MODEL SELECTION CRITERIA FOR ESTIMATING NET AND GROSS EFFECTS OF COMMERCIAL RETROFIT PROGRAMS

Steve Grover, Quantum Consulting Inc., Berkeley, CA
John Cavalli, Quantum Consulting Inc., Berkeley, CA
Mary O'Drain, Pacific Gas and Electric Company, San Francisco, CA

Introduction

When evaluating DSM measures, the statistical model specification imposes strong assumptions about the market for the energy efficient equipment being analyzed. Too often attention is placed only on the model specification without examining the market conditions necessary to support the specification. Incorrect assumptions regarding the market can introduce significant bias in the impact model results.

Careful examination of the assumptions that underlie the various model specifications is an important up-front step in specifying an impact evaluation. By focusing on the market assumptions rather than just the model specification, it is easier to select the appropriate model for the given market conditions. Once the appropriate model specification is identified, the data collection can be tailored to fit the model and market conditions.

This paper presents a systematic examination of the influence of market assumptions on the results that are obtained from standard impact analysis methods. Several important model considerations are addressed including estimating changes in energy use, correcting self-selection bias, and market transformation effects. This framework provides a simple and concise method for determining which model is appropriate for a given market condition.

The results of this analysis are summarized in matrix form at the end of this paper. The first two columns of each matrix describe the assumption. The third column describes the effect on the estimation results if the assumption does not hold. The next column indicates the plausibility of the assumption followed by a column describing the type of data necessary to check the assumption. The last four columns indicate which types of modeling and analysis methods rely on the given assumption.

The body of the paper is devoted to addressing some of the more commonly used assumptions given in the matrix. Each section is headed by an assumption from the matrix and is followed by a detailed discussion. To illustrate the implications of each assumption, the market conditions needed to support the assumption are compared with the model specification and model results.

I. Explaining Changes in Energy Usage

Examining changes in energy use between pre- and post-program periods is a common method for evaluating the effect of retrofit programs. By controlling for non-program factors, the effect of the program on energy use can be estimated. Several key assumptions are highlighted below.

> *Assumption:* _A forecast of next year's energy consumption can be modeled to account for program and non-program induced changes._

The possibility of forecasting future energy use is the primary assumption underlying statistical analysis of energy impacts. By accurately predicting energy consumption, the model can be used to predict energy use with and without the program. A typical model specification is

$$\text{POSTKWH} = \alpha + \beta'\text{PREKWH} + \beta'\text{MARKET} + \beta'\text{WEATHER} + \beta'\text{ENG}$$

where POSTKWH is the post-period energy use, PREKWH is the pre-period energy use, MARKET contains variables reflecting such responses to market conditions as equipment purchases and employment changes, WEATHER contains changes in weather conditions during the analysis period, and ENG contains engineering estimates of energy impacts of program technologies.

The condition that supports this specification is that all major market factors that affect changes in energy use can be measured by the researcher and included in the model or accounted for in the intercept or slope parameters. The specification described above illustrates the broad range of data needed to estimate the model. Factors such as market conditions, price changes, and weather that will affect energy usage during the sample time frame must be included in the model. Customer specific changes such as changes in employment and equipment purchases that affect energy use should also be included.

Omitting relevant variables due to a lack of data will reduce the explanatory power of the model and can potentially bias the estimation results. This illustrates the importance of extensive data collection on both participants and nonparticipants at the start of the analysis process. If one or more important pieces of information can be identified up-front as uncollectable, the researcher may wish to explore a different modeling technique. Only after this stage is correctly modeled can more complex modeling problems such as self-selection bias be addressed.

II. Determining the Gross Baseline for Energy Usage

The gross baseline provides a comparison level to determine the impact of the program measure. The gross baseline is the energy usage that is expected with the standard, non-program technology. By comparing the baseline with the energy usage from the program technology, the energy impact can be determined. As discussed below, determining the gross baseline as well as estimating other model components such as program participation requires accurate information on purchases made outside the program.

> *Assumption:* *Can determine if nonparticipants adopted high efficiency or standard efficiency equipment outside the program.*

One of the assumptions required to determine the gross baseline is that the researcher can obtain information on the energy efficiency of equipment purchased outside the program. In addition, this information is also useful for estimating program participation. For example, if a program rebate is one of the variables used to explain program participation, the importance of the rebate can be determined by comparing high efficiency equipment purchases both in and outside the program. Information on non-program equipment purchases can also be used in a model to determine selection into the program and hence the Mills Ratio. If this information is not accurate, then the Mills Ratio cannot be reliably estimated.

A recent assessment of this issue relates to PG&E's 1995 Commercial HVAC Retrofit program and provides a good example of the implications of this assumption. In the HVAC evaluation, information on the efficiency of technologies adopted outside the program were unavailable. In an attempt to estimate a net-to-gross ratio, several model specifications were used relying on different assumptions regarding the efficiency of non-program adoptions. Given these assumptions, a logit model is used to estimate the likelihood of purchasing high efficiency HVAC equipment. Explanatory variables include awareness of the program, rebate amount, energy savings, and customer characteristics such as size and building type.

Using the logit purchase model, different model specifications were created using different assumptions regarding the efficiency of the equipment purchased outside the program. In each model, all adoptions within the program are for high efficiency equipment. Model 1 assumes that all non-program adoptions were for standard efficiency. Model 2 assumes that half of those outside of the program and aware of the program adopt high efficiency measures. Model 3 has half of those outside the program and unaware of the program purchasing high efficiency equipment. Model 4 has half of the non-program adoptions as high efficiency equipment, regardless of awareness of the program. Outside the program, just under half (46 percent) of those sampled were aware of the program. These four models are estimated to cover the range of awareness and efficiency scenarios for nonparticipant actions.

A net-to-gross ratio is calculated using the estimation results from the logit purchase model. First, the model is used to determine the expected energy impacts with the program. This is done by multiplying the estimated probability of a high efficiency purchase by the energy impact of the technology. Next, the expected impact without the program is estimated. This is done by recalculating the probability of a high efficiency purchase when the program variables are set to zero. The probability is then multiplied by the energy impact to get the expected energy impact in absence of the program. The difference in the expected impacts with and without the program is the impact that is attributed to the program. This net impact due to the program is divided by the expected impact with the program to get the net-to-gross ratio.

The estimated net-to-gross ratios from the models are shown in Table 1. The model results are weighted to the population using the weights given at the bottom of Table 1. The weights are designed to reflect the participant and nonparticipant populations.

As shown in Table 1, the net-to-gross estimation results are quite sensitive to assumptions made regarding equipment purchases outside the program, with ratios ranging from 0.49 to 2.88. Differences in the net-to-gross ratios are due in part to the different population weights used in the analysis. Since the nonparticipant population is much greater than the participant population, any changes in high efficiency adoptions outside the program result in large changes in the net-to-gross ratios because they indicate either a substantial change in the spillover effect or large changes in the rate of naturally occurring adoptions. As the wide range of net-to-gross estimates indicates, information on the efficiency of nonparticipant adoptions is critical for accurate estimation results.

III. Market Transformation Effects

The market transformation effects discussed in the matrix focus on the effect of free riders and free drivers. Free riders are program participants who would have adopted the high efficiency measure anyway in absence of the program. Free drivers are those outside the program that implement high efficiency measures due in part to the effect of the program. For example, a customer may learn about high efficiency equipment through program advertising but decide to purchase outside to avoid the hassle of going through the program.

The identification of free riders and free driver effects is critical for determining the impact of a retrofit program. However, these effects also remain the most elusive to identify. They require extensive data collection, particularly outside the program. In addition, survey and telephone questions are required to elicit motivations for purchase decisions, decisions that are often made several years ago.

Table 1: Estimated Net-To-Gross Ratios Using Different Nonparticipant Purchase Scenarios

Model	Scenario	Net-to-Gross Ratio
Model 1:	Nonparticipants purchase standard efficiency	0.76
Model 2:	Half of NP's aware of program purchase high efficiency equipment	2.88
Model 3:	Half of NP's unaware of program purchase high efficiency equipment	0.49
Model 4:	Half of NP's (both aware and unaware) purchase high efficiency equipment	1.31

WEIGHTING

	Weight	# in Sample
Participants	2.6	322
Nonparticipants	79	102

Assumption: No Free Drivers.

Because information on the effect of the program on nonparticipants is so difficult to obtain, it is tempting to assume that there are no free driver effects. This has the advantage of simplifying the analysis, since the program impact is limited to participants. However, this assumption is restrictive as programs are likely to have at least some effect on nonparticipants. More importantly, assuming that there are no free drivers when in reality they do exist can seriously bias the estimation results.

Table 2 gives an example of how the assumption of no free drivers can affect the impact estimates.

In this example, only the 11 SEER and 12 SEER HVAC options are eligible for the program. Without the program, 800 10 SEER units, 100 11 SEER units, and 100 12 SEER units are purchased giving a market baseline of 10.30 SEER. With the program there are 400 10 SEER, 150 11 SEER and 150 12 SEER units purchased outside the program. The existence of the program results in an increase of the market baseline to 10.64.

For simplicity assume that all of the 11 and 12 SEER units purchased outside the program are the result of free drivership and should therefore be included in the program impacts. Excluding these free drivers has two effects on the estimated impact of the program. First, the number of high efficiency HVAC purchases attributed to the program is reduced. Purchases of 11 SEER units fall from 250 to 200 while 12 SEER units fall from 150 to 100. The second effect is that excluding the free drivers as program impacts increases the market baseline from 10.30 to 10.64 which also decreases the impact of the program.

The results of these two effects on the net-to-gross ratio is given at the bottom of Table 2. Using the market baseline of 10.64, the impact of an 11 SEER unit is 0.36 and the impact of a 12 SEER unit is 1.36. Multiplying the individual impacts by the number of purchases in the program gives a total net impact of 207. The gross impact is 400 which results in a net-to-gross ratio of 0.52.

When free drivers are included in the impact analysis, the market baseline is 10.30. Using the same calculations as before, the net impact increases to 430 and the estimated net-to-gross ratio is 1.08.

As this example illustrates, the assumption of no free drivers has potentially serious consequences on the estimation results if violated. The assumption is reasonable, however, in a situation with a new technology or a pilot program. With a new technology, awareness of the technology may be low resulting in few adoptions outside the program.

Assumption: All nonparticipants are in the market for the measures being promoted by program.

Researchers often fail to recognize that they are making this assumption. In this situation, all of the nonparticipants in the sample are assumed to be in the market for program measures. Whenever a nonparticipant sample is constructed using information on any customers that do not participate in the program, this assumption is being made.

The reasonableness of this assumption depends upon the technology in question as well as the sample of nonparticipants. For basic technologies such as lighting, everyone can be assumed to be in the market. For more specialized technologies such as refrigeration, nonparticipants in the market will be a much smaller subset of the nonparticipant population.

Table 2: Effect of Assuming No Free Drivers on Impact Estimates

SEER	Purchases W/O Program	Purchases W/ Program		
		Outside	Inside	W/ Free Driver
10	800	400		
11	100	150	200	250
12	100	150	100	150
Market Baseline	10.3	10.6		

	With Free Drivers	Without Free Drivers	Gross Baseline
Impact	207	430	400
Net-to Gross	0.52	1.08	

Suppose that the nonparticipant sample includes customers that have no need for the technology promoted by the program. Then, no amount of program incentives will encourage these customers to purchase the product. The result is a sample that contains too many nonparticipants and underestimates the effect of the program. Including these customers in the sample is likely to reduce the explanatory power of variables included in the model.

A simple example illustrates another potential problem created by this assumption. Suppose that in a representative sample of participants, 50 customers have an average electricity use of 50,000 kWh. The sample also contains 100 nonparticipants that are in the market for the technology with an average use of 10,000 kWh per month. Given the difference in usage, it appears that electricity use might be a factor in determining participation.

However, suppose that the nonparticipant sample is expanded to include an additional 100 customers outside the market who are more similar to the participants in total energy use but it is recognized that they are outside the market. In this example, average electricity use among nonparticipants rises from 10,000 kWh to 40,000 kWh. Because these additional nonparticipants are not in the market for the technology, the difference in electricity use between participants and nonparticipants is artificially reduced to 10,000 kWh instead of the true difference of 40,000 kWh. As a result, the value of electricity use as a variable to determine program participation is reduced, since there is less difference in usage between participants and nonparticipants.

IV. Self-Selection Bias

Self-selection bias occurs when the sample used to estimate a regression equation is not randomly determined. Since program participation is voluntary, self-selection bias is always a concern in impact analysis. Much attention has focused on how to deal with self-selection bias in impact analysis, a few of the more general issues are discussed below.

Assumption: *Nonparticipants are an adequate control group in terms of energy usage and premise characteristics.*

If nonparticipants are an adequate control group, then this group can be used to estimate energy usage in absence of the program. That is, nonparticipants can be used to predict what the energy use of participants would have been had they not entered the program. This assumption underlies those models where self-selection is not corrected for, such as bill comparison or comparison of SAE realization rates across separate models for participants and nonparticipants.

The market condition under which this assumption is valid is that there are no significant differences between participants and nonparticipants. However, since participation is voluntary, participants will likely vary systematically from nonparticipants. For example, in the commercial sector large customers have a greater incentive to participate in a retrofit program. While size can be controlled for, factors such as business strategies that might vary with size and affect the participation decision cannot be captured in the model. In this case, the nonparticipant group consists of smaller customers that do not provide a good indication of how participants would behave in absence of the program.

This assumption is plausible only in the case where the technology in question is new. In cases of mature technologies, differences between participants and nonparticipants are likely due to fundamental differences between the two groups. With new technologies, it is more plausible that participants and nonparticipants are similar, perhaps differentiated only by their awareness of the new product.

> *Assumption:* *Unobserved factors influencing participation can be controlled for in the model, thereby correcting for self-selection bias.*

Self-selection bias is often corrected for using an inverse Mills Ratio term, a procedure developed by Heckman (1979). Goldberg and Train (1996) provide a very good discussion of how the Mills Ratio method can be used in energy applications. The Mills Ratio is estimated from the parameters of a separate logit or probit model of program participation. Once calculated, the Mills Ratio is included in an SAE model to control for the effect of unobserved factors that influence participation in the program. If the model is specified so that net savings is reflected in the coefficient estimates, net savings will be constant across participants due to the fixed coefficient. This is the basic assumption of the SAE model using the standard Mills Ratio correction.

One market condition that is required for this technique is that the net impact of the program must be constant across all participants. The reasonableness of this assumption is the topic of some debate. Goldberg and Train (1996) point out that those factors that determine participation are also likely to influence the amount of savings among participants. In this case, not only is participation correlated with the unobserved factors, but net savings is correlated as well. For example, since larger customers are in general more likely to participate, net savings may also be correlated with participation since large customers will save more through the program. They show that ignoring this correlation between participation and net savings results in a biased estimate of net savings, with the direction of the bias depending upon the direction of correlation. They propose the addition of a second Mills Ratio interacted with a participation dummy variable to allow net savings to vary across participants. This technique also carries with it assumptions about the market and are discussed below.

> *Assumption:* *Unobserved factors influencing participation also affect the amount of savings resulting from the program.*

This assumption supports the use of an additional Mills Ratio in the net savings estimation model. As discussed, this specification assumes that participation and net savings are correlated. By interacting the second Mills Ratio with participation or estimated program impact, an estimate of net savings is obtained that varies across participants (Goldberg and Train (1996)).

The market condition required for this model specification is that net savings must be distributed normally across participants. This is in contrast to the single Mills Ratio specification which results in constant net impact across participants. The consequences of violating this assumption depend upon just how far from normal net savings is actually distributed. Goldberg and Train simulate a variety of situations where the double Mills Ratio is used when net savings is not normally distributed. They find that the model still performs well as along as savings is not distributed “too far from normal” (p. 4-6). However, when the net savings distribution is substantially different from normal, they find that the double Mills ratio technique performs worse than when only a single Mills ratio correcting for selection is used.

Conclusion

This paper provides a concise analytic framework for selecting the appropriate model given particular market conditions. This is done by highlighting the assumptions underlying a given model specification and emphasizing the market conditions necessary to support these assumptions. If market conditions do not hold, potentially serious bias can be introduced into the estimation results. By examining the necessary market conditions and the likelihood that these conditions exist, the researcher can tailor the analysis to fit the given conditions.

References

1. Evaluation of PG&E’s 1995 Nonresidential Energy Efficiency Incentives Program for Commercial Sector HVAC Technologies. Volume I: Analysis Appendices. Prepared for PG&E by Quantum Consulting, Inc., Berkeley, CA, March, 1997.
2. Goldberg, Miriam and Kenneth Train. “Net Savings Estimation: An Analysis of Regression and Discrete Choice Approaches.” Prepared for CADMAC by Xenergy, Inc., Madison, WI, March, 1996.
3. Greene, William. Econometric Analysis, Macmillan Publishing Co., New York, NY, 1990.
4. Heckman, J. “Sample Selection Bias as a Specification Error”, Econometrica, Vol. 47, pp. 153-162.

Table 1A: Changes in Energy Use

ASSUMPTIONS USED IN IMPACT ANALYSIS					METHODS RELYING ON ASSUMPTION			
Assumption	**Description of Assumption**	**What Happens if Assumption is Violated**	**Reasonableness of Assumption**	**Data Required to Check Assumption**	**Bill Comparison**	**Difference of SAE Realization Rate on P and NP Replacement**	**NET SAE Model Using Mills Ratio Term**	**Gross SAE, Market Change Analysis, NP Canvass**
EXPLAINING CHANGES IN ENERGY USAGE								
SAE Model can accurately forecast energy changes.	A forecast of next year's energy consumption can be modeled to account for program and non-program induced changes which are both customer specific (with non-program change vars.) and market-wide factors (with the intercept and slope on pre-kWh)	Gross impact estimate will be biased.	Easier to model changes than level	Requires a sufficient samples. Outliers can be identified possibly with on-sites.		*	*	*
Analysis Period pre/post is typical weather, and non-program changes same for P and NP	The difference in Participant and Nonparticipant energy usage accounts for all possible changes occurring in the same proportion for both samples and the weather is the same as well as typical for both periods.	Gross impact estimate will be biased.	VERY UNLIKELY	Weather data, Telephone survey data	*			
CDA model can be run on pre and post data	Whole premise usage can be successfully modeled in a CDA model. i.e., It requires info. about all equip., not just those affected by the DSM measures.	Gross impact estimate will be biased.	Very difficicult to model energy usage, rather than just year to year changes	Requires a large sample of on-sites				

Table 1B: Self-Selection Bias

ASSUMPTIONS USED IN IMPACT ANALYSIS					METHODS RELYING ON ASSUMPTION			
Assumption	**Description of Assumption**	**What Happens if Assumptiion is Violated**	**Reasonableness of Assumption**	**Data Required to Check Assumption**	**Bill Comparison**	**Difference of SAE Realization Rate on P and NP Replacement**	**NET SAE Model Using Mills Ratio Term**	**Gross SAE, Market Change Analysis, NP Canvass**
SELF-SELECTION (SS) BIAS								
NP Adequate kWh/kW control	NP must be an adequate control group in terms of energy usage and premise characteristics and have the same response to year-specific changes.	Difference in kWh usage and premise characteristics (e.g., size) between Ps and NPs will introduce bias in the intercept and pre-kwh parameters, possibly introducing bias into the SAE change parameter.	Reasonable, especially for new programs or techs.	Billing Data, Telephone surveys	*	*		*
SS Bias can be controlled for using an Inverse Mills Ratio	Unobserved factors influencing participation can be controlled for with a Mills Ratio, thereby correcting for Self-Selection Bias.	Impact estimate will be biased	Reasonable	No Data Collection Strategy			*	
Unobserved factors affecting Participation also affect net savings	Unobserved factors correlated w/ participation are also correlated w/ savings. Additional Mills Ratio interacted with participation corrects for correlation.	Requires that net savings be normally distributed across participants. If distribution substantially different from normal, net impact estimate will be biased.	Reasonable	No Data Collection Strategy			*	
Not all NPs impl. significant changes	All NPs did not make signiciant changes in the analysis period.	If all NPs made changes, the multicollinearity in the model will result in imprecise coefficient estimates.	Can achieve easily in NP sample	Billing Data, Telephone Surveys		*		*

Table 1C: Determining the Gross Baseline

ASSUMPTIONS USED IN IMPACT ANALYSIS					METHODS RELYING ON ASSUMPTION			
Assumption	**Description of Assumption**	**What Happens if Assumptiion is Violated**	**Reasonableness of Assumption**	**Data Required to Check Assumption**	**Bill Comparison**	**Difference of SAE Realization Rate on P and NP Replacement**	**NET SAE Model Using Mills Ratio Term**	**Gross SAE, Market Change Analysis, NP Canvass**
DETERMINING THE GROSS BASELINE								
Pre-existing system is baseline, NP and P similar	Pre-existing system is the baseline, NP similar to P.	Gross impact estimate will be biased depending on whether the NPs baseline is greater than P. If it is true that a standard is the baseline, then the impacts will be biased higher.	Reasonable only for programs where no standard exists, esp. for new techs. or pilots	Self-Reports on pre-existing from tele, mailers, or on-sites	*		*	
Pre-existing system observed, known	Pre-existing system is observed for P and NPs, if unobserved, they are the same or estimated using vintage of pre-existing system.	Gross impact estimate will be biased depending on whether the NPs baseline is greater than P and/or vintage estimate is biased.	Reasonable, especially for new programs or techs.	Self-Reports on pre-existing from tele, mailers, or on-sites		*		*
NP adoptions id. as baseline or high efficiency	You can determine if NPs adopted high eff. or baseline measures. (The Mills Ratio and RP model should be restricted to these actions).	Can't reliably estimate Mills Ratio.	Reasonable	Self-Reports on pre-existing from tele, mailers, or on-sites		*	*	*
RR on change = Gross, or 2 RRs can be estimated.	Realization Rate in Unadjusted Engin. Energy Change Estimate used in Gross Impact Model can be used to adjust Unadjusted Engineering Gross Impact Estimate. Or, Two RRs can be estimated, pre- to BL and Hieff.	Bias will be introduced if Realization Rate is not transferable from a change estimate to a baseline adjusted estimate. Only applies where the baseline for a gross impact is not the participants pre-existing system.	Reasonable	Self-Reports on pre-existing from tele, mailers, or on-sites		*	*	*

Table 1D: Market Transformation Effects

ASSUMPTIONS USED IN IMPACT ANALYSIS					METHODS RELYING ON ASSUMPTION			
Assumption	**Description of Assumption**	**What Happens if Assumptiion is Violated**	**Reasonableness of Assumption**	**Data Required to Check Assumption**	**Bill Comparison**	**Difference of SAE Realization Rate on P and NP Replacement**	**NET SAE Model Using Mills Ratio Term**	**Gross SAE, Market Change Analysis, NP Canvass**
MARKET TRANSFORMATION EFFECTS -- FREE RIDERSHIP/FREE DRIVERSHIP								
No free drivers	No free drivers or all changes implemented by nonparticipants are attributable to natural conservation.	Underestimate net impact by 2*FD. This occurs not only because the direct of free-riders is ignored, but because free-drivers raise the baseline estimate.	Reasonable only for new techs or pilots.	No Data Collection Strategy	*	*	*	
FR rate in P sample = NC rate in NP sample	Free rider rate in P sample is the same as the natural conservation rate in NP sample.	Overestimate net impacts if FR>NC and underestimate net if FR<NC.	Reasonable only for new techs or pilots.	No Data Collection Strategy	*	*		
All NPs in the market	All NPs are in the market for the specific measures being promoted by the program.	Need to restrict the model to the market subsample, if zero is used for nonparticipants who were not in the market the MIlls and RP model results are biased.	Unreasonable	Self-Reports on pre-existing from tele, mailers, or on-sites			*	
Free Drivers can id. in Aware NP sample.	Education Effects can be derived from aware nonparticipant actions. New programs can use trade ally, stocking survey, and other market data.	If either of these assumptions fail, then free drivership must be estimated using a conjoint study where the "no action" case is used to model market dynamics.	A cheap unreliable method for mature programs, possibly effectives in the early stages for pilots or new techs	Canvass Survey Methods				
Changes in Total High-Efficiency Adoptions can be estimated.	Using Trade Ally, Manufacturer, and nonparticipant responses, the total increase in high-efficiency adoptions can be estimated. By incorporating a Free-Rider Estimate, Free-Drivership can be determined.	Bias in the free-drivership and free-ridership rates. This method tends to produce an upperbound for both of these effects.	Reasonable for programs less than 2-3 years old. Mature programs require out of service territory canvass	Canvass Survey Methods				*