

“Top Down” Estimation of DSM Program Impacts on Natural Gas Usage

February 2010

Larry Kaufmann, Ph.D.
Senior Advisor

Lullit Getachew, Ph.D.
Senior Economist

Matt Makos
Economist

PACIFIC ECONOMICS GROUP RESEARCH

22 East Mifflin, Suite 302
Madison, Wisconsin USA 53703
608.257.1522 608.257.1540 Fax

The views expressed in this report are those of Pacific Economics Group Research and do not necessarily represent the views of, and should not be attributed to, the Ontario Energy Board, any individual Board member, or OEB staff.

Executive Summary

Enbridge Gas Distribution (EGD) and Union Gas are subject to a demand-side management (DSM) framework that was implemented in 2006 and reviewed in 2008-2009. Calculating the reductions in gas usage due to these DSM programs is a complex and cumbersome process. This computation depends on a number of assumptions about specific DSM measures, and it involves judgments on free riders, spillovers, and the attribution of benefits from a given program. Stakeholders will have differing opinions (and material interests) regarding each of the many elements that enter into these calculations, which naturally makes these calculations more contentious.

Some customer groups have suggested that the DSM framework can be improved by developing a “top down” estimate of gas usage reductions resulting from DSM efforts. Rather than starting with individual measures and programs, this approach would use econometric methods to estimate gas usage per customer given a variety of factors that influence gas consumption. One of these factors could be gas distributors’ DSM expenditures.

Pacific Economics Group Research (PEG) was asked to advise Ontario Energy Board (OEB) Staff on whether a top-down, econometric approach to estimating gas savings is feasible for EGD and Union Gas. We were asked to evaluate the current data and gas demand models used in Ontario and see whether they could be adapted to “top down” measurement of changes in gas consumption resulting from utility DSM programs. This work would include empirical investigation of “top down” econometric models. Based on these results, we would evaluate the merits of a potential “top down” approach compared with the bottom-up methods that are currently in use.

PEG is not aware of any “top-down” econometric approaches to measuring energy savings that are derived using data for all customers on a tariff. California has used a variant of econometric, “ex post” measurement of DSM savings, particularly in the 1990s. However, the econometric techniques in California use customer-specific data and distinguish between the energy consumption of customers who are participating in company DSM programs from those that are not. This is a much more data-intensive econometric approach than the “top down” methods PEG was asked to investigate.

Both EGD and Union Gas currently use gas demand models for regulatory purposes. However, these models are used for forecasting gas usage and not estimating DSM-related energy savings. PEG has examined both companies' econometric gas demand models, and we believe they have some appealing features. The extensive demand modeling in the Province also makes a wealth of information available that can provide a foundation for the current, "top down" research. At the same time, there are statistical issues with the EGD and Union Gas models that can reduce the efficiency of estimates and bias inference. These issues will be more problematic in a DSM-measurement than forecasting application.

PEG investigated several different approaches for developing "top down" estimates. The first builds on recent work in the economic literature and examines the link between DSM spending by Ontario gas distributors and subsequent changes in gas consumption. We use a two-stage econometric technique, where the first stage regress monthly volume data on monthly values of heating degree days (HDD) and prices by revenue class. We then insert monthly values for HDD and price into the fitted regressions to obtain normalized, monthly consumption volumes. The second stage uses the percentage change between actual and normalized consumption as the dependent variable. Changes in this dependent variable are regressed on DSM spending and other variables. The coefficient on DSM spending would measure the direct and spillover effects on consumption from customers participating in utility DSM programs, net of free riders, which would be an appropriate "top down" measure of gas savings to use in TRC calculations.

We also estimated updated, but somewhat modified, versions of the EGD and Union gas demand models which included estimates of monthly DSM spending as an explanatory variable. Both companies cautioned about the quality of the monthly gas demand spending data since, among other reasons, DSM costs are not necessarily booked in the same month in which actual program costs are incurred. While it is important to keep these limitations on data quality in mind, this approach is nevertheless a straightforward extension of the gas demand work that is already presented in OEB proceedings, and it may provide some indicative evidence on the relationship between DSM spending and gas consumption for different revenue classes.

Finally, PEG investigated whether there are statistically significant differences between actual and predicted changes in gas consumption, where predictions are based on econometric gas demand models that do not include DSM spending as an explanatory variable. Any statistically significant differences between actual and predicted gas usage using these models could be interpreted as an indicator, at least, of the impact of DSM programs on gas consumption.

For the first approach, PEG's first-stage regression results were generally sensible. The coefficients on HDD and prices had the expected signs and were highly significant for all revenue classes. The second-stage results were also generally sensible for the residential revenue classes, but less so for commercial customers. However, in the dozens of models we estimated, PEG was never able to identify a statistically significant relationship between changes in gas consumption (for residential or commercial customers) and DSM spending in the previous year.

The results using monthly DSM spending as an explanatory variable in updated Company demand models were more promising. We estimated that there was a statistically significant and negative relationship between DSM spending and gas consumption for all residential revenue classes and for two of the five commercial revenue classes for EGD and Union. Our models show that a 10% increase in DSM spending for residential customers will lead to a 0.6% to 1.0% decline in gas consumption. For commercial customers, our models show that a 10% increase in gas DSM spending will lead to a 0.3% to 0.8% decline in gas consumption. This provides some indicative, but not definitive, evidence of the impact of the Companies' DSM spending on gas consumption.

Our third approach evaluated the relationship between actual and predicted gas consumption by revenue class. We could never identify a year in which actual gas usage was below the predicted value and outside of the confidence intervals. Thus, this approach was not successful in identifying the impact of DSM programs on gas usage.

Overall, PEG's research did not provide any "top down" evidence that can substitute for the bottom-up methods currently used in Ontario. Our strongest results came from integrating DSM spending into variants of the gas demand models the Companies currently use to forecast gas demand. Monthly data on gas DSM spending

are not reliable, however, so these results can at best provide supplementary or supporting evidence on the impact of DSM programs on gas consumption. Our econometric models that used more reliable measures of gas DSM spending were never able to identify a significant relationship between DSM activity and gas consumption.

PEG's analysis could likely be improved if better data were available. One improvement would be more accurate data on DSM spending by revenue class and (for EGD) geographic zone. It could also be helpful to have information on when (in a given year) particular DSM measures were installed, in addition to having more accurate data on DSM spending.

More appropriate estimates of DSM savings could also be developed if demand models are estimated separately for participating and non-participating customers. A relatively small share of customers in a revenue class is likely to be participating in utility DSM programs in any given year. The behavioral characteristics of participating and non-participating customers may be so different that they effectively constitute different populations with, accordingly, different underlying demands for natural gas. However, developing detailed customer-specific data would likely entail significant costs, and it would take years for enough sample data to be available to facilitate statistical analysis. There is also no guarantee that this approach will be successful and yield statistically significant and robust results.

TABLE OF CONTENTS

1. INTRODUCTION AND SUMMARY	1
1.1 INTRODUCTION	1
1.2 SUMMARY OF RESULTS.....	4
2. EXPERIENCE IN INDUSTRY	7
3. EXISTING GAS DEMAND MODELS IN ONTARIO.....	9
3.1 UNION GAS DEMAND MODELS	10
3.2 ENBRIDGE GAS DEMAND MODELS	17
4. DEVELOPING “TOP DOWN” M&V MODELS.....	22
4.1 PREVIOUS WORK	22
4.2 PEG’S APPROACH AND SPECIFICATIONS	25
5. ECONOMETRIC ESTIMATION OF “TOP DOWN” MODELS	30
5.1 DATA	30
5.2 ECONOMETRIC METHODS	33
5.3 ECONOMETRIC RESULTS.....	34
5.3.1 <i>New Gas Demand Models</i>	34
5.3.2 <i>Econometric Results Using Monthly DSM Data</i>	47
5.3.3 <i>Differences Between Actual and Predicted Consumption</i>	57
6. CONCLUSION.....	58
APPENDIX: CALIFORNIA EXPERIENCE WITH ECONOMETRIC M&V	60
REFERENCES.....	72

1. INTRODUCTION AND SUMMARY

1.1 Introduction

Enbridge Gas Distribution (EGD) and Union Gas are subject to a demand-side management (DSM) framework that was implemented in 2006 and reviewed in 2008-2009. Well-designed DSM policies encourage customers to implement energy conservation measures that reduce their energy usage over a multi-year time horizon. Effective DSM programs lead to net total resource cost (TRC) savings which is a (discounted) stream of reductions in energy and other resource costs that more than offset the DSM equipment and program costs.

Reductions in gas usage are the main source of TRC net savings resulting from the DSM programs of EGD and Union. These reductions in gas usage are also used to compute the revenues lost from DSM measures that EGD and Union are allowed to recover through the lost revenue adjustment mechanism (LRAM), as well as the incentive-based earnings they can earn under the shared savings mechanism (SSM). An accurate measure of the reductions in gas consumption is therefore critical for ensuring that gas distribution DSM plans create appropriate incentives to pursue cost-effective energy conservation.

However, calculating changes in gas usage from EGD's and Union's DSM measures has proven to be controversial. The framework uses a "bottom up" approach for calculating these benefits, based on an assumed reduction in annual gas usage for each particular measure. Annual savings associated with a measure are calculated by multiplying savings in gas consumption per unit of the DSM technology by the number of units installed. A discounted value of the flow of these benefits is then obtained by assuming the number of years for which the unit will be in service (*i.e.* an asset life), and discounting these future benefits back to the present time by using a selected discount rate.

The basic gas savings calculation therefore hinges on a number of assumptions, including gas savings for each unit of technology, the years each installed unit will be in

service before it is replaced, and the appropriate discount rate. This calculation is made even more complex by the need to include only those gas reductions that result from the utilities' own behavior. For example, "free riders" are participants in programs who would have installed the measure even in the absence of a utility DSM program. Ontario's TRC calculation excludes both the benefits and costs of measures for all program participants who are deemed to be "free riders." On the other hand, the TRC calculation should include gas savings that result from customers who decide to adopt energy efficiency measures because of the utilities' marketing efforts even if those customers do not participate in the utilities' DSM programs. These are often referred to as "spillover" DSM benefits. In addition, the calculation of TRC net savings in Ontario depends on how the benefits stemming from a measure are attributed to a utility vis-à-vis third parties who are also promoting DSM. Only benefits attributed to utilities are included in the TRC calculation, and 100% of the benefits of a program will be attributed to a utility only if that utility can demonstrate that its role was "central" to the program.

In sum, calculating the reductions in gas usage due to utility DSM programs is a complex and cumbersome process. This computation depends on a number of assumptions about specific DSM measures, and it involves judgments on free riders, spillovers, and the attribution of benefits from a given program. Stakeholders will have differing opinions (and material interests) regarding each of the many elements that enter into the TRC calculation, which naturally makes these calculations more contentious.

Some ratepayer groups have suggested that the DSM framework can be improved by developing a "top down" estimate of gas usage reductions resulting from DSM efforts. Rather than starting with individual measures and programs, this approach would use econometric methods to estimate gas usage per customer given a variety of factors that influence gas consumption. One of those variables could be measures of utilities' DSM efforts.

There are a number of advantages with such a "top down" approach in principle. Once the gas forecasting methodologies are in place, calculating reductions in gas usage would be straightforward. The process for computing gas savings would therefore be greatly streamlined and less costly. A more rule-based and less discretionary framework

could also strengthen utilities' incentives to pursue DSM and ensure that it is undertaken in the most cost-effective manner.

Researchers have long recognized the potential value of using statistical methods in the estimation of DSM savings. For example, in an early paper examining utilities' experience with conservation programs, Joskow and Marron wrote:

“Most serious analysts recognize that it is quite difficult to measure accurately the energy savings resulting from utility conservation efforts. These difficulties arise because of diversity in customer utilization patterns, changes in these patterns over time, the limited information a utility has about both the base level of and changes in the utilization of individual participants, differences in characteristics between participants and the population upon which “typical customer” utilization data are based, changes in behavior induced by conservation etc.

In some cases it is possible to obtain good savings estimates by using statistical methods to compare utilization patterns of participating customers with those of similar non-participating customers. Such an approach requires, however, the careful identification of control groups, collection of data on all relevant customer characteristics, and careful monitoring of consumption and changes in customer characteristics for the treatment and control groups over a sufficient period of time to capture all relevant behavioral changes. In other applications, especially when there are significant idiosyncratic customer specific characteristics, it may be very difficult to make accurate measurements of savings. What is clear is that measurement of savings requires careful thought, extensive data collection, careful analysis, time, and (probably) a lot of money.”¹

As this passage indicates, while top-down approaches to estimation of savings are appealing in principle, implementing such a method involves significant challenges. In addition to the issues highlighted above, another fundamental issue is simply developing an appropriate econometric model for forecasting gas consumption. In Ontario, EGD and Union have developed gas demand models and used them in regulatory applications. However, these models differ in important respects, and have not been used directly for estimation of gas savings nor they have been designed for this purpose.

Pacific Economics Group Research (PEG) was retained to assess whether a top-down, econometric approach to estimating gas savings is feasible for EGD and Union

¹ Joskow, P. and D. Marron (1992), “What Does a Negawatt Really Cost? Evidence From Utility Conservation Programs,” *Energy Journal*, Vol 13: 4, p. 54.

Gas. We were asked to evaluate the current data and gas demand models used by the two gas distributors in Ontario and see whether they could be adapted to “top down” estimation of changes in gas consumption resulting from utility DSM programs. This work would include empirical investigation of “top down” estimation of gas savings using econometric techniques. Based on these results, we would evaluate the merits of a potential “top down” approach compared with the bottom-up methods that are currently in use. This report presents the results of PEG’s work.

1.2 Summary of Results

Our results can be briefly summarized. We are not aware of any “top-down” econometric approaches to estimating savings that are applied for data that are aggregated for all customers in a revenue class. California has used a variant of econometric, “ex post” evaluation, but it focuses specifically on the experience of program participants.

Both EGD and Union Gas currently use gas demand models for regulatory purposes. However, these models are used for forecasting gas usage and not for estimation of energy savings. PEG has examined both companies’ econometric gas demand models, and we believe they have many appealing features. The extensive demand modeling undertaken by the two utilities also makes a wealth of information available that can provide a foundation for the “top down” research. At the same time, there are some statistical issues with the EGD and Union Gas models that can reduce the efficiency of estimates and bias inference. These issues will be more problematic in an estimation of savings than a forecasting application.

PEG investigated three different approaches for developing “top down” estimates. The first builds on recent work in the economic literature and examines the link between DSM spending by Ontario gas distributors and subsequent changes in gas consumption. We used a two-stage econometric technique, where the first stage regresses monthly volume data on monthly values of heating degree days (HDD) and prices by revenue class. We then insert monthly values for HDD and price into the fitted regressions to obtain normalized, monthly consumption volumes. The normalized monthly consumption is then aggregated into annual consumption. The second stage uses the difference between actual and normalized annual consumption as the dependent variable.

Changes in this dependent variable are regressed on DSM spending and other variables. The coefficient on DSM spending would measure the direct and spillover effects on consumption from customers participating in utility DSM programs, net of free riders, which would be an appropriate “top down” measure of gas savings.

The second approach involved estimation of updated versions of the EGD and Union gas demand models which included estimates of monthly DSM spending as an explanatory variable. These models were necessarily modified to some extent; one reason is that EGD did not collect DSM spending on a regional basis, which made it impossible to include DSM as an independent variable in EGD’s gas demand models for different geographic zones. Both companies also cautioned about the quality of the monthly gas demand spending data since, among other reasons, DSM costs are not necessarily booked in the same month in which actual program costs are incurred. While it is important to keep these limitations on data quality in mind, this approach is nevertheless a straightforward extension of the gas demand modeling that is already presented in OEB proceedings, and it may provide some indicative evidence on the relationship between DSM spending and gas consumption for different revenue classes.

The third approach investigated whether there are statistically significant differences between actual and predicted changes in gas consumption, where predictions are based on econometric gas demand models that do not include DSM spending as an explanatory variable. Any statistically significant differences between actual and predicted gas usage using these models could be interpreted as an indicator, at least, of the impact of DSM programs on gas consumption.

For the first approach, PEG’s first-stage regression results were generally sensible. The coefficients on HDD and prices had the expected signs and were highly significant for all revenue classes. The second-stage results were also generally sensible for the residential revenue classes, but less so for commercial customers. However, in the dozens of models we estimated, PEG was never able to identify a statistically significant relationship between changes in gas consumption (for residential or commercial customers) and DSM spending in the previous year.

The results of the second approach that is using monthly DSM spending as an explanatory variable in updated Company demand models were more promising. We

estimated that there was a statistically significant and negative relationship between DSM spending and gas consumption for all residential revenue classes and for two of the five commercial revenue classes for EGD and Union. Specifically the results showed that a 10% increase in DSM spending could lead to a 0.6% to 1% decrease in gas consumption for residential customers and a 0.3% to 0.8% decline in gas consumption for commercial customers. This provides some indicative, but not definitive, evidence of the impact of the Companies' DSM spending on gas consumption.

Our third approach evaluated the relationship between actual and predicted gas consumption by revenue class using the models developed above under approaches one and two. We could never identify a year in which actual gas usage was below the predicted value and outside of the confidence intervals. Thus, this approach was not successful in identifying the impact of DSM programs on gas usage.

Overall, PEG's research did not provide any "top down" evidence that is definitive enough to substitute for the bottom-up methods currently used in Ontario's gas DSM framework. Our strongest results came from integrating DSM spending into variants of the gas demand models the Companies currently use to forecast gas demand. Monthly data on gas DSM spending are not reliable, however, so these results can at best provide supplementary or supporting evidence on the impact of DSM programs on gas consumption. Our econometric models that used more reliable measures of gas DSM spending were never able to identify a significant relationship between DSM activity and gas consumption.

Our report is organized as follows. The following section briefly discusses the industry's experience with top down estimates. Section three describes and analyzes the current gas demand models in Ontario. Section four discusses some implications of this research on an appropriate specification of "top down" econometric models for estimating gas savings. Section five presents PEG's econometric estimation of these top down models. Section six presents concluding remarks. There is also an Appendix that discusses the California experience with econometric estimation of savings from conservation programs.

2. EXPERIENCE IN INDUSTRY

One of the issues PEG examined was utilities' experience with "top down," econometric estimation of energy savings from approved DSM programs. The DSM programs of interest were those where econometric estimation methods were applied *ex post* (i.e. after DSM programs had been implemented) to aggregate billing data (e.g. energy consumption for all customers on a specific revenue class), rather than data for individual customers who were known to be participating in utility DSM programs.

Unfortunately, it was not possible to undertake a comprehensive survey of this issue in the short time available to prepare this report. In fact, compiling such a survey would be a formidable task even if time and resources were unlimited. There is no centralized database or library for DSM regulatory decisions, and the heyday for DSM programs was in the early and mid-1990s, before electronic copies of such files were accessible more easily through the web. Assembling the basic regulatory documents would therefore be a labor-intensive process requiring many hours of identifying, locating, and copying paper reports that outline utilities' specific measurement and verification (M&V) procedures.

Time constraints notwithstanding, PEG's review did not identify any jurisdictions that undertake the kind of "top down" econometric estimations of savings we were asked to explore in this project. The dominant approach used in the industry is clearly the *ex ante* (i.e. savings projected in advance), engineering-based approach that is used in Ontario. This view was confirmed in conversations with DSM specialists at the Edison Electric Institute (which closely monitors and compiles information on DSM programs at US investor-owned utilities), Northeast Energy Efficiency Partnerships (which facilitates a M&V Forum and conducts research on M&V practices in the industry), and M&V professionals.

It should be noted, however, that California has extensive experience with econometric estimation of savings that applies to customer-specific data differentiated by whether customers are participating in utility DSM programs. There was an explicit move away from engineering-based estimates of energy savings to econometrically-

derived estimates using customer-specific data in California in the mid-1990s. Data available at this level of detail naturally facilitates econometric studies that can identify the impact of DSM programs *per se*. Researchers can analyze customers' consumption patterns before and after specific DSM measures are installed. Differences between their "pre-" and "post-installation" normalized volumes can then be compared with changes in normalized consumption for non-participating customers over the same period to estimate net-to-gross (NTG) ratios.

After 2000, however, there has been a movement back towards engineering-based estimation of energy savings in DSM programs. The rationale for this change is not explained in detail in California Public Utility Commission documents but, according to people involved in California M&V, one critical factor was the meltdown in California's retail electricity market in 2000-01. The failure of this policy led to sweeping, systemic reforms. One such reform included greater emphasis on market transformation (rather than program specific) DSM programs, and market transformation programs are less amenable to econometric estimations. Nevertheless, econometric estimation of energy savings (using customer-specific data) remains an option in California, although M&V in the State is primarily engineering-based. The Appendix of this Report presents a more detailed discussion of California's experience with econometric estimations of savings from conservation programs.

3. EXISTING GAS DEMAND MODELS IN ONTARIO

EGD and Union both currently use econometric models to forecast their customers' natural gas usage. Both develop predictions of normalized average gas consumption for different customer groups/rate classes. The models are also regionally differentiated *i.e.* estimated for customer groups in different portions of their service territories.

Both models share some features. For example, both include some measure of the price of natural gas and weather (heating degree days) as independent variables, which is of course standard in natural gas demand models. Both also include independent variables on customer characteristics that impact the demand for natural gas, although the choices for these variables differ. EGD uses “vintage” variables that reflect the share of customers added since 1991, which was the year that Ontario’s Energy Efficiency Act increased efficiency standards for gas furnaces. Customers of a more recent “vintage” would therefore be using more energy efficient equipment, which all else equal would reduce their natural gas consumption. Union has constructed an alternate index of furnace efficiency based on estimates of the share of its customer base that still uses older, less efficient furnaces. This variable is updated annually based on assumed furnace replacement rates.

Both companies also employ diagnostic tools to assess the quality of their models, but EGD’s tests are more extensive. EGD employs the Breush-Godfrey test for autocorrelation; the autoregressive conditional heteroskedacity (ARCH) test to test for heteroskedasticity (*i.e.* non-constant variance in the residual error terms across observations); the Chow test to test whether model parameters are stable across time; and the Ramsey Regression Equation Specification Error Test (RESET) to test the overall specification of the model. Union uses the Durbin-Watson test to test for autocorrelation and an F test on the overall model specification. While Union does not employ any tests for heteroskedasticity, in 2004 they commissioned a review of their model by R.J. Rudden, and Rudden’s review undertook heteroskedasticity tests on the results from

Union models. Rudden's review also generally affirmed the reasonableness of Union's gas demand model.

It should be noted, however, that neither the EGD nor Union Gas demand models have been used directly for DSM savings calculations. In addition, neither model was specifically designed for this purpose. The standards that apply for evaluating whether econometric models can identify DSM-related energy savings may differ from those that are relevant for assessing their ability to predict overall gas consumption. Different econometric specifications may also be warranted for econometric estimation of energy savings.

This chapter will briefly review the econometric gas demand models that are currently used in Ontario. We begin with the Union Gas model and the associated Rudden Review report. We then turn to EGD's gas demand model.

3.1 Union Gas Demand Models

Union has gas demand models are used for forecasting total throughput volumes for residential and commercial general service customers. For each sector, there are separate econometric estimations of the total number of customers (*i.e.* the demand for access to gas distribution service) and normalized average gas use (NAC) per customer. The total demand forecast by sector is calculated by multiplying the estimated number of customers in each month (from the first model) by the monthly estimates of average use per customer (from the second model).

Union's econometric models that forecast customer numbers are not really relevant for this project. A "top down" estimation model would be focused on identifying the volume of gas savings due to DSM programs, for a given (and known) number of customers. Since the customer number models are not relevant for this purpose, we do not consider them in this report but instead examine only the NAC demand models.

Union has developed NAC demand models for five different customer groups: 1) residential M2 tariff volumes per customer; 2) residential R01 volumes per customer; 3) commercial M2 volumes per customer; 4) commercial R01 volumes per customer; and 5)

commercial R10 volumes per customer.² In all cases, Union actually develops two separate NAC forecasts using two separate econometric models. The first estimate is based on a regression where use per customer is the dependent variable in the regression. Union refers to this as the “Use” equation. The second estimate is based on a regression where total volume is the dependent variable. This equation is then used to project total volumes, and these volumes are divided by the forecast for customer numbers to produce a use per customer forecast. Union then averages the NAC forecasts from these two approaches to produce its use per customer forecast for each customer group.

Union uses the following independent variables in these equations:

Residential Use Per Customer

Use Per Customer Regression

- Heating degree days (HDD): a measure of weather severity which affects the demand for space heating; HDD coefficients are actually estimated separately for nine separate months (HDD are zero in the months of June through August)
- Residential furnace efficiency: Union has constructed an index of residential furnace efficiency for its customers; it is based on estimates of the current fraction of Union’s residential customer base that still has a conventional furnace, and an assumption that 6% of these customers will replace their conventional furnace with a high efficiency furnace each year. This variable reflects the expected decline in gas consumption that results when customers replace their older furnaces with higher efficiency models.
- Number of persons per household: Union has noticed that average use per customer has declined during the summer months, when the main residential use for natural gas is for water heating. The demand for water heating will be impacted by the number of persons living in a household, and customer surveys indicate that the number of persons per household have declined over time. This variable is designed to capture the reduction in demand for water heating that results from a fewer number of persons per household, on average.

² Union also developed forecasts for industrial customers, but the scope of our work applies only to residential and commercial DSM programs, so the industrial demand equations are not discussed for either Union or EGD.

- Total bill: this variable is designed to reflect the impact of changes in natural gas prices on the demand for natural gas. Union had used a price variable in earlier gas demand models but found that the t statistic on the total bill variable was greater, so total bill was substituted for a price measure. The total bill variable is lagged by a month, in most instances.

Total Volume Regression

- Heating degree days
- Total bill
- Total Customers: designed to reflect the fact that the total demand for gas volumes will naturally rise as the number of customers increases.

Commercial Use Per Customer

Use Per Customer Regression

- Heating degree days
- Segmentation and efficiency index: this is analogous to the furnace efficiency index used in the residential use per customer regression. It reflects two main trends: 1) the fact that retail and office commercial customers have lower annual NAC levels compared with other commercial customers, and these segments have grown relatively more rapidly than other segments; and 2) changes in energy utilization efficiency per unit of floor space.
- Total bill

Total Volume Regression

- Heating degree days
- Total customers
- Total bill: unlike the other equations, however, the total bill is lagged four months in this regression.

All equations are estimated using ordinary least squares. The residential equations are, in most instances, estimated using sample data for the January 1994 to

March 2005 period.³ The commercial equations are estimated using sample data from May 1990 through March 2005. The results from the regressions are presented in Tables 1 and 2 in Appendix B of Paul Gardiner's December 2005 testimony (EB-2005-0520, Exhibit C1, Tab 1).

For the "Use Equations," all of the parameter estimates are statistically significant at the 5% level or better (with the exception of the intercept on the residential M2 equation). The Durbin-Watson (DW) statistics are also good, with the exception of the Commercial 01 regression. The DW tests suggest that autocorrelation is generally not a problem in these regressions.

With respect to the "Volume Equations," all of the parameter estimates are statistically significant at the 5% level or better (with the exception of the intercept on the commercial 01 equation). The DW statistics are not as good as in the "use" equation, with the worst reported DW value again for the Commercial 01 regression. Still, in most instances the DW values are in the indeterminate range, which suggests that autocorrelation is only somewhat more problematic in these regressions.

Union also reports on the forecast accuracy of their models. The volume (demand) forecast accuracy results are presented in Table 4 of Appendix B of the testimony, and the NAC forecast accuracy results are presented in Table 5. The volume forecast errors varied from 1.1% to 3%, depending on the customer group or whether the forecasts were evaluated in-sample (*i.e.* within the same sample period used to estimate the model) or out-of-sample (which Union performs using a shortened sample period for the 2004 year; they refer to this as "Ex Post Error"). The NAC forecast errors (without DSM variances) are reported to be 1.1% to -2.8%.

RJ Rudden was commissioned to analyze the Union gas demand models. They wrote that their "objective...was to evaluate the Union Gas Forecast Models applicable to general service customers from the following perspectives: forecast accuracy; logical construction; and statistical goodness of fit."⁴ These were natural criteria, since the models were constructed to forecast Union's gas volumes and, in Rudden's words, "(f)or

³ The Use equation for the residential rate M2 was estimated using data from January 1994 to December 2004.

⁴ *RJ Rudden Review of the Union Gas Demand Forecast Methodology*, December 2005, Attached as Appendix C to EB-2005-0520, Exhibit C1 Tab 1; p. 1.

models designed to forecast in the short term, the best indicator of forecasting success is the accuracy achieved by the forecasting process.”⁵ Rudden also writes that:

Statistical issues (e.g., autocorrelation, multicollinearity, and heteroskedasticity) that could render long-term models unreliable/unstable are less of an issue in a short-term structure. The reason for this is that short-term forecasts progress only a short time distance (in term of time periods ahead) from the end point of the history of the estimated model....Thus, such structural problems, if they do exist, have less of an absolute influence on the forecast results. Autocorrelation, multicollinearity and heteroskedasticity actually increase their influence in a compounding fashion, the longer the forecast horizon. Thus, the shorter the forecast period, the less the overall period.⁶

Largely because of Rudden’s practical concern of forecast accuracy, they concluded that “Union’s forecasts and underlying methodologies are reasonable and produce accurate results. Union’s Volume Forecasts for the Residential M2, 01 and Commercial M2, 01 and 10 classes are logical and statistically credible forecasting methodologies that produce accurate results sufficient for reliable 12-24-month ahead projections.”⁷ Rudden also found that “(c)ritics of the Union forecasts appear to have a focus on statistical “perfection,” perhaps at the expense of a good forecast.”⁸

PEG does not dispute Rudden’s conclusions with respect to the forecast accuracy of Union’s econometric demand models (which, of course, is their purpose). We should note, however, that the statistical issues that Rudden de-emphasizes will be more important in any “top down” econometric model that is focused on estimating the amount of gas savings that result from utility DSM programs. If autocorrelation and heteroskedasticity exist, they will bias inferences on the statistical significance of individual explanatory variables, as well as on the hypothesis of whether the difference between actual gas consumption and the gas consumption predicted by the econometric model is statistically significant.⁹ Biased inferences on individual parameter estimates, as

⁵ *RJ Rudden Review, op cit*, p. 2.

⁶ *RJ Rudden Review, op cit*, p. 3.

⁷ *RJ Rudden Review, op cit*, p. 14.

⁸ *RJ Rudden Review, op cit*, p. 12.

⁹ Multicollinearity, or strong correlation among the independent variables used in an econometric model, is a feature of the sample data that tends to increase the standard errors of parameter estimates and therefore increase the probability that the regression will not produce statistically significant estimates of

well as on overall model predictions, could be very problematic in such a top-down model. For example, a top down model could use the coefficient on DSM expenditures as an estimate of the overall relationship between DSM spending and changes in consumption (as in the Loughran-Kulick paper, or Cicchetti book). Heteroskedasticity and autocorrelation would bias inferences on whether the DSM variable is statistically significant and could thereby lead to incorrect conclusions on the usefulness of such an approach. Another “top down” approach could be testing whether there is a statistically significant difference between utilities’ actual gas volumes and the volumes predicted by a gas demand model that does not reflect utility DSM behavior. If this difference is statistically significant, some portion of the “residual” volumes (*i.e.* those gas volumes that are not explained by the econometric model) could be interpreted as a measure of utilities’ gas DSM programs. Obviously, this approach requires an unbiased inference on whether actual and predicted gas volumes are significantly different (in a statistical sense), but these inferences will be biased if autocorrelation or heteroskedasticity are present.

Overall, PEG believes there are some positive attributes to Union’s gas volume econometric models. They are simple, straightforward and transparent, and generally perform well on the basic statistical tests. The explanatory variables are largely intuitive and, in the case of the commercial segmentation/efficiency index, creative. There is also some merit in using an average from two forecasting models to develop a “consensus” Company forecast.

We have five main concerns with the Union econometric models, particularly as a potential starting point for a top-down, M&V econometric model. First, while using total bills rather than gas prices as an explanatory variable may improve the models’ forecasting ability, it is suspect both economically and statistically. Economic theory clearly links changes in quantities to changes in prices, not changes in the total amount paid for service.¹⁰ In fact, because total bills will reflect both the total quantity of gas

the parameters on the independent variables. While this may be undesirable, it does not bias either the parameter estimates or inferences on statistical significance.

¹⁰ Price elasticity is related to the relationship between changes in quantities and changes in revenues/total bills, but price elasticity will generally not be measured by regressing quantities on total bills.

consumed and the price for service, this variable may not be completely independent of the dependent variable, which is gas consumption.

This concern is mitigated somewhat by the fact that the total bill variable in the Union models is lagged by a month, but this raises a second concern: the gas demand models that include total bill also effectively include a lagged value of gas volumes as an explanatory variable. In other words, if gas volumes in month t is a function of total bills in month $t-1$, the model is effectively regressing gas volumes in month t on gas price prices in month $t-1$ and gas volumes in month $t-1$, since total bills will reflect both the prices paid for natural gas and the total volumes that were billed in the month. Including “lagged dependent variables” as explanatory variables is not necessarily a problem, but it is well known that the Durbin Watson statistic is not a valid test of autocorrelation when one of the independent, right-hand side variables is a lagged value of the dependent variable.¹¹ This implies that the generally good DW statistics reported by Union cannot be taken as conclusive evidence that their models do not exhibit autocorrelation. As discussed, autocorrelation will be more of a problem in a top-down econometric M&V model than in Union’s forecasting model.

Third, the furnace efficiency index is a valid but relatively narrow explanatory variable. Residential volumes can also be impacted by other factors that tend to be incorporated into new construction, such as better insulation and thermal windows. Union’s furnace efficiency index does not reflect these potential impacts on residential gas consumption.

Fourth, Union’s gas demand models do not include any variables that reflect overall economic activity. Some correlation between economic activity and gas consumption would be expected, particularly for commercial customers. While economic activity variables may be less important for Union’s forecasting purposes, it is

¹¹ For example, see Greene, W. (2000), *Econometric Analysis*, Prentice Hall: Upper Saddle River, NJ, pp. 542; Nerlove, M. and K. Wallis (1966), “Use of the Durbin-Watson Statistic in Inappropriate Situations,” *Econometrica* 34: 235-238; Durbin, J., (1970), “Testing for Serial Correlation in Least Squares Regression When Some of the Regressors Are Lagged Dependent Variables,” *Econometrica* 38: 410-421; and Dezhbaksh, H. (1990), “The Inappropriate Use of Serial Correlations Tests in Dynamic Linear Models,” *Review of Economics and Statistics* 72: 126-132.

important for a top down, M&V model to include such variables so that the estimates of the explanatory variables that are included do not exhibit omitted variable bias.

Finally, Union measures the impact of weather on gas demand with a series of HDD variables that differ by month. These variables are, in reality, a combination of a dummy variable for the month multiplied by measured HDD in that month. The interpretation of these coefficients is therefore different from the interpretation of HDD in a conventional gas demand equation; it does not measure the impact of weather on gas volumes, but rather the impact of both weather and unspecified fixed effects that are specific to a month. Intuitively, this approach assumes that the impact of a given value of heating degree days in, say, January has a different impact on gas consumption than would result if the same heating degree days were experienced in February or March. While this specification may improve Union's forecasting accuracy, it is not appropriate for isolating the impact of HDD on gas consumption and should not be implemented in any "top down" model.

3.2 Enbridge Gas Demand Models

Enbridge uses a two-step estimation procedure and an Error Correction Model (ECM) that was developed by Engle and Granger.¹² Engle and Granger describe the motivation for this model as follows:

An individual economic variable, viewed as a time series, can wander extensively and yet some pairs of series may be expected to move so that they do not keep such series apart. Typically economic theory will propose forces which tend to keep such series together. Examples might be short and long term interest rates, capital appropriations and expenditures, household income and expenditures, and prices of the same commodity in different markets or close substitutes in the same market. A similar idea arises from considering equilibrium relationships, where equilibrium is a stationary point characterized by forces which tend to push the economy back toward equilibrium whenever it moves away...In this paper, these ideas are put onto a firm basis and it is shown that a class of models, known as error correcting, allows long-run components of variables to obey equilibrium constraints while short-run components have a flexible dynamic specification.¹³

¹² Engle, R. and C.W.J. Granger (1987), "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, Vol 55: 2, 251-276.

¹³ Engle and Granger, *op cit*, pp. 251-252.

Accordingly, EGD has estimated both short-run and long-run models for one residential class (Rate 1 Revenue Class 20 customers) and two commercial revenue classes (Rate 6 Revenue Class 12 customers and Rate 6 Revenue Class 48 customers)¹⁴ In all cases, separate models are estimated for customers in different geographic groups. Rate 1 Revenue Class 20 is divided into six such segments: 1) Metro Region – Central Weather Zone; 2) Western Region – Central Weather Zone; 3) Central Region – Center Weather Zone; 4) Northern Region- Central Weather Zone; 5) Eastern Weather Zone; and 6) Niagara Weather Zone. The Rate 6 Revenue Classes 12 and 48 are each divided into three groups: Central, Eastern, and Niagara. The independent variables in these models are the following:

Rate 1 Revenue Class 20

- Heating degree days
- Real residential price of natural gas (*i.e.* expressed relative to the CPI)
- A vintage variable, which reflects the share of customers added since 1991, which was the year that Ontario’s Energy Efficiency Act increased efficiency standards for gas furnaces. Customers of a more recent “vintage” would therefore be using more energy efficient equipment, which all else equal would reduce their natural gas consumption.
- Central employment zone employment (for some regressions), which reflects economic activity in the region
- A linear time trend, to reflect changes in gas consumption over time that are not captured by the other explanatory variables; these factors can include the impact of DSM programs.

Rate Class 6

The Rate 6 regressions employ a larger array of variables, although they are not all employed in every Rate 6 regression:

- Heating degree days

¹⁴ EGD has also estimated gas demand equations for Rate Class 6 Revenue Class 73 industrial customers, but our work is intended to examine residential and commercial “top down” models only, so these demand equations are not considered here.

- Employment (in either the central, eastern or Niagara weather zones)
- Real commercial price of natural gas (in the central, eastern or Niagara weather zones)
- Ontario real Gross Domestic Product
- Greater Toronto Area (GTA) commercial vacancy rate
- A linear time trend
- In some regressions, a lagged dependent variable
- The Central Revenue Class 12 regression also contains dummy variables for 2005 and 2006 and a dummy to reflect customer migration; this is also the only customer class where only a single regression is estimated.

It should also be noted that the short-run regressions are specified in “first difference” form (*i.e.* changes in variables) and include error correction mechanism terms resulting from the ECM procedure.

Like Union, the main purpose of the EGD demand models is forecasting. EGD says that the main criteria they use for evaluating their models’ predictive ability is forecast accuracy. Forecast accuracy is measured using in-sample and out-of-sample average variance.

EGD uses a variety of diagnostic tests to evaluate their econometric results. The Bruesch-Godfrey test is used to test for whether there is autocorrelation among the error terms. The autoregressive conditional heteroskedasticity (ARCH) test is used to test for the presence of heteroskedasticity, or non-constant variance in the error terms. The Chow test is used to test whether the parameter estimates are stable across different time periods. The Ramsey RESET tests for a variety of specification errors, including omitted variables, incorrect functional forms and correlation between the independent variables and the error term. The null hypothesis for this test is that the error term is normally distributed with a zero expected value; if the null is rejected, the error term is normally distributed with a non-zero expected value.

The regression results for the Rate 1 customer groups are presented in Table 5 of the Denomy testimony (EB-2008-0219, Exhibit B, Tab 2, Schedule 2). Most but not all of the estimates have the expected sign and are statistically significant at the 5% level or better (with the exception of the constant terms). The main exception is the natural gas

price term, which is significant in five of the six long run regressions but in only three of the six short run regressions (it is statistically significant at the 10% level in four of the six short run regressions).

Table 6 presents the results of the diagnostic tests for Rate 1. If these results are taken at face value, they suggest every one of the Rate 1 models is characterized by autocorrelation, heteroskedasticity, parameter instability, and specification error. The null hypothesis for no autocorrelation, no heteroskedasticity, no parameter instability, and no specification error is rejected for all six regressions.

The regression results for Rate 6 customers are presented in Table 8. The results for the commercial revenue classes are qualitatively similar to those reported for Rate 1 customers. Most but not all of the parameters have the expected sign and are statistically significant at the 5% level. The main exception again is the natural gas price, although this variable is not included in all the regressions. In fact, there is a considerable amount of heterogeneity in the independent variables included in the commercial demand equations for different regions and revenue classes.

The diagnostic tests on the Rate 6 regressions are presented in Table 9. Again, if they are taken at face value, they indicate that the regression results for commercial customers exhibit autocorrelation, heteroskedasticity, parameter instability and specification error. The only null hypothesis that is accepted is the Chow test (no parameter instability) for the Revenue Class 12 Central Weather Zone regression.

Tables 2 and 3 report the in-sample forecast variance for the Rate 1 and Rate 6 models, respectively, in each year from 2001 through 2007. The sample period used to estimate the models was 1985 through 2007. These results show that the forecast errors for the econometric models tend to be relatively small. For Rate 1, the difference between actual normalized usage per customer and the model's predicted usage per customer ranges from -1.33% (2005) to 1.15% (2007). These forecast errors also seem to be randomly distributed over the years, with no apparent trend. For Rate 6, the difference between actual normalized usage per customer and the model's predicted usage per customer ranges from -0.86% (2001) to 0.55% (2006), again with no discernible trend. These results indicate that the models appear to generate reasonable forecasts, in spite of the statistical flaws indicated by the diagnostic tests.

There are some positive elements of the Enbridge econometric models. The vintage variable is similar to Union's furnace efficiency index but somewhat more comprehensive, since it is constructed using data on residential dwellings and not the age distribution of furnaces only. EGD also uses a variety of variables that reflect economic activity. These variables are usually statistically significant, which is evidence that they have a substantial impact on gas consumption.

There are also some obvious concerns with the EGD models. Most importantly, while the EGD methodology employs a battery of tests on the output of their models, their models actually do poorly on these tests. EGD's diagnostic tests show that, in nearly all cases, their econometric models exhibit autocorrelation, heteroskedasticity and parameter instability over time. For the reasons discussed earlier, these traits will be more problematic for developing "top down" estimates of savings than is likely to be the case in a forecasting application. The EGD results suggest that improved econometric results can be obtained through generalized least squares procedures that are directed towards addressing problems with autocorrelation and heteroskedasticity.

We also do not believe there would be any value in using an ECM when investigating the viability of top-down, econometric estimation methods. Our work is not focused on investigating instances where variables in the short-run may diverge from their long-run equilibrium values. In fact, the type of equilibrating tendencies that Engle and Granger describe as motivating the ECM appear to be entirely absent for utility DSM programs.

4. DEVELOPING “TOP DOWN” M&V MODELS

4.1 Previous Work

The econometric research on energy conservation and utility DSM programs is extensive, and this report was not designed to survey this literature. Nevertheless, there are recent econometric articles that could prove helpful for developing a feasible, “top down” econometric estimation model for Ontario’s gas distributors. This section will briefly review this econometric work.

The most noteworthy article is “Demand Side Management and Energy Efficiency in the United States,” by David Loughran and Jonathan Kulick.¹⁵ This article examined whether DSM expenditures have increased the electricity efficiency of the US economy. The authors tested this issue using panel data on DSM expenditures for 324 US utilities over the 1989-99 period. These data are available from the US Energy Information Administration on EIA Form 861.

There are two intriguing elements of the Loughran-Kulick methodology. First, they use DSM spending itself as an explanatory variable in their econometric model. The main focus of their work is to examine the relationship between this variable and energy consumption. This approach therefore provides direct evidence on the relationship between spending on DSM programs and the resulting impact on energy consumption. This evidence would be more relevant to developing “top down” estimates of DSM-related savings than, say, the INSTALL dummy variables that were previously discussed in the econometric models in California.

Second, Loughran-Kulick employ a “first difference” econometric specification that examines the relationship between *changes* in energy consumption and changes in independent variables, including DSM spending. They claim that this simple approach will lead to generally appropriate estimates of the net energy impacts from utility DSM programs. They argue that a

¹⁵ Loughran, D. and J. Kulick, “Demand Side Management and Energy Efficiency in the United States,” *Energy Journal*, 25: 1, 19-43.

“first-differenced specification...controls for any fixed effects between utilities in the level of electricity sales. This is important since these fixed differences across utilities could be correlated with DSM expenditures generating spurious correlation between these expenditures and electricity sales. Utilities with large DSM programs may serve regions with particularly strong sentiments for conservation, for instance. Such regions may be more inclined to adopt building codes and appliance and equipment standards that emphasize electricity efficiency or just generally be more conservative in their use of electricity.”¹⁶

Thus, Loughran and Kulick believe a first-differenced specification can control to at least some extent for a variety of factors that come into play when determining net-to-gross ratios.

Loughran and Kulick find that utility DSM programs have a more modest impact on energy consumption than utilities typically estimate. The estimated impacts depend on the specification, but in their most credible models DSM expenditures reduce electricity sales by 0.4% to 0.6% per annum. The authors write that they “suspect utility estimates of DSM program effects are higher than our estimates because utilities generally do not control for selection bias.”¹⁷

This paper has been widely cited and was recently criticized in the same journal. Auffhammer, Blumstein and Fowlie (ABF) claim that the test statistic that Loughran and Kulick use is not appropriate for the hypothesis they are testing.¹⁸ Their main concern is that Loughran and Kulick examine the relationship between DSM expenditures and unweighted, rather than weighted, changes in energy consumption. This is relevant since utilities with lower electricity sales tend to spend less on DSM programs and report lower energy savings (in percentage terms). Since observations for these smaller utilities are treated the same (*i.e.* they are not weighted any differently) as observations for larger utilities, ABF claim that the Loughran-Kulick analysis puts too much emphasis on the experience of small utilities. ABF also claim that energy savings are unusually relative to expenditures in the first year of reporting, especially for utilities with small DSM programs. ABF develop a new test statistic that controls for these purported flaws, and

¹⁶ Loughran and Kulick, *op cit*, pp. 27-28.

¹⁷ Loughran and Kulick, *op cit*, p. 39.

¹⁸ Auffhammer, M., C. Blumstein and M. Fowlie, “Demand-Side Management and Energy Efficiency Revisited,” *Energy Journal*, 29: 3, 91-104.

their estimated savings from utility DSM programs are considerably higher than those estimated by Loughran and Kulick.

Regardless of the merits of the ABF critique, it would not appear to be relevant in an Ontario gas DSM application. Auffhammer et al. essentially find fault with the fact that Loughran and Kulick do not adequately control for differences in the size of companies in their cross sectional dataset. This issue would be far less important if a variant of Loughran-Kulick approach was applied to Ontario's gas distribution industry, since there are only two gas distributors in the Province, and they are both large and of (roughly) similar size.

It should be noted that other economists have also recently examined the relationship between spending on DSM programs and energy savings. In his 2009 book *Going Green and Getting Regulation Right: A Primer for Energy Efficiency*, PEG Senior Advisor Charles Cicchetti used EIA 861 data to examine the impact of energy efficiency spending on reported energy savings. He finds a highly significant positive relationship between these variables.¹⁹ This work provides further support for the use of gas DSM expenditures as an explanatory variable in econometric M&V models.

Some variant of the Loughran-Kulick model, which includes measures of Union and EGD DSM spending as explanatory variables, could potentially be applied to the Ontario environment. The regulatory applications of gas forecasting models in the Province make a fair amount of data on gas consumption and explanatory variables available to researchers, including data on DSM spending for certain classes of customers. This information provides a solid (if not ideal) foundation for further investigation into the "top down" M&V issue. However, as discussed in the previous chapter, we believe the econometric methods that have been used to date in Ontario can and should be enhanced to deal with autocorrelation and heteroskedasticity. It is more important for these statistical problems to be addressed in an econometric model used to measure energy savings from DSM programs than in a gas forecasting model.

¹⁹ For example, see Tables 23-2 and 23-3 on pp. 253-54 in Cicchetti, C., *Going Green and Getting Regulation Right: A Primer for Energy Efficiency*, Public Utilities Reports Inc., Vienna VA.

4.2 PEG's Approach and Specifications

Building on the Loughran-Kulick model, PEG's main approach for developing "top down" econometric estimates of gas savings examines whether there is a statistically significant link between DSM spending by Ontario gas distributors and subsequent changes in gas consumption. A supplementary approach investigated whether there are statistically significant differences between actual and predicted changes in gas consumption, where predictions are based on econometric gas demand models that do not include DSM spending as an explanatory variable. Any statistically significant differences between actual and predicted gas usage using these models could be interpreted as an indicator, at least, of the impact of DSM programs on gas consumption.

One issue that is important for evaluating an "ideal" top-down econometric specification in Ontario is the frequency of the data to be used in the analysis. Three main factors are relevant for evaluating this issue. The first is simply the amount of the sample data. In statistical analysis, more information is almost always preferred to less. All else equal, larger samples increase confidence in the statistical estimates. Monthly data naturally lead to larger sample sizes than annual data and will be preferred on this criterion.

Another factor is the accuracy of the data. It is clearly important for all data to be accurately recorded and measured. Data errors can lead to biased estimates of regression parameters. In Ontario, neither EGD nor Union has customarily reported data on its DSM expenditures on a monthly basis. Both companies also claim that the quality of any monthly DSM spending data will be suspect since, among other reasons, DSM costs are not necessarily booked in the same month in which actual program costs are incurred. Because annual expenditures will provide a more accurate measure of the companies' DSM programs over the selected interval than more frequently reported (*i.e.* monthly) data, annual data are preferred to monthly data on this criterion.

The third factor is the varying temporal pattern of gas consumption throughout the year. Clearly, for most residential and commercial customers, the pattern of gas consumption varies substantially over the course of a year, and the time pattern of gas DSM expenditures may vary as well. This implies that the quantitative relationship between expenditures on DSM measures and, say, changes in gas consumption in the

following month may also vary throughout the year. For example, expenditures on DSM measures that are installed in June may have less of an impact on gas consumption in the following month than the same measures would have if they were installed in November. The estimated coefficient on DSM expenditures in a demand equation would represent the impact of DSM on consumption in an ‘average’ month, and if this relationship varies over the course of the year there would be more variance and, all else equal, larger standard errors associated with estimates developed using monthly rather than annual data. This would reduce the likelihood of developing statistically significant estimates of the impact of DSM programs. Thus, this factor supports the use of annual rather than monthly data.

Given these factors, PEG believes the ideal “top down” specification in Ontario would utilize annual data on DSM expenditures when analyzing the relationship between DSM spending and changes in gas consumption. At the same time, it should be recognized that there is a temporal pattern for gas consumption during the course of the year that is driven primarily by weather (and, to a lesser but related extent, prices for natural gas, which often increase during high-use periods during the year). The impact of weather (and to a lesser extent price) factors can be estimated more precisely if monthly data are used, since these data will track changes in gas consumption within the year due to changes in heating degree days and delivered prices for natural gas.

PEG’s first econometric approach uses the existing data and gas demand models in Ontario as a starting point, but reflects these ideas on the “ideal” specification as well. In particular, we supplemented the existing gas demand models in Ontario with data on DSM spending and other customer and economic conditions as explanatory variables. The ideal frequency of some independent variables – particularly heating degree days and delivered natural gas prices – is monthly, while the ideal frequency of some other explanatory variables (especially DSM spending) is annual.

Because of this difference in the preferred frequency of different explanatory variables, our main top down approach used a two-stage econometric approach. The first stage regresses monthly data on gas consumption per customer monthly values of heating degree days (HDD) and prices (Pr). These regressions are done by revenue class (*i.e.* the same five revenue classes for Union, and the same three revenue classes for Enbridge,

that the companies use in their gas demand models) Thus, for each revenue class j for each company i , and in each month t , PEG estimates

$$\frac{V_{i,t}^j}{N_{i,t}^j} = f(HDD_{i,t}, Pr_{i,t}^j) \quad (1)$$

After this first stage regression was estimated for each revenue class, we inserted monthly values for HDD and price into the fitted regressions to obtain normalized, monthly consumption volumes per customer. These normalized average use per customer values were then multiplied by the associated actual customer numbers for the month to yield total normalized volumes by month

$$\hat{V}_{i,t}^j = \hat{f}(HDD_{i,t}, Pr_{i,t}^j) \bullet N_{i,t}^j \quad (2)$$

These monthly values were then aggregated over all months in a year, to compute annual values $\tilde{V}_{i,T}^j$ (T=1991 through 2008) of normalized consumption for the year.

The second stage regression uses a measure of the difference between actual and normalized consumption as the dependent variable. More specifically, the dependent variable in the second stage regression is the logarithm of actual gas consumption divided by $\tilde{V}_{i,T}^j$ in each year, which is mathematically equivalent to the log of actual consumption minus the log of $\tilde{V}_{i,T}^j$. The dependent variable therefore reflects change in annual gas consumption that cannot be attributed to changes in heating degree days or delivered natural gas prices. This dependent variable is then regressed on a constant term, DSM spending (DSM) in the previous year, a vector of economic variables (**EC**) that can impact gas consumption (*e.g.* total employment and Ontario GDP), and a vector of customer specific variables (**Cust**) that can impact demand (*e.g.* customer “vintage” or the number of persons per household). In general terms, we regress

$$\ln\left(\frac{V_{i,T}}{\hat{V}_{i,T}}\right) = b_0 + b_1 DSM_{T-1} + b_2 \Delta \mathbf{EC} + b_3 \Delta \mathbf{Cust} \quad (3)$$

With this specification, the constant term b_0 will measure changes in gas consumption for the revenue class (or classes) that are independent of changes in heating degree days and natural gas prices (both of which are reflected directly in the dependent

variable), DSM spending, and changes in economic conditions and customer characteristics that drive consumption. It is reasonable to believe that “free riders” will be reflected in b_0 , since free rider effects by definition reflect actions that customers are taking independent of utility DSM programs. On the other hand, the b_1 coefficient on the DSM variable would reflect *all* the effects of DSM spending on energy consumption net of the other independent variables. Thus, b_1 would capture both the direct effects on consumption from customers participating in utility DSM programs as well as any indirect or spillover effects. The b_1 coefficient can therefore be interpreted as the impact that a unit (*e.g.* dollar) of DSM expenditures has on the change in normalized gas consumption, independent of all other factors that drive gas usage. This would in theory be an appropriate “top down” measure of gas savings to use in TRC calculations in the Enbridge and Union DSM programs.

PEG’s second approach added DSM as an explanatory variable to variants of the gas demand models that are currently used by EGD and Union Gas. We also added the EcoEnergy variable to these models, since these programs may have had a significant impact on residential gas consumption in 2007-08 but have not been included in either of the Companies’ models to date. PEG retained the estimation procedure that we used in earlier regressions, since this corrects for ARCH and serial correlation, and it is important to correct for these influences to obtain the most efficient statistical estimates and improve statistical inference. PEG’s estimation procedure therefore differs from that used by Union and EGD.

PEG’s third approach is to regress gas usage per customer on heating degree days, prices and the vectors of economic and customer characteristics discussed above, but **not** include DSM as an independent variable. Thus for each revenue class j for each firm i , and for monthly observations t , PEG estimates

$$\frac{V_{i,t}^j}{N_{i,t}^i} = g(HDD_{i,t}, Pr_{i,t}^j, \mathbf{EC}_E^j, \mathbf{Cust}_{i,t}^j) \quad (4)$$

The **EC** and **Cust** vectors are defined as above, so the actual variables to be explored will differ by revenue class.

We then insert values for each of the independent variables above to compute predicted values for gas consumption, by revenue class by year. We then construct 95%

confidence intervals around this predicted value and compare actual gas consumption to the predicted value and the confidence intervals. If actual consumption is below the predicted value and outside of the confidence interval, we can conclude that actual gas consumption is significantly below its predicted value. The difference between actual and predicted consumption may therefore also be seen as an indicator of the impact of DSM programs on gas consumption, which are excluded from the regression, or at least evidence that either supports or fails to support the evidence from the first approach.

In all of our regressions, PEG tests for autoregressive conditional heteroskedasticity (ARCH) and serial correlation in the residuals. As discussed in the previous chapter, these statistical problems are more problematic when trying to identify the independent impact of DSM expenditures on gas consumption than they would be in a forecasting application. If we cannot reject the hypothesis of either ARCH or serial correlation, we will use a generalized least squares estimation procedure that corrects for these problems. Further details of these corrections are discussed in the following chapter.

5. ECONOMETRIC ESTIMATION OF “TOP DOWN” MODELS

This chapter presents the results of PEG’s econometric research on “top down” estimation of gas demand models for EGD and Union Gas. We begin by describing our data sources and some data issues encountered in our work. We then briefly discuss our econometric methods, particularly the tests and corrections for ARCH and serial correlation. Finally, we present our econometric results.

5.1 Data

We used the existing gas demand models and data sources in Ontario as a starting point for our work. Much of the data that we used was therefore provided by EGD and Union Gas. In particular, in our first stage regression, we used gas consumption, customer numbers, heating degree days (HDD) and delivered natural gas price data for eight revenue classes that was provided by EGD and Union Gas and previously used in their demand models (three for EGD and five for Union Gas).

Some of the companies’ gas demand models were geographically disaggregated into more than one region. This presented more geographic detail than PEG would be able to use in our subsequent work. To keep our analysis tractable, we therefore estimated a single first stage regression for each revenue class. Where gas consumption, price, or HDD data provided by EGD or Union were geographically disaggregated, PEG computed customer-weighted averages of these variables for each revenue class. The revenue-share level variables were then used in our first stage regressions. More precisely, using 1991-2008 data, the first stage econometric models regressed monthly observations of average gas consumption (total volumes divided by total customers) on monthly HDD and delivered prices for each revenue class. We estimated three such regressions for EGD (for residential revenue class 20 and commercial revenue classes 12 and 48) and five first-stage regressions for Union Gas (for residential 01 customers and residential M2 customers; and for commercial 01 customers, commercial M2 customers, and Commercial revenue class 10 customers).

The available DSM data were also not perfectly aligned with our volume data. For EGD, we only had DSM spending for Rate 1 and Rate 6 tariffs. For the purposes of our second stage regressions, we took all of Rate 1 DSM spending as a measure of DSM activity for the Revenue 20 class within Rate 1. We also added the volumes (*i.e.* the difference between actual and normalized gas volumes) for EGD's Revenue classes 12 and 48 of Rate 6 and linked this to DSM spending for all of Rate 6. This introduced a degree of imprecision, and possible bias, into the econometric estimates but was unavoidable given available data.

PEG also had a limited time series of data available for the second stage regressions for each revenue class. Data on DSM spending were available only from 1998 through 2008. Thus, there were no more than 10 observations for each revenue class for our second stage regressions, since these equations use the logarithmic *change* in (actual minus normalized) volumes as the dependent variable and a lagged value of DSM spending as an independent variable.

There would be very few degrees of freedom, and little chance of obtaining statistically significant results, using such a small sample for each revenue class. PEG's second-stage regressions for residential customers therefore "stacked" the data for all of EGD's and Union's residential revenue classes (*i.e.* for EGD revenue class 20, Union residential 01 and Union residential M2). Stacking the data for different revenue classes triples the number of observations to 30, thereby greatly increasing the likelihood of obtaining statistically significant results. The coefficient on the DSM variable in this regression would measure the impact of a dollar of DSM spending on residential gas savings for residential gas customers in the Province.

Similarly, our second-stage regressions for commercial customers stacked the data for EGD's and Union's commercial revenue classes (*i.e.* for EGD revenue class 12 plus revenue class 48, Union commercial 01, commercial M2 and revenue class 10). Compared to the 10 observations available for each commercial revenue class, this quadrupled the number of observations to 40 for estimating a commercial second-stage regression and hence increased the likelihood of obtaining statistically significant results. The coefficient on the DSM variable in this regression would measure the impact of a dollar of DSM spending on gas savings for commercial gas customers in the Province.

However, our residential and commercial second stage regressions both have separate constant terms for each revenue class. There were accordingly three constants or fixed effects in the residential regression, and four fixed effects in the commercial regression. Having different constants for different revenue classes allows for differences in free ridership and similar unmeasured factors across companies and revenue classes. Only the coefficients on the DSM, economic and customer characteristics are assumed to be the same for Union and EGD in our second-stage regressions.

PEG considered a variety of different economic and customer characteristic variables. The economic variables we explored for residential customers were the unemployment rate, personal income, and personal income per capita in Ontario (all obtained from StatsCanada). Gas usage should be positively correlated with economic activity, so we would expect the coefficients on the latter two variables to be positive and negative on the unemployment rate.

For commercial customers, we considered the Ontario unemployment rate, the Greater Toronto Area (GTA) commercial vacancy rate and Ontario GDP (the latter two provided by EGD). The coefficient on Ontario GDP is expected to be positive. Higher GTA vacancy rates signal a decline in commercial economic activity, so we expect the coefficient on this variable to be negative.

For residential customer characteristics, we considered the number of people per household and the number of households with school age kids in Ontario (from StatsCanada). Gas usage is expected to increase in line with the total number of persons in a household as well as with the presence of school age children. The expected coefficients on these variables are therefore positive.

PEG also considered a variant of the “vintage” variable that was used in EGD’s gas demand models. However, since it was necessary to use both EGD and Union data in the same regression, we needed to develop comparable vintage measures for both companies. It was not possible to replicate the methodology that EGD used to construct its vintage variable for Union. We therefore constructed a simplified vintage variable for both firms, which was calculated in each year as the number of residential customers in 1991 divided by the number of residential customers for the year. Declining values for this variable indicate a newer customer base, on average. This should be associated with

declines in gas usage due to the installation and use of more energy efficient gas-using equipment. The expected sign of the vintage coefficient is therefore positive.

For commercial customers, we included Union's segmentation index as an independent variable. It was also not possible to replicate this methodology for EGD given available data, and in this instance there was no straightforward alternative that could be constructed for both companies. We therefore assumed that EGD had the same "segmentation" values as Union in the sample years.

In addition, for the residential second stage regression, we included a variable that reflected the extent of EcoEnergy residential DSM programs in 2007 and 2008. We used the number of post-retrofit evaluations that occurred under the EcoEnergy programs. These are the evaluations that trigger a government DSM grant. Because these programs are intended to reduce energy consumption, the expected sign of the EcoEnergy variable is negative.

5.2 Econometric Methods

Our estimation procedures tested for autoregressive conditional heteroskedasticity (ARCH) and serial correlation. If we could not reject the hypothesis of either ARCH or serial correlation, we used generalized least squares (GLS) procedures to address the problems. GLS estimates will lead to more efficient estimates and more accurate inferences on whether a given variable has a statistically significant effect on (changes in) gas consumption.

ARCH arises when the variance of an error term in a given period is a function of the variance of previous periods. For example, in the model $y_t = \beta_0 + \beta_1 x_t + \mu_t$, the variance of the error term in period t can be dependent on the variance of the error in period $t-1$ and given by $E(\mu_t^2) = \alpha_0 + \alpha_1 \mu_{t-1}^2$. We test for ARCH by running a least square regression of the squared residual errors on the squared residual error in the previous time period. The Lagrange multiplier statistic computed from this regression is $(t-1) \cdot R^2$, has a chi-square distribution with one degree of freedom and can be used to test the null

hypothesis of no ARCH. If we cannot reject the null, we re-estimate the model using a GLS procedure that corrects for unequal variances across observations.²⁰

Serial correlation arises when the error terms of different time periods are correlated. The most common form of serial correlation in error terms is an autoregressive process of order 1 (AR(1)), where the error term in period t is given by $\mu_t = \rho\mu_{t-1} + e_t$. We test for the presence of AR(1) using the Durbin-Watson statistic. If we cannot reject the null of no serial correlation, then we use a Prais-Winsten procedure to estimate the degree of serial correlation and transform the data to correct for AR (1).²¹

5.3 Econometric Results

5.3.1 New Gas Demand Models

We begin by summarizing our two-stage results. The results from the first stage regressions are presented in Tables One through Eight. Tables One through Three present results for Enbridge Revenue Classes 20, 12, and 48, respectively. Tables Four through Eight present results for Union Revenue Classes 01 Residential, 01 Commercial, M2 Residential, M2 Commercial, and Commercial class 10, respectively. Table Nine presents second-stage regression results for residential gas customers in Ontario, and Table Ten presents second-stage regression results for commercial gas customers in Ontario.

²⁰ More precisely, we obtain initial OLS parameter estimates and estimates of residual variances for each observation, and transform both the dependent and independent matrices by multiplying them by the variance associated with the observation divided by the sum of the variances across all observations.

²¹ This procedure uses OLS to obtain an estimate of ρ , by regressing the error term on the error of the previous period, transforming the data by multiplying the dependent (y) and independent (x) variables by $(1 - \hat{\rho}^2)^{1/2}$ y_1 in the first period and by $(y_t - \rho y_{t-1})$ and $(x_t - \rho x_{t-1})$ in all subsequent periods.

Table 1

First Stage Regression: Average Gas Use Per Customer Enbridge Revenue Class 20

VARIABLE KEY

HDD= Heating Degree Days for Revenue Class 20

P= Residential Total Delivery Price for Revenue Class 20

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
HDD	0.288	189.775	P	-0.129	-2.826
Constant	5.672	453.120	Trend	-0.005	-2.512

Other Results

System Rbar-Squared	0.721
Durbin-Watson Statistic	1.783
F Statistic	139.976
Sample Period	1991-2008
Number of Observations	216

Table 2

First Stage Regression: Average Gas Use Per Customer Enbridge Revenue Class 12

VARIABLE KEY

HDD= Heating Degree Days for Revenue Class 12

P= Commercial Total Delivery Price for Revenue Class 12

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
HDD	0.278	143.320	P	-0.263	-6.947
Constant	8.842	894.565	Trend	0.031	15.107

Other Results

System Rbar-Squared	0.686
Durbin-Watson Statistic	2.077
F Statistic	118.41
Sample Period	1991-2008
Number of Observations	216

Table 3

First Stage Regression: Average Gas Use Per Customer Enbridge Revenue Class 48

VARIABLE KEY

HDD= Heating Degree Days for Revenue Class 48

P= Commercial Total Delivery Price for Revenue Class 48

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
HDD	0.344	355.379	P	-0.055	-2.812
Constant	7.324	889.740	Trend	0.012	11.511

Other Results

System Rbar-Squared	0.753
Durbin-Watson Statistic	1.818
F Statistic	164.96
Sample Period	1991-2008
Number of Observations	216

Table 4

First Stage Regression: Average Gas Use Per Customer Union Revenue Class 01 Residential

VARIABLE KEY

HDD= Heating Degree Days for Northern Region

P= Total Delivery Price for Revenue Class 01 Residential

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
HDD	0.556	79.588	P	-0.629	-17.665
Constant	5.385	362.058	Trend	0.010	6.099

Other Results

System Rbar-Squared	0.917
Durbin-Watson Statistic	1.764
F Statistic	591.97
Sample Period	1991-2008
Number of Observations	216

Table 5

First Stage Regression: Average Gas Use Per Customer Union Revenue Class 01 Commercial

VARIABLE KEY

HDD= Heating Degree Days for Northern Region

P= Total Delivery Price for Revenue Class 01 Commercial

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
HDD	0.683	82.750	P	-0.604	-21.069
Constant	6.561	831.924	Trend	0.012	9.109

Other Results

System Rbar-Squared	0.881
Durbin-Watson Statistic	1.629
F Statistic	398.48
Sample Period	1991-2008
Number of Observations	216

Table 6

First Stage Regression: Average Gas Use Per Customer Union Revenue Class M2 Residential

VARIABLE KEY

HDD= Heating Degree Days for Southern Region

P= Total Delivery Price for Revenue Class M2 Residential

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
HDD	0.296	87.312	P	-1.325	-48.956
Constant	5.138	408.929	Trend	0.037	25.273

Other Results

System Rbar-Squared	0.822
Durbin-Watson Statistic	1.892
F Statistic	248.52
Sample Period	1991-2008
Number of Observations	216

Table 7

First Stage Regression: Average Gas Use Per Customer Union Revenue Class M2 Commercial

VARIABLE KEY

HDD= Heating Degree Days for Southern Region

P= Total Delivery Price for Revenue Class M2 Commercial

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
HDD	0.406	169.847	P	-0.058	-58.389
Constant	7.240	1419.319	Trend	0.016	17.251

Other Results

System Rbar-Squared	0.802
Durbin-Watson Statistic	1.575
F Statistic	218.74
Sample Period	1991-2008
Number of Observations	216

Table 8

First Stage Regression: Average Gas Use Per Customer Union Revenue Class 10

VARIABLE KEY

HDD= Heating Degree Days for Northern Region

P= Total Delivery Price for Revenue Class 10

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
HDD	0.595	252.715	P	-0.248	-22.629
Constant	9.035	1383.267			

Other Results

System Rbar-Squared	0.873
Durbin-Watson Statistic	1.647
F Statistic	370.45
Sample Period	1991-2008
Number of Observations	216

Beginning with the first-stage results, it can be seen that the results are generally sensible. In every regression, the coefficient on HDD is positive and the coefficient on price is negative, as expected. Both estimates are also highly significant (at the 1% level) in every regression. We also included a trend term in seven of the eight regressions. While this term was also statistically significant, in most regressions it was positive, which was a surprising and perhaps counterintuitive (but nevertheless robust) result.

It is also notable that the R^2 values in the Union first-stage regressions are quite high and larger than those for comparable EGD regressions. The R^2 values in the Union regressions ranged from .802 to .917, which shows that from 80% to just over 90% of the variation in Union's monthly gas sales can be accounted for by changes in HDD, delivered natural gas prices, and the temporal trend in consumption. R^2 values in the EGD regressions range from 0.686 to 0.753, showing that approximately 70% of the variation in EGD monthly gas sales for these revenue classes can be attributed to changes in HDD, natural gas prices, and the trend.

We explored a variety of second-stage regressions, with varying results and degrees of success. It would be more distracting than illuminating to present the full range of these results in this report. In Tables Nine and Ten, we present what we believe are the most sensible and “best” regression results for Ontario residential and commercial customers, respectively.

The first column in Table Nine presents the fixed effects for the three revenue classes in the residential customer equation. All three are negative, statistically significant at the 5% level, and have similar magnitudes. Recall that the dependent variable in the second stage regression is gas consumption, net of changes due to HDD and prices. These negative fixed effect terms show there are significant reductions in residential gas consumption that are not due to weather, prices or any of the variables included in the second-stage regression. These reductions could reflect, in whole or in part, energy conservation actions that customers are undertaking at their own initiative. It is also interesting that the magnitudes of these fixed effects are almost identical for EGD revenue class 20, Union's 01 residential customers, and Union's M2 residential customers. This may suggest that such “free rider” effects tend to be fairly uniform across EGD and Union residential customers.

Table 9

Second Stage Regression: Change in 'Normalized' Gas Use Residential Revenue Classes

VARIABLE KEY

ID1= Constant for Enbridge Revenue Class 20
 ID2= Constant for Union Revenue Class 01 Residential
 ID3= Constant for Union Revenue Class M2 Residential
 ECOE= EcoEnergy dummy variable
 DSM= DSM Spending in previous year
 UR= Unemployment Rate
 VIN= Customer Vintage
 NPHH= Number of People per Household

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
ID1	-2.641	-2.545	ECOE	-0.047	-4.315
ID2	-2.494	-2.509	DSM	-0.009	-0.496
ID3	-2.576	-2.510	UR	-0.005	-3.223
			VIN	0.259	2.026
			NPHH	2.385	2.550

Other Results

System Rbar-Squared	0.388
Durbin-Watson Statistic	2.86
F Statistic	3.18
Sample Period	1999-2008
Number of Observations	30

Table 10

Second Stage Regression: Change in 'Normalized' Gas Use Commercial Revenue Classes

VARIABLE KEY

ID1= Constant for Enbridge Rate Class 6
 ID2= Constant for Union Revenue Class 01 Commercial
 ID3= Constant for Union Revenue Class M2 Commercial
 ID4= Constant for Union Revenue Class 10
 BUC= 2005, 2006 dummy variables
 DSM= DSM Spending in previous year
 SEGM= Segmentation index
 CVR= Commercial Vacancy Rate

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
ID1	-0.402	-1.117	BUC	-0.001	-0.188
ID2	-0.315	-0.831	DSM	0.017	1.320
ID3	-0.331	-0.877	SEGM	0.343	0.900
ID4	-0.083	-0.876	CVR	-0.006	-3.307

Other Results

System Rbar-Squared	0.188
Durbin-Watson Statistic	2.354
F Statistic	2.00
Sample Period	1999-2008
Number of Observations	40

Regarding the other variables, it can be seen that the EcoEnergy variable is negative and significant. This has the expected sign, and it provides evidence that the EcoEnergy programs in 2007-2008 have led to a significant decline in residential gas usage, independent of other energy conservation activities. The unemployment rate has the expected negative sign and is highly significant, showing that increases in unemployment are correlated with declines in residential gas usage. The vintage variable has the expected positive sign and is significant at the 5% level. This provides evidence that relatively newer “vintages” of customers and associated gas-using equipment are associated with declines in residential gas usage. The coefficient on the number of people per household is also positive and significant, showing that changes in the composition of residential households also have a significant impact on gas usage.

The main coefficient of interest in this regression is on the DSM variable. It is negative but not statistically significant. This result was robust in all of the specifications we explored. PEG was never able to identify a statistically significant relationship between changes in residential gas consumption and DSM spending in the previous year.

Finally, we note that the R^2 value in the residential regression was 0.388. This may appear relatively low, but recall that the first stage residential regressions explained more than 80% of the variation in residential gas consumption. This regression focuses on the approximately 20% of changes in gas consumption that were not explained in the first stage. The R^2 for this regression shows that about 40% of gas consumption that cannot be explained by HDD, prices and the secular trend is explained by the variables in this model.

Table 10 presents results on the commercial second stage regression. The first column presents results on the fixed effects for the individual revenue classes. Each of these fixed effects is negative, which is consistent with intuition, but none are statistically significant. In fact, the only variable in this regression that is statistically significant is the commercial vacancy rate (it also has the expected negative sign). The DSM variable is insignificant and, as in the residential regressions, this result is robust; PEG was never able to identify a statistically significant relationship between changes in commercial gas consumption and DSM spending in the previous year.

5.3.2 Econometric Results Using Monthly DSM Data

PEG also experimented with regressions using monthly observations on DSM expenditures as an independent variable. As discussed in Section 4.2, there are acknowledged problems with the monthly DSM expenditure data. In fact, EGD and Union monthly data on DSM expenditures for different revenue classes are spotty, at best. There can also be discrepancies between when DSM costs are incurred and when actual DSM measures are installed. For these and related reasons, the Companies cautioned against the use of monthly DSM expenditures.

While it is important to keep these limitations on data quality in mind, there may also be some value in exploring the use of monthly DSM expenditures in gas demand modeling. The use of monthly DSM data will greatly expand sample sizes and increase the likelihood of obtaining statistically significant results. Monthly DSM data can also be added as an explanatory variable to existing demand models that are already being used by EGD and Union. This approach is therefore a straightforward extension of the gas demand work that is already presented in OEB proceedings. While the concerns about data quality reduce the reliability of any results based on these data, these results may still provide some indicative - but not definitive - evidence on the relationship between DSM spending and gas consumption for different revenue classes.

PEG therefore added DSM as an explanatory variable to variants of the gas demand models that are currently used by EGD and Union Gas. We also added the EcoEnergy variable to these models, since these programs may have had a significant impact on residential gas consumption in 2007-08 but have not been included in either of the Companies' models to date. PEG retained the estimation procedure that we used in earlier regressions, since this corrects for ARCH and serial correlation, and it is important to correct for these influences to obtain the most efficient statistical estimates and improve statistical inference. PEG's estimation procedure therefore differs from that used by Union and EGD.²²

These econometric results are presented in Tables 11 through 18. Tables 11 through 13 show the results for the three EGD revenue classes. In all cases, the

coefficients on price and HDD have the expected sign and are statistically significant. The EcoEnergy variable also has the expected negative sign and is significant in the residential regression (revenue class 20). Two variables are statistically significant but have the incorrect, or unexpected, sign: the furnace efficiency variable in revenue class 20, and the commercial vacancy rate in revenue class 48. For two of the three revenue classes (residential class 20 and commercial class 12), the coefficients on DSM variable have a negative sign and are statistically significant; the estimate on DSM is not significant for revenue class 48. The values of the two, statistically significant coefficients are -0.105 and -0.084, which indicates that a 1% increase in DSM expenditures will be associated with contemporaneous declines in gas consumption of 0.105% for revenue class 20 and 0.084% for revenue class 12.

The econometric results for Union's five revenue classes are presented in Tables 14 through 18. In all cases, the coefficient on price is negative and statistically significant and, in nearly every instance, the coefficient on Union's monthly HDD variable is positive and significant. The EcoEnergy variable is not significant in either of the two residential revenue class regressions. Two variables also have an unexpected sign and are statistically significant: number of people per household (for M2 residential customers), and the segmentation index (for commercial revenue class 10). The coefficient on DSM is negative and statistically significant on both of the residential revenue classes and for one of the three commercial revenue classes (Commercial 01 customers). These DSM coefficients are -0.077 for the 01 Residential Class, -0.056 for the M2 Residential Class, and -0.034 for 01 Commercial Customers. These values indicate that a 1% increase in DSM expenditures will be associated with contemporaneous declines in gas consumption of 0.077% for Residential 01 customers, 0.056% for Residential M2 customers, and 0.034% for Commercial 01 customers.

²² For reasons that were explained in Chapter Three, we also substituted the delivered price of natural gas for the total bill in the Union regressions that used total bill as an explanatory variable.

Table 11

Alternate Regression: Monthly DSM Data Enbridge Revenue Class 20

VARIABLE KEY

P= Total Delivery Price for Revenue Class 20
HDD= Heating Degree Days for Revenue Class 20
FE= Furnace Efficiency Index
ECOE= Eco Energy dummy variable
DSM= DSM Cost for Rate Class 1

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
P	-0.172	-1.596	ECOE	-0.023	-5.266
HDD	0.335	34.769	DSM	-0.105	-18.717
FE	-0.328	-11.045			
Constant	5.691	74.654	Trend	0.027	4.112

Other Results

System Rbar-Squared	0.735
Durbin-Watson Statistic	1.811
F Statistic	52.69
Sample Period	1998-2008
Number of Observations	132

Table 12

Alternate Regression: Monthly DSM Data Enbridge Revenue Class 12

VARIABLE KEY

P= Total Delivery Price for Revenue Class 12
HDD= Heating Degree Days for Revenue Class 12
BUC1= Building Code 2005 dummy variable
BUC2= Building Code 2006 dummy variable
RM= Rate Migration dummy variable
DSM= DSM Cost for Rate Class 6

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
P	-0.263	-27.714	BUC2	-0.015	-10.238
HDD	0.251	134.269	RM	-0.005	-5.012
BUC1	-0.015	-4.370	DSM	-0.084	-5.232
Constant	8.237	174.744	Trend	0.083	18.350

Other Results

System Rbar-Squared	0.68
Durbin-Watson Statistic	1.792
F Statistic	35.69
Sample Period	1998-2008
Number of Observations	132

Table 13

Alternate Regression: Monthly DSM Data Enbridge Revenue Class 48

VARIABLE KEY

P= Total Delivery Price for Revenue Class 48
HDD= Heating Degree Days for Revenue Class 48
ONTGDP= Ontario GDP
CVR= GTA Commercial Vacancy Rate
DSM= DSM Cost for Rate Class 6
Trend= Time Trend

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
P	-0.684	-4.342	CVR	0.137	1.815
HDD	0.301	47.370	DSM	-0.021	-1.176
ONTGDP	2.266	4.250	Trend	0.003	0.169
Constant	7.328	45.301			

Other Results

System Rbar-Squared	0.734
Durbin-Watson Statistic	1.702
F Statistic	52.63
Sample Period	1998-2008
Number of Observations	132

Table 14

Alternate Regression: Monthly DSM Data Union Revenue Class 01 Residential

VARIABLE KEY

P= Total Delivery Price for Revenue Class 01 Residential
HDD1= January Heating Degree Days for Northern Region
HDD2= February Heating Degree Days for Northern Region
HDD3= March Heating Degree Days for Northern Region
HDD4= April Heating Degree Days for Northern Region
HDD5= May Heating Degree Days for Northern Region
HDD9= September Heating Degree Days for Northern Region
HDD10= October Heating Degree Days for Northern Region
HDD11= November Heating Degree Days for Northern Region
HDD12= December Heating Degree Days for Northern Region
NPHH= Number of Persons Per Household
FE= Furnace Efficiency Index
ECOE= Eco Energy dummy variable
DSM= DSM Cost for Rate 01Residential

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
P	-0.339	-16.540	HDD10	0.050	22.368
HDD1	0.131	50.066	HDD11	0.102	42.915
HDD2	0.130	54.414	HDD12	0.136	60.714
HDD3	0.117	49.760	NPHH	-0.107	-0.885
HDD4	0.080	38.278	FE	0.155	7.748
HDD5	0.038	16.079	ECOE	0.001	0.900
HDD9	0.010	3.217	DSM	-0.077	-10.367
Constant	4.124	1563.138			

Other Results

System Rbar-Squared	0.973
Durbin-Watson Statistic	1.893
F Statistic	318.58
Sample Period	1998-2008
Number of Observations	132

Table 15

Alternate Regression: Monthly DSM Data Union Revenue Class M2 Residential

VARIABLE KEY

P= Total Delivery Price for Revenue Class M2 Residential
HDD1= January Heating Degree Days for Southern Region
HDD2= February Heating Degree Days for Southern Region
HDD3= March Heating Degree Days for Southern Region
HDD4= April Heating Degree Days for Southern Region
HDD5= May Heating Degree Days for Southern Region
HDD9= September Heating Degree Days for Southern Region
HDD10= October Heating Degree Days for Southern Region
HDD11= November Heating Degree Days for Southern Region
HDD12= December Heating Degree Days for Southern Region
NPHH= Number of Persons Per Household
FE= Furnace Efficiency Index
ECOE= Eco Energy dummy variable
DSM= DSM Cost for Rate M2 Residential

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
P	-0.214	-10.319	HDD10	0.053	29.520
HDD1	0.137	43.947	HDD11	0.096	61.445
HDD2	0.138	81.708	HDD12	0.142	109.293
HDD3	0.126	69.459	NPHH	-0.038	-0.204
HDD4	0.084	53.539	FE	0.112	9.645
HDD5	0.037	28.495	ECOE	-0.001	-1.515
HDD9	-0.007	-4.970	DSM	-0.056	-8.587
Constant	4.129	773.148			

Other Results

System Rbar-Squared	0.984
Durbin-Watson Statistic	1.684
F Statistic	544.56
Sample Period	1998-2008
Number of Observations	132

Table 16

Alternate Regression: Monthly DSM Data Union Revenue Class 01 Commercial

VARIABLE KEY

P= Total Delivery Price for Revenue Class 01 Commercial
HDD1= January Heating Degree Days for Northern Region
HDD2= February Heating Degree Days for Northern Region
HDD3= March Heating Degree Days for Northern Region
HDD4= April Heating Degree Days for Northern Region
HDD5= May Heating Degree Days for Northern Region
HDD9= September Heating Degree Days for Northern Region
HDD10= October Heating Degree Days for Northern Region
HDD11= November Heating Degree Days for Northern Region
HDD12= December Heating Degree Days for Northern Region
SEGM= Segmentation Index
DSM= DSM Cost for Rate 01 Commercial

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
P	-0.125	-3.948	HDD9	0.009	5.950
HDD1	0.181	91.066	HDD10	0.095	71.405
HDD2	0.175	145.277	HDD11	0.135	79.792
HDD3	0.162	274.763	HDD12	0.174	131.892
HDD4	0.114	129.495	SEGM	0.164	0.949
HDD5	0.060	78.135	DSM	-0.034	-3.825
Constant	5.100	736.159			

Other Results

System Rbar-Squared	0.956
Durbin-Watson Statistic	1.991
F Statistic	221.43
Sample Period	1998-2008
Number of Observations	132

Table 17

Alternate Regression: Monthly DSM Data Union Revenue Class M2 Commercial

VARIABLE KEY

P= Total Delivery Price for Revenue Class M2 Commercial
HDD1= January Heating Degree Days for Southern Region
HDD2= February Heating Degree Days for Southern Region
HDD3= March Heating Degree Days for Southern Region
HDD4= April Heating Degree Days for Southern Region
HDD5= May Heating Degree Days for Southern Region
HDD9= September Heating Degree Days for Southern Region
HDD10= October Heating Degree Days for Southern Region
HDD11= November Heating Degree Days for Southern Region
HDD12= December Heating Degree Days for Southern Region
SEGM= Segmentation Index
DSM= DSM Cost for Rate M2 Commercial

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
P	-0.128	-6.992	HDD9	0.014	22.459
HDD1	0.163	114.384	HDD10	0.082	219.577
HDD2	0.160	102.119	HDD11	0.129	113.005
HDD3	0.150	221.101	HDD12	0.158	198.070
HDD4	0.106	98.564	SEGM	0.515	4.625
HDD5	0.060	143.988	DSM	-0.002	-0.268
Constant	5.949	1229.972			

Other Results

System Rbar-Squared	0.968
Durbin-Watson Statistic	1.891
F Statistic	309.49
Sample Period	1998-2008
Number of Observations	132

Table 18

Alternate Regression: Monthly DSM Data Union Revenue Class 10 Commercial

VARIABLE KEY

P= Total Delivery Price for Revenue Class 10 Commercial
HDD1= January Heating Degree Days for Northern Region
HDD2= February Heating Degree Days for Northern Region
HDD3= March Heating Degree Days for Northern Region
HDD4= April Heating Degree Days for Northern Region
HDD5= May Heating Degree Days for Northern Region
HDD9= September Heating Degree Days for Northern Region
HDD10= October Heating Degree Days for Northern Region
HDD11= November Heating Degree Days for Northern Region
HDD12= December Heating Degree Days for Northern Region
SEGM= Segmentation Index
DSM= DSM Cost for Rate 10 Commercial

EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC	EXPLANATORY VARIABLE	PARAMETER ESTIMATE	T-STATISTIC
P	-0.070	-3.425	HDD9	-0.002	-0.504
HDD1	0.152	124.821	HDD10	0.083	105.458
HDD2	0.140	453.624	HDD11	0.117	458.747
HDD3	0.137	175.920	HDD12	0.136	319.824
HDD4	0.090	57.349	SEGM	-0.208	-2.139
HDD5	0.043	61.816	DSM	-0.003	-0.467
Constant	7.842	3622.985			

Other Results

System Rbar-Squared	0.94
Durbin-Watson Statistic	1.825
F Statistic	157.93
Sample Period	1998-2008
Number of Observations	132

Again, these results must be interpreted cautiously in light of the problems with the monthly DSM expenditure data. However, they do provide some indicative evidence on the impact of DSM spending on changes in gas consumption for several revenue classes for EGD and Union Gas. This information could potentially be used as supplementary, or supporting, evidence in DSM proceedings in Ontario, but it is certainly not definitive enough to substitute for the bottom-up approach that is currently used in the Province.

5.3.3 Differences Between Actual and Predicted Consumption

Finally, we briefly discuss the results from PEG's third approach towards top-down M&V of gas savings from utility DSM programs. This approach constructed gas demand models for each of the eight revenue classes, generated predicted values and 95% confidence intervals for gas usage for each revenue class in each year, and compared this to actual gas consumption. None of these gas demand models included DSM spending as an explanatory variable. If actual gas consumption was below the predicted value, and outside the confidence interval, this could provide more indirect evidence of the impact of DSM spending on gas consumption.

PEG investigated dozens of such models, and none of them identified a year in which actual gas usage was below the predicted value and outside of the confidence intervals. Thus, this approach did not identify any negative and statistically significant differences between actual and predicted gas consumption. Like the main approach detailed on Tables One through Ten, this secondary approach therefore does not provide "top down" evidence that can be used to substitute for the bottom-up methods currently used in Ontario's gas DSM programs. Because the econometric results from the secondary approach add little or nothing of value to Tables One through Ten, they are not presented in this report.

6. CONCLUSION

This project attempted to develop “top down” estimates of gas savings from utility DSM programs by applying econometric methods to the aggregate billing data of EGD and Union Gas. This is an approach that can lead to substantial benefits, although it has rarely (if ever) been used to measure energy savings in approved DSM programs. PEG carefully examined the economic literature and developed a number of econometric models and techniques that we believed were appropriate and which could be feasibly implemented using available data in Ontario.

However, our efforts were largely unsuccessful. PEG explored scores of econometric models and variants of our preferred “top down” econometric specification, which used monthly values of HDD and prices but annual values for DSM expenditures and other economic and customer characteristic variables. None of them produced results that were suitable for generating “top down” estimates of gas savings from DSM programs that could substitute, in whole or part, for the M&V methods currently used in the Province.

The results using monthly DSM spending as an explanatory variable in updated Company demand models were more promising. Notwithstanding the acknowledged deficiencies of the monthly DSM expenditure data, we found a statistically significant and negative relationship between DSM spending and gas consumption for all residential revenue classes and for two of the five commercial revenue classes for EGD and Union.

Our models show that a 10% increase in DSM spending for residential customers will lead to a 0.6% to 1.0% decline in gas consumption. For commercial customers, our models show that a 10% increase in gas DSM spending will lead to a 0.3% to 0.8% decline in gas consumption.

Overall, PEG’s research did not provide any “top down” evidence that is definitive enough to substitute for the bottom-up methods currently used in Ontario’s gas DSM programs. Our strongest results came from integrating DSM spending into variants

of the gas demand models the Companies currently use to forecast gas demand. Monthly data on gas DSM spending are unreliable, however, so these results can at best provide supplementary or supporting evidence on the impact of DSM programs on gas consumption. Our econometric models that used more reliable measures of gas DSM spending were never able to identify a significant relationship between DSM activity and gas consumption.

PEG's analysis could likely be improved if better data were available. One improvement would be more accurate data on DSM spending by revenue class and (for EGD) geographic zone. It could also be helpful to have information on when (in a given year) particular DSM measures were installed, in addition to having more accurate data on DSM spending.

More appropriate estimates of DSM savings could also be developed if demand models are estimated separately for participating and non-participating customers. A relatively small share of customers in a revenue class is likely to be participating in utility DSM programs in any given year. The behavioral characteristics of participating and non-participating customers may be so different that they effectively constitute different populations with, accordingly, different underlying demands for natural gas. However, developing detailed customer-specific data would likely entail significant costs, and it would take years for enough sample data to be available to facilitate statistical analysis. There is also no guarantee that this approach will be successful and yield statistically significant and robust results.

APPENDIX: CALIFORNIA EXPERIENCE WITH ECONOMETRIC M&V

California is clearly the leading DSM jurisdiction in North America. California has implemented utility conservation programs more or less continuously since the 1970s, which is the longest DSM experience in North America. California's programs also tend to be sizable and include ample budgets devoted to M&V. Even in 2007 (the last available data), California's electric utilities' accounted for over half of spending on electricity energy efficiency programs by all investor-owned utilities in the US.²³ California is also widely acknowledged to be a leading jurisdiction in technical matters, including innovations used in M&V. Many jurisdictions incorporate these techniques and/or designs into their own DSM programs. Because of its leading position and rich DSM history, California's M&V experience is especially relevant for this project.

California's energy utilities have administered DSM programs since the 1970s, with shareholder incentive mechanisms first approved in the late 1980s. In June 1990, the Division of Ratepayer Advocates (DRA, later the Office of Ratepayer Advocates or ORA) of the California Public Utilities Commission (CPUC) was first provided a budget to review utilities' DSM programs. Energy savings from the 1970s through the early 1990s were mainly calculated using engineering methods, similar to the "bottom up" approach currently used in Ontario's gas DSM framework. Because these savings were projected in advance of the programs – rather than calculated after the programs had been in effect – these were often referred to as "ex ante" estimates.

This approach changed significantly in 1993. In Decision D.93-05-063, the CPUC approved the *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*. These protocols were jointly developed by the four main investor-owned utilities in the State (Pacific Gas and Electric (PG&E), San Diego Gas and Electric (SDG&E), Southern California Edison (SCE), and Southern California Gas (SoCalGas)), the ORA, the Natural Resources

²³ Data from the US Energy Information Administration, EIA Form 861. In 2007, spending on electricity energy efficiency programs by Pacific Gas & Electric, Southern California Electric, and San

Defense Council (NRDC), and the California Energy Commission (CEC), a State government agency responsible for energy planning and promoting energy efficiency. The 1993 Decision also established the California DSM Measurement Advisory Committee (CADMAC, later renamed the California Measurement Advisory Committee or CALMAC) to develop these protocols. Collectively, these institutional changes launched what is sometimes referred to as “the Protocol Era.”

The Protocol era created greater consistency and rigor in how the costs and benefits of DSM programs were measured, as well as in how shareholder incentive mechanisms were implemented. A consolidated annual hearing was held for the four utilities to evaluate earnings claims based on DSM programs.²⁴ Importantly, the Protocols changed the dominant “ex ante” method for measuring savings to an “ex post” evaluation of the energy savings actually achieved by the programs. These ex post evaluations were to be conducted using regression analysis and utilities’ actual billing data.

The Protocols established what was called “The General Approach to Load Impact Measurement.” This general approach was to be applied to specific utility DSM programs, which is more program-focused than the more aggregated “top down” methods proposed in Ontario for, say, entire customer classes. The Protocols also distinguished between the estimation of gross and net energy impacts (where those terms are analogous to how they are used in Ontario).

Regarding the estimation of gross energy impacts, the Protocols stated:

The statistical estimation of gross energy impacts requires billing data and explains changes in energy use as a function of other variables in order to estimate the gross load impacts attributable to a DSM program. A variety of model types - including conditional demand analysis (CDA), statistically adjusted engineering (SAE), fixed effects, and other linear and nonlinear regression models - may qualify as acceptable load impact regression models (LIRMs), depending upon the circumstances.

The LIRMs used to estimate gross energy savings should have the following characteristics:

Diego Gas & Electric totaled \$351.2 million; the comparable figure for all US investor-owned utilities was \$690 million. Unfortunately, the EIA does not provide data on DSM programs for gas utilities.

²⁴ This was known as the Annual Earnings Assessment Proceeding, or AEAP.

- a. The model is an econometric or statistical model, embodying accepted or thoroughly defensible empirical techniques for measuring impacts of policies, programs and measures.
- b. The models employ billing and weather data, pooled by customer, for multiple time periods, as well as customer-specific attributes, and/or other measured or observed data to estimate energy impacts.
- c. The model produces diagnostics and test statistics that allow others to assess the robustness of its estimates and/or simulations.
- d. The model specification is developed in consideration of the issues identified in the Protocols in Section D.5 of Table 7. That is, the model specification should follow from an accurate conceptualization of the energy consumption process, and should use compatible econometric and statistical techniques. *The estimates of energy impacts should flow from a statistical model rather than a deterministic engineering model*, while perhaps relying to some extent on engineering information. For example, an SAE model exhibits an acceptable blend of statistical and engineering models. Confounding effects on energy consumption should be controlled for. The use of a comparison group and the inclusion of social, political and economic changes, are acceptable methods (emphasis added).²⁵

It can be seen that the protocols clearly called for an econometric, ex post approach to measuring the load impacts and benefits from utility DSM programs. This represents a clear and explicit change in direction from the engineering-based, ex ante approach that had been mainly used up to that time. Engineering estimates could be incorporated as inputs, or independent variables, into statistical models, but billing data and econometric methods were both primary and mandatory. The statistical models were also required to control for a variety of other factors that could impact gas and electricity demand (“confounding effects on energy consumption should be controlled for”), in order to isolate the impact of DSM programs *per se*.

One allowable approach for identifying the gross impact of DSM on consumption is the use of “comparison groups.” In practice, this often took the form of comparing the energy consumption of customers after they installed DSM measures with those same customers’ energy consumption before the measures were installed. For example, monthly data on participating customers’ pre-DSM consumption could be regressed on

²⁵ D. 93-05-063, Table 5, p. 12.

weather, price and other economic variables that affect demand. Normalized measures for participating customers' consumption could then be computed to the post-DSM period by inserting values for these independent variables into the estimated gas demand model.²⁶ The difference between pre- and post-DSM normalized consumption could then be calculated. Since other factors that could impact consumption were controlled for in the statistical model, the difference between pre- and post-DSM energy consumption could be interpreted as the gross energy impact resulting from the conservation measures.

The Protocols specified even more detailed rules regarding the estimation of net energy impacts, as evident in the following:

The estimation of net energy impacts can also involve the use of a statistical model that does not use energy consumption as the dependent variable but rather uses the observed decisions of customers to participate in DSM programs and to install efficient equipment as the dependent variables. The purpose of these models is to control for free ridership or to derive a net-to-gross savings adjustment. The models may also be used to estimate an adjustment factor to control for self-selection bias.

The LIRMs used to estimate net energy savings should have the following characteristics:

- a. The model is an econometric or statistical model, embodying accepted or thoroughly defensible empirical techniques for measuring impacts of policies, programs and measures.
- b. The model utilizes comparisons between participants and nonparticipant behavior in a discrete choice, difference-of-differences, or other statistical modeling context to isolate net from gross load impacts.

The model produces diagnostics and test statistics that allow others to assess the robustness of its estimates and/or simulations.

3. If the methodology involves comparing participants and nonparticipants with respect to energy consumption, then the following framework can be used:

$$\text{Net Load Impacts} = \text{Participant Group Load Impacts}$$

²⁶ The Protocols refer to these types of econometric models as “conditional demand analysis” or CDA models, since customer demand is “conditional” on a variety of factors. Since controlling for such independent variables is standard in good econometric practice, the CDA term used in the Protocols is synonymous with the term “gas demand model” that PEG uses throughout this report.

minus

Comparison Group Load Impacts

plus or minus

Effects of Uncontrolled Differences between Participant and Comparison Groups

OR

- Net load impacts = Participant Group Load Impacts minus Comparison Group Load Impacts (referred to as the difference of differences method);
- Participant Group Load Impacts = Participant Group Base Usage minus Participant Group Usage in the Impact Year
- Comparison Group Load Impacts = Comparison Group Base Usage minus Comparison Group Usage in the Impact Year
- Participant Group Base Usage = Participant Group Pre-Installation Usage
- Comparison Group Base Usage = Comparison Group Pre-Installation Usage
- Pre-Installation Usage (Participant and Comparison Group) = measured consumption or proxies for consumption of the energy using equipment or building prior to installation of the measure(s) intended to change energy use, adjusted (when applicable) to reflect the minimum efficiency level of the equipment or building that would have been installed without the utility assistance.²⁷

To isolate net impacts, the protocols require customer data for both participating and non-participating customers. The “difference of differences method” is one approach for controlling for the free ridership reflected in the estimated gross energy impacts. The basic idea is that environmental and behavioral factors that are independent of utility DSM programs can encourage customers to adopt energy conservation measures. These factors would likely be reflected in the normalized energy consumption patterns for non-participating customers (and certainly, any reductions in normalized energy consumption for non-participating customers cannot be the result of utility DSM programs). Hence, to isolate the *incremental* impact of DSM programs on energy savings, one could compute

²⁷ D. 93-05-063, Table 5, p. 13.

the difference between gross savings for participating customers and the changes in normalized energy consumption over the same period for non-participating customers. If the behavioral characteristics of the participating and non-participating customers are similar, this difference can be interpreted as the value of the net energy impacts that result from DSM programs *per se*. Alternatively, the “self-selection bias” can be quantified by explicitly modeling customers’ process for deciding whether or not to participate in utility DSM programs. Estimates from these “discrete choice models” can then be used to derive net-to-gross ratios that are used to transform gross energy impacts into net energy impacts.

The protocols also specified other rules in order to obtain the most accurate measures of net energy savings. Participants were defined as those who received financial assistance in connection with an energy conservation measure or received services under an appropriately authorized DSM program (such as an energy audit). Pre-installation usage was to be based on 12 months billing data before a measure was installed; usage in the initial “impact” year was based on a minimum of nine months of billing data.²⁸ In addition, there were detailed rules on how to determine appropriate samples for participating customers and controls designed to ensure data quality and processing.

These concepts (including alternative applications of the basic ideas) can be made more concrete by considering specific examples of CDA models used to measure energy savings in California. One early model applied to SCE’s Energy Management Services and Hardware Rebate Program Evaluation. This program actually took effect in 1990, before the Protocols required econometric M&V. The model used to evaluate energy savings was the following:²⁹

²⁸ In addition to the initial year impact study, utilities were required to undertake follow-up “persistence” studies done (usually in years four and nine after the measure was installed) to see whether the measure was retained and hence determine effective useful lives and long-run savings from specific energy conservation measures.

²⁹ For more details, see Pacific Consulting Services (1994), “An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Impacts,” prepared for the California Demand Side Management Advisory Committee: The Subcommittee on Modeling Standards for End Use Consumption and Load Impact Models.

$$E_{it} = \alpha + \beta_1 \text{INSTALL}_{it} + \beta_2 \text{PRICE}_{it} + \beta_3 \text{SQFT}_i + \beta_4 \text{HOURS}_i + \beta_5 \text{INFLATION}_t + \varepsilon_{it}$$

where

- E_{it} = electricity consumption for the i th customer at time t
- α = a constant term
- INSTALL_{it} = installation of the ECM by the i th customer at time t
- PRICE_{it} = price of electricity faced by the i th customer at time t
- SQFT_i = square footage of the i th customer
- HOURS_i = operating hours of the i th customer
- INFLATION_t = the rate of inflation at time t
- ε_{it} = the error terms for the i th customer at time t

This model is estimated using a panel data set, based on the monthly electricity consumption for a cross section of customers who have installed specific energy conservation measures. The data are therefore customer-specific, and the time subscript t references the month. The INSTALL variable is a “dummy variable” that takes a value of 0 before the efficiency measure is installed and a value of 1 afterwards. The coefficient estimated on the “ INSTALL ” variable can thus be interpreted as the gross impact of the energy conservation measure on an average customers’ consumption, independent of the other variables in the model.

A more complex variant of this model was used to evaluate a PG&E DSM program during the Protocol era. This model used both “pre” and “post” consumption data and a variable to reflect potential behavioral differences between the composition of participating and non-participating customers. The general form of this model was:

$$E_{i,\text{Post}} = \alpha + \beta_1 \text{INSTALL}_i + \beta_2 E_{i,\text{Pre}} + \beta_3 \text{Inverse Mills Ratio} + \sum \beta_k X_{ik} + \varepsilon_i$$

where

- $E_{i,\text{Post}}$ = electricity consumption for the i th customer after the energy efficiency measure was installed
- α = a constant term
- INSTALL_i = installation of the energy efficiency measure for the i th customer

- $E_{i,Pre}$ = electricity consumption for the i th customer prior to the installation of the energy efficiency measure
- Inverse Mills Ratio = a term designed to control for selection bias, reflecting differences in behavioral characteristics between participating and non-participating customers
- X = a vector of other economic variables, such as changes in the price of energy, square feet, operating hours, and the rate of inflation for the i th customer
- ε_i = the error term

This model reflects the use of pre- and post-installation consumption data for participating customers, consistent with the Protocols. It also includes a comparison group of non-participating customers. Because there are both participating and non-participating customers in the model, the INSTALL dummy variable now takes a value of zero for non-participating customers and a value of one for participating customers.

The “Inverse Mills Ratio” is an explanatory variable that is specifically designed to control for selection bias, or the fact that there may be behavioral differences between customers who “select” utility DSM programs and those who do not. It is important to control for these factors when isolating the impact of utility DSM programs *per se*. A good explanation for why this is the case is presented in an earlier article by Raymond Hartman:

In many evaluations, program-induced effects have been estimated by comparing the observed savings of program participants and non-participants. This comparison is appropriate *only if* participants and non-participants are identical in all respects except program participation. However, participants and nonparticipants usually differ in observed economic and demographic characteristics, which in turn induce differences in unobserved preferences for energy consumption *and* program participation. Attribution of the observed difference in energy savings to the programs *alone* ignores these other differences. The result can be an upwardly-biased estimate of program effectiveness because the demographic and economic characteristics of program participants would have induced some conservation *in the absence* of the programs (all italics in original).³⁰

³⁰ Hartman, R. (1988), “Self-Selection Bias in the Evolution of Energy Conservation Programs,” *The Review of Economics and Statistics*, 70: 3, p. 448.

As this explanation indicates, the “self selection” problem relates directly to the issue of “free riders” and the difference between the estimated gross and net energy impacts from DSM programs. Including an inverse Mills ratio is one means of controlling for customer characteristics that would tend to exaggerate the estimated impact of utility programs *per se*.³¹ However, there are other (econometric and sample design) methods for controlling for this type of selection bias. The best techniques for doing so, and therefore for computing net-to-gross ratios, remains a hotly contested issue.³²

The initial Protocol Era ran from 1994 through 1997, with new protocols established in 1998.³³ There were some similarities between the updated and initial protocols. Both used participant and comparison groups to determine gross and net savings, and both had carefully detailed rules for specific DSM programs. However, in the updated protocols, statistical and econometric methods were no longer mandatory for estimating gross energy impacts. Companies were now given the choice of estimating load impacts through conditional demand analysis (*i.e.* econometric modeling of energy demand) or a calibrated engineering (CE) model. The new protocols did not discuss why econometrics were no longer mandatory, but the head of the company overseeing the

³¹ The procedure for using the inverse Mills ratio to control for sample selection bias was recommended by Nobel Laureate Econometrician James Heckman. The process is: 1) run a Probit regression on individuals’ binary decisions to participate (*i.e.* 1 = participation, 0 = no participation) against a set of explanatory variables **Z**; 2) obtain the estimated coefficients Γ on **Z**, and compute the inverse Mills ratio as (the standard normal density function Γ^*Z) divided by (the cumulative normal density function Γ^*Z); 3) include the inverse Mills ratio as an additional variable in an ordinary least squares regression of dependent variable **Y** on a vector of independent variables **X**. For more details, see section 13.13.1 in Johnston and Dinardo, *Econometric Methods*.

³² A detailed review of this literature goes well beyond the scope of this project, and would probably not be fruitful in any case, since the econometric methods require customer specific data on participating and non-participating customers. PEG was told that neither EGD nor Union collects data at this level. However, interested parties can find a review of this literature in Cook, G. “Attribution Methodology Wars: Self-Report Methods Versus Statistical Number Crunching – Which Should Win?”

³³ Many other institutional changes also took place beginning in 1998, including more emphasis on funding for “market transformation” DSM programs and less scope for incentive-based awards for traditional utility DSM programs. There was also a significant increase in spending on M&V studies that attempted to quantify the market effects and indirect benefits attributable to market transformation programs. For further details, see TecMarket Works Framework Team, *The California Evaluation Manual*, Prepared for Southern California Edison, to conduct a joint study supported by Pacific Gas and Electric, San Diego Gas and Electric, Southern California Edison and Southern California Gas, as mandated by the California Public Utilities Commission; June 2004: p. 40.

current master contract for M&V programs in California indicated that an important consideration was the difficulty in controlling for customer characteristics and their impact on behavior in econometric studies.³⁴

In 2001, California adopted new Energy Efficiency Policy Rules in Decision D. 01-11-066. Some parts of this Decision reflect a further movement towards ex ante, engineering based estimates of energy impacts. For example, the Decision required utilities to create a Best Practices database that could be used to assist parties (including non-utilities) in how to design the most effective energy conservation programs. As part of this Best Practices database

“the Commission requires the utilities to undertake expert evaluation of “*ex ante*” (projected) or deemed savings estimates of energy savings associated with a set of reasonably predictable energy efficiency measures. Currently, the CEC’s Database for Energy Efficiency Resources (DEER) is the most comprehensive resource for program planners to use when projecting energy savings associated with particular program activities. This database, though updated periodically, is primarily for use by technical experts. In developing a set of deemed savings values for the state, the Commission seeks to simplify the assumptions used to project energy savings into a user-friendly format assessable to a wider audience. The goal of this effort would be to produce an Internet-accessible, searchable tool containing best-available deemed savings values for all regions of the state, grouped by sector, building type, end-use, and climate zone (where applicable).”³⁵

Although the purpose of this mandate is to assist program planners in advance (ex ante) of undertaking programs, it is notable that the “best practices” database relies overwhelmingly on engineering rather than econometric evidence. It is also notable that one of the aims of this effort was to “simplify the assumptions used to project energy savings” since this is also one of the objectives motivating Ontario’s interest in “top down” econometric M&V models. In California, where there had been considerable experience with econometric M&V, regulators decided to move in a different direction when attempting to simplify the estimation of energy savings.

In 2004, the CPUC mandated a study that led to the creation of *The California Evaluation Framework*. This was a detailed policy document which, among other things,

³⁴ Telephone conversation between Larry Kaufmann and Nick Hall, President of TecMarket Works, November 18, 2009.

³⁵ Decision 01-11-066, *Interim Opinion Adopting Energy Efficiency Policy Rules*, November 29 2001, p. 22.

contained a discussion of the merits of using econometric or engineering analyses to measure energy savings from DSM programs. According to the *Framework*:

Billing analysis (*i.e.* econometric analysis of billing data) will tend to be preferred when:

- Both pre and post-retrofit billing data are available
- Expected program impacts can be expected to be observed in a billing analysis (e.g., at least 10% of total consumption, depending upon method used, cleanliness of billing data, and accuracy of measured variables in analysis)
- The analysis is of a program with larger numbers of participants that are more homogenous

Engineering analysis will tend to be preferred when:

- No pre-measure billing data is available, e.g., new construction
- Expected impacts are too small to likely be observed in a billing analysis (e.g., less than 10% of total consumption)
- The programs have a small number of participants or unique measures, e.g., with industrial process improvements
- The programs have significant investments in engineering methods within the program that can provide cost savings for a similar evaluation, e.g., programs that include substantial engineering M&V or building energy simulation modeling³⁶

As this passage indicates, in the judgment of the *Framework* team, the California experience indicates that econometric methods are the preferred M&V approach when energy efficiency measures are expected to have relatively large (10% or more) impacts on consumption, and when pre- and post-installation data are available for relatively homogeneous customer groups. The “observed impact” criterion is no doubt motivated by the likelihood that, all else equal, it is easier to identify statistically significant impacts of variables when the impact of those variables tends to be large relative to random or non-quantifiable effects. Pre- and post-installation data and customer homogeneity are valuable for isolating the impact of DSM measures per se and transforming gross into net energy savings. If these criteria are not satisfied, the *Framework* finds that engineering-based estimates of energy savings are preferred. However, the main author of the *Framework* emphasizes that this recommendation applies to the calculation of *gross* energy savings only; net-to-gross analysis is inherently concerned with individual behavior and not technology or engineering assessments, yet many policymakers and

³⁶ TecMarket Works Framework Team, *The California Evaluation Framework*, *op cit*, p. 100.

even M&V professionals continue to approach NTG calculations from a technological or engineering perspective.³⁷

Since the *Framework* was published, some other, relatively comprehensive reference documents have been produced that address the measurement and verification of DSM savings. These include the April 2007 *International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings* and the November 2007 *Model Energy Efficiency Program Impact Evaluation Guide*. Both of these documents emphasize the use of engineering M&V methods.

³⁷ E-mail communications from Nick Hall to Larry Kaufmann, November 30, 2009.

REFERENCES

- Auffhammer, M., C. Blumstein and M. Fowlie, (2008), “Demand-Side Management and Energy Efficiency Revisited,” *Energy Journal*, 29(3), 91-104.
- California Public Utilities Commission (1993), *Decision No. 93-05-063*.
- California Public Utilities Commission (2001), “Interim Opinion Adopting Energy Efficiency Policy Rules,” *Decision No. 01-11-066*.
- Cicchetti, C. (2009), *Going Green and Getting Regulation Right: A Primer for Energy Efficiency*, Vienna, VA: Public Utilities Reports.
- Cook, G. (2008). “Attribution Methodology Wars: Self-Report Methods Versus Statistical Number Crunching –Which Should Win?” *Presented at the ACEEE Summer Study of Energy Efficiency in Buildings*, Pacific Grove, CA.
- Dezhbaksh, H. (1990), “The Inappropriate Use of Serial Correlations Tests in Dynamic Linear Models,” *Review of Economics and Statistics* 72: 126-132.
- Durbin, J., (1970), “Testing for Serial Correlation in Least Squares Regression When Some of the Regressors Are Lagged Dependent Variables,” *Econometrica* 38: 410-421.
- Engle, R. and C.W.J. Granger (1987), “Co-Integration and Error Correction: Representation, Estimation and Testing,” *Econometrica*, Vol 55: 2, 251-276.
- Efficiency Valuation Organization (2007), *International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings* (www.evo-world.org)
- Greene, W. (2000), *Econometric Analysis*, Prentice Hall: Upper Saddle River, NJ, pp. 542.
- Hartman, R. (1988), “Self-Selection Bias in the Evolution of Energy Conservation Programs,” *The Review of Economics and Statistics*, 70(3), 448.
- Johnston, J. and J. DiNardo (1998), *Econometric Methods*, Oxford: McGraw Hill Book Company.
- Joskow, P. and D. Marron, (1992), “What Does a Negawatt Really Cost? Evidence From Utility Conservation Programs,” *Energy Journal*, 13(4), 54.

Loughran, D. and J. Kulick (2004), “Demand Side Management and Energy Efficiency in the United States,” *Energy Journal*, 25(1), 19-43.

National Action Plan for Energy Efficiency (2007), *Model Energy Efficiency Program Impact Evaluation Guide*, Prepared by Steven R. Schiller, Schiller Consulting, Inc. (www.epa.gov/eeactionplan)

Nerlove, M. and K. Wallis (1966), “Use of the Durbin-Watson Statistic in Inappropriate Situations,” *Econometrica* 34: 235-238.

Ridge, R., R. Violette, and D. Dohrman, (1994). “An Evaluation of Statistical and Engineering Models for estimating Gross Energy Impacts,” Prepared for the *California Demand Side Management Advisory Committee: The Subcommittee on Modeling Standards for End Use Consumption and Load Impact Models*.

RJ Rudden Review of the Union Gas Demand Forecast Methodology, December 2005, Attached as Appendix C to EB-2005-0520, Exhibit C1 Tab 1

TecMarket Works Framework Team (2004), *The California Evaluation Framework*, Prepared for the California Public Utilities Commission.

US Energy Information Administration (2007), *Form EIA-861 Final Data File for 2007*, Retrieved from: <http://www.eia.doe.gov/cneaf/electricity/page/eia861.html>