Why Discrete-Continuous Billing Models Mis-Estimate Net Savings of DSM Programs

R. Eric Paquette, Cambridge Systematics, Inc.

A great deal of attention has been devoted to developing a direct billing analysis technique to measure net savings of DSM programs. In the most basic approach, billing models are estimated for a combined group of participants and non-participants. Two indicator variables are employed to capture the program savings effect, and the naturally occurring conservation effect. The approach is flawed because participants and non-participants are behaviorally distinct groups. Participants are more likely to install measures without program incentives than non-participants. Thus, the non-participant group underestimates the participant naturally occurring conservation, and overestimates net savings.

An approach called the Discrete-Continuous method has been touted as a way to address this bias. In this approach, two models are developed. First, the decision to participate in a program is estimated using choice models for a group of eligible customers. The probability of participation is simulated for all customers. Second, billing models are estimated with an indicator variable to capture naturally occurring savings, and the participation probability variable to capture the program savings. The intended effect is to remove the self-selection bias that corrupts the basic participant/non-participant comparison.

In this paper, I demonstrate why the Discrete-Continuous approach fails to account for the bias, and is actually more problematic than the basic participant/non-participant comparison.

INTRODUCTION

For many Demand-Side Management (DSM) programs, shareholder incentives are tied directly to the verification of program impacts. In most instances, net program savings, or the savings that would not have occurred if the program did not exist, are used to measure the program impact. In order to properly measure net savings, it is necessary to estimate what savings occurred with the program, and what savings would have occurred without the program.

Oftentimes, a billing analysis is used to estimate the savings that occurred with the program. In this approach, the energy consumption of the program participants is compared for a time period prior to program participation and a time period after program participation. The difference in consumption, after controlling for weather and other outside influences on consumption, is the estimated gross savings, or the savings that occurred among participants for measures installed as part of participating in the program. In rebate programs, gross savings are those savings that occur from the installation of rebated measures. This approach does not truly measure gross savings, because it also captures any participant spillover impacts, or savings that were induced by the program, but did not occur from measures installed as part of program participation. In a rebate program, spillover savings are from measures that were adopted because of the program, but were not rebated through the program. Since spillover impacts are not the primary problem associated with estimating net savings with billing analysis techniques, I will ignore spillover impacts in this paper.[1]

Using program participant billing data only, the billing models are typically estimated using an a model similar to the following:

$$kWh_{ij} = \beta_0 + \beta_1 x W_{ij} + \beta_2 x X_{ij} + \beta_{gross} x Part_{ij}, \text{ where}$$

kWh_{ij}	=	the kWh consumed by customer i in billing period j;
β_0	=	the estimated base kWh consumed in a billing period;
β_1	=	the estimated reaction to weather;
W_{ij}	=	the weather faced by customer i in billing period j;
β_2	=	the estimated effect of other influences on kWh consumption;
X_{ij}	=	the other influential factors on kWh consumption for customer i in time period j;
β_{gross}	=	the estimated gross savings; and,
$Part_{ij}$	=	1 in the post-program billing periods for program participants, 0 otherwise

Gross savings are captured by measuring the difference in kWh consumed between the pre-program period, and the post-program period while controlling for the other outside factors.

It is much more difficult to measure the savings that would have occurred without the program. In order to do this for the participant group, it is necessary to remove the savings that participants would have achieved without the programs, or the free rider savings. A common approach has been to assume that the savings from non-participants are an estimate of what participants would have done if not for the program. The rationale is that non-participants have not received rebates or other incentives for installing measures, so participants should behave similarly when rebates or other incentives are removed. With this assumption, the estimation of net program savings becomes relatively easy.

Billing models are developed for a sample that includes both participants and non-participants. The models have a general form similar to the following:

$$kWh_{ij} = \beta_0 + \beta_1 xW_{ij} + \beta_2 xX_{ij} + \beta_{fr} xPost_{ij} + \beta_{net} xPart_{ij}, \text{ where}$$

kWh_{ij} = the kWh consumed by customer i in billing period j;
β_0 = the estimated base kWh consumed in a billing period;
β_1 = the estimated reaction to weather;
W_{ij} = the weather faced by customer i in billing period j;
β_2 = the estimated effect of other influences on kWh consumption;
X_{ij} = the other influential factors on kWh consumption for customer i in time period j;
β_{fr} = the estimated free rider savings;
$Post_{ij}$ = 1 in the post-program period for all customers, 0 otherwise;
β_{net} = the estimated net savings; and,
$Part_{ij}$ = 1 in the post-program billing periods for program participants, 0 otherwise.

In these models, the gross savings from the participant only model is split into two components, free rider savings, and net savings. Because of the nature of the $Post_{ij}$ and $Part_{ij}$ variables, the free rider savings are estimated entirely by the average reduction in consumption among non-participants. The net savings are estimated as the average reduction in consumption among participants minus the average reduction in consumption among non-participants.[2]

Unfortunately, the assumption that non-participants estimate participant behavior in the absence of a program is incorrect. The most obvious indication that the two groups are behaviorally different is the fact that one group chooses to participate in a program, while the other does not. In most instances, participants are more likely to adopt program measures than non-participants in the absence of a program. It is this propensity that may draw them to the program in the first place. This effect has been referred to as a self-selection effect. It is more appropriately termed to have bi-directionality of causation. In other words, someone who participates in a program has an increased chance of installing program measures, and someone who would choose to install program measures if the program did not exist, has an increased chance in participating in the program if it is available. As a result, using non-participant behavior as a proxy for participant behavior in the absence of the program will lead to underestimation of the savings that participants would have achieved without the program, and overestimation of the resulting net savings.

DISCRETE-CONTINUOUS BILLING ANALYSIS

Discrete-Continuous billing analysis has been suggested as an alternative approach to the basic participant/non-participant billing analysis approach. Proponents of the approach believe that it captures and controls the effects of self-selectivity bias, and provides appropriate estimates of net savings.

In the most basic form, the Discrete-Continuous approach utilizes two modeling steps. First, a binary choice model of program participation is estimated using available customer data. The purpose of this model is to estimate the probability of a customer participating in the DSM program. These models can be estimated using standard binary logit or probit models that are available in many statistical packages. Once the choice model is estimated, it is used to simulate the likelihood that each customer contained in the participant/non-participant billing analysis would participate in the program.

In the second step, a billing model is estimated for a sample of program participants and non-participants. The model is similar to the participant/non-participant billing analysis described above, but the $Part_{ij}$ variable is replaced with $P(Part)_{ij}$, where $P(Part)_{ij}$ is the probability that customer i has participated in the program in billing period j, as estimated by the binary choice model. The full model specification would be similar to the following:

$$kWh_{ij} = \beta_0 + \beta_1 xW_{ij} + \beta_2 xX_{ij} + \beta_{fr} xPost_{ij} + \beta_{net} xP(Part)_{ij}.$$

As a result of replacing $Part_{ij}$ with $P(Part)_{ij}$, the consumption decrease of all modeled customers is used to estimate the free rider savings, and net program savings, instead of only using the consumption decrease in non-participants to establish free rider savings, and in participants for the net savings. The approach yields higher estimates of free rider savings, and lower estimates of net savings than the basic participant/non-participant billing models. This is a result that is expected if the bias from the basic participant/non-participant billing models is controlled. Unfortunately, the approach does not necessarily yield the correct estimates of free rider and net savings. The shift in the estimates are due solely to the nature of the participant and non-participant savings, not the fact that the bias has been properly controlled.

FLAWS WITH THE DISCRETE-CONTINUOUS APPROACH

While the Discrete-Continuous approach yields results that are closer to the expected results of proper net savings estimation, the approach is fundamentally flawed. The principle problem with the approach is that it is still using non-participant behavior to help establish how participants would have behaved in the absence of the program. When the $P(Part)_{ij}$ variable replaces the $Part_{ij}$ variable in the billing analysis, it blurs the distinction between participants and non-participants, and similarly blurs the allocation of savings into free rider savings, and net savings. In the Discrete-Continuous model, free rider savings are estimated based on some portion of the savings of all customers in the billing analysis, not just the non-participants. Similarly, the supposed net savings are based on the remainder of the savings for all customers in the billing analysis. Free rider savings are still estimated as a function of non-participant savings, when non-participant savings has absolutely nothing to do with how a participant would save in the absence of a program. What needs to be estimated is the program effect on participants, and we already know that the program effect is different for participants and non-participants as is demonstrated by one group's participation.

Perhaps the easiest way to realize that the approach is flawed is to look at the ''best case'' scenario for the Discrete-Continuous model estimation. Assume that there is sufficient information to perfectly estimate the decision to participate using the binary choice participation model. This is something that a modeler would strive to attain. In this case, the probability of participating would be 1 for all participants, and the probability of participating would be 0 for all non-participants. Under this scenario of a perfect choice model, the Discrete-Continuous billing model becomes identical to the original participant/non-participant billing model which is known to be flawed. In the ''best case'' scenario, we are right back where we started with a model that we know does not work. In a more realistic scenario where participation is not perfectly modeled, the only difference between the participant/non-participant model and the Discrete-Continuous model is the error that is included through estimating the probability of participation. As that error is diminished or removed, the model moves towards a model that is known to be incorrect. This approach assumes that the error introduced by the choice model is in the exact amount to correctly apportion the total savings into free rider savings, and net savings. Obviously this is not appropriate.

The results of the Discrete-Continuous approach differ from the basic participant/non-participant model as would be expected if it were properly capturing net savings. This is entirely because of the nature of participant and non-participant savings. As long as participants are saving more than non-participants, net savings will always be lower in the Discrete-Continuous approach than in the participant/non-participant approach.[3] In the Discrete-Continuous model, net savings are estimated as a weighted average of participant savings and non-participant savings, while in the participant/non-participant model, it is estimated only by participant savings. Any weighted average of participant and non-participant savings must be lower than participant savings, so the estimated net savings are always lower.

Slight modifications of the Discrete-Continuous approach have been applied, including the use of a Mills' ratio in the estimation process. In a Mills' ratio model, a function of the probability of participating (Mills' ratio) is included as an independent variable to account for the self-selection bias. As Train pointed out (Train, 1994), this approach also fails to account for the type of bias encountered in net savings estimation. While these approaches may yield different estimates of net savings, they do not yield correct estimates of net savings, because they continue to rely on non-participant behavior to estimate participant behavior.

CONCLUSION

The Discrete-Continuous modeling approach to net savings estimation does not accurately capture the self-selection bias that corrupts participant/non-participant billing analysis approaches. In participant billing data, it is impossible to directly model two different savings estimates (net and free rider), because both savings components occur at the same time, and are realized in only one reduction in consumption for participants. The effects are computationally inseparable for participants alone. To properly measure net savings, it is essential to know what participants would have done in the absence of the program. From a statistical modeling perspective, it is necessary to explicitly capture (1) why customers choose to install or not install conservation measures, and (2) what effect the program has on that installation decision. It is these two components that ultimately effect the amount that a customer will finally save through their installation choice. If models were developed which explicitly capture these effects, it would be possible to use the model to simulate what would have occurred with the program effect removed.

Discrete-Continuous billing models explain why customers choose to participate. They do not explain why customers choose to install or not install measures, and they do not explain what effect the program has on the decision to install measures. Basically the Discrete-Continuous approach answers the question, ''Why do customers participate?'' when it needs to answer the questions, ''Why do customers install conservation measures?'' and, ''What effect does the program have on the installation decision?''. The Discrete-Continuous approach introduces uncertainty into an issue that is not relevant to the estimation of net program savings. This has the effect of shifting estimates of net savings and free rider savings in the directions that would be expected when compared to a basic participant/non-participant comparison. This gives a false sense of security that the approach is effective in capturing the self-selection bias. In fact, the approach is almost arbitrarily shifting the estimates as a result of needless introduction of error into the billing models. As a result, there is potential that the results of a Discrete-Continuous model are less accurate than the results of a basic participant/non-participant comparison.

RECOMMENDATIONS

In order to accurately estimate the net savings, it is necessary to model the impact of the program on participants outside of a billing analysis. Several approaches have been developed to estimate Net-to-Gross ratios outside of billing data. The modeling approaches which produce unbiased estimates of the Net-to-Gross ratios involve discrete choice modeling approaches (Train, et al., 1994, Train, 1994, and Train, Paquette, 1995). As an alternative to discrete choice modeling approaches, survey based approaches can be applied to estimate the Net-to-Gross ratio as well (Cambridge Systematics, Inc., and Freeman, Sullivan and Company, 1994). Once the Net-to-Gross ratios are estimated, it is not necessary to include them in any way in the billing analysis. Instead, they should simply be applied to the gross savings estimated in a participant only billing analysis.

There is no sense in utilizing a Discrete-Continuous modeling approach. There is additional work required over a basic participant/non-participant comparison, and all that is introduced is some random effect due to the error in estimating participation. While Discrete-Continuous approaches derive different estimates of net savings than basic participant/non-participant billing analyses, they do not necessarily offer better estimates of net savings.

ENDNOTES

1. The presence of participant spillover effects do not change anything discussed in this paper. However, non-participant spillover effects act to further corrupt estimation of net savings in participant/non-participant comparisons. In most participant/non-participant comparisons, non-participant savings are assumed to estimate the naturally occurring savings level. In fact, if a program is inducing non-participants to install measures outside of the program then this naturally occurring savings estimate is overestimated by the amount of the spillover impact.

2. This becomes obvious when we realize that $Post_{ij}$ is equal to 1 for both participants and non-participants in the post-program billing periods, while $Part_{ij}$ is equal to 1 only for participant in the post-program billing periods. $Part_{ij}$ always equals zero for non-participants. Thus, no net savings are ever estimated in the non-participant group. Any reduction in consumption must be captured by β_{fr}. In contrast, every participant has two sources of savings, β_{fr} and β_{net}, to be estimated from only one reduction in consumption. The model uses the pure estimate of β_{fr} from the non-participant group as the estimate for the participant group, and allocates the remainder of the reduction in consumption to net savings, or β_{net}.

3. This will always be the case unless a program convinces customers to not install energy saving measures as part of participation. While this is theoretically possible, it is unlikely.

REFERENCES

Cambridge Systematics and Freeman, Sullivan, and Company 1994. *DSM Free Ridership Study*, (Empire State Electric Energy Research Corporation Research Report EP 92–65).

Train, K. 1994. Estimation of Net Savings from Energy Conservation Programs. *Energy: The International Journal*, *19*, 423–441.

Train, K., Buller, S., Mast, B., Paquette, E., & Parikh, K. 1994. Estimation of Net Savings for Rebate Programs: A Three-Option Nested Logit Approach. *Resource Planning Methodologies: Vol. 7. Proceedings at the 1994 ACEEE Summer Study on Energy Efficiency in Buildings*. (pp. 7.239–7.247). Washington, DC: American Council for an Energy Efficient Economy.

Train, K., and Paquette, E. 1995. A Discrete Choice Method to Estimate Free-Ridership, Net-to-Gross Ratios, and the Effect of Program Advertising. *Energy Services Journal, Vol. 1, Num. 1* (pp. 21–33).