A WHITE PAPER:

Mitigating Self-Selection Bias in Billing Analysis for Impact Evaluation

Pacific Gas and Electric Company

CALMAC Study ID: PGE0401.01

Date: 8-4-2017

Prepared by:

Miriam Goldberg and Ken Agnew, DNV GL

Kenneth Train, NERA and the University of California, Berkeley

Meredith Fowlie, University of California, Berkeley

Prepared for:

Pacific Gas and Electric Company

Brian Arthur Smith, project sponsor

Acknowledgements

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Purpose

This paper describes methods to estimate the net savings of energy efficiency programs using customer-level consumption data analysis, also known as billing analysis for net savings. The specific focus is on mitigating self-selection bias.

Not addressed in this paper is the effect of nonparticipant spillover, that is, of nonparticipants adopting program measures because of the program but outside of the program. Addressing such effects is beyond the scope of this paper.

This paper is intended for use by evaluators who want to understand the techniques better, as well as by program administrators, regulators, and other stakeholders who want to understand what is and isn't possible. The opening sections offer a discussion of conceptual issues and approaches. Technical details for the interested readers in the later sections and appendices. A shorter discussion that includes key points from this paper is in Goldberg et al. (2017).

While the primary thrust of this paper is on net savings, many of the same issues and methods apply to gross savings estimation. A key point of this discussion is that the use of a comparison group, and even the use of a randomly assigned comparison group under some designs, is often not sufficient to identify net savings. Depending on the study design, the result of the comparison group analysis may represent net savings, gross savings, or neither.

1.2 Approach

The paper considers alternative assumptions about customers' decision to participate or not in a program, and describes analytic methods that can be used to avoid self-selection bias for each of these situations. We start by describing two related research designs: randomized control trials (RCT) and random encouragement designs (RED). We identify situations under which these designs can be used to estimate the net savings of interest, and delineate why they cannot always be used. We then describe a new alternative approach to address self-selection when the random assignment procedures are inapplicable.

A key element in this work is the use of a model of program participation, which can be enhanced by use of an RED. The new estimation procedure is both simpler and more robust compared to an earlier approach that used similar terms from a participation model[1]. Importantly, we show how, in situations where the RED design with a standard analysis does not by itself provide the net savings of interest, this quantity can be estimated conditional on additional assumptions about the process that determines program participation.

1.3 Background

1.3.1 Renewed interest in billing analysis

The use of consumption data regression analysis for program net savings estimation is of increasing interest in California with the adoption of AB802, which emphasizes normalized metered usage data as the basis for savings estimates. Additional interest in these estimation approaches has been generated by the recent publication of the Uniform Methods Project Chapter 8, (National Renewable Energy Laboratory.

2013) the use of random assignment methods as the basis for ongoing savings estimation from Home Energy Reports programs, (e.g., Applied Energy Group 2014) as well as the increased use of random assignment methods for pilot programs and special studies (e.g. DNV GL 2015).

1.3.2 Gross and Net savings

Net program savings is the difference between participants' consumption with versus without the *program* in place. As noted, nonparticipant spillover is not addressed in this paper and is assumed for discussion purposes to be zero. The effect of the program on the participant consumption includes the effect of the program on the measure adoption, along with any incidental effect of the program on adoption of other measures or behavioural modifications outside the program (participant spillover) as well as any economic takeback effects.

Gross program savings is the difference between participants' consumption with versus without the *measures* targeted by the program in place. To the extent the program measure itself induces a household to adopt other measures or to alter energy-using behaviour in other ways, these effects are also part of the gross program savings. These are effects of the measure, regardless of how the program influenced its adoption.

1.3.3 Why self-selection matters

Self-selection is a challenge for comparison group methods whenever customers are not randomly assigned to participate or not participate in the program. Self-selection means that, even starting from a pool of customer with similar characteristics and program/measure applicability, those who choose to join a program or adopt a measure are different from those who don't, in ways that could affect changes in energy consumption apart from the participation choice. As a result, the analysis cannot separate the program or measure effect from the effect of being in the "inclined to join/adopt" group.

Terms like "self-selection bias mitigation" and the associated analysis techniques are sufficiently arcane to make both evaluation practitioners and their audiences often regard these issues as nuances and fine points not of general interest. However, the effects of self-selection in comparison group analyses can be substantial and meaningful. As one program administrator has put it, "Self-selection is the point of programs. We can't assume it's not there."

When we talk about the need for the comparison group to be similar to the participant group, we usually consider factors such as premise characteristics, equipment, and demographics/firmographics. In practice we often use prior consumption to represent their combined effects. While these can all be important, a key concern for net savings estimation is how well the comparison group represents the "natural adoption" rate among the participants. Natural adopters are those who would have adopted the program measure on their own if the program didn't exist. Participants who are natural adopters, also called free riders, contribute zero to net savings. For many programs, however, natural adopters who are aware of the program will be more likely to become participants than to stay outside the program. As a result, the proportion of natural adopters among the comparison group will tend to be lower than the proportion among participants. Thus, even accounting for other customer characteristics, the comparison group will not by itself "net out" the effect of free ridership.

Examples of self-selection effects include the following:

- A high-efficiency HVAC program is well known to local contractors, who facilitate customer applications. As a result, a high proportion of those who would adopt high efficiency equipment on their own obtain a rebate from the program. A comparison group of non-participant equipment replacers is identified by phone, and savings are estimated as the difference between the average change in consumption for program participants and that of non-program replacers. The comparison group doesn't reflect the natural adoption of high-efficiency equipment, because most natural adopters of high efficiency join the program. In this case, free ridership isn't accounted for by the comparison group. On the other hand, the comparison group also does not represent average change in consumption with adoption of standard efficiency equipment, because at least some of the nonparticipants might have adopted high efficiency equipment but not obtained a rebate. Thus, the analysis produces neither gross nor net savings, but something in between.

- A whole-house retrofit program is available to the general residential population, and tends to be joined by higher income households at a time when they are having other work done in their homes. The effect of the other home upgrade activities in conjunction with the program distorts the savings estimated by the analysis, unless a comparison group can be identified of similar demographics, who are doing similar work on their homes but not also participating in the program.

1.4 Organization of the paper

Section 2 summarizes the key results of this paper and briefly describes a new method for estimating net savings with billing data. This section provides high-level guidance, without technical detail. It may be of interest in particular to funders and users of evaluation results who want perspectives on the strengths and limitations of alternative methods.

Section 3 establishes notation and terminology for net savings estimation using regression analysis.

Section 4 describes two random assignment procedures that have been applied to estimate net savings on billing data: randomized control trials (RCT) and random encouragement designs (RED). We show conditions under which each of these methods provides a valid estimate of net savings for all participants. We also describe situations where these methods must be augmented with additional assumptions or additional methods are needed in order to identify the net savings of interest.

Section 5 presents regression-based methods for net savings estimation that include corrections for self-selection. These methods are potentially useful when random assignment procedures are not applicable or valid. To explain the need and form of the corrections, we begin with assumptions under which a standard regression is accurate without any need for correction terms; we then relax (generalize) the assumptions to account for various types of self-selection. Importantly, we provide a new method that addresses self-selection in its most general, and most common, form.

Section 6 describes simulation results using the new method. The simulations confirm that the approach works as intended when the required assumptions are true. The simulations also investigate the robustness of the method under departures from those assumptions, as well as the effect of increased sample size on the method accuracy.

Section 7 describes how the methods can be used to estimate gross savings.

Section 8 gives a summary of key findings and practical considerations, with somewhat more technical detail than is in Section 2.

Section 9 provides references

Appendix A describes the instrumental variables interpretation of random encouragement designs, and Appendix B provides the formal derivation our new method.

Appendix C describes the extension of the method to a statistically adjusted engineering (SAE) framework.

Appendix D compares the new method introduced in this paper to a previous "Double Inverse Mills Ratio" approach.

2 KEY LESSONS

2.1 An improved method for controlling for self-selection: the IV-IMR method

This paper introduces a method for controlling for self-selection that addresses key sources of bias that can confound net savings estimates from billing analysis. The new method appears to be more robust than prior methods, without adding more complexity. The method incorporates a model of the probability of participation. The predicted probability and the Inverse Mills Ratio, which is derived from the same estimated probability function, are both included in the regression analysis of customers' consumption. Inclusion of participation probability in a billing analysis regression is not by itself new, but is a basic Instrumental Variables (IV) approach. Our discussion shows that the IV approach alone provides net savings only in special circumstances, while our new method combining the IV and IMR terms can provide net savings under more realistic assumptions.

For ease of exposition, we describe the IV-IMR method starting from a simple regression model as follows. Let Δ_j denote the change in consumption for customer j and let D_j be a dummy variable equal to 1 if customer j is a participant and equal to 0 if customer j is a non-participant. The regression is then:

$$\Delta_j = a - bD_j + \nu_j$$

The coefficient b is intended to capture the average net savings of the program.[2] Estimation of this coefficient will be biased relative to the true net savings if the comparison group are not a good representation of participants absent the program. More precisely, bias is introduced when the change in consumption that would have occurred without the program is different for participants and non-participants. For example, customers who would have adopted the measure on their own may be more likely to join the program than those who would not adopt on their own, resulting in a more negative change on average for participants than nonparticipant, even without any effect of the program. In terms of the regression, this means that the participation dummy variable D_j is correlated with the residual term ν_j. This correlation violates a fundamental requirement for unbiased estimation, that the explanatory variables be uncorrelated with the error terms.

To address this bias, we enhance the regression by taking the following steps:

1. Fit a model of the probability of participation as a function of available explanatory variables. A common model form for this purpose is a probit.
2. Using the estimated participation model, calculate for each participant and each nonparticipant
   a. The predicted participation probability from the fitted model $\widehat{D}_{\mathbf{j}}$
   b. The "Inverse Mills Ratio" IMR_j, if the prediction model is a probit, or the analogous term if the prediction model is based on other assumptions. (The IMR formula is given in Section 5.3.)
3. Estimate the primary regression equation with these two changes:

[3] This effect is sometimes referred to as the effect of encouragement. However, it is important to recognize that it is the effect of the program, for those who would not otherwise have joined. It is not the effect of encouragement alone. Thus, we emphasize that it is a program effect, not an encouragement effect.

a. Replace the participation dummy D_j in the primary regression equation by the predicted participation probability $\widehat{D}_{\mathbf{j}}$ from the estimated participation model.

b. Include an extra term in the primary regression that is the product of the predicted participation probability $\widehat{D}_{\mathbf{j}}$ and the IMR_j.

That is, fit the resulting regression model

$$\Delta_j = a - b\widehat{D}_{\mathbf{j}} - c\widehat{D}_j IMR_j + \nu_j^*$$

4. Calculate average net savings per participant from the fitted model as the average over all participants of the estimated participation terms. That is,

$$\overline{net_P} = \hat{b} + \hat{c}\overline{IMR_P}$$

where $\overline{IMR}_P$ is the average of IMR_j over participants. We call the procedure IV-IMR because it uses both instrumental variables (replacing the participation dummy with the probability), and an inverse Mills ratio. The participation model (and the associated IMR parameters) are most credibly estimated when there is one or more variable that affects the participation decision but does not otherwise affect (change in) energy consumption. Reasons that both terms are needed in general are described in the later sections.

The specific method presented here, using a probit model and the IMR, assumes a normal distribution of underlying drivers of consumption change and of participation. Alternative model forms can be used for different distributional assumptions. There are theoretical reasons to believe the method will not be highly sensitive to departures from normality. Simulation results presented in the paper support this conjecture. The paper provides extensions to allow more explanatory variables that affect the change in consumption and net savings. The principles and the key steps remain the same. The new method remains to be tested for practical trade-offs between biases of different sources and variance.

2.2 A partial correction: IV only, without IMR

A simple method that is sometimes used to address the self-selection problem is the same as above, but without the additional term involving the IMR. The method then is simply use of an instrumental variable $\widehat{D}_{\mathbf{j}}$ in place of the participation variable D_j. This IV-only method produces an unbiased estimate of net savings per participant <u>only</u> if it is reasonable to assume that participation is unrelated to the net savings a customer would obtain if they join the program. This assumption probably does not hold for most programs.

For those customers who would adopt the program measure without the program, net savings is zero, while for those who would not otherwise adopt the program measure net savings is the gross savings of the measure. Thus, to assume that participation is unrelated to the net savings that will be obtained is to assume that participation is unrelated to the natural adoption tendency. A more likely assumption for most programs is that natural adopters will be more attracted to the program than those who will need to take additional action and incur additional costs.

2.3 The importance of good predictors of participation

The IV-IMR procedure described in section 2.1 provides an unbiased net savings estimate (subject to the assumed normal distribution) regardless of what variables are available to explain participation probability. However, if the participation model is not very informative, the estimated net savings coefficients will have high variance. Obtaining well-determined net savings by either IV-only or IV-IMR methods requires a model of participation that itself has good predictive power. Moreover, if the participation predictors are also direct explanatory variables for the change in consumption, the regression estimates become more sensitive to the assumed probit (or other) distribution assumption to separate the effect of participation from the direct consumption effects. Thus, the ideal situation is one where there are strong predictors of participation that are not also direct drivers of consumption, or close correlates of such drivers.

2.4 The challenge of obtaining data for key participation drivers

A key factor affecting program participation is the applicability of the program measures. If the program measure would make no sense for a group of customers, it's hard to make a case that those customers can account for what participants would have looked like without the program, including the effect of natural adoption of the measure. For example, if savings for a furnace replacement program are to be calculated relative to standard efficiency equipment, the comparison group would ideally consist only of nonparticipating customers who are replacing their furnaces. If savings are calculated relative to existing equipment, the ideal comparison group is other customers whose furnaces are close to needing replacement.

The framework developed in this paper is built from specific representations of participation probability, naturally occurring savings, and potential net savings. In theory, if these models are sufficiently informative then measure applicability is in principle reflected in these models. In practice, however, these models tend to be fairly blunt tools. Pre-screening on applicability could be a more direct way to establish a suitable comparison group, but requires conducting a survey and relying on respondent recall to collect such information from a large pool of nonparticipants. As a result, most post-hoc comparison group selection methods are unlikely to directly account for measure applicability. In the absence of this information, the participation model is left to account for such effects via other variables. A key driver of participation probability—such as the need for the replacement equipment the measure applies to—will be omitted from the prediction model. The result can be a weak participation model, with the associated poorly determined net savings coefficients.

2.5 Use of Randomized control trials (RCT)

If customers can be assigned randomly to be program participants or not, there is no role for self-selection and no potential for self-selection bias in a simple difference-in-differences design. However, randomization of program eligibility is not consistent with the way most programs are delivered. Usually, customers cannot be forced to participate in a program. And even when participation can be required for some customers, denying participation to other customers is often politically or ethically difficult. Situations where net savings estimation is a challenge are precisely those situations where program participation is voluntary.

2.6 Use of Random encouragement design (RED)

A related "random encouragement design" (RED) manipulates the probability of participation, versus participation status directly. The program is available to all customers, but a randomly assigned subset of customers receive extra encouragement which increases the probability that these customers will elect to participate. A RED provides an unbiased estimate of the net savings of customers who were induced to join the program because of the encouragement.

A key point that is sometimes overlooked in RED analysis is that this framework does not in general provide an estimate of the net savings for participants who did not need the encouragement to join (i.e., the participants who did not receive encouragement and those who received encouragement but would have joined anyway.) This point means that adding an RED to an existing program will not ordinarily provide net saving for the existing program. However, RED can be useful in two ways for obtaining net savings for all participants:

- For some programs, net savings can be assumed to be the same for all participants, whether or not the encouragement induced them to join. In this case, the net savings for the participants who joined because of the encouragement, which RED estimates, is applicable to all participants.

- RED creates variation in customers' probability of participating, with customers who were encouraged having a higher probability of participating than those who were not encouraged. This variation is useful for estimating the IV and IV-IMR models described above. Specifically, RED creates variation in $\widehat{D}_{\mathbf{j}}$ and IMR_j, which improves the estimation of the corrected regression equation.

In principle, the participation model could predict the participation decision solely as a function of the encouragement assignment indicator. However, unless net savings is the same for both encouraged and not-encouraged participants, it is important to have good explanatory variables for participation *in addition to* the randomly assigned encouragement. If only the RED indicator is available to explain participation, the participation model cannot be informative as to the relationship between net savings and participation absent encouragement. This is the self-selection relationship that needs to be addressed to obtain net savings for the not-encouraged participants—that is, for the base existing program.

If the RED is used to obtain net savings for an ongoing program, it is the net savings absent the encouragement that is of interest. Since it is unlikely that the customers who required extra encouragement to join have the same natural adoption rate as those who join without extra encouragement, the net savings for the encouraged and not-encouraged groups will typically be different. Thus, for example, if without the RED we would want the participation model to include variables such as income and education, or neighbourhood averages of these from Census data, those variables should still be included if we do have an RED.

2.7 Billing analysis for gross savings

This paper focuses on estimation of net savings. However, all the methods described here are applicable to estimation of gross savings, with the dummy variable defined as indicating measure adoption rather than program participation.

3 SPECIFICATION FOR NET SAVINGS ESTIMATION

To establish concepts and terminology, we begin by considering a standard regression framework. A general approach to estimating the causal effect of an energy efficiency program or intervention on energy consumption is to regress consumption or change in consumption on a set of explanatory variables including program participation. Data from both participants and a comparison group of nonparticipants are included in the regression. The regression is often structured as a panel or pooled time series-cross-sectional regression, where each observation corresponds to a customer and a time period. For expositional clarity, we consider the cross-sectional analog to this kind of panel data analysis. Further, we start with a simple explanatory form, describing change in consumption as a function of participation only.

Regardless of the details of the structure, the maintained assumption is that energy consumption among the comparison group (or the model parameters estimated using this comparison group) provides an unbiased estimate of what energy consumption (or change in consumption) among participants would have looked like absent the program.

In its simplest form, the regression for net savings analysis using a participant/non-participants comparison is

(1) $\Delta_j = a - b\, D_j + \nu_j$

where

Δ_j = change in annual energy consumption for household j

D_j = 0/1 indicator variable for whether household j participated in the program.

ν_j = residual error.

The change Δ_j is the difference between consumption in the later year (post) and the earlier year (pre) so that a positive value of Δ_j corresponds to an increase and a negative value to a decrease.

To explore the potential effects of self-selection into participant and comparison groups, we consider a decomposition of the consumption change for customer j, namely

(2) $\Delta_j = noc_j - net_j D_j$

where

noc_j = naturally occurring change for customer j.

net_j = the net savings customer j will have IF customer j participates in the program, which we call the **potential net savings**.

If customer j would adopt the measures offered by the program even if the program didn't exist, noc_j includes the effect of the measure adoption, and $net_j = 0$. If customer j would adopt only if they participate in the program, net_j = the gross savings customer j will have if they adopt the measure. If customer j would take some energy-saving actions without the program, but less than their full gross savings when they do participate, net_j is something in between 0 and full gross savings. Importantly, the "potential net savings" -- net_j -- exists for both participants and comparison group customers, but is realized only by participants. For non-participants, it is the net savings that they *would have* obtained if they had chosen to participate.

As indicated, the potential net savings net_j is not necessarily discrete. There could be a range of net effects from 0 through full gross savings, and gross savings itself may have a range of values. For purposes of this paper, both potential net savings and gross savings for customer j include any rebound effect, as well as any participant spillover that results from adopting the measure, but as noted, nonparticipant spillover is assumed to be zero.

To flesh out the possibilities, we represent the naturally occurring savings for customer j as the average over the entire eligible population, including both participants and nonparticipants, plus the difference between customer j's value and the population average, and similarly for potential net savings. That is

(3) $noc_j = a + \varepsilon_j$

$net_j = b + \psi_j$

where a and b are the respective unknown population averages, and the deviations from the averages ε_j and ψ_j are random with zero mean in the population of all customers. At this point, we make no assumptions about the distribution of the random elements ε_j and ψ_j. Whatever the distribution of naturally occurring change and of potential net savings, the unknown parameters a and b are the corresponding averages over the population of customers (both participants and non-participants), and the random components are simply the difference between a given customer's value and the corresponding population mean. We use the convention that positive savings represents a reduction in consumption, so that the coefficient b is assumed to be positive, and consumption would be reduced by this amount on average if all customers in the population participated.

With this framework, Eq. (2) becomes

(4) $\Delta_j = a - bD_j + \varepsilon_j - \psi_j D_j$.

We rewrite Eq. (4) as Eq. (1) (copied here for convenience)

(5) $\Delta_j = a - bD_j + \nu_j$

with

$\nu_j = \varepsilon_j - \psi_j D_j$.

The coefficient b is the average potential net savings over participants *and* nonparticipants. The average realized net savings among all participants is the average of $(b + \psi_j)D_j$ over all participants.

(6) $$\overline{\boldsymbol{net}_P} = \overline{\boldsymbol{b} + \boldsymbol{\psi D}} = \boldsymbol{b} + \overline{\boldsymbol{\psi}}_{\boldsymbol{P}}$$

That is, the average savings among participants, which is what we want to estimate, is the same as the coefficient b only if the random component ψ_j is zero on average over all participants. (Note that, by the definition in (3), ψ_j is zero on average over all customers in the population, including participants and non-participants. It need not be zero on average over participants.) As will be seen, under certain circumstances, simple analysis will provide an unbiased estimate of $\overline{net_P}$ whether or not it is the same as the population mean potential net savings b. Under other circumstances, additional steps are required.

4 PREVIOUS APPROACHES TO NET SAVINGS ESTIMATION

In this section, we describe several prominent approaches that have been used for estimating net savings. We have two purposes here. First, we want to show the conditions under which these earlier approaches provide valid estimates of net savings. Under these conditions, the new method that we describe in this paper is not needed: the earlier method can be used instead. Usually, the earlier methods are easier, and so it is advantageous to use them whenever possible. Second, we want to describe the conditions under which these earlier methods do not provide a valid estimate of net savings. The new method is useful in these situations.

4.1 Difference in Differences

The regression estimate of net savings from Eq. (1) using ordinary least square regression is the same algebraically as the Difference of Differences (DID) estimate

(7) $$\Delta\Delta = \overline{\Delta}_P - \overline{\Delta}_C$$

where the subscripts P and C, respectively, denote the participant and comparison groups, and the bar over the term indicates average over the indicated group. That is, both the regression formula (1) and the DID estimator (7) will yield the same estimate of net savings.

We can rewrite the DID estimate as:

(8) $$\begin{aligned}\Delta\Delta &= \overline{\Delta}_P - \overline{\Delta}_C \\ &= (\overline{noc}_P - \overline{net}_P) - \overline{noc}_C \\ &= -\overline{net}_P + (\overline{noc}_P - \overline{noc}_C)\end{aligned}$$

The program net savings is the potential net savings for those who do in fact participate, whose average is $\overline{net_P}$. Equation **(8)** yields the desired quantity $\overline{net_P}$ only when the average naturally occurring savings is the same for participants and non-participants. However, Eq. **(8)** does not require that potential net savings be the same for participants and nonparticipants.

We can now consider two ways that the DID estimate has been used: research designs that implement random assignment, those that rely on naturally occurring variation in program participation.

4.1.1 DID with RCT

Consider a randomized control trial (RCT) in which customers are randomly assigned into two groups, a group of participants and a group of non-participants. This form of assignment constitutes the classic design for controlled scientific experiments. Since the two groups are determined randomly, it is reasonable to expect that the naturally occurring savings is the same on average for the two groups. In this case, the difference in differences provides an unbiased estimate of the net savings of program participants: $\overline{noc}_P = \overline{noc}_C$ such that the DID in eq. (**8**) becomes $-\overline{net}_P$.

Note that this RCT design requires a mechanism for ensuring compliance with random assignment, such that customers who are assigned to be participants are actually participate and customers who are assigned to be non-participants do not participate. For some programs, assignment can be straightforward. Programs that send home energy reports to customers are an important example: customers who are sent a report are participants, and those who are not sent a report are non-participants. The utility creates an RCT design by sending reports to some randomly-selected customers and not sending reports to other randomly-selected customers. The only difficulty that might arise is that customers who were assigned to the non-participant group must be denied a report even if they request it. For other programs, however, some action by the customer is required in order for the customer to be a participant. In these cases, assignment to participation and non-participation groups can be difficult or even conceptually impossible. For example, to create an RCT design for a rebate program, the customers in the participant group must be somehow required to adopt (and pay for) a measure that would qualify for a rebate. And customers in the non-participant group must not obtain rebates. If they adopted a measure that qualifies for a rebate, they would not be allowed to obtain the rebate and, importantly, they must know beforehand that they would not obtain the rebate. Even if these conditions could be enforced, the RCT design would provide an estimate of the net savings from a program that forces customers to take specified actions, rather than the net savings from a program that promotes action.

4.1.2 DID without non-random assignment

Usually, participation in efficiency programs is voluntary: customers decide themselves whether they want to participate or not, or, more directly, whether they want to take the actions that qualify them for participation. With voluntary participation, it is doubtful that the naturally occurring change would be the same for participants and non-participants. For example, customers who are planning to buy high efficiency appliances even without a program are probably more likely to join a rebate program, in order to get the rebate, than customers who were not planning to buy any high efficiency appliances. Their naturally occurring change in consumption is therefore lower -- more negative -- reflecting the savings from the high efficiency measures that they would have taken without the program.

Since naturally occurring change of the two groups is not the same, the DID estimate in equation (**8**) does not equal $-\overline{net}_P$. Without random assignment, the DID estimate, and likewise the regression estimate of net savings, has error equal to the difference in naturally occurring change between participants and non-participants. The DID or simple regression estimator is biased by the amount of the expected difference in naturally occurring savings, which includes the differential rate of natural adoption between the two groups. Our new method corrects for this bias.

4.2 Random Encouragement Design

Another evaluation strategy that makes use of random assignment is the Random Encouragement Design (RED). Customers are randomly assigned to receive or not receive special encouragement to participate.

For convenience, we use the terms "encouraged group" and "not-encouraged group." The specific form that this encouragement takes can vary significantly across settings. It could be as simple as an informative phone call, or it could involve a much more effortful campaign to encourage targeted consumers to participate. The program participation rate is compared for the two groups, with the expectation that the encouragement induced more customers to participate, such that the participation rate is higher in the encouraged group than the not-encouraged group. The change in consumption of customers for the two groups is compared. Any difference is attributable to the encouragement-induced increase in program participation, since the two groups are randomly assigned and hence otherwise the same.[3] This information can be used, as shown below, to estimate the net program savings for the customers who were induced by the encouragement to participate.

In section 4.1 above, we considered a standard DID estimator that compares participants with non-participants. For RED, a DID estimator is also used, but with different groups being compared. In particular, rather than considering participants versus nonparticipants, RED compares the encouraged customers with the not-encouraged customers. The DID estimator becomes:

(9) $$\Delta\Delta = \overline{\Delta}_E - \overline{\Delta}_0$$

where the subscript E denotes the encouraged group and the subscript 0 indicates the group not-encouraged group. For each of these two groups of customers, the average consumption change is the sum of the average of the two components of Eq. (2), such that:

(10) $$\begin{aligned}\Delta\Delta &= (\overline{noc}_E - \overline{net}_E\overline{D}_E) - (\overline{noc}_0 - \overline{net}_0\overline{D}_0) \\ &= -(\overline{net}_E\overline{D}_E - \overline{net}_0\overline{D}_0) + (\overline{noc}_E - \overline{noc}_0)\end{aligned}$$

where $\overline{D}_E$ is the share of customers in the encouraged group who participated, $\overline{net}_E$ is the average net savings for participants in the encouraged group, and similarly for the not-encouraged group with subscript 0. In the second line of Eq. (10) the first term in parentheses is the difference in average realized net savings between the encouraged and not encouraged groups. Because of the random assignment, the second term in parentheses -- the difference in average naturally occurring change for the two groups – can be expected to be zero. With this zero difference in naturally occurring change, the DID estimator for the RED becomes the difference in average realized net savings between the encouraged and the not encouraged group:

(11) $$\Delta\Delta = -(\overline{net}_E\overline{D}_E - \overline{net}_0\overline{D}_0)$$

This is the impact of the encouragement on the average change in consumption.

The impact of the encouragement on the share of customers who participate is the difference in participation rates between the encouraged and non-encouraged groups: $R_E = \overline{D}_E - \overline{D}_0$. This is an estimate of the share of customers in the encouraged group who were induced by the encouragement to participate (and would not have participated without the encouragement.) The average savings of these extra participants (that is, of the customers who were induced by the encouragement to participate) is the extra

[3] This effect is sometimes referred to as the effect of encouragement. However, it is important to recognize that it is the effect of the program, for those who would not otherwise have joined. It is not the effect of encouragement alone. Thus, we emphasize that it is a program effect, not an encouragement effect.

savings induced by the encouragement divided by the share of customers who were induced by the encouragement:

(12) LATE = $\Delta\Delta$ / R_E.

This is the average net savings of the customers who were induced by the encouragement to participate. In the statistics literature, this is called the "Local Average Treatment Effect" (LATE), but the term needs to be translated appropriately to be meaningful in the current context. The word "local" refers to observations on the margin. In our context, "local" refers to the customers who were induced by the encouragement to participate and would not have participated without the encouragement. The word "treatment" refers to the program, not the encouragement. So, the "local average treatment effect" is the average effect of the program ("treatment") on the customers who were induced by the encouragement to join the program (the "local" customers).

The LATE from RED is an unbiased estimator of the average net savings per participant who was induced by the encouragement to participate. The question arises: when is this a valid estimate of the net savings of the program?

From Eq. (11), we see that the LATE calculation (12) gives us

(13) LATE $= \Delta\Delta\ /\ R_E = -(\overline{net}_E \overline{D}_E - \overline{net}_0 \overline{D}_0) / R_E$

There is an important situation for which the RED LATE estimator provides net savings for the program. Suppose that net savings is the same for participants who participated because of the encouragement as for participants who would have participated without encouragement. In this case, Eq. (11) becomes:

(14) $\Delta\Delta = \overline{net}_P (\overline{D}_E - \overline{D}_0)$

and the LATE estimator becomes:

(15) LATE $= -\overline{net}_P (\overline{D}_E - \overline{D}_0) / (\overline{D}_E - \overline{D}_0) = -\overline{net}_P$.

That is, when the average net savings per participant is the same for the participants who were induced by the encouragement as for those who would have participated without encouragement, the RED's standard LATE calculation provides an unbiased estimate for this uniform net savings per participant.

Low-income home weatherization programs are an important example of this situation. Low-income households might not be able or willing to incur the expense of weatherizing their homes without the program. Then the net savings of the program are simply the gross savings of the weatherizations, since no customers would have weatherized without the program. Furthermore, weatherization perhaps provides the same savings for households who joined the program because of the encouragement as for those who did not need the encouragement in order to join.

For most programs, however, it is unlikely that participants who did not need the encouragement to participate would have the same net savings as participants who were induced by the encouragement to participate. In particular, the free-ridership rate is unlikely to be the same among participants who would have joined even without the encouragement as among participants who needed to be additionally encouraged to join. A far more likely situation is that those who would install measures on their own would be more likely to participate in the first place without extra encouragement, and the encouragement would be needed for those who are less likely to install the measure on their own. In this situation, the RED estimate from Eq. (12) would overstate the net savings of the program.[4] In Section 5 we will consider methods to address these challenges.

One additional note maybe needed before moving to the next section. We have described above how the DID estimator takes a different form with an RED design than the participant-nonparticipant difference given by Eq. (7). The corresponding regression formulation also takes a different form. The RED estimator can be interpreted as an instrumental variables (IV) estimator of the regression equation. We give details of this interpretation in the appendix A. The IV interpretation is useful in our discussion below.

5 NEW PROCEDURE: IV-IMR FOR NET SAVINGS WITHOUT RANDOM ASSIGNMENT

As stated above, program participation is usually voluntary, which means that customers self-select into the participant group. The issue that determines the appropriate method of analysis is: what factors affect customers' decision to participate in the program? We use the specification for net savings described in Section 3.1, and consider three increasingly challenging situations.

1. Whether or not a customer participates **is not** related either to the customer's naturally occurring savings ***nor*** to the net savings the customer will get if the customer participates.
2. Whether or not a customer participates **is** related to the customer's naturally occurring savings, but **not** to the net savings the customer will get if the customer participates.
3. Whether or not a customer participates **is** related **both** to the customer's naturally occurring savings **and** to the net savings the customer will get if the customer participates.

5.1 Participation is unrelated to the customer's naturally occurring savings and potential net savings

If there's no relation between participation and naturally occurring savings, and no relation between participation and potential net savings, the participant-nonparticipant DID estimator of average net savings, or the corresponding regression estimate from Eq. (1), is unbiased. This is the condition that says the non-participant group is essentially the same as the participant group apart from participation itself, aside from random differences that are zero on average. Thus, there are no self-selection effects to be controlled or corrected for. As discussed in Section 3.1, for most programs these assumptions are difficult to justify outside of RCT assignment.

[4] There is another consideration with RED that warrants mentioning. In particular, participants who did not need the encouragement to participate might nevertheless be induced by the encouragement to take more actions than they would have without the encouragement. In this case, the realized savings for those who would participate anyway is affected by the encouragement. The DID with RED gives an estimate of the overall encouragement effect, including the extra savings of customers who were *not* induced by the encouragement to participate but took more measures because of the encouragement. As a result, eq. (12) overestimates the average effect on the customers who participated because of the program.

5.2 Participation is related to the customer's naturally occurring savings but not to their potential net savings

5.2.1 Instrumental Variables Correction

Now we suppose that the participation decision is related to naturally occurring savings but is independent of potential net savings. The assumption that participation is independent of potential net savings means that the participation decision does not depend on potential net savings directly, and also that potential net savings is not correlated with naturally occurring savings. Participation related to naturally occurring savings would arise, for example, if customers who tend to conserve energy year over year in other ways are more likely to join the program than other customers. In that case, we would expect a greater reduction in consumption to be associated with those who choose to participate than with those who do not, apart from any effects of participation itself. The result would be to overstate net savings. The opposite direction of bias would be expected if people who take more energy efficiency actions on their own are less likely to participate.

To develop an unbiased estimator for this situation, we first fit a model that predicts participation as a function of a set of observable customer characteristics $\mathbf{z_j}$, where those characteristics are uncorrelated with the variable component ε_j of naturally occurring savings in Eq. (3). Denote by $\widehat{D}(\mathbf{z_j})$ the predicted participation probability for customer j based on this estimated model. We then fit, in place of Eq. (4) or Eq. (1), the regression equation

(16) $\Delta_j = a - b\widehat{\boldsymbol{D}}(z_j) + \nu_j^*$

where the error becomes

$$\nu_j^* = \nu_j - b(D_j - \widehat{D}(\mathbf{z_j})) = \varepsilon_j - D_j\psi_j - b(D_j - \widehat{D}(\mathbf{z_j}))$$

This procedure is called instrumental variables (IV) regression where the variables in $\mathbf{z_j}$ are instruments that explain participation.

If the instruments in $\mathbf{z_j}$ are uncorrelated with ε_j, as stated above, then the participation residual $D_j - \widehat{D}j$ has zero conditional mean by construction. We are assuming, in the current situation, that net savings are unrelated to participation and hence is uncorrelated with the participation residual. As a result, the predictor $\widehat{D}$ is uncorrelated with all the components of the residual ν_j^* and the regression will give an unbiased estimate of the average net savings *b*. Because net savings are unrelated to participation (by assumption), the net saving for participants are the same regardless of participation probability $\widehat{D}(\mathbf{z_j})$).

The regression equation (16) thus provides an unbiased estimate of the average net savings per participant. It is important to re-iterate two important caveats underlying this result. First, this interpretation requires the strong assumption that participation in the program is unrelated to the magnitude of net savings that will be realized if the customer does participate. As discussed earlier, this assumption will not be satisfied in many contexts. Second, for this approach to provide meaningful results, we need good explanatory variables for the participation decision. This requirement is discussed further in relation to adding other explanatory variables to the primary equation (16).

Importantly, if an RED was implemented, this provides a potentially useful instrument to use in the estimation of the participation equation. A dummy variable indicating assignment to the encouragement will be uncorrelated with ε_j by design. The variation in $\widehat{D}(\mathbf{z_j})$ that is induced by the encouragement can be used to estimate the coefficients in equation (16). Even absent a RED, the participation probabilities predicted using the participation equation can vary over customers in a way that supports the identification of average net savings in Equation (16). The key challenge is isolating variation in participation that is independent of ε_j.

5.2.2 Adding other explanatory variables

It may seem counterintuitive to assume as in Eq. (16) that there are variables $\mathbf{z_j}$ that explain participation, but are uncorrelated with naturally occurring savings, while naturally occurring savings is itself correlated with participation. The assumption becomes more plausible if we replace the simple population mean of naturally occurring savings in Eq. (3) by a more informed model

(17) $noc_j = x_j\alpha + \varepsilon_j$

where $\mathbf{x_j}$ is a set of explanatory variables that is also included in the set $\mathbf{z_j}$, α is a corresponding set of coefficients, and ε_j is still a random deviation with zero mean. Thus, $\mathbf{z_j}$ is related to noc_j through the common explanatory variables $\mathbf{x_j}$, but there is no relationship between the noc residual ε_j and either the participation predictors $\mathbf{z_j}$ or the corresponding predicted participation $\widehat{D}(\mathbf{z_j})$. The expanded expression for the observed change corresponding to Eq. (4) then becomes

(18) $\Delta_j = \mathbf{x_j}\alpha - bD_j + \nu_j$

and the unbiased regression equation corresponding to Eq. (16) becomes

(19) $\Delta_j = \mathbf{x_j}\alpha - b\widehat{D}(\mathbf{z_j}) + \nu_j^*$

With this structure, OLS regression will provide an unbiased estimate of the coefficient b, net savings per participant, *provided* that participation is unrelated to potential net savings and provided that valid instruments for participation are available. The ideal instrument is a variable that is a good predictor of participation, but does not otherwise affect energy consumption. In practice, it may be difficult to find variables that belong in the participation predictor set **z** and do not also belong in the primary equation's explanatory variables **x**. If the only available predictors of participation are also included in the primary equation's explanatory variables, the estimated coefficient b will be sensitive to the particular functional form used to specify the participation model $\widehat{D}(\mathbf{z_j})$. That is, the estimation result will rely on the probit or other distributional assumption to make it possible to distinguish between effects on participation and direct effects on consumption. In such a case, it would be important to at least test for the robustness of the estimated net savings under alternative distributional assumptions.

5.3 Participation is related to both the customer's naturally occurring savings and their potential net savings

A less restrictive set of assumptions is that a customer's decision to participate or not is related to both naturally occurring savings noc_j and the potential net savings net_j. With these assumptions, Eq. **(19)** no longer provides an unbiased estimate of average net savings.

The reason is that replacing the participation dummy by predicted participation $\widehat{D}(\mathbf{z}_j)$ isolates the variation in participation that is uncorrelated with the residual variation ε_j in noc_j, but expected participation $\widehat{D}$ is still correlated with $\psi_j D_j$, which is part of the error ν_j^* in the regression. The correction in this case depends on the distribution of the error terms, including the error term in the underlying driver of participation.

The structure we consider is model (17) for naturally occurring change, together with net savings in the simple form of Eq. (3). We will relax the latter assumption later. We also assume that customer j participates in the program if the unobservable program attractiveness $U_j > 0$, and otherwise not, where U_j is itself a function of net savings, naturally occurring savings, and other observable variables **z**, plus a random error term η_j.

5.3.1 IMR Correction with normally distributed error terms

A relatively straightforward estimation is available under the assumption that all the error terms are independent, normally distributed with zero means. That is, all of the following are assumed to be normally distributed:[5]

- the deviation ε_j in Eq. (17)
- the deviation ψ_j in Eq. (3)
- the deviation η_j in U_j.

With this assumption, the regression becomes

(20) $\Delta_j = \mathbf{x_j}\alpha - b\widehat{D}(\mathbf{z_j}) - c\widehat{D}(z)\,\phi(\mathbf{z_j}\gamma)/\Phi(\mathbf{z_j}\gamma) + \nu_j^{**}$

Where

$\mathbf{x_j}$ is the set of linear predictors of naturally occurring change from Eq. (17)

$\mathbf{z_j}$ is the set of predictors of participation in an estimated probit model

$\widehat{D}(\mathbf{z_j})$ = the predicted participation probability from the fitted probit model

$= 1 - \Phi(-\mathbf{z_j}\gamma)$

α = coefficients determined by the regression (20)

γ = coefficients determined by the probit fit

$\phi(\,)$ denotes the probability density function of the standard normal distribution

$\Phi()$ denotes the cumulative distribution function of the standard normal distribution

The ratio

$IMR(\mathbf{z_j}\,|D_j = 1) = \phi(\mathbf{z_j}\gamma)/\Phi(\mathbf{z_j}\gamma))$

$IMR(\mathbf{z_j}\,|D_j = 0) = -\phi(\mathbf{z_j}\gamma)/\Phi(-\mathbf{z_j}\gamma)$.

is known as the Inverse Mills Ratio. Thus, the regression (20) can be written as

(21) $\Delta_j = \mathbf{x_j}\alpha - b\widehat{D}(\mathbf{z_j}) - c\,\widehat{D}(\mathbf{z_j})\,IMR(\mathbf{z_j}, D_j = 1) + \nu_j^{**}$

Eq. **(21)** is the same as Eq. (19), with the addition of the additional term involving the product of the predicted participation and the IMR. The derivation of this formula is provided in Appendix B. To simplify the notation, we use the shorthand

$IMR_j = IMR(\mathbf{z_j}\,|D_j = 1)$.

The use of predicted participation by itself in place of the participation dummy breaks the correlation between the participation term and naturally occurring savings. Predicted participation interacted with the IMR corrects for the correlation between participation and net savings. And, again, a RED is useful in providing variation in the participation probability and the inverse Mills ratio, which assists in estimation of the coefficients b and c.

---

[5] For each set of error terms ε_j, ψ_j and η_j, a fixed variance is also assumed—that is, each error component is a set of independent identically distributed variables. The 3 variances are not assumed to be the same.

5.3.2 Relation to Double Mills Ratio

Earlier work several years ago (XENERGY, 1996) addressed self-selection in billing analysis for net saving with a "Double Inverse Mills Ratio" method. That approach, like the IV-IMR method, included terms involving both predicted participation and the IMR in the regression. The IV-IMR method introduced here has advantages over the earlier method. The IV-IMR method has a somewhat simpler form, and directly utilizes the IV term as a predictor. This term has a familiar interpretation for those experienced with RED and other IV applications. Second, the IV-IMR method incorporates the IMR term in a different way that turns out to be more robust. Appendix D describes the relation between the two methods in more detail.

5.3.3 Extension to Statistically Adjusted Engineering Estimates

An alternative formulation of the regression for net savings uses a Statistically Adjusted Engineering (SAE) approach. The SAE approach uses estimated gross savings in place of a simple participation dummy in the regression equation (1) or **(17)**. The coefficient of the gross savings can be interpreted as a net-to-gross ratio, rather than as average net savings. Appendix C describes the basic extension of the IV-IMR to the SAE structure,

6 SIMULATION RESULTS FOR THE IV-IMR APPROACH

As discussed, the regression formula (20) with the combined correction terms including the participation probability by itself and interacted with the IMR is unbiased under the assumptions of normally distributed errors. We describe below simulations that adhere to the assumptions for this approach. We report the effects of varying the magnitude and direction of correlations. Then, to explore the sensitivity of the method to the distributional assumption, we simulated naturally occurring and potential net savings under non-normal distributions.

The general structure follows Eqs. (2) and (3) above

$$\Delta_j = noc_j - net_j D_j$$

$$noc_j = a + \varepsilon_j$$

$$net_j = b + \psi_j$$

The participation decision is based on the program attractiveness, an unobservable quantity denoted by U. Customer j participates or not, meaning $D_j = 1$ or $D_j = 0$, for $U_j > 0$ or $U_j \leq 0$, respectively. It's this underlying structure of the participation decision that gives rise to the distributions that determine the form of the correction terms, as described further in Appendix B.

6.1 Simulations with distributions matching the regression assumptions

Our results show that when the assumptions for the method are true, the regression method (20) does produce close to the true net savings, for sufficiently large sample size. By varying the sample size, the correlations, and the relative magnitude of the random terms, we see the effects of these factors on the stability of the estimates.

In all the scenarios examined, the propensity to participate (program attractiveness) is negatively related to naturally occurring savings and positively related to potential net savings.

Case A: Positive relation between net savings and participation, net savings uncorrelated with naturally occurring change

Data are generated according to the following specifications:

Mean naturally occurring savings noc	a = -1
Mean net savings	b = 1
Unobservable program attractiveness	$U_j = -4 - 1 \times noc_j + 1 \times net_j + 1 \times z_j + \eta j$
Participation driver	z_j ~ Uniform between -2 and 2
Random variation of noc distribution	$\varepsilon j \sim N(0; 1)$
Random variation of net distribution	$\psi j \sim N(0; 1)$
Random variation of U distribution	$\eta j \sim N(0; 1)$
Correlation of noc and net	$E(\varepsilon_j \psi_j) = 0$

With a = -1, the mean naturally occurring change noc is negative, meaning that customers reduce their consumption outside of the program on average. Net savings has a mean of 1 over all customers, which implies that, if all customers participated in the program, the average change in consumption would be -2 (noc of -1, with average net savings = 1). The attractiveness of the program, U_j , depends on noc_j with a negative coefficient, such that customers who reduce their consumption more outside of the program (i.e., noc_j is more negative) are more likely to join the program. Potential net savings enters U_j with a positive coefficient, meaning that customers with larger potential net savings are more likely to join the program. With these parameters, the share of customers who participate in the program is about 17 percent (varying with the draws of the error terms.)

For each simulation:

- Data were generated by the specification above.
- A binary probit model of participation was estimated by maximum likelihood.
- The probability of participating and the inverse Mills ratio were calculated from the probit model;
- A regression following Eq. (21) was estimated by OLS. That is, the regression had Δ_j as the dependent variable, and explanatory variables a constant, the negative of the probability of participating, and the inverse Mills ratio times the negative probability of participating.
- Net savings was estimated as the sum of the latter two terms multiplied by their estimated coefficients.

To see the effect of sample size, simulations were run with 1000, 10,000, and 100,000 sampled customers in the estimation dataset. For each sample size, the data were generated and net savings was estimated a total of 100 times, using different draws of the random terms.

The results are given in the first panel of Table 1. With a sample size of 1000, true net savings per participant, averaged over the 100 runs, is 1.6988. As specified above, the mean net savings in the population (i.e., if all customers participated) is 1. Customers with higher net savings chose to participate in the program, such that the mean net savings among participants is 1.6988 rather than 1.000. The estimated net savings per participant, averaged over the 100 runs, is 1.6995, which differs from the true net savings by only 0.007.

The RMSE is the square root of the average, over the 100 runs, of the squared difference between true and estimated net savings in each run; this statistic is a measure of the expected magnitude of difference between true and expected net savings in any one run. The value of 0.4790 means that, with a sample of 1000 customers, the net savings estimated by the procedures is expected to differ from the true net savings by 0.4790. This RMSE is fairly large with only 1000 customers in the sample, implying that the estimate is expected to be “off” by 28 percent in any given application. With larger sample sizes, the RMSE drops, as expected, such that with 100,000 customers in the sample, the procedure is “off” by only 3.3 percent on average (0.0574/1.7146).

Table 1: Simulation Results: Data Generation Process Matching Assumptions

Participation positively related to net savings and negatively related to naturally occurring change

Sample size	Net Savings per participant			
	True	Estimated	Difference	RMSE
A: Net savings uncorrelated with naturally occurring change				
1,000	1.6988	1.6995	0.0007	0.4790
10,000	1.7153	1.7366	0.0214	0.1541
100,000	1.7146	1.7042	-0.0104	0.0574
B: Net savings negatively correlated with naturally occurring change				
1,000	0.6426	0.6554	0.0128	0.2256
10,000	0.6467	0.6533	0.0066	0.0737
100,000	0.6467	0.6442	-0.0025	0.0246
C: Net savings uncorrelated with naturally occurring change, Less variation in observable participation driver				
1,000	1.7909	1.5197	-0.2712	8.1034
10,000	1.8009	2.1137	0.3127	2.4434
100,000	1.8027	1.6628	-0.1398	0.7782

Case B: Positive relation between net savings and participation, negative relation between net savings and naturally occurring change

The same data generating process is specified, except that net savings is correlated with naturally occurring savings. The attractiveness of the program still depends on noc_j and net_j as before, with the customer finding the program more attractive when they would have reduced consumption anyway without the program and when the potential net savings of the program are larger. But in the current situation, net savings are smaller for customers who would have reduced their consumption more without the program.

The data generation process (dgp) is the same as for example A but with these two changes:

$$E(\varepsilon_j \psi_j) = 0.9$$

$$U_j = -3 - 1 \times noc_j + 0.5 \times net_j + 1 \times z_j + \eta_j$$

In this set of simulations, net savings is very highly correlated (0.90) with the naturally occurring change in consumption. In U_j, the coefficient of net_j is changed from 1 to 0.5, which means that the attractiveness of the program depends less on potential net savings than on the naturally occurring change in consumption. Also, to keep the participation share about the same as in example A, the constant in U_j is changed from -4 to -3, which gives an average participation rate of about 19 percent.

The second panel of Table 1 gives the results. With sample size of 1000, the true net savings per participant is 0.6426, which is lower than the average net savings that would occur if all customers participated, because customers with a negative naturally occurring change in consumption and hence (given the correlation) low net savings tend to join the program. The estimated net savings is close to the true net saving, on average over the 100 runs. The RMSE is smaller in this situation than previously, and decreases, as expected, with sample size.

Case C: Less variation in the participation drivers

The procedure is highly dependent on variation over customers in the probability of participating. To examine this issue, the same data generation process as in example A is used, except that now the variable $\mathbf{z_j}$ is specified to range only from -0.5 to 0.5, rather than from -2 to 2. The unobservable components of program attractiveness ε_j, ψ_j, and η_j still have the same level of variability. To maintain a similar participation share with this change in z_j, the constant in U_j is changed from -4 to -3.5, which gives an average participation share of about 19 percent.

As shown in the third panel of Table 1, the procedure is far less accurate than in example A. These results indicate the importance in the procedure of variables that explain participation such that the participation probability varies considerably over customers. As discussed in Section 4.4.3, RED can assist in providing such variation, but it is still important to have variables that explain participation apart from the RED.

6.2 Data Generation Process Does Not Match the Model Assumptions

Case D: NOC and NET are distributed logistically

Instead of being normally distributed, noc_j and net_j are assumed to be distributed logistically. All other aspects of example A are the same. The simulation results are shown in the top part of Table 2. The average difference between true and estimated net savings is fairly small. RMSE is higher than when the distributional assumptions are met (part A of Table 1), but with 100,000 sampled customers, RMSE is less than 10 percent of the true savings.

Table 2: Simulation Results: Data Generation Process Not Matching Assumptions

Sample size	**Net Savings per participant**			
	True	**Estimated**	**Difference**	**RMSE**
D: NOC and NET are logistically distributed				
1,000	2.3887	2.4616	0.0730	1.9258
10,000	2.3847	2.2909	-0.0938	0.5194
100,000	2.3854	2.2750	-0.1104	0.2081
E: NOC and NET are uniformly distributed				
1,000	1.8605	2.1047	0.2442	0.7890
10,000	1.8563	1.9053	0.0490	0.2092
100,000	1.8532	1.9564	0.1033	0.1256
F: NOC is discrete and NET is zero				
1,000	0	0.0089	0.0089	0.1170
10,000	0	0.0038	0.0038	0.0377
100,000	0	0.0027	0.0027	0.0131

Case E: NOC and NET are distributed uniformly

In this scenario, noc_j and net_j are assumed to be distributed uniformly between -2 and 2. The results are similar to those with logistic errors: RMSE is larger than when the NOC and NET are normally distributed, but, with 100,000 sampled customers, RMSE is less than 10 percent of true net savings.

Case F: NOC is discrete and NET is zero

In this scenario, half of the customers, chosen randomly, have noc_j = -2 while the other half have noc_j = 0. Net savings are zero for all customers. Customers whose noc_j = -2 are more likely to join the program than customers with noc_j = 0, as might happen if the program offered rebates for measures the customers would have taken anyway. To maintain a similar participation share as in example A, the constant in U_j is changed from -4 to -2. As shown in the bottom panel of table 2, the true net savings are zero, as specified. The estimated net savings are close to zero, on average, for all sample sizes. RMSE is fairly small, especially with large sample sizes.

6.3 Summary of Simulation Results

The results displayed in Table 1, where the data follow the assumed normal distributions, indicate the following:

- The method does produce unbiased estimates.
- In Cases A and B, average net savings, average naturally occurring change, the range of variation in the observable participation driver z and the ranges of all the sources of unobservable variation are all of similar magnitude. In these cases, the RMSE is on the order of 1/3 of the estimate for a sample size of 1000.
- The relative accuracy of the method (ratio of RMSE to true value) is similar for Case A with no correlation between net savings and naturally occurring change as for Case B where these are very highly correlated.
- When less of the variation in program attractiveness is explained by the observable variable z_j Case C), the estimates still appear to be unbiased but the RMSE is much worse for a given sample size.
- For all the scenarios, the RMSE changes in proportion to the square root of the sample size, as expected.

The results displayed in Table 2, with various departures from the assumed normal distributions, indicate the following:

- The method still appears to have low bias for logistical and uniform distributions of net savings and naturally occurring change (Cases D and E), but the RMSE is higher for a given sample size compared to the corresponding case (A) where the data actually follow the normal distribution assumed by the method.
- When net savings is zero, the method correctly estimates very low net savings, even for naturally occurring change very far from a normal distribution (Case F).

Thus, overall

- As conjectured, the method appears to be valid even with departures from the assumed normal distribution.
- A mismatch between the actual distribution and the assumed distribution leads to higher standard errors.
- As suggested on conceptual grounds, having good predictors of participation is critical to the success of the method.

7 GROSS SAVINGS

Consumption analysis to estimate net savings as described above compares the change in consumption for program participants with that of a comparison group of nonparticipants. Entirely analogously, the same methods may be used to estimate gross savings, comparing customers who install a particular measure with those who do not. Thus, for gross savings estimation, the dummy variable D_j indicates that customer j installed the measure. All the same results apply as described above.

Note that the estimation of gross savings by this approach requires that information be available on what customers installed in or outside of a program. Note also that the same challenge of self-selection bias applies for estimation of gross savings as for net.

DID for gross savings estimation is the difference in consumption change between those who do and do not install the measure. If those who do and those who don't install are otherwise similar, this difference is gross savings.

Analogous to Eq. (2) we write

(22) $\Delta_j = base_j - gross_j D_j$

where

D_j = 0/1 dummy for customer j adopting the measure.

$base_j$ = change in consumption for customer j apart from how consumption changes with the measure in place.

$gross_j$ = the gross savings customer j will have IF they adopt the measure.

The goal of the regression or DID analysis in this case is to estimate average gross savings. Without random assignment, the savings estimate has bias equal to the difference in the average base change apart from the measure, for those who do and don't choose to install it. With RCT the average gross savings estimate is unbiased.

With an RED design, the LATE savings estimate is an unbiased estimate of the average incremental gross savings per incremental adopter. If the average gross savings per adopter is the same for the incremental adopters as for those who adopt without encouragement, Eq. (16) gives an unbiased estimate for this uniform gross savings per adopter.

If the average gross savings is different for those who adopt only with encouragement than for those who would adopt anyway, but the realized gross savings for those who would adopt anyway is unaffected by the encouragement, Eq. (16 gives an unbiased estimate of the savings per adopter for those customers who adopt with encouragement but not otherwise.

If the average gross savings is different for those who adopt only with encouragement than for those who would adopt anyway, and the realized gross savings for those who would adopt anyway is affected by the encouragement, Eq. (16 includes both the realized gross savings of the incremental adopters and the incremental gross savings due to encouragement for those who would have adopted anyway. The IV-IMR formulation of Eq. (20) provides the gross savings for the natural adopters as well as for the incrementally encouraged adopters, with adoption rather than participation as the modelled dummy variable D_j.

A key difference between the application of methods described here to gross savings estimation and application to net savings estimation is the underlying reason to expect the action (participation or adoption) to be related to a customer's naturally occurring or base savings, or to their conditional net or gross savings, respectively. In the case of net savings, we are concerned that customers who are natural adopters, hence have zero potential net savings, will be more likely to participate. By contrast, for the parallel case with gross savings, we expect customers with 0 conditional gross savings to be unlikely to adopt. Thus, the underlying assumptions that drive the need for the IV-IMR method for net savings in most situations may not apply as often for gross savings estimation.

On the other hand, we do expect that if the measures have limited potential to provide savings for a particular customer, that customer will have negligible likelihood of either adopting the measure or participating in the program. Thus, if applicability is not available as an explicit screening variable, both participation likelihood and adoption likelihood will be related to the conditional net or gross savings. In these cases, if good proxy variables for measure applicability can be found, the methods described here may be useful for gross as well as for net savings estimation. If no such proxy variables are available, the methods here may not be effective. As noted, a finding that the IV or IV-IMR terms are not statistically significant does not mean that there is no self-selection problem, only that we don't have the information needed to identify and adjust for it.

A study design sometimes used is to take later participants as a comparison group for the participants in a particular year. This design has the advantage that similar information is known from program records for both the earlier participants and their comparison group. Also, the later participants can be identified as having particular program measures applicable based on the measures they implemented at the time of their participation. Since all the customers in the analysis are program participants, though at different times, the difference between early and later participants is that the earlier participants have adopted the measure(s) and the later ones have not. Thus, the dummy variable for this design is often best viewed as an adoption variable, and the analysis as providing an estimate of gross savings.

In the case where the measures are effectively unavailable outside the program, all the installations are program-attributable, so that net and gross savings are the same. This situation is sometimes assumed, for example, for some income-targeted weatherization programs. If net savings is assumed to be the same as gross, this assumption should be stated explicitly.

8 CONCLUSIONS

8.1 Key Findings and Practical Considerations

Outside of random assignment contexts, all regression analysis requires assumptions that are only approximately true. The method validity then depends on how reasonable the assumptions are, and how sensitive the results are to the assumptions.

8.1.1 No Self-Selection Correction

To use Eq. (1), or its SAE equivalent, with no self-selection correction terms implicitly assumes that the participant and comparison groups can be treated as if they were essentially randomly assigned. For this to be a reasonable assumption would require

1. That the program measures are applicable to virtually all customers included in the comparison group.
2. That customers who are interested on their own in adopting the program-eligible measures are no more likely to learn about the program than those who have no such interest.
3. That customers who are not interested on their own in adopting program-eligible measures are just as likely to participate as those who do have an interest in the measures, once they learn about the program.

8.1.1 Correction using IV only

One strategy that has been used to address the self-selection problem is to find an "instrument" for participation. That is, a variable that significantly determines participation, but affects energy consumption only through the effect on program participation. With such an instrument, one can estimate the probability of participation and use this probability in place of the actual participation. This IV-only method produces an unbiased estimate of incremental net savings per increment of the instrument. These IV-only estimates provide an unbiased estimate of net savings for all participants only if it is reasonable to assume that participation is unrelated to the net savings a customer will have if they join the program. This assumption is not easily justified in most cases.

In terms of the framework develop in this paper, to use the participation instrument but not the Mills Ratio term, as described in Section 5.2, requires the assumption that participation is related to naturally occurring change noc but not to the randomly varying component of potential net savings. This assumption is more palatable when the regression includes explanatory variables for noc as in Eq. (19), rather than leaving all variability in noc in the random term ε_j as in Eq. (16).

Even so, it is necessary to believe that customers who, for unobservable reasons, tend to have higher or lower net savings if they participate will participate at the same rates. For example, consider two customers of similar size, income, and house structure. One has already decided to implement a measure, so that their potential net savings $net_j = 0$. The other will not implement without program assistance, so that their potential net savings is equal to the measure gross savings, $net_j = gross_j$. The value of participation to the first customer is the value of the program rebate less the "hassle cost" of filling out the application. The value of participation to the second customer is the value of the program rebate plus the value of the gross savings, less the direct and hassle costs of installing the measure, and less the hassle cost of filling out the application. It's hard to imagine that these two customers would tend to have similar propensity to participate, unless the hassles, implementation costs, and savings are all trivial.

Use of the participation instrument without the IMR term does not require any distributional assumptions as does the IMR, but does assume particular functional forms. For example, the participation model may be a probit or logistic model with a simple linear form. It's possible to test the sensitivity of results to alternative participation model forms and to identify the most appropriate structural form using GLM approaches.

Use of the predicted participation instrument does require that available variables for predicting participation account for the strongest participation drivers that are related to naturally occurring or potential net savings. The challenge is that key drivers of participation are likely to include attitudes toward energy efficiency, and measure applicability. Typically, the only way to obtain these variables is by surveys of the customers to be included in the regression analysis. In that case, the study trades survey nonresponse bias for the structural bias in the regression equation.

Once the participation model is estimated, constructing the IMR (or analogous term for alternative distribution) is little extra work. In practice, then, the IV-IMR variant would be favoured, or at least tested, unless the uncertain distributional assumptions are considered too problematic.

8.1.1 Correction using the IV-IMR method

This paper introduces a method for controlling for self-selection that addresses key sources of bias in net savings estimates from billing analysis. This IV-IMR method combines an instrumental variable for participation with an Inverse Mills Ratio. The new method appears to be more robust than prior methods to departures from the normality assumption of the IMR, without adding more complexity.

Using the participation instrument plus Inverse Mills Ratio (or corresponding factor for an alternative distribution) as in Section 5.3 allows for unobservable drivers of both naturally occurring and potential net savings to affect participation. The required assumption is that the distribution of the unobservable factors, including those that drive program attractiveness to a customer, are all normal (or alternative specified form). Again, this assumption becomes more palatable when more explicit structure is incorporated for both net and noc, as in Eq. **(17**(16), rather than leaving all variability in the unobservable residuals as in Eq. (3).

A key to the success of any of these methods is having variables that are strong predictors of participation, but can be safely excluded from the primary equation. If all of the variables that predict program participation also affect energy consumption directly, the adjustment steps will add to the variance of the estimate with little mitigation of the biases that are of concern.

8.1.2 The importance of good participation prediction

The IV-IMR method can provide an unbiased net savings estimate. In order for either the IV-only or the IV-IMR approach to work well, we need a strong and valid instrument for participation. As above, a valid instrument is a variable that significantly determines program participation, but does not affect energy consumption directly. That is, in Eq. (18) or (21), there should be terms included in the participation predictors **z** that are not in the direct consumption predictors **x.** With such an instrument, this approach can generate unbiased estimates even, to a certain extent, under departures from assumed normality.

The IV-only or IV-IMR method can be applied if some or all of the terms in the participation model also affect energy consumption directly, However, the more overlap there is between the participation drivers and the direct drivers of (change in) energy consumption, the less the model can reliably distinguish between direct consumption effects and participation effects.

If we have a good participation prediction model but its explanatory variables are all included as direct predictors of (change in) consumption in the primary equation, (all the terms in **z** are also included in **x or** should be) the estimated coefficients will depend heavily on the assumed distributional form. Thus, in this situation the IV-IMR method is less likely to be robust to departures from the normal distribution assumption underlying the specific IMR form.

Having a good participation model means not just that the model coefficients are statistically significant, but that the model can separate customers with higher versus lower participation likelihood. If the range of predicted participation probability is low over the observations in the regression, the participation instrument will have limited variability and will be closely correlated with the IMR term. For both IV-only and IV-IMR methods, limited variability in predicted participation leads to high variance in the estimated coefficients and in the net savings estimate. A finding that the terms $\widehat{D}$ or IMR have insignificant coefficients does not prove that there was no self-selection bias, only that we don't have the variation needed to mitigate it.

8.1.3 The importance of measure applicability

A key factor affecting program participation is the applicability of the program measures. If we have sufficiently informative models of participation probability, naturally occurring savings, and potential net savings, measure applicability is in principle reflected in these models. In practice these models tend to be fairly blunt tools. Pre-screening on applicability is a more reliable way to establish the comparison group, but requires additional information. If measure applicability, or other key drivers of participation, cannot be accounted for by available variables, the result can be a weak participation model, with the associated poorly determined net savings coefficients.

As an example, consider an HVAC replacement program, where savings are defined as the difference in consumption with the high efficiency equipment in place versus with standard new equipment in place. The ideal comparison group is a set of customers who replaced their HVAC equipment in the same span of time as the program participants. If we're able to identify a set of nonparticipant replacers, and restrict the analysis to this group, the participation model has only to identify factors affecting participation for those who replace. If we're not able to identify the replacers, it's unlikely that the estimated participation model will correctly assign a negligible probability to non-replacers and a higher probability to replacers. The result will be that the participants, who went from old HVAC equipment to high efficiency new equipment, are compared with customers who mostly didn't change out their equipment at all, and the savings estimate will be substantially overstated.

If the savings for the HVAC program are to be calculated relative to existing equipment, the bias is less dramatic but still present. The customers that should be included in the comparison group—or identified by modelling as having non-negligible participation probability—are those who need to replace their equipment but haven't yet. Without this restriction, or identification via the participation modelling, the participants are compared with customers whose HVAC equipment spans a range from almost new to past the end of its expected life, rather than being compared with customers whose equipment was at a similar stage, some of whom might have replaced equipment on their own. Analysis of change in consumption rather than consumption itself, as in the framework described here, mitigates this problem, but doesn't eliminate it. A similar situation arises for measures such as shell improvements or controls that are "add-ons" to existing equipment rather than replacements. For customers that already have climate-appropriate shell features, or already have the offered control measures, the program measures are inapplicable, but the participation model is unlikely to identify them as low participation probability unless the explicit applicability information is available. The participants end up being compared with a pool of nonparticipants who had no reason to be adding the measure, rather than with those who could have benefited from it and may have installed it on their own.

8.1.4 RCT

If customers can be assigned randomly to be program participants or not, no self-selection exists and therefore no self-selection bias in simple regression or Difference of Difference methods. However, randomized control trials assignment is not the way most programs are delivered. Situations where net savings estimation is a challenge are precisely those situations where program participation is voluntary.

8.1.5 RED

In a random encouragement design (RED), customers are randomly assigned to receive or not receive supplemental encouragement to join a program. This randomly assigned encouragement, assuming it significantly increases the probability of participation and does not directly affect energy consumption, can serve as an instrument for participation. The RED design therefore provides a true instrument for the application of the IV-only method described above. The RED can be used with or without additional variables contributing to the participation model. A model that includes more variables will tend to produce a wider range of variation in participation probability, which can facilitate estimation of the primary equation.

With the RED, the instrumental variables approach (IV-only) provides an unbiased estimate of the average net savings of the customers who participated because of the encouragement. Such information can be very useful for testing alternative program designs. However, the quantity of interest in most cases is the net savings for all participants, including those who join the program without special encouragement. The standard RED analysis does not provide this estimate unless we assume that net savings are the same for those who require extra encouragement to join as for those who join without that.

8.1.6 RED together with IV-IMR

Applying the IV-IMR approach together with an RED helps address weakness in each method. Adding the IMR term to a RED analysis addresses the limitation of the standard RED analysis, that the result is average net savings only for the incrementally encouraged participants, not for those who participate without special encouragement. The use of the RED provides the foundation for an informative participation model necessary to obtaining well determined results from either IV alone or the IV-IMR method.

However, even with an RED, it is still important to have good explanatory variables for participation apart from encouragement. If only the RED indicator is available to explain participation, the participation model can't be informative as to the relationship between net savings and participation absent encouragement.

This is the self-selection relationship that needs to be addressed. In fact, if the encouragement indicator is the only participation predictor, and the design includes only one type of encouragement, there are only two possible combinations of predicted participation and the IMR, one for encouraged customers and one for non-encouraged. In this case the IV and IMR terms are completely collinear, and the IV-IMR model cannot be estimated.

8.1.7 Billing analysis for gross savings

This paper focuses on estimation of net savings. However, all the methods described here are equally applicable to estimation of gross savings, if the dummy variable is measure adoption rather than program participation.

8.1.8 Spillover

The methods described here assume there is no nonparticipant spillover. A comprehensive net savings analysis including nonparticipant spillover would use the methods here, together with other methods to address nonparticipant spillover. The comprehensive analysis would need to address whether and to what extent the nonparticipant spillover is likely to affect the apparent net savings from the billing analysis.

8.2 Next Steps

The IV-IMR method is a promising approach for net savings analysis, particularly when combined with an RED. Further work is needed to assess the performance of the method in practice.

This work includes

- applications to existing data sets
- simulations using parameters based on particular real-world examples
- exploration of alternative models of the underlying participation drivers
- further development of procedures to implement the method in a statistically adjusted engineering form
- extension of the methods to pooled time series cross-sectional models.

Key questions to be addressed by additional applications with simulated and existing data sets include

- How effectively can we obtain variables that account well for measure applicability and other natural drivers of participation?
- Under a realistic simulation process that generates potential net savings and participation, what is the difference in bias and variance using no correction, IV-only, or IV-IMR?

9 REFERENCES

Applied Energy Group. 2014. SCE's Home Energy Report Program Savings Assessment: Ex-Post Evaluation Results, Program Year 2013. Prepared for Southern California Edison. http://www.calmac.org/publications/SCE_2013_HER_Evaluation_Final_Report_v10-24-14.pdf

Churchwell, C. A. et al. 2013. "Default Critical Peak Pricing for Non-residential Customers: Do Demand Reductions Persist? Are the Reductions Reliable?" *Getting It Done! Evaluation Today, Better Programs Tomorrow*, Chicago, IL: International Energy Program Evaluation Conference, August 2013.

Department of Energy. 2014. Smart Pricing Options Final Evaluation The final report on pilot design, implementation, and evaluation of the Sacramento Municipal Utility District's Consumer Behavior Study. Prepared by J. Potter, L. Jimenez Sacramento, Municipal Utility District, and Stephen S. George, Ph.D., Nexant. https://www.smartgrid.gov/files/SMUD_SmartPricingOptionPilotEvaluationFinalCombo11_5_2014.pdf

DNV GL 2017, *A White Paper: Mitigating Self-Selection Bias in Billing Analysis for Impact Evaluation*, Prepared for Pacific Gas and Electric Company, San Francisco (*to appear*)

DNV GL, 2016. Impact Evaluation of 2014 San Diego Gas & Electric Home Energy Reports Program (Final Report). Prepared for the California Public Utilities Commission. http://www.calmac.org/publications/DNVGL_SDGE_HERs_2014_FINAL_TO_Calmac.pdf

DNV GL, 2015. Impact Evaluation of PSE Web-Enabled Thermostat Program. https://conduitnw.org/_layouts/Conduit/FileHandler.ashx?rid=2965

Gandhi, R. et al., 2016, "Running Randomized Field Experiments for Energy Efficiency Programs: A Practitioner's Guide," *Economics of Energy & Environmental Policy*, Vol. 5, No. 2.

Goldberg, M. et al. 2017, "Not Just Another Pretty Formula: Practical Methods for Mitigating Self-Selection Bias in Billing Analysis Regressions," *Making Ambitious-Reductions Real: Accurate and Actionable Evaluation, Baltimore, Maryland: International Energy Program Evaluation Conference*, August, 2017

National Renewable Energy Laboratory. 2013. "Chapter 8: Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol," *The Uniform Methods Project: Methods for Determining Energy Efficient Savings for Specific Measures.* Prepared by K. Agnew and M. Goldberg, DNV GL. https://energy.gov/sites/prod/files/2013/11/f5/53827-8.pdf

Nexant. 2016. *Behavioral Demand Response Study - Load Impact Evaluation Report*. Prepared for Pacific Gas & Electric. http://www.calmac.org/publications/Behavioral_Demand_Response_Study_Final_Report_CALMAC.pdf

State and Local Energy Efficiency Action Network. 2012. *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkeley National Laboratory. http://behavioranalytics.lbl.gov.

State and Local Energy Efficiency Action Network. 2012b. *Energy Efficiency Program Impact Evaluation Guide*. DOE/EE-0829

XENERGY, Rev. 1996. *Net Savings Estimation: An Analysis of Regression and Discrete Choice Approaches.* Prepared for CADMAC by M. Goldberg, XENERGY, and Kenneth Train, NERA Economic Consulting.

APPENDIX A: IV INTERPRETATION OF RED

We use the same decomposition of change into naturally occurring and potential net savings given by Eqs. (2) and (3), so that Eq. (4) is still a correct description of the consumption change. The challenge of fitting Eq. (4) as an ordinary least squares (OLS) regression is that the participation variable D may be correlated with the error term ν. This correlation violates a key assumption of OLS regression, leading to biased estimates. The correlation will be present if, as is likely, participation D_j is related either to naturally occurring savings noc_j or to potential net savings net_j.

To break this correlation, the RED regression replaces the participation dummy D_j by predicted participation $\widehat{D}j$. In the simplest form, predicted participation is a function solely of whether or not the customer was in the encouragement group ENC:

(A1) $$\widehat{D}j = \begin{cases} \bar{D}_E, j \in ENC \\ \bar{D}_0, j \in \sim ENC \end{cases}$$

Eq. (4) then becomes

(A2) $$\Delta_j = a - b\widehat{D}j - b(D_j - \widehat{D}j) + \varepsilon_j - \psi_j D_j.$$

$$= a - b\,\widehat{D}j + \nu_j^*$$

with

$$\nu_j^* = -b(D_j - \widehat{D}j) + \varepsilon_j - \psi_j D = \nu_j - b(D_j - \widehat{D}j)$$

The corresponding estimated coefficient is the LATE estimate given by Eq. (12)

(A3) $$\boldsymbol{\hat{b}} = (\bar{\Delta}_{\boldsymbol{E}} - \bar{\Delta}_{\boldsymbol{o}})/(\bar{\boldsymbol{D}}_{\boldsymbol{E}} - \bar{\boldsymbol{D}}_{\boldsymbol{o}})$$

This is the same as the LATE estimator given by Eq. (10(12), with expectation given by Eq. (11).

In Eq. (A2), the predictor $\widehat{D}j$, being a scalar multiple of the encouragement dummy, is uncorrelated with either of the random elements ε_j or ψ_j, and is also by construction uncorrelated with the residual $D_j - \widehat{D}j$. However, if the participation decision, with or without encouragement, depends on potential net savings, so that D_j is correlated with ψ_j, the residual term ν_j^* in Eq. (A2) does not have zero expectation. The estimated coefficient $\hat{b}$ is an unbiased estimate of the average incremental net savings per incrementally encouraged participant, as described in Section 4.2. However, the expected value of $\hat{b}$ is not the same as the average potential net savings of all participants as b is defined in Eq. (3). Further, as described in Section 4.2, only in the special and rare case where the average savings per participant is the same for encouraged and non-encouraged participants is $\hat{b}$ an unbiased estimate of net savings per participant absent encouragement.

APPENDIX B: THE IV-IMR METHOD

Underlying Model of Program Attractiveness

To develop the IV and IV-IMR methods, we begin by modelling the decision process of the customer. Let U_j be a measure of the attractiveness of the program to person j, which depends on the customer's naturally occurring change, as well as on other observed and unobserved factors:

(B1) $U_j = \lambda noc_j + \theta net_j + \zeta \mathbf{z_j} + \eta_j$

where $\mathbf{z_j}$ is a vector of observed variables and η_j captures the impact of other unobserved factors. The customer chooses to participate if doing so provides a positive level of attractiveness:

(B2) $D_j = \begin{cases} 1 \ if \ Uj > 0 \\ 0 \ if \ Uj \ \leq 0 \end{cases}$

Substituting for noc_j and net_j gives

(B3) $U_j = \lambda a + \theta b + \zeta z_j + \eta_j^*$

where the unobserved term is

(B4) $\eta_j^* = \lambda \varepsilon_j + \theta \psi_j + \eta_j.$

If the variables in $\mathbf{z_j}$ are exogenous, i.e., independent of ε_j, ψ_j, and η_j, then equation (B3) becomes the basis for a choice model. The probability that customer j chooses to participate is:

(B5) $P(\mathbf{z_j}) = \text{Prob}\,(\lambda a + \theta b + \zeta z_j + \eta_j^* > 0)$

$= \text{Prob}(\eta_j^* > -(\lambda a + \theta b + \zeta \mathbf{z_j}))$

$= 1 - F(-(\lambda a + \theta b + \zeta \mathbf{z_j}))$

where F is the cumulative distribution function of η_j^*. If ε_j, ψ_j and η_j are normally distributed, then η_j^* is normally distributed and the participation decision is represented by a probit model.

Consider now Eq. (4) in the main text for Δ_j, which we reproduce here:

(4) $\Delta_j = a - bD_j + \varepsilon_j - \psi_j D_j.$

We describe the bias in this equation, and the technique to address it, starting from more restrictive and moving to less restrictive assumptions, as in Section 5.

B.1 No correlation between the participation decision and naturally occurring savings or potential net savings

In Section 5.1, we assume that the participation decision is not related either to customer j's naturally occurring change noc_j nor to the customer's potential net savings net_j. Under those assumptions, $\lambda = \theta = 0$ in Eq. (B3) and $\eta_j^* = \eta_j$ in Eq. (B4). As a result, there is no correlation between the participation dummy D_j, determined by the value of the unobservable program attractiveness U_j, and the unexplained errors ε_j and ψ_j in Eq. (4).. Eq. (1) estimated by OLS therefore gives an unbiased estimate of b.

B.2 Correlation between the participation decision and naturally occurring savings but not with potential net savings

In Section 5.2, we assume that the participation decision is related to customer j's naturally occurring change but not to the customer's net savings. Thus, in Eq. (B3) $\theta = 0$ but $\lambda \neq 0$. In this case, the participation dummy D_j is correlated with ε_j, which is a component of the overall residual ν_j in Eq. (1), because the participation decision depends on the customer's naturally occurring change in consumption. The expectation of Δ_j conditional on D_j is:

(B6) $E(\Delta_j \mid D_j) = a - bD_j + E(\varepsilon_j \mid D_j) - E(\psi_j D_j \mid D_j)$

where $E(\varepsilon_j \mid D_j) \neq 0$. However, we can take the expectation conditional on the customer's probability of participating rather than on whether the customer actually participated:

(B7) $E(\Delta_j \mid P(\mathbf{z_j})) = a - bE(D_j \mid P(\mathbf{z_j})) + E(\varepsilon_j \mid P(\mathbf{z_j})) - E(\psi_j D_j \mid P(\mathbf{z_j}))$

$= a - bP(\mathbf{z_j}) + 0 - E(\psi_j \mid D_j = 1, P(\mathbf{z_j}))E(D_j \mid P(\mathbf{z_j}))$

$= a - bP(\mathbf{z_j})$.

The result (B7) follows because

$P(\mathbf{z_j}) = E(D_j \mid P(\mathbf{z_j}))$ by definition,

$E(\varepsilon_j \mid P(\mathbf{z_j})) = 0$ since $\mathbf{z_j}$ is independent of ε_j, and

$E(\psi_j \mid D_j = 1, P(\mathbf{z_j})) = 0$ since the participation decision is independent of net savings.

The regression equation (4) then becomes

(B8) $\Delta_j = a - bP(\mathbf{z_j}) + \nu_j^*$

where $\nu_j^* = \varepsilon_j - \psi_j D_j - b(D_j - P(\mathbf{z_j}))$ has zero expectation conditional on $P(\mathbf{z_j})$. OLS on equation (B8) is unbiased since the error term has zero mean conditional on the explanatory variables. Eq. (16) in Section 5.2 is Eq. (B8), with $\hat{D}_j$ denoting the estimated probability $P(\mathbf{z_j})$.

This is the instrumental variables formulation. It is valid only if the participation decision is unrelated to the potential net savings. As discussed in Section 3, this assumption is difficult to justify in most cases.

B.2.1 How to apply the method

The researcher obtains a sample of customers, including participants and non-participants. The researcher estimates a binary probit model of whether the customer participated, using exogeneous explanatory variables $\mathbf{z_j}$. The probability of participating $P(\mathbf{z_j})$ is calculated for each sampled customer j using the estimated probit model. Then a regression is estimated where the dependent variable is the pre-to-post program change in consumption Δ_j and the explanatory variables are a constant a and the probability of participating $P(\mathbf{z_j})$. The estimated net savings per participant for the program is $\hat{b}$, the negative of the estimated coefficient of the probability of participating.

Note that the variables in $\mathbf{z_j}$ are independent of noc_j, even though the customer's decision to participate depends on its naturally occuring change in consumption. As described in Section 5.2.2, this seeming paradox arises only because we have not parameterized the naturally occurring change in consumption as a function of observed variables. Usually, the researcher will specify the naturally occurring change with its mean expressed as a vector of explanatory variables and corresponding coefficients $\mathbf{x_j}\alpha$. In this case, the $\mathbf{x_j}$ variables that relate to naturally occurring change enter the choice model and the regression, in place of the constant a.

B.3 Participation is related to both naturally occurring and net savings

In Section 5.3, we present the most general case, where the participation decision may be related to both a customer's naturally occurring change and to their net savings. Customers might consider, when deciding whether to participate in a program, the net savings that they would obtain from it. Or net savings might be correlated with naturally occurring savings, such that participation becomes related to net savings indirectly.

The choice model is the same as in the previous section. The attractiveness U of the program is again given by Eq. (B3), and now neither λ nor $\theta = 0$. Eq. (B3) can be written as

(B9) $U_j = a^{**} + \zeta z_j + \eta_j^{*}$

where $a^{**} = \lambda a + \theta b$.

If each of the error terms in Eq. (B4) is normally distributed, then the model is a probit. The expectation of Δ_j conditional on $P(\mathbf{z_j})$ becomes:

(B10) $E(\Delta_j \mid P(\mathbf{z_j})) = a - bP(\mathbf{z_j}) - E(\psi_j \mid P(\mathbf{z_j}))$

$= a - bP(\mathbf{z_j}) - E(\psi_j \mid D_j = 1, P(\mathbf{z_j}))P(\mathbf{z_j})$

The term $E(\psi_j \mid D_j = 1, P(\mathbf{z_j}))$ does not equal zero because now the participation decision depends on net savings. To obtain a useable regression equation, we need to calculate the value of $E(\psi_j \mid D_j = 1, P(\mathbf{z_j}))$ for each customer in the sample.

If all error terms are normally distributed, then the formula for $E(\psi_j \mid D_j = 1, P(\mathbf{z_j}))$ is the inverse Mills ratio:

(B11) $E(\psi_j \mid D_j = 1, P(\mathbf{z_j})) = c\ \phi(-(a^{**} + \zeta \mathbf{z_j})/\sigma))/[1-\Phi\ (-(a^{**} + \zeta \mathbf{z_j})/\sigma))]$

where c is a parameter to be estimated, σ is the standard deviation of η_j^{**} , and the inverse Mills ratio is defined as

(B12) $IMR \equiv \phi(-(a^{**} + \zeta \mathbf{z_j})/\sigma))/[1-\Phi\ (-(a^{**} + \zeta \mathbf{z_j})/\sigma))] = \phi(a^{**} + \zeta \mathbf{z_j})/\sigma)/\Phi\ ((a^{**} + \zeta \mathbf{z_j})/\sigma))$

Estimation of the participation model provides estimates of a^{**}/σ and b/σ, which are used to calculate IMR_j. The regression equation becomes:

(B13) $\Delta_j = a - bP(\mathbf{z_j}) - cIMR_j P(\mathbf{z_j}) + \nu_j^{**}$

Using $\hat{D}_j$ to denote the estimated probability $P(\mathbf{z_j})$, we have

(B13) $\Delta_j = a - b\hat{D}_j - cIMR_j \hat{D}_j + \nu_j^{**}$

This is the IV-IMR method given by Eq. (20) in Section 5.3.

B.3.1 How to apply the method

As for the simpler case of A.3, the researcher obtains a sample of customers, including participants and non-participants. The researcher estimates a binary probit model of whether the customer participated, using exogeneous explanatory variables $\mathbf{z_j}$. The probability of participating and the inverse Mills ratio are calculated for each sampled customer. Then a regression is estimated where the dependent variable is the pre-to-post program change in consumption Δ_j and the explanatory variables are a constant a, the probability of participating $\hat{D}_j$, and the inverse Mills ratio multiplied by the probability of participating. The estimated net savings per program participant is estimated as the average over all participants in the sample of ($\hat{b}+\hat{c}IMR_j$) where $\hat{b}$ and $\hat{c}$ are the negatives of the estimated coefficients of the participation probability and the inverse Mills ratio time this probability.

If the error terms are not normally distributed, then the choice model will not be probit and, perhaps more importantly, the inverse Mills ratio will not represent the true conditional mean $E(\psi_j|D_j=1, P(\mathbf{z_j}))$. It is possible to utilize a flexible functional form for the conditional mean; however, identifying an appropriately flexible function might be difficult.

APPENDIX C. IV-IMR FOR STATISTICALLY ADJUSTED ENGINEERING ESTIMATES OF NET SAVINGS

An alternative formulation to Eq. (1) starts with a Statistically Adjusted Engineering (SAE) approach. The general SAE framework represents the program not by a participation dummy D_j but by the estimated gross savings G_j for each participant j. Thus, analogous to Eq. (1), the SAE representation is

(C1) $\Delta_j = a - b\, G_j + \nu_j$

where G_j = program estimate of gross savings for participants, zero for nonparticipants. In the context of net savings estimation, the coefficient b is intended to be the ratio of true net savings to program estimated gross savings. If the program estimated gross savings is correct, the coefficient b is the net-to-gross ratio.

In this case, known gross savings G_j varies across participants, and the regression is designed to determine the net-to-gross ratio, rather than average net gross savings. In this case, Eq. (2) becomes

(C2) $\Delta_j = noc_j - net_j G_j$

and the underlying assumptions corresponding to Eqs. (3) and (18) are

$$noc_j = \mathbf{x_j}\alpha + \varepsilon_j$$

$$net_j = (b + \psi_j) G_j.$$

Thus,

$$\Delta_j = \mathbf{x_j}\alpha - bG_j + \varepsilon_j - \psi_j G_j.$$

The corresponding regression, including the instrument for participation and the Mills ratio, is

(C3) $\Delta_j = \mathbf{x_j}\alpha - b\widehat{D}(\mathbf{z_j})G_j + c\,\widehat{D}(z)G_j IMR_j + \nu_j^{**}$

In Eq. (C3), G_j is the gross savings that customer j would have if they participated in the program, whether or not the customer actually was a participant. Determining G_j meaningfully for nonparticipants is an obvious challenge for implementing the approach.

APPENDIX D. COMPARISON OF THE IV-IMR METHOD TO THE "DOUBLE IMR" METHOD

Eq. (21) is similar to the "Double Inverse Mills Ratio" method developed by the authors in prior work (XENERGY, 1996) but with an important difference. In the prior work, the IMR is used by itself to address the correlation between participation and both naturally occurring savings and participation, and is used interacted with the participation dummy to address the correlation between potential net savings and participation. The resulting regression equation from that work is

(D1) $\Delta_j = \mathbf{x_j}\alpha - bD_\mathbf{j} + c_1 IMR_j(\mathbf{z_j} \mid D_j) - c_2\, IMR_j(\mathbf{z_j} \mid D_j)D_\mathbf{j} + error_j$

where the IMR is defined as above.

In Eq. (D1), the second term involving the IMR is multiplied by the participation dummy D_j, so that this term becomes 0 for nonparticipants. The first IMR term, which is the IMR by itself, is positive for participants ($D_j = 1$) and negative for nonparticipants.

In the current formulation given by Eq. (21), the correlation of participation with naturally occurring savings is addressed by the instrumental variable $\widehat{D}(z)$ rather than by the IMR. The advantage of this formulation is that the validity of using predicted participation $\widehat{D}$ to mitigate the bias due to dependence of participation on noc does not depend on distributional assumptions as does the use of the IMR term in Eq. (D1). This makes the method of Eq. (21) more robust than the prior Double Mills Ratio approach.

Eq. (21) does still include the IMR to address the relationship between potential net savings and participation. The method is therefore still dependent on the normal distribution assumptions. However, the IMR term retained in Eq. (21) enters only in the form for $D_j = 1$. In Eq. (D1), by contrast, the first IMR term is constructed differently for participants than for nonparticipants, so that the way this term varies between the two groups depends strongly on the normal distribution assumption. Structurally, then, the dependency of Eq. (21) on the distributional assumption appears to be less than that of Eq. (D1).

Our conjecture is therefore that the sensitivity of the regression to departures from the assumed distribution is less using Eq. (21) where the IMR enters only once. This conjecture is partially supported by the simulations presented in Section 6.