puisqu'elle est tenue par contrat de verser les sommes en cause au moment où elles deviennent exigibles. Ce désaccord est important car le contrôle de la prudence « sans recul », sur lequel je reviendrai plus en détail, a vu le jour dans le contexte de dépenses « convenues ». Il est en effet absurde d'appliquer ce critère lorsque le service public peut encore décider, en fin de compte, de faire ou non les dépenses; la décision de convenir de ces dépenses n'a pas encore été prise. Par conséquent, lorsque l'organisme de réglementation possède un pouvoir discrétionnaire quant à la méthode à employer, la qualification des dépenses — « prévues » ou « convenues » — peut constituer une étape importante pour statuer sur le caractère raisonnable de la méthode retenue.

[84] En l'espèce, au moins une partie des dépenses de rémunération jugées excessives par la Commission était imputable à des conventions collectives qu'OPG avait conclues avant la présentation de sa demande et qui faisaient en sorte que sa masse salariale globale dépasse le 75^e percentile pour des emplois comparables dans d'autres services publics. Les conventions collectives laissaient

regarding overall compensation rates or staffing levels — OPG was required to abide by wage and staffing levels established by collective agreements, and retained flexibility only over terms outside the bounds of those agreements — and thus those portions of OPG’s compensation rates and staffing levels that were dictated by the terms of the collective agreements were committed costs.

[85] However, the Board found that OPG’s compensation costs for the test period were not entirely driven by the collective agreements, and thus were not entirely committed, because OPG retained some flexibility to manage total staffing levels in light of projected attrition of a mature workforce. The Board Decision did not, however, include detailed forecasts regarding exactly how much of the \$145 million in disallowed compensation costs could be recovered through natural reduction in employee numbers or other adjustments, and how much would necessarily be borne by the utility and its shareholder. Accordingly, the disallowed costs at issue must be understood as being at least partially committed. It is unreasonable to characterize them as entirely forecast in view of the constraints placed on OPG by the collective agreements.

[86] Having established that the disallowed costs are at least partially committed, it is necessary to consider whether the Board acted reasonably in not applying a no-hindsight prudent investment test in assessing those costs. Accordingly, I now turn to the jurisprudential history and methodological details of the prudent investment test.

E. *The Prudent Investment Test*

[87] In order to assess whether the Board’s methodology was reasonable in this case, it is necessary to provide some background on the prudent investment test (sometimes referred to as “prudence review” or the “prudence test”) in order to identify its origins, place it in context, and explore how it has

peu de marge de manœuvre quant aux barèmes de rémunération et aux niveaux de dotation dans leur ensemble, OPG devait respecter ceux établis par les conventions collectives et elle ne jouissait d’une marge de manœuvre que pour les conditions qui n’étaient pas ainsi régies. Par conséquent, les dépenses liées aux barèmes de rémunération et aux niveaux de dotation imposés par les conventions collectives étaient des dépenses convenues.

[85] La Commission conclut cependant que les dépenses de rémunération pour la période de référence ne sont pas toutes déterminées par les conventions collectives et qu’elles ne sont donc pas toutes convenues, car OPG dispose d’une certaine marge de manœuvre pour gérer globalement les niveaux de dotation en fonction du départ prévu d’employés d’âge mûr. Toutefois, la décision de la Commission ne précise pas quel pourcentage exact des 145 millions de dollars refusés au chapitre de la rémunération pourrait être recouvré grâce à la réduction naturelle du nombre d’employés ou à d’autres ajustements, ni quel pourcentage serait nécessairement assumé par le service public et son actionnaire. Par conséquent, les dépenses refusées en l’espèce doivent être considérées comme des dépenses convenues, du moins en partie. Il est déraisonnable d’y voir en totalité des dépenses prévues étant donné l’effet contraignant des conventions collectives sur OPG.

[86] Après avoir établi que les dépenses refusées sont, du moins partiellement, des dépenses convenues, il faut déterminer si la Commission a agi de façon raisonnable en appliquant le critère de l’investissement prudent sans exclure le recul. J’examine donc maintenant l’historique jurisprudentiel du critère de l’investissement prudent et les données méthodologiques y afférentes.

E. *Le critère de l’investissement prudent*

[87] Décider si la méthode de la Commission était raisonnable en l’espèce exige de se pencher sur l’historique du critère de l’investissement prudent (parfois appelé « contrôle de la prudence » ou « critère de la prudence ») pour déterminer ses origines, le situer dans le contexte et savoir quelle portée lui

been understood by utilities, regulators, and legislators.

(1) American Jurisprudence

[88] American jurisprudence has played a significant role in the history of the prudent investment test in utilities regulation. In discussing this history, I would first reiterate this Court’s observation that “[w]hile the American jurisprudence and texts in this area should be considered with caution given that Canada and the United States have very different political and constitutional-legal regimes, they do shed some light on the issue”: *ATCO Gas and Pipelines Ltd. v. Alberta (Energy and Utilities Board)*, 2006 SCC 4, [2006] 1 S.C.R. 140, at para. 54.

[89] The origins of the prudent investment test in the context of utilities regulation may be traced to Justice Brandeis of the Supreme Court of the United States, who wrote a concurring opinion in 1923 to observe that utilities should receive deference in seeking to recover “investments which, under ordinary circumstances, would be deemed reasonable”: *State of Missouri ex rel. Southwestern Bell Telephone Co. v. Public Service Commission of Missouri*, 262 U.S. 276 (1923), at p. 289, fn.1.

[90] In the decades that followed, American utility regulators tasked with reviewing past-incurred utility costs generally employed one of two standards: the “used and useful” test or the “prudent investment” test (J. Kahn, “Keep *Hope* Alive: Updating the Prudent Investment Standard for Allocating Nuclear Plant Cancellation Costs” (2010), 22 *Fordham Envtl. L. Rev.* 43, at p. 49). These tests took different approaches to determining what costs could justly and reasonably be passed on to ratepayers. The used and useful test allowed utilities to earn returns only on those investments that were actually used and useful to the utility’s operations, on the principle that ratepayers should not be compelled to pay for investments that do not benefit them.

ont attribué les services publics, les organismes de réglementation et les rédacteurs législatifs.

(1) Jurisprudence américaine

[88] La jurisprudence américaine a joué un rôle important dans l’application du critère de l’investissement prudent aux services publics réglementés. Rappelons d’abord l’observation de notre Cour selon laquelle, « [b]ien qu’il faille aborder avec circonspection la jurisprudence et la doctrine américaines dans ce domaine — les régimes politiques des États-Unis et du Canada étant fort différents, tout comme leurs régimes de droit constitutionnel —, elles éclairent la question » (*ATCO Gas and Pipelines Ltd. c. Alberta (Energy and Utilities Board)*, 2006 CSC 4, [2006] 1 R.C.S. 140, par. 54).

[89] L’application du critère de l’investissement prudent aux services publics réglementés s’origine de l’opinion concordante du juge Brandeis, de la Cour suprême des États-Unis, datant de 1923 et selon laquelle les services publics ont droit à la déférence lorsqu’ils cherchent à recouvrer [TRADUCTION] « un investissement qui, normalement, serait considéré comme raisonnable » (*State of Missouri ex rel. Southwestern Bell Telephone Co. c. Public Service Commission of Missouri*, 262 U.S. 276 (1923), p. 289, note 1).

[90] Dans les décennies qui ont suivi, les organismes de réglementation américains chargés de l’examen de dépenses déjà faites par les services publics ont généralement appliqué soit le critère axé sur [TRADUCTION] « l’emploi et l’utilité », soit le critère de « l’investissement prudent » (J. Kahn, « Keep *Hope* Alive : Updating the Prudent Investment Standard for Allocating Nuclear Plant Cancellation Costs » (2010), 22 *Fordham Envtl. L. Rev.* 43, p. 49). À chacun de ces critères correspond une approche différente pour déterminer quelles dépenses peuvent équitablement et raisonnablement être refilées aux consommateurs. Le critère de l’emploi et de l’utilité permet au service public d’obtenir un rendement, mais seulement sur l’investissement qui est réellement employé et qui se révèle utile à l’exploitation de l’entreprise, étant entendu que les consommateurs ne doivent pas être tenus de payer pour un investissement dont ils ne bénéficient pas.

[91] By contrast, the prudent investment test followed Justice Brandeis’s preferred approach by allowing for recovery of costs provided they were not imprudent based on what was known at the time the investment or expense was incurred: Kahn, at pp. 49-50. Though it may seem problematic from the perspective of consumer interests to adopt the prudent investment test — a test that allows for payments related to investments that may not be used or useful — it gives regulators a tool to soften the potentially harsh effects of the used and useful test, which may place onerous burdens on utilities. Disallowing recovery of the cost of failed investments that appeared reasonable at the time, for example, may imperil the financial health of utilities, and may chill the incentive to make such investments in the first place. This effect may then have negative implications for consumers, whose long-run interests will be best served by a dynamically efficient and viable electricity industry. Thus, the prudent investment test may be employed by regulators to strike the appropriate balance between consumer and utility interests: see Kahn, at pp. 53-54.

[92] The states differed in their approaches to setting the statutory foundation for utility regulation. Regulators in some states were free to apply the prudent investment test, while other states enacted statutory provisions disallowing compensation in respect of capital investments that were not “used and useful in service to the public”: *Duquesne Light Co. v. Barasch*, 488 U.S. 299 (1989), at p. 302. Notably, when asked in *Duquesne* to consider whether “just and reasonable” payments to utilities required, as a constitutional matter, that the prudent investment test be applied to past-incurred costs, the U.S. Supreme Court held that “[t]he designation of a single theory of ratemaking as a constitutional requirement would unnecessarily foreclose alternatives which could benefit both consumers and investors”: p. 316.

[91] Au critère de l’investissement prudent correspond l’approche retenue par le juge Brandeis et selon laquelle des dépenses peuvent être recouvrées si elles ne sont pas imprudentes compte tenu de ce qu’on sait au moment où est fait l’investissement ou la dépense (Kahn, p. 49-50). Bien qu’il puisse sembler problématique du point de vue de la protection des intérêts des consommateurs d’adopter le critère de l’investissement prudent — dans la mesure où il autorise un paiement pour un investissement qui n’a été ni employé ni utile —, ce critère permet aux organismes de réglementation d’atténuer les possibles effets draconiens du critère de l’emploi et de l’utilité, lequel impose un lourd fardeau au service public. Par exemple, refuser le recouvrement d’un mauvais investissement qui paraissait raisonnable au moment où il a été fait risque de compromettre la santé financière du service public et d’avoir un effet dissuasif sur l’investissement ultérieur de capitaux par ce dernier. Pareil résultat peut ensuite entraîner des conséquences négatives pour les consommateurs, dont les intérêts à long terme sont mieux servis si le secteur de l’électricité est à la fois dynamique, efficace et viable. Par conséquent, un organisme de réglementation peut recourir au critère de l’investissement prudent afin d’établir un juste équilibre entre les intérêts des consommateurs et ceux du service public (voir Kahn, p. 53-54).

[92] Les États ont eu recours à des approches différentes pour établir le fondement légal de la réglementation des services publics. Certains ont permis aux organismes de réglementation d’appliquer le critère de l’investissement prudent, alors que d’autres ont légiféré pour écarter le recouvrement de capitaux investis qui n’étaient [TRADUCTION] « ni employés ni utiles au public » (*Duquesne Light Co. v. Barasch*, 488 U.S. 299 (1989), p. 302). Fait à signaler, dans cette affaire où on lui demandait si des paiements « justes et raisonnables » à un service public nécessitaient, sur le plan constitutionnel, que le critère de l’investissement prudent s’applique aux dépenses déjà faites, la Cour suprême des É.-U. a conclu que « [l’]élévation d’une seule méthode de tarification au rang de norme constitutionnelle écarterait inutilement d’autres avenues dont pourraient bénéficier à la fois consommateurs et investisseurs » (p. 316).

[93] American courts have also recognized that there may exist some contexts in which certain features of the prudent investment test may be less justifiable. For example, the Supreme Court of Utah considered whether a presumption of reasonableness was justified when reviewing costs passed to a utility by an unregulated affiliate entity, and concluded that it was not appropriate:

. . . we do not think an affiliate expense should carry a presumption of reasonableness. While the pressures of a competitive market might allow us to assume, in the absence of a showing to the contrary, that nonaffiliate expenses are reasonable, the same cannot be said of affiliate expenses not incurred in an arm's length transaction.

(*U.S. West Communications, Inc. v. Public Service Commission of Utah*, 901 P.2d 270 (Utah 1995), at p. 274)

[94] Treatment of the prudent investment test in American jurisprudence thus indicates that the test has been employed as a tool that may be useful in arriving at just and reasonable outcomes, rather than a mandatory feature of utilities regulation that must be applied regardless of whether there is statutory language to that effect.

(2) Canadian Jurisprudence

[95] Following its emergence in American jurisprudence, several Canadian utility regulators and courts have also considered the role of prudence review and, in some cases, applied a form of the prudent investment test. I provide a review of some of these cases here not in an attempt to exhaustively catalogue all uses of the test, but rather to set out the way in which the test has been invoked in various contexts.

[96] In *British Columbia Electric Railway Co. v. Public Utilities Commission of British Columbia*, [1960] S.C.R. 837, Martland J. observed that the statute at issue in that case directed that the regulator, in fixing rates,

[93] Les cours de justice américaines ont aussi reconnu que, dans certains contextes, des aspects du critère de l'investissement prudent peuvent se révéler moins justifiables. Par exemple, saisie du contrôle judiciaire de coûts transférés à un service public par une entreprise affiliée non réglementée, la Cour suprême de l'Utah s'est demandé s'il était justifié de présumer que les coûts étaient raisonnables et elle a conclu par la négative :

[TRADUCTION] . . . nous ne pensons pas que les dépenses de l'affiliée devraient être présumées raisonnables. Bien que la pression exercée par un marché concurrentiel puisse nous permettre de présumer, faute d'une preuve contraire, que les dépenses d'une entreprise non affiliée sont raisonnables, on ne peut en dire autant des dépenses d'une affiliée qui ne sont pas faites dans le cadre d'une opération sans lien de dépendance.

(*U.S. West Communications, Inc. c. Public Service Commission of Utah*, 901 P.2d 270 (Utah 1995), p. 274)

[94] Il appert donc de la jurisprudence américaine que le critère de l'investissement prudent s'est révélé utile pour arriver à un résultat juste et raisonnable, mais qu'il ne saurait constituer un élément obligatoire de la réglementation des services publics dont l'application s'impose même lorsqu'aucune disposition législative ne le prévoit.

(2) Jurisprudence canadienne

[95] Sous l'impulsion de la jurisprudence américaine, plusieurs organismes de réglementation et cours de justice du Canada se sont aussi penchés sur le rôle du contrôle de la prudence et ont parfois appliqué une variante du critère de l'investissement prudent. Je passerai en revue certaines de leurs décisions dans le but non pas de répertorier toutes les applications du critère, mais bien de faire état de la manière dont on l'a appliqué dans différents contextes.

[96] Dans l'arrêt *British Columbia Electric Railway Co. c. Public Utilities Commission of British Columbia*, [1960] R.C.S. 837, le juge Martland relève que, suivant la loi en cause, l'organisme de réglementation est tenu à ce qui suit lorsqu'il fixe des tarifs :

[TRADUCTION]

- (a) . . . shall consider all matters which it deems proper as affecting the rate: [and]
- (b) . . . shall have due regard, among other things, to the protection of the public from rates that are excessive as being more than a fair and reasonable charge for services of the nature and quality furnished by the public utility; and to giving to the public utility a fair and reasonable return upon the appraised value of the property of the public utility used, or prudently and reasonably acquired, to enable the public utility to furnish the service. [p. 852]

(Quoting *Public Utilities Act*, R.S.B.C. 1948, c. 277, s. 16(1)(b) (repealed S.B.C. 1973, c. 29, s. 187).)

The consequence of this statutory language, Martland J. held, was that the regulator, “when dealing with a rate case, has unlimited discretion as to the matters which it may consider as affecting the rate, but that it must, when actually setting the rate, meet the two requirements specifically mentioned in clause (b)”: p. 856. That is, the regulator, under this statute, must ensure that the public pays only fair and reasonable charges, and that the utility secures a fair and reasonable return upon its property used *or prudently and reasonably acquired*. This express statutory protection for the recovery of prudently made property acquisition costs thus provides an example of statutory language under which this Court found a non-discretionary obligation to provide a fair return to utilities for capital expenditures that were either used or prudently acquired.

[97] In 2005, the Nova Scotia Utility and Review Board (“NSUARB”) considered and adopted a definition of the prudent investment test articulated by the Illinois Commerce Commission:

. . . prudence is that standard of care which a reasonable person would be expected to exercise under the same circumstances encountered by utility management at the time decisions had to be made. . . . Hindsight is not applied in assessing prudence. . . . A utility’s decision is

- (a) . . . considérer tout élément qu’il juge susceptible d’influer sur les tarifs; [et]
- (b) . . . tenir dûment compte, notamment, de la protection du public contre les tarifs excessifs qui excèdent ce qui est juste et raisonnable en contrepartie du service de la nature et de la qualité de celui fourni et de l’obtention par le service public d’un rendement juste et raisonnable sur les biens qu’il affecte à la prestation du service ou qu’il acquiert à cette fin de manière prudente et raisonnable, selon leur valeur d’expertise. [p. 852]

(Citant *Public Utilities Act*, R.S.B.C. 1948, c. 227, al. 16(1)(b) (abrogé S.B.C. 1973, c. 29, art. 187).)

Le juge Martland conclut de ce libellé que l’organisme de réglementation [TRADUCTION] « appelé à se prononcer sur la fixation de tarifs jouit d’un pouvoir discrétionnaire absolu quant aux éléments qu’il juge susceptibles d’influer sur les tarifs, mais qu’il doit, lorsqu’il établit la tarification, satisfaire aux deux exigences expressément prévues à l’al. (b) » (p. 856). Ainsi, l’organisme de réglementation est tenu par cette loi de faire en sorte que le public ne paie que ce qui est juste et raisonnable et que le service public obtienne un rendement juste et raisonnable sur la valeur des biens qu’il a utilisés *ou acquis de manière prudente et raisonnable*. Cette protection légale expresse du recouvrement du coût des biens acquis avec prudence offre un exemple de libellé législatif sur le fondement duquel notre Cour a conclu à l’existence d’une obligation non discrétionnaire d’assurer au service public un rendement juste sur les immobilisations qu’il a utilisées ou acquises avec prudence.

[97] En 2005, la Nova Scotia Utility and Review Board (« NSUARB ») a examiné puis adopté la définition du critère de l’investissement prudent proposée par l’Illinois Commerce Commission :

[TRADUCTION] . . . la prudence est la norme de diligence qu’une personne raisonnable aurait respectée dans la situation rencontrée par la direction du service public au moment où elle a dû prendre les décisions. [. . .] Le recul est exclu lorsqu’il s’agit d’apprécier la prudence. [. . .]

prudent if it was within the range of decisions reasonable persons might have made. . . . The prudence standard recognizes that reasonable persons can have honest differences of opinion without one or the other necessarily being imprudent.

(*Nova Scotia Power Inc., Re*, 2005 NSUARB 27 (“*Nova Scotia Power 2005*”), at para. 84 (CanLII))

The NSUARB then wrote that “[f]ollowing a review of the cases, the Board finds that the definition of imprudence as set out by the Illinois Commerce Commission is a reasonable test to be applied in *Nova Scotia*”: para. 90. The NSUARB then considered, among other things, whether the utility’s recent fuel procurement strategy had been prudent, and found that it had not: para. 94. It did not, however, indicate that it believed itself to be compelled to apply the prudent investment test.

[98] The NSUARB reaffirmed its endorsement of the prudent investment test in 2012: *Nova Scotia Power Inc. (Re)*, 2012 NSUARB 227 (“*Nova Scotia Power 2012*”), at paras. 143-46 (CanLII). In that case, the utility whose submissions were under review “confirmed that from its perspective this is the test the Board should apply”: para. 146. The NSUARB then applied the prudence test in evaluating whether several of the utility’s operational decisions were prudent, and found that some were not: para. 188.

[99] In 2006, the Ontario Court of Appeal considered the meaning of the prudent investment test in *Enbridge*. This case is of particular interest for two reasons. First, the Ontario Court of Appeal endorsed in its reasons a specific formulation of the prudent investment test framework:

– Decisions made by the utility’s management should generally be presumed to be prudent unless challenged on reasonable grounds.

La décision du service public est prudente si elle fait partie des décisions qu’une personne raisonnable aurait pu prendre. [. . .] La norme de la prudence reconnaît que des personnes raisonnables peuvent sincèrement différer d’opinions sans pour autant que l’une ou l’autre soit imprudente.

(*Nova Scotia Power Inc., Re*, 2005 NSUARB 27 (« *Nova Scotia Power 2005* »), par. 84 (CanLII))

La NSUARB conclut alors que, [TRADUCTION] « [a]près examen de la jurisprudence, [. . .] la définition d’imprudence proposée par l’Illinois Commerce Commission constitue un critère raisonnable susceptible d’application en Nouvelle-Écosse » (par. 90). Elle se demande notamment si la stratégie récente d’achat de carburant du service public a été prudente, et elle répond par la négative (par. 94). Elle ne se dit cependant pas tenue d’appliquer le critère de l’investissement prudent.

[98] En 2012, la NSUARB a renouvelé son adhésion au critère de l’investissement prudent (*Nova Scotia Power Inc. (Re)*, 2012 NSUARB 227 (« *Nova Scotia Power 2012* »), par. 143-146 (CanLII)). Dans cette affaire, le service public dont les arguments faisaient l’objet de l’examen [TRADUCTION] « a confirmé que, selon lui, il s’agit du critère que la commission devrait appliquer » (par. 146). La NSUARB a ensuite appliqué le critère de la prudence pour décider si plusieurs décisions opérationnelles du service public avaient été prudentes ou non, et elle a conclu que certaines d’entre elles ne l’avaient pas été (par. 188).

[99] En 2006, dans l’arrêt *Enbridge*, la Cour d’appel de l’Ontario se penche sur la teneur du critère de l’investissement prudent. Cet arrêt revêt un intérêt particulier pour deux raisons. Premièrement, la Cour d’appel y circonscrit précisément l’application du critère :

[TRADUCTION]

– La décision de la direction du service public est généralement présumée prudente, sauf contestation pour motifs valables.

– To be prudent, a decision must have been reasonable under the circumstances that were known or ought to have been known to the utility at the time the decision was made.

– Hindsight should not be used in determining prudence, although consideration of the outcome of the decision may legitimately be used to overcome the presumption of prudence.

– Prudence must be determined in a retrospective factual inquiry, in that the evidence must be concerned with the time the decision was made and must be based on facts about the elements that could or did enter into the decision at the time. [para. 10]

[100] Second, the Court of Appeal in *Enbridge* made certain statements that suggest that the prudent investment test was a necessary approach to reviewing committed costs. Specifically, it noted that in deciding whether Enbridge’s requested rate increase was just and reasonable,

the [Board] was required to balance the competing interests of Enbridge and its consumers. That balancing process is achieved by the application of what is known in the utility rate regulation field as the “prudence” test. Enbridge was entitled to recover its costs by way of a rate increase only if those costs were “prudently” incurred. [para. 8]

The Court of Appeal also noted that the Board had applied the “proper test”: para. 18. These statements tend to suggest that the Court of Appeal was of the opinion that prudence review is an inherent and necessary part of ensuring just and reasonable payments.

[101] However, the question of whether the prudence test was a required feature of just-and-reasonable analysis in this context was not squarely before the Court of Appeal in *Enbridge*. Rather, the parties in that case “were in substantial agreement on the general approach the Board should take to reviewing the prudence of a utility’s decision” (para. 10), and the question at issue was whether

– Pour qu’elle soit prudente, la décision doit être raisonnable eu égard aux circonstances que connaissait ou qu’aurait dû connaître le service public au moment où il l’a prise.

– Le recul est exclu de l’appréciation de la prudence, même lorsque les conséquences de la décision peuvent légitimement servir à réfuter la présomption de prudence.

– La prudence est appréciée dans le cadre d’une analyse factuelle rétrospective en ce que la preuve doit porter sur le moment où la décision a été prise et reposer sur des faits quant aux éléments qui ont pu entrer en ligne de compte ou qui sont effectivement entrés en ligne de compte dans la décision. [par. 10]

[100] Deuxièmement, elle donne plusieurs fois à entendre que le recours au critère de l’investissement prudent est nécessaire pour se prononcer sur les dépenses convenues. Plus précisément, elle signale que pour décider du caractère juste et raisonnable de l’augmentation des tarifs demandée par Enbridge,

[TRADUCTION] la [Commission] était tenue de soupeser les intérêts opposés d’Enbridge et des consommateurs. Pour ce faire, elle devait appliquer ce qu’on appelle dans le domaine de la réglementation des tarifs des services publics le critère de la « prudence ». Enbridge était en droit de recouvrer ses coûts au moyen d’une augmentation de ses tarifs, mais seulement si la décision derrière ces coûts était « prudente ». [par. 8]

La Cour d’appel ajoute que la Commission a appliqué le [TRADUCTION] « bon critère » (par. 18). Ces affirmations tendent à indiquer que, selon la Cour d’appel, le contrôle de la prudence est fondamental et nécessaire afin que les paiements soient justes et raisonnables.

[101] Or, dans cette affaire, la Cour d’appel n’était pas directement saisie de la question de savoir si, dans ce contexte, l’application du critère de la prudence était nécessaire à l’appréciation du caractère juste et raisonnable des paiements. En fait, les parties s’entendaient [TRADUCTION] « pour l’essentiel sur la démarche qui devait être celle de la Commission pour apprécier la prudence d’une décision d’un

the Board had reasonably applied that agreed-upon approach. In this sense, *Enbridge* is similar to *Nova Scotia Power 2012*: both cases involved the application of prudence analysis in contexts where there was no dispute over whether an alternative methodology could reasonably have been applied.

(3) Conclusion Regarding the Prudent Investment Test

[102] The prudent investment test, or prudence review, is a valid and widely accepted tool that regulators may use when assessing whether payments to a utility would be just and reasonable. While there exist different articulations of prudence review, *Enbridge* presents one express statement of how a regulatory board might structure its review to assess the prudence of utility expenditures at the time they were incurred or committed. A no-hindsight prudence review has most frequently been applied in the context of capital costs, but *Enbridge* and *Nova Scotia Power* (both 2005 and 2012) provide examples of its application to decisions regarding operating costs as well. I see no reason in principle why a regulatory board should be barred from applying the prudence test to operating costs.

[103] However, I do not find support in the statutory scheme or the relevant jurisprudence for the notion that the Board should be *required* as a matter of law, under the *Ontario Energy Board Act, 1998*, to apply the prudence test as outlined in *Enbridge* such that the mere decision not to apply it when considering committed costs would render its decision on payment amounts unreasonable. Nor is the creation of such an obligation by this Court justified. As discussed above, where a statute requires only that the regulator set “just and reasonable” payments, as the *Ontario Energy Board Act, 1998* does in Ontario, the regulator may make use of a variety of analytical tools in assessing the justness and reasonableness of a utility’s proposed payment

service public » (par. 10). La question en litige était celle de savoir si la Commission avait eu recours à cette démarche de manière raisonnable. En ce sens, l’affaire *Enbridge* s’apparente à *Nova Scotia Power 2012* : les deux concernent l’application du critère de la prudence lorsqu’aucune des parties ne soutient qu’une autre démarche aurait pu raisonnablement s’appliquer.

(3) Conclusion sur le critère de l’investissement prudent

[102] Le critère de l’investissement prudent — ou contrôle de la prudence — offre aux organismes de réglementation un moyen valable et largement reconnu d’apprécier le caractère juste et raisonnable des paiements sollicités par un service public. Il existe certes des formulations différentes du contrôle de la prudence, mais l’arrêt *Enbridge* précise en détail quelle peut être la démarche d’un organisme de réglementation appelé à décider si, au moment où le service public les a faites ou en a convenu, les dépenses étaient prudentes ou non. Le plus souvent, le contrôle de la prudence excluant le recul s’applique aux coûts en capital, mais l’arrêt *Enbridge* et les décisions *Nova Scotia Power* (2005 et 2012) montrent qu’il s’applique aussi aux dépenses d’exploitation. Je ne vois aucune raison de principe d’interdire à un organisme de réglementation d’appliquer le critère de la prudence aux dépenses d’exploitation.

[103] Toutefois, aucun élément du régime législatif ou de la jurisprudence applicable ne me paraît appuyer l’idée que la Commission devrait être *tenue* en droit, suivant la *Loi de 1998 sur la Commission de l’énergie de l’Ontario*, d’appliquer le critère de la prudence énoncé dans l’arrêt *Enbridge*, de sorte que la seule décision de ne pas l’appliquer pour apprécier la prudence de dépenses convenues rendrait déraisonnable sa décision sur les paiements. Notre Cour n’est pas non plus justifiée de créer pareille obligation. Je le répète, lorsqu’un texte législatif — telle la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* en Ontario — exige seulement qu’il fixe des paiements « justes et raisonnables », l’organisme de réglementation peut avoir recours à divers

amounts. This is particularly so where, as here, the regulator has been given express discretion over the methodology to be used in setting payment amounts: O. Reg. 53/05, s. 6(1).

[104] To summarize, it is not necessarily unreasonable, in light of the particular regulatory structure established by the *Ontario Energy Board Act, 1998*, for the Board to evaluate committed costs using a method other than a no-hindsight prudence review. As noted above, applying a presumption of prudence would have conflicted with the burden of proof in the *Ontario Energy Board Act, 1998* and would therefore not have been reasonable. The question of whether it was reasonable to assess a particular cost using hindsight should turn instead on the circumstances of that cost. I emphasize, however, that this decision should not be read to give regulators *carte blanche* to disallow a utility's committed costs at will. Prudence review of committed costs may in many cases be a sound way of ensuring that utilities are treated fairly and remain able to secure required levels of investment capital. As will be explained, particularly with regard to committed capital costs, prudence review will often provide a reasonable means of striking the balance of fairness between consumers and utilities.

[105] This conclusion regarding the Board's ability to select its methodology rests on the particulars of the statutory scheme under which the Board operates. There exist other statutory schemes in which regulators are expressly required to compensate utilities for certain costs prudently incurred: see *British Columbia Electric Railway Co.* Under such a framework, the regulator's methodological discretion may be more constrained.

moyens d'analyse pour apprécier le caractère juste et raisonnable des paiements sollicités par le service public. Cela est particulièrement vrai lorsque, comme en l'espèce, l'organisme de réglementation se voit accorder expressément un pouvoir discrétionnaire quant à la méthode à appliquer pour fixer les paiements (règlement 53/05, par. 6(1)).

[104] En résumé, il n'est pas nécessairement déraisonnable, à la lumière du cadre réglementaire établi par la *Loi de 1998 sur la Commission de l'énergie de l'Ontario*, que la Commission se prononce sur les dépenses convenues en employant une autre méthode que l'application d'un critère de prudence qui exclut le recul. Comme nous l'avons vu, présumer la prudence serait incompatible avec le fardeau de preuve que prévoit la *Loi de 1998 sur la Commission de l'énergie de l'Ontario* et, de ce fait, déraisonnable. Qu'il soit raisonnable ou non d'apprécier certaines dépenses avec le recul devrait plutôt dépendre des circonstances de la décision dont s'originent ces dépenses. Je précise toutefois que la présente décision ne doit pas être interprétée de façon à permettre aux organismes de réglementation de refuser à leur guise d'approuver des dépenses convenues. Le contrôle de la prudence de dépenses convenues peut, dans bien des cas, constituer un bon moyen de faire en sorte que les services publics soient traités équitablement et demeurent aptes à obtenir les investissements de capitaux requis. Comme je l'explique plus loin, en ce qui a trait plus particulièrement aux coûts en capital convenus, le contrôle de la prudence offre le plus souvent un moyen raisonnable d'établir un équilibre entre les intérêts du consommateur et ceux du service public.

[105] Cette conclusion sur le pouvoir de la Commission de décider de sa démarche découle du régime législatif qui régit son fonctionnement. D'autres régimes législatifs prévoient expressément que l'organisme de réglementation en cause est tenu d'indemniser le service public de certaines dépenses découlant de décisions prudentes (voir l'arrêt *British Columbia Electric Railway Co.*). Selon ces autres cadres législatifs, le pouvoir discrétionnaire qui permet à l'organisme de réglementation de décider de sa démarche peut être plus restreint.

(4) Application to the Board's Decision

[106] In this case, the Board disallowed a total of \$145 million in compensation costs associated with OPG's nuclear operations, over two years. As discussed above, these costs are best understood as at least partly committed. In view of the nature of these particular costs and the circumstances in which they became committed, I do not find that the Board acted unreasonably in not applying the prudent investment test in determining whether it would be just and reasonable to compensate OPG for these costs.

[107] First, the costs at issue are operating costs, rather than capital costs. Capital costs, particularly those pertaining to areas such as capacity expansion or upgrades to existing facilities, often entail some amount of risk, and may not always be strictly necessary to the short-term ongoing production of the utility. Nevertheless, such costs may often be a wise investment in the utility's future health and viability. As such, prudence review, including a no-hindsight approach (with or without a presumption of prudence, depending on the applicable statutory context), may play a particularly important role in ensuring that utilities are not discouraged from making the optimal level of investment in the development of their facilities.

[108] Operating costs, like those at issue here, are different in kind from capital costs. There is little danger in this case that a disallowance of these costs will have a chilling effect on OPG's willingness to incur operating costs in the future, because costs of the type disallowed here are an inescapable element of operating a utility. It is true that a decision such as the Board's in this case may have the effect of making OPG more hesitant about committing to relatively high compensation costs, but that was precisely the intended effect of the Board's decision.

(4) Application à la décision de la Commission

[106] En l'espèce, la Commission refuse à OPG le recouvrement au total de 145 millions de dollars au titre des dépenses de rémunération dans le secteur nucléaire, sur deux ans. Rappelons qu'il faut considérer que ces dépenses constituent, du moins en partie, des dépenses convenues. Compte tenu de la nature de ces dépenses en particulier et des circonstances dans lesquelles le service public en a convenu, je ne saurais conclure que la Commission a agi déraisonnablement en n'appliquant pas le critère de l'investissement prudent pour décider s'il était juste et raisonnable d'indemniser OPG à leur égard.

[107] Premièrement, il s'agit de dépenses d'exploitation, et non de coûts en capital. Les coûts en capital, en particulier ceux qui se rapportent par exemple à l'accroissement de la capacité ou à l'amélioration des installations actuelles, comportent souvent un risque et peuvent ne pas être nécessaires, à strictement parler, à la production à court terme du service public. Ces coûts peuvent néanmoins constituer un investissement judicieux pour le bon fonctionnement et la viabilité ultérieurs de ce dernier. Dès lors, le contrôle de la prudence, qui exclut le recul (et présume ou non la prudence, selon les dispositions législatives applicables), peut jouer un rôle particulièrement important pour faire en sorte que le service public ne soit pas dissuadé d'investir de manière optimale dans le développement de ses installations.

[108] Les dépenses d'exploitation, comme celles visées en l'espèce, diffèrent des coûts en capital. Il est peu probable que le refus de les approuver dissuade OPG d'en faire à l'avenir, car les dépenses de la nature de celles qui ont été refusées sont inhérentes à l'exploitation d'un service public. Certes, une décision comme celle rendue par la Commission en l'espèce peut faire hésiter OPG à convenir de dépenses relativement élevées au chapitre de la rémunération, mais tel était précisément l'effet voulu par la Commission.

[109] Second, the costs at issue arise in the context of an ongoing, “repeat-player” relationship between OPG and its employees. Prudence review has its origins in the examination of decisions to pursue particular investments, such as a decision to invest in capacity expansion; these are often one-time decisions made in view of a particular set of circumstances known or assumed at the time the decision was made.

[110] By contrast, OPG’s committed compensation costs arise in the context of an ongoing relationship in which OPG will have to negotiate compensation costs with the same parties in the future. Such a context supports the reasonableness of a regulator’s decision to weigh all evidence it finds relevant in striking a just and reasonable balance between the utility and consumers, rather than confining itself to a no-hindsight approach. Prudence review is simply less relevant when the Board’s focus is not solely on compensating for past commitments, but on regulating costs to be incurred in the future as well. As will be discussed further, the Board’s ultimate disallowance was not targeted exclusively at committed costs, but rather was made with respect to the total compensation costs it evaluated in aggregate. Though the Board acknowledged that OPG may not have had the discretion to reduce spending by the entire amount of the disallowance, the disallowance was animated by the Board’s efforts to get OPG’s ongoing compensation costs under control.

[111] Having already given OPG a warning that the Board found its operational costs to be of concern (see Board 2008-2009 Decision, at pp. 28-32), it was not unreasonable for the Board to be more forceful in considering compensation costs to ensure effective regulation of such costs going forward. The Board’s statement that its disallowance was intended “to send a clear signal that OPG must take responsibility for improving its performance” (Board Decision, at para. 350) shows that it had the ongoing effects of its disallowance squarely in mind in issuing its decision in this case.

[109] Deuxièmement, les dépenses en cause découlent d’une relation continue entre OPG et ses employés. Le contrôle de la prudence tire son origine de l’examen de décisions d’effectuer certains investissements, notamment pour accroître la capacité; il s’agit souvent de décisions isolées prises à la lumière d’un ensemble de données alors connues ou supposées.

[110] À l’opposé de celles issues de telles décisions, les dépenses de rémunération convenues d’OPG découlent d’une relation continue dans le cadre de laquelle OPG devra négocier ultérieurement les barèmes de rémunération avec les mêmes parties. Pareil contexte milite en faveur du caractère raisonnable de la décision de l’organisme de réglementation de soupeser toute preuve qu’il juge pertinente aux fins d’établir un équilibre juste et raisonnable entre le service public et les consommateurs, au lieu de s’en tenir à une approche excluant le recul. Le contrôle de la prudence se révèle tout simplement moins indiqué lorsque la Commission n’entend pas seulement indemniser le service public des engagements déjà pris, mais aussi réguler les dépenses qui seront faites dans l’avenir. En fin de compte, le refus de la Commission ne vise pas que des dépenses convenues, mais bien la totalité des dépenses de rémunération considérées globalement. Même si la Commission reconnaît qu’OPG n’avait peut-être pas de pouvoir discrétionnaire lui permettant de réduire ses dépenses à raison du montant total refusé, le refus de la Commission vise à inciter OPG à la maîtrise constante de ses dépenses de rémunération.

[111] Après que la Commission eut signifié à OPG que ses dépenses d’exploitation lui paraissaient préoccupantes (voir la décision 2008-2009 de la Commission, p. 28-32), il n’était pas déraisonnable qu’elle se montre plus stricte dans l’examen des dépenses de rémunération du service public afin d’en assurer la régulation réelle à l’avenir. Le fait que la Commission dit refuser l’approbation [TRADUCTION] « afin de signifier clairement à OPG qu’il lui incombe d’accroître sa performance » (décision de la Commission, par. 350) montre qu’elle a bel et bien conscience des répercussions actuelles de son refus.

[112] The reasonableness of the Board’s decision to disallow \$145 million in compensation costs is supported by the Board’s recognition of the fact that OPG was bound to a certain extent by the collective agreements in making staffing decisions and setting compensation rates, and its consideration of this factor in setting the total disallowance: Board Decision, at para. 350. The Board’s methodological flexibility ensures that its decision need not be “all or nothing”. Where appropriate, to the extent that the utility was unable to reduce its costs, the total burden of such costs may be moderated or shared as between the utility’s shareholders and the consumers. The Board’s moderation in this case shows that, in choosing to disallow costs without applying a formal no-hindsight prudence review, it remained mindful of the need to ensure that any disallowance was not unfair to OPG and certainly did not impair the viability of the utility.

[113] Justice Abella, in her dissent, acknowledges that the Board has the power under prudence review to disallow committed costs in at least some circumstances: para. 152. However, she speculates that any such disallowance could “imperil the assurance of reliable electricity service”: para. 156. A large or indiscriminate disallowance might create such peril, but it is also possible for the Board to do as it did here, and temper its disallowance to recognize the realities facing the utility.

[114] There is no dispute that collective agreements are “immutable” between employees and the utility. However, if the legislature had intended for costs under collective agreements to also be inevitably imposed on consumers, it would not have seen fit to grant the Board oversight of utility compensation costs. The existence both of collective bargaining for utility employees and of the Board’s power to fix payment amounts covering compensation costs indicates neither regime can trump the other. The Board cannot interfere with the collective agreement by ordering that a utility break its

[112] Le caractère raisonnable du refus de la Commission d’approuver des dépenses de 145 millions de dollars au titre de la rémunération tient à ce qu’elle reconnaît qu’OPG était liée dans une certaine mesure par les conventions collectives dans sa prise de décisions en matière de personnel et dans la fixation des barèmes de rémunération, et à ce qu’elle en tient compte pour déterminer la somme totale refusée (décision de la Commission, par. 350). La souplesse méthodologique dont bénéficie la Commission lui permet d’éviter les extrêmes. Lorsque le service public ne peut réduire ses dépenses, la prise en charge de celles-ci peut, si le dossier s’y prête, être modérée ou répartie entre les actionnaires du service public et les consommateurs. La modération opérée par la Commission en l’espèce montre que, en refusant d’approuver les dépenses sans recourir formellement à un contrôle de la prudence excluant le recul, elle ne perd pas de vue la nécessité de veiller à ce que tout refus ne soit pas injuste envers OPG ni, assurément, à ce qu’il ne nuise pas à sa viabilité.

[113] Dans ses motifs de dissidence, la juge Abella reconnaît que, lors du contrôle de la prudence, la Commission peut, du moins dans certaines circonstances, refuser des dépenses convenues (par. 152). Elle dit toutefois craindre qu’un tel refus puisse « mettre en péril la garantie d’un service d’électricité fiable » (par. 156). Le refus d’une somme importante ou opposé sans discernement pourrait exposer à un tel risque, mais il se peut aussi que l’organisme de réglementation fasse ce que la Commission fait en l’espèce, c’est-à-dire modérer son refus en tenant compte des réalités auxquelles fait face le service public.

[114] Nul ne conteste que les conventions collectives intervenues entre le service public et ses employés sont « immuables ». Toutefois, si le législateur avait voulu que les dépenses qui en sont issues se répercutent inévitablement sur les consommateurs, il n’aurait pas jugé opportun d’investir la Commission du pouvoir de surveiller les dépenses de rémunération d’un service public. La coexistence du droit à la négociation collective des employés du service public et du pouvoir de la Commission de fixer le montant des paiements pour les dépenses de rémunération indique que ni l’un ni l’autre n’a

obligations thereunder, but nor can the collective agreement supersede the Board’s duty to ensure a just and reasonable balance between utility and consumer interests.

[115] Justice Abella says that the Board’s review of committed costs using hindsight evidence appears to contradict statements made earlier in its decision. The Board wrote that it would use all relevant evidence in assessing forecast costs but that it would limit itself to a no-hindsight approach in reviewing costs that OPG could not “take action to reduce”: Board Decision, at para. 75. In my view, these statements can be read as setting out a reasonable approach for analyzing costs that could reliably be fit into forecast or committed categories. However, not all costs are amenable to such clean categorization by the Board in assessing payment amounts for a test period.

[116] With regard to the compensation costs at issue here, the Board declined to split the total cost disallowance into forecast and committed components in conducting its analysis. As Hoy J. observed, “[g]iven the complexity of OPG’s business, and respecting its management’s autonomy, [the Board] did not try to quantify precisely the amount by which OPG could reduce its forecast compensation costs within the framework of the existing collective bargaining agreements”: Div. Ct. reasons, at para. 53. That is, the Board did not split all compensation costs into either “forecast” or “committed”, but analyzed the disallowance of compensation costs as a mix of forecast and committed expenditures over which management retained some, but not total, control.

préséance. La Commission ne peut empiéter sur les conventions collectives en ordonnant au service public de manquer aux obligations qu’elles lui imposent, mais les conventions collectives ne priment pas l’obligation de la Commission d’assurer un équilibre juste et raisonnable entre le service public et les consommateurs.

[115] La juge Abella affirme que l’examen des dépenses convenues auquel se livre la Commission à partir d’éléments de recul paraît contredire ce que l’organisme affirme précédemment dans sa décision. La Commission écrit en effet qu’elle prendra en compte tout élément de preuve pertinent pour apprécier les dépenses prévues, mais qu’elle s’en tiendra à un examen sans recul pour ce qui concerne les dépenses à l’égard desquelles OPG [TRADUCTION] « ne pouvait prendre de mesures de réduction » (décision de la Commission, par. 75). À mon sens, on peut en conclure qu’elle recourt à une démarche raisonnable pour l’analyse de dépenses que l’on peut assimiler avec assurance soit à des dépenses prévues, soit à des dépenses convenues. Cependant, toutes les dépenses ne sont pas susceptibles d’une distinction aussi nette par la Commission lorsqu’il s’agit d’apprécier le montant des paiements pour une période de référence.

[116] En ce qui a trait aux dépenses de rémunération en cause, la Commission refuse de préciser quelle partie de la somme totale refusée correspond à des dépenses prévues et quelle partie correspond à des dépenses convenues pour les besoins de son analyse. Le juge Hoy fait observer que, [TRADUCTION] « [v]u la complexité de l’activité d’OPG et l’autonomie de gestion dont elle jouit, [la Commission] n’a pas tenté de déterminer avec précision le montant dont les dépenses de rémunération prévues d’OPG auraient pu être réduites dans le contexte des conventions collectives en vigueur » (motifs de la C. div., par. 53). En somme, la Commission ne départage pas les dépenses de rémunération totales entre celles qui sont « prévues » et celles qui sont « convenues ». Elle considère plutôt que les dépenses de rémunération refusées se composent à la fois de dépenses prévues et de dépenses convenues sur lesquelles la direction conservait une certaine maîtrise, mais non une maîtrise totale.

[117] It was not unreasonable for the Board to proceed on the basis that predicting staff attrition rates is an inherently uncertain exercise, and that it is not equipped to micromanage business decisions within the purview of OPG management. These considerations mean that any attempt to predict the exact degree to which OPG would be able to reduce compensation costs (in other words, what share of the costs were forecast) would be fraught with uncertainty. Accordingly, it was not unreasonable for the Board to adopt a mixed approach that did not rely on quantifying the exact share of compensation costs that fell into the forecast and committed categories. Such an approach is not inconsistent with the Board’s discussion at paras. 73-75, but rather represents an exercise of the Board’s methodological discretion in addressing a challenging issue where these costs did not fit easily into the categories discussed in that passage.

[118] Justice Abella emphasizes throughout her reasons that the costs established by the collective agreements were not adjustable. I do not dispute this point. However, to the extent that she relies on the observation that the collective agreements “made it *illegal* for the utility to alter the compensation and staffing levels” of the unionized workforce (para. 149 (emphasis in original)), one might conclude that the Board was in some way trying to interfere with OPG’s obligations under its collective agreements. It is important not to lose sight of the fact that the Board decision in no way purports to force OPG to break its contractual commitments to unionized employees.

[119] Finally, her observation that the Canadian Nuclear Safety Commission (“CNSC”) “has . . . imposed staffing levels on Ontario Power Generation to ensure safe and reliable operation of its nuclear stations” (para. 127) is irrelevant to the issues raised in this case. While the regime put in place by the CNSC surely imposes operational and staffing restraints on nuclear utilities (see OPG record, at

[117] Il n’est pas déraisonnable que la Commission considère que la prévision du taux d’attrition du personnel constitue en soi une entreprise incertaine et qu’elle n’est pas en mesure de microgérer les décisions d’affaires qui relèvent des dirigeants d’OPG. Dès lors, toute tentative de prédire la mesure exacte dans laquelle OPG pourrait abaisser ses dépenses de rémunération (autrement dit, quelle partie de ces dépenses est prévue) serait empreinte d’incertitude. Il n’est donc pas déraisonnable que la Commission opte pour une démarche hybride qui ne se fonde pas sur la répartition exacte des dépenses de rémunération entre celles qui sont prévues et celles qui sont convenues. Pareille démarche est compatible avec l’analyse de la Commission figurant aux par. 73-75 de sa décision et correspond à un exercice du pouvoir discrétionnaire de la Commission sur le plan méthodologique lorsqu’elle est appelée à se prononcer sur une question épineuse et que les dépenses en cause ne sont pas aisément assimilables à l’une ou l’autre des catégories mentionnées dans cette analyse.

[118] Tout au long de ses motifs, la juge Abella rappelle que les dépenses découlant des conventions collectives ne peuvent être rajustées. Je n’en disconviens pas. Cependant, lorsqu’elle opine que les conventions collectives « rend[ent] *illégal* la modification par le service public [. . .] des barèmes de rémunération et des niveaux de dotation » à l’égard de son personnel syndiqué (par. 149 (en italique dans l’original)), d’aucuns pourraient en conclure que la Commission tente de quelque manière de s’immiscer dans l’exécution des obligations d’OPG suivant les conventions collectives. Il importe de ne pas oublier que la Commission n’entend pas, par sa décision, contraindre OPG à se soustraire à ses engagements contractuels envers ses employés.

[119] Enfin, la remarque de ma collègue selon laquelle la Commission canadienne de sûreté nucléaire (« CCSN ») « [a] impos[é] [. . .] des niveaux de dotation à Ontario Power Generation afin de garantir l’exploitation sûre et fiable de ses installations nucléaires » (par. 127) importe peu quant aux questions soulevées en l’espèce. Bien que le régime établi par la CCSN impose sûrement des conditions

pp. 43-46), there is nothing in the Board's reasons, and no argument presented before this Court, suggesting that the Board's disallowance will result in a violation of the provisions of the *Nuclear Safety and Control Act*, S.C. 1997, c. 9.

[120] I have noted above that it is essential for a utility to earn its cost of capital in the long run. The Board's disallowance may have adversely impacted OPG's ability to earn its cost of capital in the short run. Nevertheless, the disallowance was intended "to send a clear signal that OPG must take responsibility for improving its performance" (Board Decision, at para. 350). Such a signal may, in the short run, provide the necessary impetus for OPG to bring its compensation costs in line with what, in the Board's opinion, consumers should justly expect to pay for an efficiently provided service. Sending such a signal is consistent with the Board's market proxy role and its objectives under s. 1 of the *Ontario Energy Board Act*, 1998.

VI. Conclusion

[121] I do not find that the Board acted improperly in pursuing this matter on appeal; nor do I find that it acted unreasonably in disallowing the compensation costs at issue. Accordingly, I would allow the appeal, set aside the decision of the Court of Appeal, and reinstate the decision of the Board.

The following are the reasons delivered by

[122] ABELLA J. (dissenting) — The Ontario Energy Board was established in 1960 to set rates for the sale and storage of natural gas and to approve pipeline construction projects. Over time, its powers and responsibilities evolved. In 1973, the Board became responsible for reviewing and reporting to the Minister of Energy on electricity rates. During this period, Ontario's electricity market was lightly regulated, dominated by the government-owned

d'exploitation et de dotation aux installations nucléaires (voir dossier OPG, p. 43-46), nul élément des motifs de la Commission et nulle plaidoirie devant notre Cour n'indiquent que le refus de la Commission entraînera le non-respect des dispositions de la *Loi sur la sûreté et la réglementation nucléaires*, L.C. 1997, c. 9.

[120] Je rappelle qu'il est essentiel qu'un service public obtienne à long terme l'équivalent du coût du capital. Le refus de la Commission a pu nuire à la possibilité qu'OPG obtienne à court terme l'équivalent de son coût du capital. Toutefois, il vise à [TRADUCTION] « signifier clairement à OPG qu'il lui incombe d'accroître sa performance » (décision de la Commission, par. 350). L'envoi d'un tel message peut, à court terme, donner à OPG l'impulsion nécessaire pour rapprocher ses dépenses de rémunération de ce que, selon la Commission, les consommateurs devraient à bon droit s'attendre à payer pour la prestation efficace du service. L'envoi d'un tel message est conforme au rôle de substitut du marché de la Commission et à ses objectifs selon l'article premier de la *Loi de 1998 sur la Commission de l'énergie de l'Ontario*.

VI. Conclusion

[121] Je conclus que la Commission n'a pas agi de manière inappropriée en se pourvoyant en tant que partie en appel; elle n'a pas non plus agi déraisonnablement en refusant d'approuver les dépenses de rémunération en cause. Par conséquent, je suis d'avis d'accueillir le pourvoi, d'annuler la décision de la Cour d'appel et de rétablir celle de la Commission.

Version française des motifs rendus par

[122] LA JUGE ABELLA (dissidente) — La Commission de l'énergie de l'Ontario a été mise sur pied en 1960. Son mandat était alors d'établir les tarifs applicables à la vente et au stockage de gaz naturel et d'autoriser les projets de construction de pipelines. Au fil du temps, ses compétences et ses fonctions ont évolué. En 1973, le législateur lui a confié la responsabilité d'examiner les tarifs d'électricité puis de faire rapport au ministre de l'Énergie. Pendant cette

Ontario Hydro, which owned power generation assets responsible for about 90 per cent of electricity production in the province: Ron W. Clark, Scott A. Stoll and Fred D. Cass, *Ontario Energy Law: Electricity* (2012), at p. 134; *2011 Annual Report* of the Office of the Auditor General of Ontario, at pp. 5 and 67.

[123] A series of legislative measures in the late 1990s were adopted to transform the electricity industry into a market-based one driven by competition. Ontario Hydro was unbundled into five entities. One of them was Ontario Power Generation Inc., which was given responsibility for controlling the power generation assets of the former Ontario Hydro. It was set up as a commercial corporation with one shareholder — the Province of Ontario: Clark, Stoll and Cass, at pp. 5-7 and 134.

[124] As of April 1, 2008, the Board was given the authority by statute to set payments for the electricity generated by a prescribed list of assets held by Ontario Power Generation: *Ontario Energy Board Act, 1998*, S.O. 1998, c. 15, Sch. B, s. 78.1(2); O. Reg. 53/05, *Payments Under Section 78.1 of the Act*, s. 3. Under the legislative scheme, Ontario Power Generation is required to apply to the Board for the approval of “just and reasonable” payment amounts: *Ontario Energy Board Act, 1998*, s. 78.1(5). The Board sets its own methodology to determine what “just and reasonable” payment amounts are, guided by the statutory objectives to maintain a “financially viable electricity industry” and to “protect the interests of consumers with respect to prices and the adequacy, reliability and quality of electricity service”: O. Reg. 53/05, s. 6(1); *Ontario Energy Board Act, 1998*, paras. 1 and 2 of s. 1(1).

[125] Ontario Power Generation remains the province’s largest electricity generator. It was unionized by the Ontario Hydro Employees’ Union (the predecessor to the Power Workers’ Union) in

période, en Ontario, le marché de l’électricité était peu réglementé. Il était dominé par la société d’État Ontario Hydro, qui possédait des installations de production d’énergie fournissant plus de 90 p. 100 de l’électricité dans la province (Ron W. Clark, Scott A. Stoll et Fred D. Cass, *Ontario Energy Law : Electricity* (2012), p. 134; *Rapport annuel 2011*, Bureau du vérificateur général de l’Ontario, p. 1 et 72).

[123] À la fin des années 1990, une série de mesures législatives a été adoptée en vue d’axer le secteur de l’électricité sur le marché et de le soumettre à la concurrence. Ontario Hydro a été scindée en cinq entités. L’une d’elles, Ontario Power Generation Inc., s’est vu confier l’actif de production d’électricité de l’ancienne société Ontario Hydro. Elle a été constituée en société commerciale dont le seul actionnaire est la province d’Ontario (Clark, Stoll et Cass, p. 5-7 et 134).

[124] Depuis le 1^{er} avril 2008, la Commission est légalement investie du pouvoir de fixer les paiements pour l’électricité produite par les installations prescrites que possède Ontario Power Generation (*Loi de 1998 sur la Commission de l’énergie de l’Ontario*, L.O. 1998, c. 15, ann. B, par. 78.1(2); règlement 53/05 de l’Ontario (*Payments Under Section 78.1 of the Act*) (« règlement 53/05 », art. 3). Suivant le régime législatif, Ontario Power Generation est tenue de faire une demande à la Commission pour obtenir l’approbation de paiements « justes et raisonnables » (*Loi de 1998 sur la Commission de l’énergie de l’Ontario*, par. 78.1(5)). La Commission établit sa propre méthode pour déterminer ce qui constitue des paiements « justes et raisonnables » au regard des objectifs législatifs qui consistent à maintenir une « industrie de l’électricité financièrement viable » et à « protéger les intérêts des consommateurs en ce qui concerne les prix, ainsi que la suffisance, la fiabilité et la qualité du service d’électricité » (règlement 53/05, par. 6(1); *Loi de 1998 sur la Commission de l’énergie de l’Ontario*, dispositions 1 et 2 du par. 1(1)).

[125] Ontario Power Generation demeure le plus grand producteur d’électricité de la province. L’Ontario Hydro Employees’ Union (auquel a succédé le Syndicat des travailleurs et travailleuses du secteur

the 1950s, and by the Society of Energy Professionals in 1992: Richard P. Chaykowski, *An Assessment of the Industrial Relations Context and Outcomes at OPG* (2013) (online), at s. 6.2. Today, Ontario Power Generation employs approximately 10,000 people in its regulated businesses, 90 per cent of whom are unionized. Two thirds of these unionized employees are represented by the Power Workers' Union, and the rest by the Society of Energy Professionals.

énergétique) a été accrédité comme agent négociateur auprès de l'entreprise dans les années 1950, alors que Society of Energy Professionals l'a été à son tour en 1992 (Richard P. Chaykowski, *An Assessment of the Industrial Relations Context and Outcomes at OPG* (2013) (en ligne), art. 6.2). Le personnel d'Ontario Power Generation affecté à ses activités réglementées se compose aujourd'hui d'environ 10 000 personnes, dont 90 p. 100 sont syndiquées. Deux tiers de ces employés syndiqués sont représentés par le Syndicat des travailleurs et travailleuses du secteur énergétique, un tiers par Society of Energy Professionals.

[126] Both the Power Workers' Union and the Society of Energy Professionals had collective agreements with Ontario Hydro before Ontario Power Generation was established. As a successor company to Ontario Hydro, Ontario Power Generation inherited the full range of these labour relations obligations: Ontario *Labour Relations Act*, 1995, S.O. 1995, c. 1, Sch. A, s. 69. Ontario Power Generation's collective agreements with its unions prevent the utility from unilaterally reducing staffing or compensation levels.

[126] Le syndicat des travailleurs et travailleuses du secteur énergétique et Society of Energy Professionals avaient tous deux conclu des conventions collectives avec Ontario Hydro avant la création d'Ontario Power Generation. Lorsqu'elle a succédé à Ontario Hydro, Ontario Power Generation a hérité de la totalité des obligations issues de ces conventions (*Loi de 1995 sur les relations de travail* de l'Ontario, L.O. 1995, c. 1, ann. A, art. 69), qui la lient et l'empêchent de réduire unilatéralement les niveaux de dotation ou les barèmes de rémunération.

[127] The Canadian Nuclear Safety Commission, an independent federal government agency responsible for ensuring compliance with the *Nuclear Safety and Control Act*, S.C. 1997, c. 9, has also imposed staffing levels on Ontario Power Generation to ensure safe and reliable operation of its nuclear stations.

[127] La Commission canadienne de sûreté nucléaire, un organisme fédéral indépendant chargé de faire respecter la *Loi sur la sûreté et la réglementation nucléaires*, L.C. 1997, c. 9, impose également des niveaux de dotation à Ontario Power Generation afin de garantir l'exploitation sûre et fiable de ses installations nucléaires.

[128] On May 26, 2010, Ontario Power Generation applied to the Board for a total revenue requirement of \$6,909.6 million, including \$2,783.9 million in compensation costs — wages, benefits, pension servicing, and annual incentives — to cover the period from January 1, 2011 to December 31, 2012: EB-2010-0008, at pp. 8, 49 and 80.

[128] Le 26 mai 2010, Ontario Power Generation a demandé à la Commission d'approuver des recettes nécessaires se chiffrant à 6 909,6 millions de dollars pour la période allant du 1^{er} janvier 2011 au 31 décembre 2012, dont 2 783,9 millions devaient être affectés à la rémunération du personnel — salaires, avantages sociaux, prestations de retraite et incitatifs annuels (EB-2010-0008, p. 8, 49 et 80).

[129] In its decision, the Board explained that it would use "two types of examination" to assess the utility's expenditures. When evaluating forecast costs — costs that the utility has estimated for

[129] Dans sa décision, la Commission dit soumettre à [TRADUCTION] « deux types d'examen » les dépenses du service public. En ce qui concerne les dépenses prévues — par le service public, pour une

a future period and which can still be reduced or avoided — the Board said that Ontario Power Generation bears the burden of showing that these costs are reasonable. On the other hand, when the Board would be evaluating costs for which “[t]here is no opportunity for the company to take action to reduce”, otherwise known as committed costs, it said that it would undertake “an after-the-fact prudence review . . . conducted in the manner which includes a presumption of prudence”, that is, a presumption that the utility’s expenditures are reasonable: p. 19.

[130] The Board made no distinction between those compensation costs that were reducible and those that were not. Instead, it subjected all compensation costs to the kind of assessment it uses for reducible, forecast costs and disallowed \$145 million because it concluded that the utility’s compensation rates and staffing levels were too high.

[131] On appeal, a majority of the Divisional Court upheld the Board’s order. In dissenting reasons, Aitken J. concluded that the Board’s decision was unreasonable because it did not apply the proper approach to the compensation costs which were, as a result of legally binding collective agreements, fixed and not adjustable. Instead, the Board “lumped” all compensation costs together and made no distinction between those that were the result of binding contractual obligations and those that were not. As she said:

First, I consider any limitation on [Ontario Power Generation’s] ability to manage nuclear compensation costs on a go-forward basis, due to binding collective agreements in effect prior to the application and the test period, to be costs previously incurred and subject to an after-the-fact, two-step, prudence review. Second, I conclude that, in considering [Ontario Power Generation’s] nuclear compensation costs, as set out in its application, the [Board] in its analysis (though not necessarily in its final number) was required to differentiate between such earlier incurred liabilities and other aspects of the nuclear compensation cost package that were truly projected and

période ultérieure et qu’il est toujours possible de réduire ou d’éviter —, la Commission soutient qu’il incombe à Ontario Power Generation de démontrer leur caractère raisonnable. En revanche, pour ce qui est des dépenses à l’égard desquelles « [l]a société ne pouvait prendre de mesures de réduction », à savoir les dépenses convenues, la Commission dit qu’elle effectuera « un contrôle de la prudence après coup, [. . .] comportant l’application d’une présomption de prudence », c’est-à-dire une présomption selon laquelle les dépenses du service public sont raisonnables (p. 19).

[130] La Commission ne fait aucune distinction entre les dépenses de rémunération qui sont réductibles et celles qui ne le sont pas. Elle soumet plutôt toutes les dépenses de rémunération à l’appréciation qu’elle réserve aux dépenses prévues réductibles et elle refuse d’approuver les paiements demandés à raison de 145 millions de dollars au motif que les barèmes de rémunération et les niveaux de dotation sont trop élevés.

[131] En appel, les juges majoritaires de la Cour divisionnaire confirment l’ordonnance de la Commission. Dans ses motifs dissidents, la juge Aitken conclut que la décision de la Commission est déraisonnable, car elle n’applique pas la bonne approche aux dépenses de rémunération, lesquelles constituent, par l’effet de conventions collectives contraignantes en droit, des dépenses fixes et non ajustables. Selon elle, la Commission [TRADUCTION] « regroupe » plutôt toutes les dépenses de rémunération et ne fait aucune distinction entre celles qui découlent d’obligations contractuelles obligatoires et celles qui n’en découlent pas. Comme elle l’affirme :

[TRADUCTION] Premièrement, j’estime que les dépenses de rémunération du secteur nucléaire [d’Ontario Power Generation], pour une période ultérieure, assujetties à une contrainte en raison de conventions collectives qui s’appliquaient avant la demande et la période de référence, constituent des dépenses déjà faites qui doivent faire l’objet d’un contrôle de la prudence après coup, en deux étapes. Deuxièmement, dans l’analyse (mais pas nécessairement dans l’appréciation finale) des dépenses de rémunération du secteur nucléaire dont fait état la demande, la [Commission] était tenue de faire une distinction entre les dépenses déjà effectuées et d’autres

not predetermined. Third, in my view, the [Board] was required to undergo a prudence review in regard to those aspects of the nuclear compensation package that arose under binding contracts entered prior to the application and the test period. In regard to the balance of factors making up the nuclear compensation package, the [Board] was free to determine, based on all available evidence, whether such factors were reasonable. Fourth, had a prudence review been undertaken, there was evidence upon which the [Board] could reasonably have decided that the presumption of prudence had been rebutted in regard to those cost factors mandated in the collective agreements. Unfortunately, I cannot find anywhere in the Decision of the [Board] where such an analysis was undertaken. The [Board] lumped all nuclear compensation costs together. It dealt with them as if they all emanated from the same type of factors and none reflected contractual obligations to which the [Ontario Power Generation] was bound due to a collective agreement entered prior to the application and the test period. Finally, I conclude that, when the [Board] was considering the reasonableness of the nuclear compensation package, it erred in considering evidence that came into existence after the date on which the collective agreements were entered when it assessed the reasonableness of the rates of pay and other binding provisions in the collective agreements. [para. 75]

[132] The Court of Appeal unanimously agreed with Aitken J.'s conclusion, finding that "the compensation costs at issue before the [Board] were committed costs" which should therefore have been assessed using a presumption of prudence. As they both acknowledged, it was open to the Board to find that the presumption had been rebutted in connection with the binding contractual obligations, but the Board acted unreasonably in failing to take the immutable nature of the fixed costs into consideration.

[133] I agree. The compensation costs for approximately 90 per cent of Ontario Power Generation's regulated workforce were established through legally binding collective agreements which obligated the utility to pay fixed levels of compensation, regulated staffing levels, and provided unionized employees with employment security. Ontario Power Generation's compensation costs

réellement prévues, mais non préétablies. Troisièmement, à mon avis, la [Commission] devait soumettre à un contrôle de la prudence la partie des dépenses de rémunération du secteur nucléaire qui découlait de contrats obligatoires conclus avant la demande et la période de référence. Pour ce qui est des autres facteurs présidant à la rémunération globale du secteur nucléaire, la [Commission] pouvait, en se fondant sur toute la preuve disponible, décider s'ils étaient raisonnables ou non. Quatrièmement, si un contrôle de la prudence avait été effectué, des éléments de preuve auraient pu raisonnablement permettre à la [Commission] de conclure à la réfutation de la présomption de prudence en ce qui a trait aux éléments issus des conventions collectives qui influaient sur les dépenses. Malheureusement, je constate que nulle part dans sa décision la [Commission] ne se livre à une telle analyse. Elle regroupe sans distinctions toutes les dépenses de rémunération du secteur nucléaire. Elle considère qu'elles ont toutes la même origine et qu'aucune ne découle d'obligations contractuelles auxquelles [Ontario Power Generation] était tenue par une convention collective conclue avant la demande et la période de référence. Enfin, j'estime que, lorsqu'elle se penche sur le caractère raisonnable de la rémunération globale du secteur nucléaire, la [Commission] commet l'erreur de tenir compte d'éléments de preuve ayant vu le jour après la conclusion des conventions collectives pour apprécier le caractère raisonnable des barèmes de rémunération et d'autres dispositions contraignantes des conventions collectives. [par. 75]

[132] La Cour d'appel souscrit à l'unanimité à la conclusion de la juge Aitken et statue que [TRA-DUCTION] « les dépenses de rémunération en cause devant la [Commission] étaient des dépenses convenues » qu'il aurait donc fallu apprécier en présupposant leur prudence. Elles reconnaissent toutes deux qu'il était loisible à la Commission de conclure que la présomption était réfutée en ce qui concerne les obligations contractuelles obligatoires, mais qu'elle a agi déraisonnablement en ne tenant pas compte de la nature immuable des coûts fixes.

[133] Je suis d'accord. Les dépenses de rémunération visant environ 90 p. 100 de l'effectif obligatoire d'Ontario Power Generation étaient établies par des conventions collectives contraignantes en droit qui imposaient des barèmes de rémunération fixes, qui déterminaient les niveaux de dotation et qui garantissaient la sécurité d'emploi des employés syndiqués. Les dépenses de rémunération

were therefore overwhelmingly predetermined and could not be adjusted by the utility during the relevant period. These are precisely the type of costs that the Board referred to in its decision as costs for which “[t]here is no opportunity for the company to take action to reduce” and which must be subjected to “a prudence review conducted in the manner which includes a presumption of prudence”: para. 75.

[134] In my respectful view, failing to acknowledge the legally binding, non-reducible nature of the cost commitments reflected in the collective agreements and apply the review the Board itself said should apply to such costs, rendered its decision unreasonable.

Analysis

[135] Pursuant to s. 78.1(5) of the *Ontario Energy Board Act, 1998*, upon application from Ontario Power Generation, the Board is required to determine “just and reasonable” payment amounts to the utility. In the utility regulation context, the phrase “just and reasonable” reflects the aim of “navigating the straits” between overcharging a utility’s customers and underpaying the utility for the public service it provides: *Verizon Communications Inc. v. Federal Communications Commission*, 535 U.S. 467 (2002), at p. 481; see also *Northwestern Utilities Ltd. v. City of Edmonton*, [1929] S.C.R. 186, at pp. 192-93.

[136] The methodology adopted by the Board to determine “just and reasonable” payments to Ontario Power Generation draws in part on the regulatory concept of “prudence”. Prudence is “a legal basis for adjudging the meeting of utilities’ public interest obligations, specifically in regard to rate proceedings”: Robert E. Burns et al., *The Prudent Investment Test in the 1980s*, report NRRI-84-16, The National Regulatory Research Institute, April 1985, at p. 20. The concept emerged in the early 20th century as a judicial response to the “mind-numbing complexity” of other approaches being

d’Ontario Power Generation étaient donc en très grande partie préétablies et ne pouvaient être rajustées par l’entreprise au cours de la période considérée. Il s’agit précisément du type de dépenses que la Commission qualifie, dans sa décision, de dépenses à l’égard desquelles [TRADUCTION] « [l]a société ne pouvait prendre de mesures de réduction » et qui doivent faire l’objet d’un « contrôle de la prudence comportant l’application d’une présomption de prudence » (par. 75).

[134] Soit dit tout en respect, la Commission rend une décision déraisonnable en ne reconnaissant pas le caractère contraignant en droit et non réductible des dépenses auxquelles le service public s’était engagé lors de la signature des conventions collectives et en omettant de soumettre ces dépenses au contrôle qui s’imposait pourtant selon elle à leur égard.

Analyse

[135] Conformément au par. 78.1(5) de la *Loi de 1998 sur la Commission de l’énergie de l’Ontario*, sur demande d’Ontario Power Generation, la Commission fixe le montant des paiements « justes et raisonnables » auxquels a droit le service public. Dans le contexte de la réglementation des services publics, l’expression « justes et raisonnables » traduit l’objectif qui consiste à [TRADUCTION] « naviguer entre les récifs » que sont, d’une part, les tarifs excessifs imposés au consommateur et, d’autre part, la rétribution insuffisante du service public (*Verizon Communications Inc. c. Federal Communications Commission*, 535 U.S. 467 (2002), p. 481; voir aussi *Northwestern Utilities Ltd. c. City of Edmonton*, [1929] R.C.S. 186, p. 192-193).

[136] La méthode retenue par la Commission pour déterminer le montant des paiements « justes et raisonnables » auxquels a droit Ontario Power Generation prend en partie appui sur la notion de « prudence ». En droit réglementaire, la prudence offre un [TRADUCTION] « fondement juridique pour se prononcer sur le respect des obligations des services publics liées à l’intérêt public, plus particulièrement en ce qui concerne le processus de tarification » (Robert E. Burns et autres, *The Prudent Investment Test in the 1980s*, rapport NRRI-84-16, The National Regulatory Research Institute, avril 1985, p. 20). Apparue

used by regulators to determine “just and reasonable” amounts, and introduced a legal presumption that a regulated utility has acted reasonably: *Verizon Communications*, at p. 482. As Justice Brandeis famously explained in 1923:

The term prudent investment is not used in a critical sense. There should not be excluded from the finding of the base, investments which, under ordinary circumstances, would be deemed reasonable. The term is applied for the purpose of excluding what might be found to be dishonest or obviously wasteful or imprudent expenditures. Every investment may be assumed to have been made in the exercise of reasonable judgment, unless the contrary is shown. [Emphasis added.]

(*State of Missouri ex rel. Southwestern Bell Telephone Co. v. Public Service Commission of Missouri*, 262 U.S. 276 (1923), at p. 289, fn. 1, per Brandeis J., dissenting)

[137] The presumption of prudence is the starting point for the type of examination the Board calls a “prudence review”. In undertaking a prudence review, the Board applies a “well-established set of principles”:

- Decisions made by the utility’s management should generally be presumed to be prudent unless challenged on reasonable grounds.
- To be prudent, a decision must have been reasonable under the circumstances that were known or ought to have been known to the utility at the time the decision was made.
- Hindsight should not be used in determining prudence, although consideration of the outcome of the decision may legitimately be used to overcome the presumption of prudence.
- Prudence must be determined in a retrospective factual inquiry, in that the evidence must be concerned with the time the decision was made and must be

au début du 20^e siècle, cette notion jurisprudentielle visait à remédier à la [TRADUCTION] « complexité paralysante » des approches différentes utilisées par les organismes de réglementation pour arrêter des montants « justes et raisonnables », et elle présumait que le service public réglementé avait agi raisonnablement (*Verizon Communications*, p. 482). Ainsi, comme l’explique le juge Brandeis dans un extrait bien connu datant de 1923 :

[TRADUCTION] L’emploi de l’expression « investissement prudent » n’est pas décisif. L’établissement de la base de tarification ne devrait pas exclure les investissements qui, dans des circonstances ordinaires, seraient considérés raisonnables. Cet emploi vise plutôt à exclure les dépenses qui pourraient être jugées malhonnêtes ou manifestement excessives ou imprudentes. On peut supposer que tout investissement considéré a été fait dans l’exercice d’un jugement raisonnable, sauf preuve du contraire. [Je souligne.]

(*State of Missouri ex rel. Southwestern Bell Telephone Co. c. Public Service Commission of Missouri*, 262 U.S. 276 (1923), p. 289, note 1, le juge Brandeis (dissident))

[137] La présomption de prudence constitue le point de départ de l’examen que la Commission appelle [TRADUCTION] « contrôle de la prudence ». Lorsqu’elle entreprend ce contrôle de la prudence, la Commission applique un « ensemble bien établi de principes » :

[TRADUCTION]

- La décision de la direction du service public est généralement présumée prudente, sauf contestation pour motifs valables.
- Pour qu’elle soit prudente, la décision doit être raisonnable eu égard aux circonstances que connaissait ou qu’aurait dû connaître le service public au moment où il l’a prise.
- Le recul est exclu dans l’appréciation de la prudence, même lorsque les conséquences de la décision peuvent légitimement servir à réfuter la présomption de prudence.
- La prudence est appréciée dans le cadre d’une analyse factuelle rétrospective en ce que la preuve doit porter sur le moment où la décision a été prise et

based on facts about the elements that could or did enter into the decision at the time.

(*Enersource Hydro Mississauga Inc. (Re)*, 2012 LNONOEB 373 (QL), at para. 55, citing *Enbridge Gas Distribution Inc. (Re)*, 2002 LNONOEB 4 (QL), at para. 3.12.2.)

[138] This form of prudence review, including a presumption of prudence and a ban on hindsight, was endorsed by the Board and by the Ontario Court of Appeal as an appropriate method to determine “just and reasonable” rates in *Enbridge Gas Distribution Inc. (Re)*, at paras. 3.12.1 to 3.12.5, aff’d *Enbridge Gas Distribution Inc. v. Ontario Energy Board* (2006), 210 O.A.C. 4, at paras. 8 and 10-12.

[139] In the case before us, however, the Board decided not to submit all costs to a prudence review. Instead, it stated that it would use two kinds of review. The first would apply to “forecast costs”, that is, those over which a utility retains discretion and can still be reduced or avoided. It explained in its reasons that it would review such costs using a wide range of evidence, and that the onus was on the utility to demonstrate that its forecast costs were reasonable:

When considering forecast costs, the onus is on the company to make its case and to support its claim that the forecast expenditures are reasonable. The company provides a wide spectrum of such evidence, including business cases, trend analysis, benchmarking data, etc. The test is not dishonesty, negligence, or wasteful loss; the test is reasonableness. And in assessing reasonableness, the Board is not constrained to consider only factors pertaining to [Ontario Power Generation]. The Board has the discretion to find forecast costs unreasonable based on the evidence — and that evidence may be related to the cost/benefit analysis, the impact on ratepayers, comparisons with other entities, or other considerations.

reposer sur des faits quant aux éléments qui ont pu entrer en ligne de compte ou qui sont effectivement entrés en ligne de compte dans la décision.

(*Enersource Hydro Mississauga Inc. (Re)*, 2012 LNONOEB 373 (QL), par. 55, citant *Enbridge Gas Distribution Inc. (Re)*, 2002 LNONOEB 4 (QL), par. 3.12.2.)

[138] Dans *Enbridge Gas Distribution Inc. (Re)*, par. 3.12.1 à 3.12.5, conf. par *Enbridge Gas Distribution Inc. c. Ontario Energy Board* (2006), 210 O.A.C. 4, par. 8 et 10-12, la Commission et la Cour d’appel de l’Ontario considèrent ce contrôle — qui comporte l’application d’une présomption de prudence et exclut le recul — comme la méthode appropriée pour fixer des tarifs « justes et raisonnables ».

[139] Toutefois, dans la présente affaire, la Commission choisit de ne pas soumettre toutes les dépenses à un contrôle de la prudence. Elle dit plutôt recourir à deux examens. Le premier s’appliquerait aux « dépenses prévues », soit celles à l’égard desquelles le service public conserve un pouvoir discrétionnaire et qu’il peut toujours réduire ou éviter. Dans ses motifs, la Commission explique qu’elle examine ces dépenses au regard d’une vaste gamme d’éléments de preuve et qu’il incombe au service public de démontrer le caractère raisonnable de ses dépenses :

[TRADUCTION] Lors de l’examen des dépenses prévues, il incombe à la société d’établir le bien-fondé de sa demande et d’étayer son allégation selon laquelle ces dépenses sont raisonnables. Elle doit fournir un large éventail d’éléments de preuve en ce sens, notamment des analyses de rentabilité et de tendances, des données de référence, etc. Le critère applicable n’est pas celui de la malhonnêteté, de la négligence ou de la perte menant au gaspillage, mais bien celui du caractère raisonnable. Et dans l’appréciation du caractère raisonnable, la Commission n’est pas tenue d’examiner uniquement les données qui intéressent [Ontario Power Generation]. Elle a le pouvoir discrétionnaire de conclure que les dépenses prévues sont déraisonnables au vu de la preuve, laquelle peut se rapporter à l’analyse coût/bénéfice, à l’incidence sur les consommateurs, aux comparaisons avec d’autres entités ou à autre chose.

The benefit of a forward test period is that the company has the benefit of the Board's decision in advance regarding the recovery of forecast costs. To the extent costs are disallowed, for example, a forward test period provides the company with the opportunity to adjust its plans accordingly. In other words, there is not necessarily any cost borne by shareholders (unless the company decides to continue to spend at the higher level in any event). [paras. 74-75]

[140] A different approach, the Board said, would be applied to those costs the company could not "take action to reduce". These costs, sometimes called "committed costs", represent binding commitments that leave a utility with no discretion about whether to make the payment. The Board explained that it evaluates these costs using a "prudence review", which includes a presumption that the costs were prudently incurred:

Somewhat different considerations will come into play when undertaking an after-the-fact prudence review. In the case of an after-the-fact prudence review, if the Board disallows a cost, it is necessarily borne by the shareholder. There is no opportunity for the company to take action to reduce the cost at that point. For this reason, the Board concludes there is a difference between the two types of examination, with the after-the-fact review being a prudence review conducted in the manner which includes a presumption of prudence. [para. 75]

[141] In *Enersource Hydro Mississauga Inc. (Re)*, for example, the Board concluded that it had to conduct a prudence review when evaluating the costs that Enersource had already incurred:

This issue concerns expenditures which have largely already been incurred by the company. . . . Given that the issue concerns past expenditures which are now in dispute, the Board must conduct a prudence review. [para. 55]

[142] As the Board said in its reasons, the prudence review makes sense for committed costs because disallowing costs Ontario Power Generation cannot avoid, forces the utility to pay out of pocket

L'avantage d'une période de référence ultérieure est qu'elle permet à la société de connaître à l'avance la décision de la Commission concernant le recouvrement de dépenses prévues. Par exemple, lorsque des dépenses sont refusées, la société peut modifier ses plans en conséquence. Autrement dit, l'actionnaire n'a pas nécessairement à assumer un coût (à moins que la société ne décide, en tout état de cause, de maintenir les dépenses jugées excessives). [par. 74-75]

[140] Selon la Commission, une démarche différente serait suivie pour les dépenses à l'égard desquelles la société ne pouvait [TRADUCTION] « prendre de mesures de réduction ». Ces dépenses, parfois appelées « dépenses convenues », résultent d'obligations contractuelles qui excluent tout pouvoir discrétionnaire permettant au service public de ne pas les acquitter. La Commission explique qu'elle jauge ces dépenses en se livrant à un « contrôle de la prudence » qui comporte l'application d'une présomption selon laquelle les dépenses ont été faites de manière prudente :

[TRADUCTION] Des considérations quelque peu différentes entreront en jeu lors d'un contrôle de la prudence après coup. La dépense que la Commission refusera alors d'approuver sera nécessairement assumée par l'actionnaire. La société ne pourra plus prendre de mesures de réduction à son égard. C'est pourquoi la Commission estime qu'il existe une différence entre les deux types d'examen, le contrôle après coup constituant un contrôle de la prudence assorti d'une présomption de prudence. [par. 75]

[141] À titre d'exemple, dans *Enersource Hydro Mississauga Inc. (Re)*, la Commission conclut qu'elle doit effectuer un contrôle de la prudence pour apprécier les dépenses qu'Enersource a déjà faites :

[TRADUCTION] Le présent dossier porte sur des dépenses que la société a déjà faites en grande partie. [. . .] Comme il est question de dépenses antérieures qui sont aujourd'hui contestées, la Commission doit effectuer un contrôle de la prudence. [par. 55]

[142] Comme le dit la Commission dans ses motifs, il est logique de soumettre à un contrôle de la prudence des dépenses convenues, car refuser d'approuver des dépenses auxquelles Ontario

for expenses it has already incurred. This could negatively affect Ontario Power Generation's ability to operate, leading the utility to restructure its relationships with the financial community and its service providers, or even lead to bankruptcy: see Burns et al., at pp. 129-65. These outcomes would "increase capital costs and utility rates above the levels that would exist with a limited prudence penalty", forcing Ontario consumers to pay higher electricity bills: Burns et al., at p. vi.

[143] The issue in this appeal therefore centres on the Board assessing *all* compensation costs in Ontario Power Generation's collective agreements as adjustable forecast costs, without determining whether any of them were costs for which "[t]here is no opportunity for the company to take action to reduce" (para. 75). The Board did not actually call them forecast costs, but by saying that "collective agreements may make it difficult to eliminate positions quickly" and that "changes to union contracts . . . will take time" (paras. 346 and 352), the Board was clearly treating them as reducible in theory. Moreover, the fact that it failed to apply the prudence review it said it would apply to non-reducible costs confirms that it saw the collectively bargained commitments as adjustable.

[144] The Board did not explain why it considered compensation costs in collective agreements to be adjustable forecast costs, but the effect of its approach was to deprive Ontario Power Generation of the benefit of the Board's assessment methodology that treats committed costs differently. In my respectful view, the Board's failure to separately assess the compensation costs committed as a result of the collective agreements from other compensation costs, ignored not only its own methodological template, but labour law as well.

Power Generation ne peut se soustraire oblige le service public à acquitter sur ses propres deniers des dépenses déjà faites. Cela pourrait nuire au bon fonctionnement d'Ontario Power Generation et l'amener à restructurer ses liens avec les milieux financiers et ses fournisseurs de services, voire à faire faillite (voir Burns et autres, p. 129-165). Dès lors, [TRADUCTION] « les coûts en capital et les tarifs seraient supérieurs à ce qu'ils auraient été si une sanction modérée avait résulté de l'application du principe de prudence », de sorte que le consommateur ontarien serait contraint de payer des tarifs d'électricité plus élevés (Burns et autres, p. vi).

[143] Le présent pourvoi a donc pour objet la décision de la Commission de considérer *toutes* les dépenses de rémunération issues des conventions collectives d'Ontario Power Generation comme des dépenses prévues ajustables, sans se demander s'il s'agit en partie de dépenses pour lesquelles [TRADUCTION] « [l]a société ne pouvait prendre de mesures de réduction » (par. 75). La Commission ne les qualifie pas à proprement parler de dépenses prévues, mais lorsqu'elle affirme que « les conventions collectives peuvent rendre ardue l'élimination rapide de certains postes » et que « modifier des conventions collectives [. . .] prend du temps » (par. 346 et 352), elle considère clairement qu'il s'agit de dépenses théoriquement compressibles. De plus, l'omission de soumettre celles-ci au contrôle de la prudence qu'elle dit pourtant s'appliquer aux dépenses non réductibles confirme l'assimilation des obligations issues de négociations collectives à des obligations ajustables.

[144] La Commission ne dit pas pourquoi elle estime que les dépenses de rémunération issues des conventions collectives constituent des dépenses prévues ajustables, mais par l'adoption de son approche, elle empêche Ontario Power Generation de bénéficier de l'application de sa méthode d'appréciation qui considère différemment les dépenses convenues. À mon humble avis, en omettant d'apprécier les dépenses de rémunération issues des conventions collectives séparément des autres dépenses de rémunération, la Commission méconnaît à la fois son propre cadre méthodologique et le droit du travail.

[145] Ontario Power Generation was a party to binding collective agreements with the Power Workers' Union and the Society of Energy Professionals covering most of the relevant period. At the time of the application, it had already entered into a collective agreement with the Power Workers' Union for the period of April 1, 2009 to March 31, 2012.

[146] Its collective agreement with the Society of Energy Professionals, which required resolution by binding mediation-arbitration in the event of contract negotiations disputes, expired on December 31, 2010. As a result of a bargaining impasse, the terms of a new collective agreement for January 1, 2011 to December 31, 2012 were imposed by legally binding arbitration: *Ontario Power Generation v. Society of Energy Professionals*, [2011] O.L.A.A. No. 117 (QL).

[147] The collective agreements with the Power Workers' Union and the Society of Energy Professionals prescribed the compensation rates for staff positions held by represented employees, strictly regulated staff levels at Ontario Power Generation's facilities, and limited the utility's ability to unilaterally reduce its compensation rates and staffing levels. The collective agreement with the Power Workers' Union, for example, stipulated that there would be no involuntary layoffs during the term of the agreement. Instead, Ontario Power Generation would be required either to relocate surplus staff or offer severance in accordance with rates set out in predetermined agreements between the utility and the union: "Collective Agreement between Ontario Power Generation Inc. and Power Workers' Union", April 1, 2009 to March 31, 2012, at art. 11.

[148] Similarly, Ontario Power Generation's collective agreement with the Society of Energy Professionals severely limited the utility's bargaining power and control over compensation levels. When the contract between Ontario Power Generation and

[145] Ontario Power Generation était partie à des conventions collectives obligatoires qui étaient intervenues avec le Syndicat des travailleurs et travailleuses du secteur énergétique et Society of Energy Professionals et qui s'appliquaient pendant la plus grande partie de la période considérée. À l'époque de la demande, elle avait déjà conclu une convention collective avec le Syndicat des travailleurs et travailleuses du secteur énergétique pour la période comprise entre le 1^{er} avril 2009 et le 31 mars 2012.

[146] La convention collective intervenue avec Society of Energy Professionals et imposant la médiation-arbitrage pour le règlement des différends pendant des négociations collectives a expiré le 31 décembre 2010. Par suite d'une impasse dans les négociations, les conditions d'une nouvelle convention collective pour la période du 1^{er} janvier 2011 au 31 décembre 2012 ont été imposées par voie d'arbitrage obligatoire (*Ontario Power Generation c. Society of Energy Professionals*, [2011] O.L.A.A. No. 117 (QL)).

[147] Les conventions collectives conclues avec les deux syndicats prescrivaient les barèmes de rémunération des employés syndiqués, réglementaient rigoureusement les niveaux de dotation aux installations d'Ontario Power Generation et limitaient le pouvoir du service public de réduire unilatéralement ses barèmes de rémunération et ses niveaux de dotation. Par exemple, la convention collective conclue avec le Syndicat des travailleurs et travailleuses du secteur énergétique prévoyait qu'il n'y aurait aucun licenciement pendant la durée de son application. Bien au contraire, Ontario Power Generation serait contrainte soit de réaffecter tout employé excédentaire, soit de lui offrir une indemnité de départ selon les barèmes établis au préalable par le service public et le syndicat (« Collective Agreement between Ontario Power Generation Inc. and Power Workers' Union », 1^{er} avril 2009 au 31 mars 2012, art. 11).

[148] De même, la convention collective conclue avec Society of Energy Professionals limitait grandement le pouvoir du service public de négocier et de déterminer les barèmes de rémunération. À l'expiration de cette convention le 31 décembre 2010,

the Society of Energy Professionals expired on December 31, 2010, the utility’s bargaining position had been that its sole shareholder, the Province of Ontario, had directed that there be a zero net compensation increase over the next two-year term. The parties could not reach an agreement and the dispute was therefore referred to binding arbitration as required by previous negotiations. The resulting award by Kevin M. Burkett provided mandatory across-the-board wage increases of 3 per cent on January 1, 2011, 2 per cent on January 1, 2012, and a further 1 per cent on April 1, 2012: *Ontario Power Generation v. Society of Energy Professionals*, at paras. 1, 9, and 28.

[149] The obligations contained in these collective agreements were immutable and legally binding commitments: *Labour Relations Act, 1995*, s. 56. As a result, Ontario Power Generation was prohibited from unilaterally reducing the staffing levels, wages, or benefits of its unionized workforce. These agreements therefore did not just leave the utility “with limited flexibility regarding overall compensation rates or staffing levels”, as the majority notes (at para. 84), they made it *illegal* for the utility to alter the compensation and staffing levels of 90 per cent of its regulated workforce in a manner that was inconsistent with its commitments under the agreements.

[150] Instead, the Board, applying the methodology it said it would use for the utility’s forecast costs, put the onus on Ontario Power Generation to prove the reasonableness of its costs and concluded that it had failed to provide “compelling evidence” or “documentation or analysis” to justify compensation levels: para. 347. Had the Board used the approach it said it would use for costs the company had “no opportunity . . . to reduce”, it would have used an after-the-fact prudence review, with a rebuttable presumption that the utility’s expenditures were reasonable.

le service public défendait la position de son unique actionnaire, la province d’Ontario, à savoir l’exclusion de toute augmentation nette des salaires pendant les deux années suivantes. Les parties n’ont pu parvenir à un accord, de sorte que le dossier a été renvoyé à l’arbitrage obligatoire comme convenu lors de négociations précédentes. Dans sa décision, l’arbitre Kevin M. Burkett a ordonné une augmentation générale des salaires de 3 p. 100 le 1^{er} janvier 2011, de 2 p. 100 le 1^{er} janvier 2012 et, en sus, de 1 p. 100 le 1^{er} avril 2012 (*Ontario Power Generation c. Society of Energy Professionals*, par. 1, 9 et 28).

[149] Les obligations contractées dans ces conventions collectives constituaient des engagements immuables ayant force obligatoire (*Loi de 1995 sur les relations de travail*, art. 56). Il était donc interdit à Ontario Power Generation de réduire unilatéralement les niveaux de dotation, les salaires ou les avantages sociaux de ses employés syndiqués. Contrairement à ce qu’affirment les juges majoritaires (par. 84), ces conventions ne laissaient pas seulement « peu de marge de manœuvre quant aux barèmes de rémunération et aux niveaux de dotation dans leur ensemble », elles rendaient *illégal* la modification par le service public — d’une manière incompatible avec les engagements qu’il y prenait — des barèmes de rémunération et des niveaux de dotation quant à 90 p. 100 de son effectif obligatoire.

[150] En appliquant la méthode qu’elle a dit qu’elle utiliserait à l’égard des dépenses prévues du service public, la Commission oblige en fait Ontario Power Generation à prouver le caractère raisonnable de ses dépenses et conclut que l’entreprise n’a présenté ni [TRADUCTION] « preuve convaincante », ni « documents ou analyses » qui justifient les barèmes de rémunération (par. 347). Si elle avait eu recours à l’approche qu’elle a dit qu’elle utiliserait pour les dépenses à l’égard desquelles la société ne pouvait « prendre de mesures de réduction », la Commission aurait contrôlé la prudence des dépenses après coup et appliqué la présomption réfutable selon laquelle elles étaient raisonnables.

[151] Applying a prudence review to these compensation costs would hardly, as the majority suggests, “have conflicted with the burden of proof in the *Ontario Energy Board Act, 1998*”. To interpret the burden of proof in s. 78.1(6) of the *Ontario Energy Board Act, 1998* so strictly would essentially prevent the Board from ever conducting a prudence review, notwithstanding that it has comfortably done so in the past and stated, even in its reasons in this case, that it would review committed costs using an “after-the-fact prudence review” which “includes a presumption of prudence”. Under the majority’s logic, however, since a prudence review always involves a presumption of prudence, the Board would not only be limiting its methodological flexibility, it would be in breach of the Act.

[152] The application of a prudence review does not shield the utility’s compensation costs from scrutiny. As the Court of Appeal observed, a prudence review

does not mean that the [Board] is powerless to review the compensation rates for [Ontario Power Generation’s] unionized staff positions or the number of those positions. In a prudence review, the evidence may show that the presumption of prudently incurred costs should be set aside, and that the committed compensation rates and staffing levels were not reasonable; however, the [Board] cannot resort to hindsight, and must consider what was known or ought to have been known at the time. A prudence review allows for such an outcome, and permits the [Board] both to fulfill its statutory mandate and to serve as a market proxy, while maintaining a fair balance between [Ontario Power Generation] and its customers. [para. 38]

[153] The majority’s suggestion (at para. 114) that “if the legislature had intended for costs under collective agreements to also be inevitably imposed on consumers, it would not have seen fit to grant

[151] Contrairement à ce que soutiennent les juges majoritaires, appliquer le contrôle de la prudence à ces dépenses de rémunération serait difficilement « incompatible avec le fardeau de preuve que prévoit la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* ». Considérer que le par. 78.1(6) de la *Loi de 1998 sur la Commission de l’énergie de l’Ontario* prévoit un fardeau de preuve aussi strict a essentiellement pour effet d’empêcher totalement la Commission d’effectuer des contrôles de la prudence, alors qu’elle en a effectués sans difficulté dans le passé et qu’elle a affirmé — comme dans ses motifs en l’espèce — qu’il y a lieu de soumettre les dépenses convenues à « un contrôle de la prudence après coup, [...] comportant l’application d’une présomption de prudence ». Or, suivant le raisonnement des juges majoritaires, comme le contrôle de la prudence présume toujours la prudence, la Commission ne verrait pas seulement sa marge de manœuvre réduite sur le plan méthodologique, mais elle contreviendrait aussi à la Loi.

[152] L’application du principe de la prudence ne soustrait pas les dépenses de rémunération du service public à tout examen. Comme le fait remarquer la Cour d’appel, le contrôle de la prudence

[TRANSLATION] n’écarte pas la possibilité que la [Commission] puisse contrôler les barèmes de rémunération applicables aux employés syndiqués d’[Ontario Power Generation] ou le nombre de leurs postes. Lors d’un tel contrôle, il peut ressortir de la preuve, d’une part, que la présomption selon laquelle les dépenses ont été faites de manière prudente doit être écartée et, d’autre part, que les barèmes de rémunération et les niveaux de dotation convenus ne sont pas raisonnables; cependant, la [Commission] ne peut se prononcer avec le recul, mais doit tenir compte de ce qui était connu ou qui aurait dû l’être à l’époque. Le contrôle de la prudence admet un tel résultat et permet à la [Commission] de s’acquitter de son mandat légal et de jouer son rôle de substitut du marché tout en assurant un juste équilibre entre les intérêts d’[Ontario Power Generation] et ceux de ses clients. [par. 38]

[153] L’affirmation des juges majoritaires selon laquelle, « si le législateur avait voulu que les dépenses [...] issues [de conventions collectives] se répercutent inévitablement sur les consommateurs, il

the Board oversight of utility compensation costs”, is puzzling. The legislature did not intend for *any* costs to be “inevitably” imposed on consumers. What it intended was to give the Board authority to determine just and reasonable payment amounts based on Ontario Power Generation’s existing and proposed commitments. Neither collective agreements nor any other contractual obligations were intended to be “inevitably” imposed. They were intended to be inevitably considered in the balance. But it is precisely because of the unique nature of binding commitments that the Board said it would impose a different kind of review on these costs.

[154] It may well be that Ontario Power Generation has the ability to manage some staffing levels through attrition or other mechanisms that did not breach the utility’s commitments under its collective agreements, and that these costs may therefore properly be characterized as forecast costs. But no factual findings were made by the Board about the extent of any such flexibility. There is in fact no evidence in the record, nor any evidence cited in the Board’s decision, setting out what proportion of Ontario Power Generation’s compensation costs were fixed and what proportion remained subject to the utility’s discretion. The Board made virtually no findings of fact regarding the extent to which the utility could reduce its collectively bargained compensation costs. On the contrary, the Board, as Aitken J. noted, “lumped” all compensation costs together, acknowledged that reducing those in the collective agreements would “take time” and “be difficult”, and dealt with them as globally adjustable.

[155] Given that collective agreements are legally binding, it was unreasonable for the Board to assume that Ontario Power Generation could reduce the costs fixed by these contracts in the absence of any evidence to that effect. To use the majority’s

n’aurait pas jugé opportun d’investir la Commission du pouvoir de surveiller les dépenses de rémunération d’un service public » (par. 114), laisse perplexe. Le législateur ne voulait pas que *toute* dépense se répercute « inévitablement » sur les consommateurs. Son intention était de donner à la Commission le pouvoir d’arrêter des paiements justes et raisonnables en fonction des engagements actuels et projetés d’Ontario Power Generation. Ni les conventions collectives ni aucune autre obligation contractuelle ne devaient « inévitablement » se répercuter sur qui que ce soit. Cependant, elles devaient inévitablement peser dans la balance. Or, c’est précisément la nature unique des engagements contraignants qu’a invoquée la Commission lorsqu’elle a affirmé qu’elle soumettrait ces dépenses à un contrôle différent.

[154] Il se peut fort bien qu’Ontario Power Generation puisse modifier certains niveaux de dotation par voie d’attrition ou grâce à d’autres mécanismes qui ne vont pas à l’encontre de ses obligations suivant les conventions collectives. Il se peut fort bien aussi que les dépenses puissent donc être assimilées à juste titre à des dépenses prévues. La Commission ne tire toutefois aucune conclusion de fait sur l’étendue d’une telle marge de manœuvre. En fait, aucun élément du dossier ou de la preuve invoquée par la Commission n’indique dans quelle proportion les dépenses de rémunération d’Ontario Power Generation sont fixes et dans quelle proportion elles demeurent assujetties au pouvoir discrétionnaire du service public. La Commission ne tire pour ainsi dire aucune conclusion de fait quant à savoir dans quelle mesure l’entreprise pouvait réduire ses dépenses de rémunération issues des conventions collectives. Au contraire, comme le souligne la juge Aitken, la Commission [TRADUCTION] « regroupe » sans distinctions toutes les dépenses liées à la rémunération, reconnaît que la réduction de celles issues des conventions collectives « prend[rait] du temps » et « [serait] ardue », et considère qu’elles sont globalement ajustables.

[155] Comme les conventions collectives sont contraignantes en droit, il était déraisonnable que la Commission présume qu’Ontario Power Generation pouvait réduire les dépenses déterminées par ces contrats en l’absence de toute preuve en ce

words, these costs are “legal obligations that leave [the utility] with no discretion as to whether to make the payment in the future” (para. 82). According to the Board’s own methodology, costs for which “[t]here is no opportunity for the company to take action to reduce” are entitled to “a presumption of prudence”: para. 75.

[156] Disallowing costs that Ontario Power Generation is legally required to pay as a result of its collective agreements, would force the utility and the Province of Ontario, the sole shareholder, to make up the difference elsewhere. This includes the possibility that Ontario Power Generation would be forced to reduce investment in the development of capacity and facilities. And because Ontario Power Generation is Ontario’s largest electricity generator, it may not only threaten the “financial viability” of the province’s electricity industry, it could also imperil the assurance of reliable electricity service.

[157] The majority nonetheless assumes that the ongoing relationship between Ontario Power Generation and the unions should give the Board greater latitude in disallowing the collectively bargained compensation costs than it would have had if it applied a no-hindsight, presumption-of-prudence analysis. It also accepts the Board’s conclusion that Ontario Power Generation’s collectively bargained compensation costs may be “excessive”, and therefore concludes that the Board was reasonable in choosing to avoid the “prudence” test in order to so find. This approach finds no support even in the methodology the Board set out for itself for evaluating just and reasonable payment amounts.

[158] In my respectful view, selecting a test which is more likely to confirm an assumption that collectively bargained costs are excessive, misconceives the point of the exercise, namely, to determine

sens. Pour reprendre les propos des juges majoritaires, ces dépenses correspondent à des « obligations qui écartent tout pouvoir discrétionnaire [. . .] permettant [au service public] de ne pas acquitter la somme ultérieurement » (par. 82). Selon la propre méthode de la Commission, les dépenses à l’égard desquelles [TRADUCTION] « [l]a société ne pouvait prendre de mesures de réduction » bénéficient d’une « présomption de prudence » (par. 75).

[156] Refuser d’approuver des dépenses qu’Ontario Power Generation est juridiquement tenue d’acquitter en raison de ses conventions collectives obligerait le service public et son seul actionnaire, la province d’Ontario, à combler la différence en puisant ailleurs. Ontario Power Generation pourrait notamment être forcée de réduire ses investissements dans l’accroissement de sa capacité et dans l’amélioration de ses installations. Et, comme il s’agit du plus grand producteur d’électricité de l’Ontario, un tel refus pourrait non seulement nuire à la « viabilité financière » du secteur de l’électricité de la province, mais également mettre en péril la garantie d’un service d’électricité fiable.

[157] Les juges majoritaires tiennent cependant pour acquis que la relation continue entre Ontario Power Generation et les syndicats devrait conférer à la Commission, relativement aux dépenses de rémunération issues de négociations collectives, un pouvoir de refus plus grand que celui dont elle bénéficie dans le cadre d’une analyse qui exclut le recul et présume la prudence. Ils font droit également à la conclusion de la Commission selon laquelle les dépenses de rémunération issues de négociations collectives auxquelles Ontario Power Generation a participé pourraient être [TRADUCTION] « excessives » et concluent donc que la Commission a agi raisonnablement en écartant le principe de la « prudence » pour arriver à sa conclusion. Leur approche ne trouve aucun appui, pas même dans la méthode que la Commission établit elle-même pour déterminer le montant de paiements justes et raisonnables.

[158] En tout respect pour l’opinion contraire, en choisissant un critère éminemment susceptible de confirmer l’hypothèse que les dépenses issues de négociations collectives sont excessives, on se

whether those costs were in fact excessive. Blaming collective bargaining for what are *assumed* to be excessive costs, imposes, with respect, the appearance of an ideologically driven conclusion on what is intended to be a principled methodology based on a distinction between committed and forecast costs, not between costs which are collectively bargained and those which are not.

[159] I recognize that the Board has wide discretion to fix payment amounts that are “just and reasonable” and, subject to certain limitations, to “establish the . . . methodology” used to determine such amounts: O. Reg. 53/05, s. 6, *Ontario Energy Board Act*, 1998, s. 78.1. That said, once the Board establishes a methodology to determine what is just and reasonable, it is, at the very least, required to faithfully apply that approach: see *TransCanada Pipelines Ltd. v. National Energy Board* (2004), 319 N.R. 171 (F.C.A.), at paras. 30-32, per Rothstein J.A. This does not mean that collective agreements “supersede” or “trump” the Board’s authority to fix payment amounts; it means that once the Board selects a methodology for itself for the exercise of its discretion, it is required to follow it. Absent methodological clarity and predictability, Ontario Power Generation would be left in the dark about how to determine what expenditures and investments to make and how to present them to the Board for review. Wandering sporadically from approach to approach, or failing to apply the methodology it declares itself to be following, creates uncertainty and leads, inevitably, to needlessly wasting public time and resources in constantly having to anticipate and respond to moving regulatory targets.

[160] In disallowing \$145 million of the compensation costs sought by Ontario Power Generation on the grounds that the utility could reduce salary and staffing levels, the Board ignored the legally binding nature of the collective agreements

méprend sur l’objectif de la démarche, qui est de déterminer si ces dépenses étaient bel et bien excessives. Imputer à la négociation collective ce que l’on *suppose* constituer des dépenses excessives revient, soit dit tout en respect, à substituer ce qui a l’apparence d’une conclusion idéologique à ce qui est censé résulter d’une méthode d’analyse raisonnée qui distingue entre les dépenses convenues et les dépenses prévues, non entre les dépenses issues de négociations collectives et celles qui ne le sont pas.

[159] Je reconnais que la Commission jouit d’un vaste pouvoir discrétionnaire lui permettant de déterminer les paiements qui sont « justes et raisonnables » et, à l’intérieur de certaines limites, de [TRADUCTION] « définir la [. . .] méthode » utilisée pour établir le montant de ces paiements (règlement 53/05, art. 6; *Loi de 1998 sur la Commission de l’Énergie de l’Ontario*, art. 78.1). Cela dit, dès lors qu’elle a établi une méthode pour déterminer ce qui est juste et raisonnable, la Commission doit à tout le moins l’appliquer avec constance (*TransCanada Pipelines Ltd. c. Office national de l’Énergie*, 2004 CAF 149 (CanLII), par. 30-32, le juge Rothstein). Pour autant, les conventions collectives ne « pri-ment » pas le pouvoir de la Commission de fixer les paiements, mais une fois que la Commission a choisi une méthode pour exercer son pouvoir discrétionnaire, elle doit s’y tenir. En l’absence de clarté et de prévisibilité quant à la méthode à appliquer, Ontario Power Generation serait vouée à l’incertitude quant à la démarche à suivre pour déterminer les dépenses et les investissements à faire et quant à la manière de les soumettre à l’examen de la Commission. Passer sporadiquement d’une approche à une autre ou ne pas appliquer la méthode que l’on prétend appliquer crée de l’incertitude et mène inévitablement au gaspillage inutile du temps et des ressources publics en ce qu’il faut constamment anticiper un objectif réglementaire fluctuant et s’y ajuster.

[160] En refusant d’approuver des dépenses de 145 millions de dollars au motif qu’Ontario Power Generation pouvait réduire ses barèmes de rémunération et ses niveaux de dotation, la Commission a méconnu le caractère contraignant en droit des

and failed to distinguish between committed compensation costs and those that were reducible. Whether or not one can fault the Board for failing to use a particular methodology, what the Board can unquestionably be analytically faulted for, is evaluating all compensation costs fixed by collective agreements as being amenable to adjustment. Treating these compensation costs as reducible was, in my respectful view, unreasonable.

[161] I would accordingly dismiss the appeal, set aside the Board's decision, and, like the Court of Appeal, remit the matter to the Board for reconsideration in accordance with these reasons.

Appeal allowed, ABELLA J. dissenting.

Solicitors for the appellant: Stikeman Elliott, Toronto.

Solicitors for the respondent Ontario Power Generation Inc.: Torys, Toronto; Ontario Power Generation Inc., Toronto.

Solicitors for the respondent the Power Workers' Union, Canadian Union of Public Employees, Local 1000: Paliare Roland Rosenberg Rothstein, Toronto.

Solicitors for the respondent the Society of Energy Professionals: Cavalluzzo Shilton McIntyre Cornish, Toronto.

Solicitors for the intervener: Jay Shepherd Professional Corporation, Toronto.

conventions collectives et a omis de distinguer les dépenses de rémunération convenues de celles qui étaient réductibles. On peut reprocher ou non à la Commission de ne pas avoir appliqué une certaine méthode, mais on peut assurément lui reprocher, sur le plan analytique, d'avoir considéré toutes les dépenses de rémunération déterminées par des conventions collectives comme des dépenses ajustables. Voir dans ces dépenses des dépenses réductibles est à mon sens déraisonnable.

[161] Je suis donc d'avis de rejeter le pourvoi, d'annuler la décision de la Commission et, à l'instar de la Cour d'appel, de renvoyer l'affaire à la Commission pour qu'elle la réexamine à la lumière des présents motifs.

Pourvoi accueilli, la juge ABELLA est dissidente.

Procureurs de l'appelante : Stikeman Elliott, Toronto.

Procureurs de l'intimée Ontario Power Generation Inc. : Torys, Toronto; Ontario Power Generation Inc., Toronto.

Procureurs de l'intimé le Syndicat des travailleurs et travailleuses du secteur énergétique, Syndicat canadien de la fonction publique, section locale 1000 : Paliare Roland Rosenberg Rothstein, Toronto.

Procureurs de l'intimée Society of Energy Professionals : Cavalluzzo Shilton McIntyre Cornish, Toronto.

Procureurs de l'intervenante : Jay Shepherd Professional Corporation, Toronto.