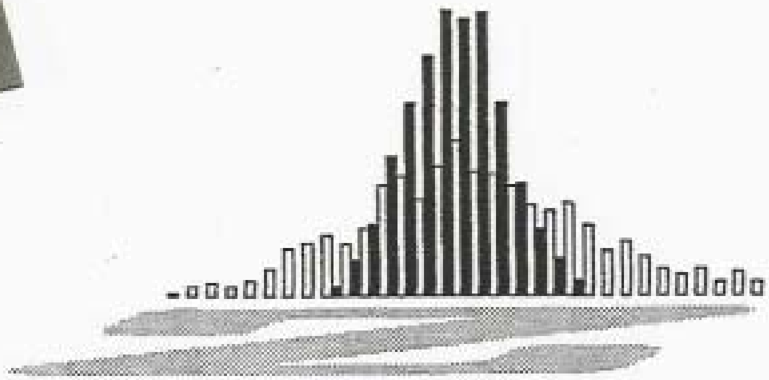
SDGE

San Diego Gas & Electric
Marketing Information & Planning
8306 Century Park Court
San Diego, Calif 92123 (619) 654-1122

Appendix Z Simulation Study

Comparison of Alternative Methods for Measuring the Gross and Net Energy Impacts of Demand-Side Management Programs (with Addendum)

August 1993



MIAP-93-P99-S01-330

Principal Investigators

Dean A. Schiffman
(SDG&E, Marketing Information and Planning, Evaluation & Analysis Section)

Robert F. Engle
(University of California at San Diego)

The authors wish to thank Dr. Kenneth Train (University of California, Berkeley) for his valuable comments on this work, as well as Kevin Mckinley, Andrew Sickels, Paula Ham-Su, and Athena Besa of the Marketing Department of SDG&E for their substantial theoretical and practical contributions to this effort.

Table of Contents

Key Words:

Simulation Study (or, Monte Carlo Study)

Measurement Protocols

Appendix Z

Net Load Impact

Base Usage

Conventional Conditional Demand Analysis (C-CDA)

Simplified Conditional Demand Analysis (S-CDA)

Discrete Choice Analysis

Introduction

Overview

The importance of accurately measuring the impact of demand-side management programs is clear. Utilities, regulatory authorities, and rate payers, as well as society as a whole, all have incentives to carry out effective demand-side management (DSM) and to avoid ineffective and expensive programs. As a result, these parties have the incentive to minimize the costs of measuring these programs and to maximize the effectiveness of measurement methods. This paper is designed to compare--in terms of their accuracy--alternative methods for measuring the impact of DSM programs. The technique that will be used in drawing the comparison is rather new to the utility industry but is well established in many other business and economic disciplines. The technique is that of *Simulation Analysis*, often referred to as *Monte Carlo Analysis*.

> Key Point: *Simulation Analysis* (often referred to as *Monte Carlo Analysis*) is the appropriate technique for evaluating the accuracy of alternative DSM measurement methods.

Regulatory Background

In 1991 the California Public Utilities Commission released an OII/OIR designed to establish a set of protocols that would govern efforts to empirically verify the energy savings from utilities' demand-side management programs. In the summer of 1992, SDG&E actively participated in the drafting of the main body of the DSM measurement protocols, along with other utilities, CPUC and CEC representatives, and other parties that are associated with the California electric and natural gas energy industry.

While SDG&E genuinely favors the overall intent of this effort, and in 1992 actively supported the construction of a uniform set of guidelines, the company became concerned over the fact that the protocols specifically prescribed the statistical techniques that would be applied in the DSM studies, and that the prescribed statistical techniques imply excessive DSM measurement expense. While SDG&E sought to support the overall effort to draft a comprehensive set of statewide measurement protocols, it looked to establish its own position with respect to these issues. In general, the company's twofold position was advanced along the following lines:

(1) The statistical techniques that are prescribed in the protocols are too narrowly defined, and the prescribed techniques may, in many cases, be inappropriate for the task of DSM measurement. As of mid-1992, SDG&E had developed a set of statistical techniques that were specifically designed to detect DSM energy savings, and the company had successfully applied the techniques to data derived from its largest DSM program (see Schiffman, et al., 1993, for published results). These statistical techniques are capable of satisfying the key elements of the protocols; yet they are incapable of satisfying only one of the protocol's particular *reporting* requirements, a reporting requirement that is not central to the DSM measurement issue. SDG&E instead seeks to employ the statistical techniques that it has developed.

(2) The data-collection costs associated with the protocol's prescribed techniques are excessive, and there is no compelling statistical justification for the additional costs. SDG&E has estimated that the additional cost of data collection and analysis that stems from the protocols' narrowly defined statistical techniques is most likely in the range of $1 million annually. SDG&E seeks to avoid this expense given that the statistical evidence indicates that there is no added benefit from the expenditure of this sum.

SDG&E's position was articulated in the company's "Appendix Z filing." Details concerning these issues and the Appendix Z filing will now be discussed.

The Protocols and Specific DSM Measurement Issues

Prescriptive Elements of the Protocols: The "Base Usage" Issue

The central issue which generated the company's Appendix Z filing is contained in the reporting requirements that are given in the protocols' Table 6: *Protocols for Reporting of Results of Impact Measurement Studies Used to Support an Earnings Claim* (Part I, p. 15). The company approves of the main body of Table 6; it is only the "Base Usage" element of the table--an element which from the point of view of effective DSM measurement is unnecessary--to which the company objects.

> Key Point: SDG&E's Appendix Z filing is based primarily on a single reporting requirement within the protocols, a reporting requirement that is not central to the measurement of DSM energy-savings.

A correct understanding of the company's objection depends on an understanding of the relationship between the protocols' Table 6 and the protocols' Table 5 (Part I, pp. 12-13). In Table 5, "Load Impact"--which is defined by end-use--is established as a concept.[1] For an indoor lighting program, for example, the lighting Load Impact is the energy impact per square foot of lighted area, per thousand hours-of-operation (or simply, the impact per square foot of lighted area).

Clearly, it is Load Impact that is of interest to all parties. An estimate of Load Impact is the natural goal of DSM measurement efforts, and it is this goal around which SDG&E has designed it DSM measurement activity. However, the protocols' Table 5 goes on to define Load Impact as follows:

(1) $$\text{Load Impact} = (\text{Base Usage}) - (\text{Usage in the Impact Year}),$$

where, conceptually, "Base Usage" is defined as energy consumption (for the end-use) prior to the program impact, and "Usage in the Impact Year" is energy consumption (for the end-use) after the program impact. An

> Key Point: The best estimator for Load Impact may be one that estimates Load Impact directly, rather than one that estimates Load Impact indirectly by estimating its two usage components.

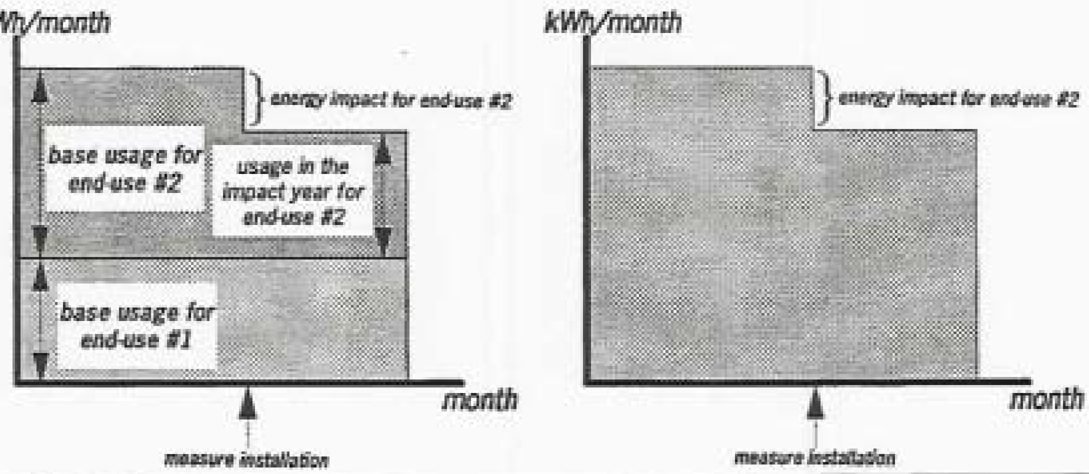
[1] See also the protocols' Table C-4.

important point is that while equation (1) is certainly conceptually correct, it has no implications for the estimation process. That is, it may be that the best estimator for Load Impact is one that estimates Load Impact directly, rather than one that estimates Load Impact indirectly by estimating its two usage components (Base Usage, and Usage in the Impact Year). It is in fact the position of SDG&E--based on the company's experience in measuring the load impact of its Commercial Lighting Retrofit program--that the best estimator for Load Impact is one that estimates Load Impact directly, and it is this position that is reflected in the company's Appendix Z filing.

This position can be supported by a simple illustration. Imagine the two end-use world of Figure 1 which shows, in two different ways, the monthly energy consumption of a hypothetical customer. On the left-hand side of Figure 1 consumption is disaggregated by end-use, while only total energy consumption is shown on the right-hand side. Consumption for both end-uses is constant, except that end-use #2 is subject to the installation of an energy-saving measure which has its energy (and load) impact. The point that can be made from Figure 1 is that the estimation of the energy and load impacts for end-use #2 can be accomplished based only on the data that is found on the right-hand side of the figure (total consumption). There is no need to attempt to estimate Base Usage and Usage in the Impact Year by end-use.

Figure 1

Is it Necessary to Disaggregate Energy Consumption by End-Use
in Order to Measure the Savings that are Associated with a Single End-Use?



The statistical technique that is designed to estimate Base Usage and Usage in the Impact Year by end-use is referred to in the company's Appendix Z filing as *Conventional Conditional Demand Analysis* (C-CDA). (As will be discussed, the statistical technique that has been developed by SDG&E to exploit data like those on the right-hand side of Figure 1 is referred to in the company's Appendix Z filing as *Simplified Conditional Demand Analysis*, or S-CDA.) The C-CDA technique was, at its inception, specifically designed as a means of estimating the level of energy usage (Base Usage and Usage in the Impact Year) by end-use (see Parti and Parti (1980)). The

C-CDA framework has been used widely for residential customers and has been surveyed by Lawrence and Parti (1984) and Sebold and Parris (1989). Particular applications which combine engineering and econometric estimates are Bartels and Fiebig (1990), Caves, Train, *et al.* (1987), and Aigner and Schonfeld (1990) while others focus on end-use load shapes as in Fiebig, Bartels, and Aigner (1991), Hendricks and Koenker (1992), Hill (1982), and Engle, Granger, and Ramanathan (1981).

However, none of these studies deals with commercial customers, and it is highly questionable whether C-CDA is the optimal technique for estimating the change in energy usage for a particular end-use, especially for the nonresidential sector. In EPRI (1988) which documents the COMMEND system, energy usage was estimated by end-use, for the commercial sector. A variety of techniques including expert judgment, direct metering, bill disaggregation, and engineering studies, were used to arrive at the estimates of usage; estimates of energy usage were actually estimated using C-CDA for only a few end-uses. In fact, RER (1991) says "Because of the complexity of commercial building systems and the diverse nature of occupant behavior, the pure econometric approach does not work for commercial sector applications."

Therefore, in general, the use of C-CDA for estimating the change in energy usage for a particular end-use is highly questionable for numerous reasons. As previously argued, it is obviously unnecessary to disaggregate energy consumption by end-use in order to estimate the change in energy usage for a particular end-use. In addition, the C-CDA technique depends crucially on the correct and complete specification of a single mathematical equation which supposedly adequately describes the energy-consumption behavior of individual customers within a large sample of customers; detailed site-specific data (data related to customers' end-uses and building characteristics) are then used in an attempt to calibrate the equation. As already discussed, the calibration (estimation) of these sorts of C-CDA equations is an extremely ambitious undertaking in the case where it is only the levels of consumption (by end-use) that are to be estimated and that this is due to the fact that it is extremely difficult to represent energy consumption behavior, to the required degree of accuracy, by a single mathematical equation. While this must be attempted if the fundamental purpose of the analysis is to estimate the level of consumption by end-use, the mathematical disaggregation of total consumption by end-use, when it is not the fundamental goal of the analysis, unnecessarily invites into the analysis all the problems that are associated with C-CDA. In fact, the main problems have to do with faulty site-specific data collection and the misspecification of the mathematics of the C-CDA equation, not to mention excessive data-collection expense. While these problems are unavoidable if C-CDA is being applied according to its original purposes, they can be avoided in DSM measurement.

It has been argued by some that there is value in the C-CDA framework for DSM measurement in that it requires detailed site-specific data; these data (the argument goes) can be used to account for changes in energy usage within the facilities of a given customer that are not associated with the DSM measure. SDG&E agrees that detailed site-specific data are very useful. However, there is certainly no need to adopt an inappropriate estimation framework (C-CDA) simply because it entails the collection of useful data. If detailed site-specific data are useful, they should indeed be collected as time and financial resources allow, and they should be utilized in the estimation framework that is optimal with respect to estimating the impact of DSM activity. In fact, detailed site-specific data can be collected with better focus if resources that would be devoted to the nonessential task of estimating energy consumption by end-use (Base Usage, and Usage in Impact Year) are made available in the effort to collect only those data that are truly useful.

Finally, there is the issue of the timely (and cost conscious) completion of DSM measurement studies. No C-CDA study can even begin without the long and costly process of site-specific data collection. And the implementation of C-CDA typically involves a prolonged sequence of attempts to improve the C-CDA energy-consumption equation. As a result, very few significant attempts at C-CDA are completed within a year's time and for less than six-figure dollar sums. This is especially alarming in the case of DSM measurement, since the large majority of cases in which C-CDA has been applied have little or nothing to do with that area; for DSM measurement, the expense and duration of the analytical activity that is required for C-CDA is likely to be greatly increased.

However, as will be shown, there exist alternative statistical techniques that can be implemented without detailed site-specific data, and which are perfectly capable of exploiting such data as time and resources make them available. These techniques (the primary one being that of *Simplified Conditional Demand Analysis*, or S-CDA) will be discussed at length later, as they are an integral part of the company's Appendix Z filing. However, at this point in the report, the framework of the protocols will be discussed further in order to further clarify the issues.

Prescriptive Elements of the Protocols: The "Net Impact" Issue

The overall goal of DSM measurement is to estimate the energy savings impact of the DSM program. As a result, it is necessary to go beyond the task of measuring the energy impact of a program's energy-saving technology, to measure savings net of the technology's impact in the absence of the program.

The protocols propose to measure *Net Load Impact* for the program as follows (see the protocols' Table 5, pp. 12-13):

(2) $$\text{Net Load Impact} = (\text{Participant Group Load Impact}) - (\text{Nonparticipant Group Load Impact})$$

The fundamental thinking here is that the estimated Nonparticipant Group Load Impact acts as a measure of the technology's impact in the absence of the program.

The main problem with this approach has to do with the so-called "self-selection" phenomenon which causes the results from this approach to be biased in the direction of an over-estimation of Net Load Impact. The phenomenon is based on the fact that energy-using customers can (obviously) make decisions concerning program participation. To simplify the argument, imagine an energy-saving technology that is clearly cost effective; suppose, for example, that the payback period for the DSM technology is only one month (that is, the payback is virtually instantaneous). As a result, most energy-using customers will adopt the technology in the absence of the program. However, in the presence of the program these same customers will choose to participate in the program (they will "select themselves" into the program), thereby removing themselves from the population of nonparticipants, the population that is--under the protocols--supposed to provide information concerning the technology adoption that would occur in the absence of the program. This obviously leads to an estimate of Participant Group Load Impact that is quite large. Moreover, the nonparticipant population will contain only those customers that have not adopted, and will not adopt, the technology (Why should any customer who is going to adopt the technology remain outside of the DSM program?), so that in line with equation (2) above, the Nonparticipant Group Load Impact would be zero. The result is an estimate of Net Load Impact that is quite large when in fact the actual Net Load Impact of the program is close to zero.

> Key Point: In the protocols, the formula for calculating the Net Load Impact of a DSM program leads to estimates of Net Load Impact that are biased in the direction of overestimation. This is due to the customer "self-selection" issue.

Based on this issue, SDG&E filed the "Discrete Choice Analysis" portion of its Appendix Z filing. The Discrete Choice approach to calculating Net Load Impact has to do with going directly to the issue of customers' decision making with respect to the adoption of DSM technology. Once this decision making behavior is characterized and empirically verified, the self-selection of customers into DSM program can be correctly accounted for in the estimation of the Net Load Impact of the program.

The Appendix Z Filing: Summary

In summary, the company's Appendix Z filing is directed at two statistical issues. First and most important is the issue of optimally estimating the gross impact of DSM programs; the corresponding element of the Appendix Z filing is couched in terms of substituting the so-called *Simplified Conditional Demand Analysis* for the *Conventional Conditional Demand Analysis* that is prescribed in the measurement protocols. The second issue is that of correcting for self-selection bias by using the Discrete Choice Analysis. Each of these issues will come into greater focus as the report continues.

Objections to the Appendix Z Filing and the Origin of the Simulation Study

Once SDG&E had advanced its position on improving the protocols through the Appendix Z filing, the DRA of the CPUC argued that SDG&E had not adequately established the fact that the statistical techniques that are contained within the Appendix Z filing lead to a comparable level of accuracy in terms of DSM measurement results. SDG&E responded with several additional points.

SDG&E pointed out that there is no real basis for the claim that the statistical techniques that are prescribed within the protocols are more accurate than other techniques. The DRA's claim that C-CDA is relatively more accurate was unsupported by either statistical theory or empirical evidence. In fact, the argument that C-CDA is more accurate can only be made (if at all) along two lines of reasoning. One argument--one which has already been alluded to--is that C-CDA requires the collection of detailed site-specific data on energy-use. However, as already mentioned, there is no need to adopt an inappropriate estimation framework (C-CDA) simply because it entails the collection of useful data, and if detailed site-specific data are useful, they should indeed be collected as time and financial resources allow. SDG&E favors the collection of detailed site-specific except for those data whose *only* purpose is to support a suboptimal estimation framework. The second argument for C-CDA accuracy is that it utilizes information concerning the overall nature of energy consumption (in that it explicitly considers all the individual end-uses), and that whenever more information is used, the resulting estimation results must be better, at least to some degree. This argument is in fact a valid one if those who are employing the C-CDA framework are successful in correctly and completely specifying a single mathematical equation that accurately describes the energy-consumption behavior of all customers within a large sample, provided that detailed site-specific data are collected accurately. The only issues that remain are the degree of relative accuracy (assuming the accurate collection of site-specific data) and the impact of faulty data collection.

To effectively address the accuracy issue, SDG&E offered to undertake a *Simulation Study*, often referred to as a *Monte Carlo Study*. A simulation study is a computer-based study that simulates the "real world" (the world that generates "real-world data"), in this case, the "world" of energy consumption and DSM. In general, simulation (Monte Carlo) methods have been used extensively in Economics, Business, and Statistics for many years. In 1984, the *Handbook of Econometrics* surveyed the applications in economics, although the list of such studies is much longer today. These methods are now becoming so computationally inexpensive that they are often advocated as a companion to econometric model building.

The computer simulated world of the simulation study provides several advantages. First, this world can be duplicated at will, so that issues such as which DSM measurement technique should be adopted can be analyzed under a variety of conditions; the "real world" takes place only once. Second, in the real world, the true numerical value for that item which is being estimated (e.g., energy savings from a DSM program) is unknown; as a result, there is no known standard by which alternative analytical tools can be judged. In a simulation study, a known value for the item which is being estimated can be built directly into the computerized framework, and alternative analytical tools can be judged in terms of their ability to detect this known value.

> Key Point: Alternative statistical techniques cannot be tested using "real" data since the object of estimation (e.g., actual DSM savings) is unknown. On the other hand, with simulated data the object of estimation is known, and the two alternative techniques can be judged accordingly.

In the specific case of the Appendix Z filing, a simulation study is the natural tool. While as competing statistical techniques, C-CDA and S-CDA could be applied to actual DSM program data, for reasons just stated there could be no answering the accuracy question in this setting due to the fact that the object of estimation (the true program savings) is in reality unknown, while this is not the case in a simulated DSM setting. Moreover, with a simulated DSM environment, the task of estimating DSM savings can be repeated hundreds or even thousands of times, so that relative accuracy of the alternative techniques can be viewed directly.

> Key Point: With a simulated environment, the estimation task can be repeated hundreds or even thousands of times, so that relative accuracy of the alternative techniques can be viewed directly.

The simulation study idea was ultimately sanctioned in D.93-05-063, dated May 19, 1993. It is the purpose of this report to present, in significant detail, the results of a two-part simulation study which addressed the issue of C-CDA and S-CDA in the estimation of the gross energy impact, as well as the issue of estimating net impacts using Discrete Choice Analysis.

The Analytical Framework

Overview

Figure 2 provides a broad schematic view of the simulation study. The study is basically composed of a large number of "iterations" (1,000 iterations for each portion of the study). At the start of each iteration, each customer (building) is assigned an entirely new set of attributes. In addition, weather is recalculated with each iteration. Most important, with each new iteration, new "random variation" is added to the model; the random variation is added to the framework in order to obfuscate its underlying structure, so that alternative estimators (such as C-CDA and S-CDA) can be judged in terms of their ability to overcome this element of the model. The model then generates the simulated data that are required by the measurement protocols and the Appendix Z filing, and the alternative estimators are applied to these data. The results are stored for each of 1,000 iterations so that the relative accuracy of the alternative estimators (e.g., C-CDA versus S-CDA) can be determined.

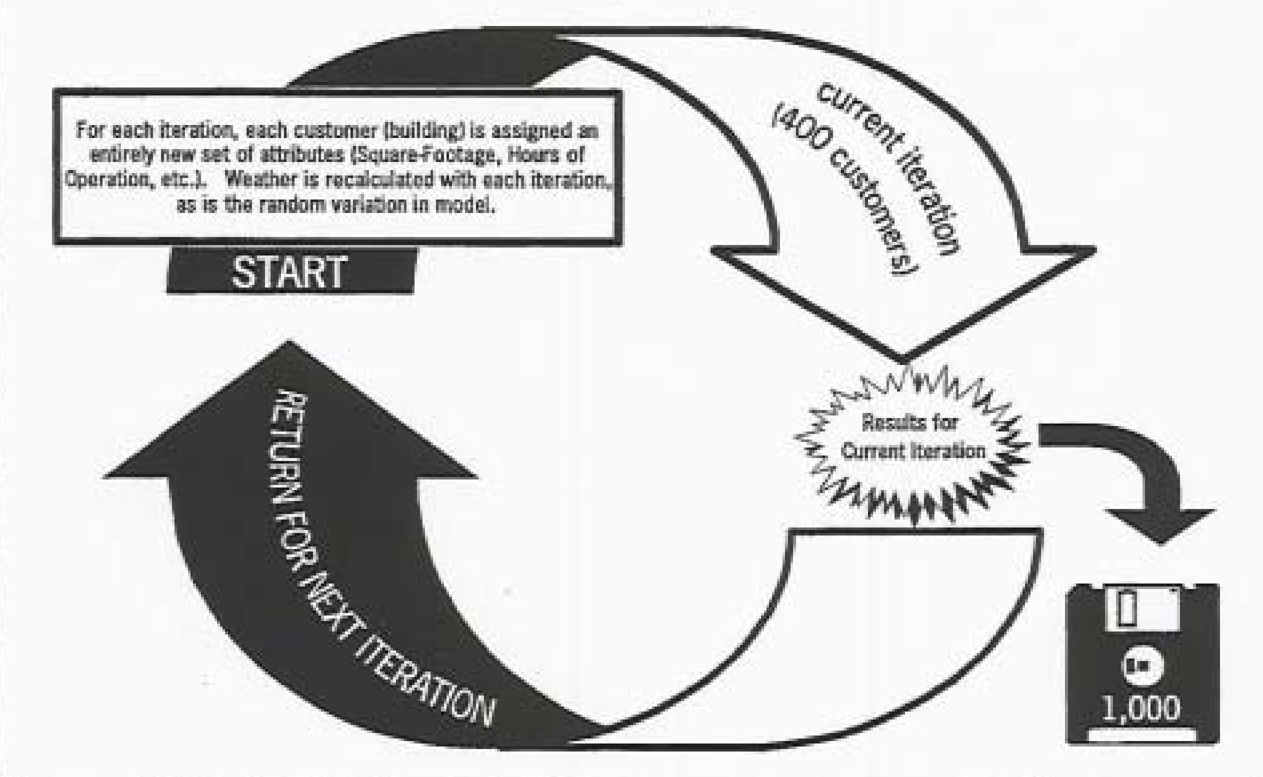
This fundamental technique was applied to the issue of C-CDA versus S-CDA, and to the DSM measurement adoption issue that was previously discussed having to do with Discrete Choice Analysis. In addition to studying the fundamental properties of alternative estimators under the best conditions for data collection and mathematical specification, the properties of the alternative estimators were studied under less than optimal conditions; each issue of this sort entailed an additional 1,000 iterations of the computerized model. Details will now be provided.

Investment Decision Model

Preliminary Specifications

The exact details of the model will be fully explained in forthcoming paragraphs. However, several of the model's elements must be presented on a preliminary basis, so that the overall framework can be understood correctly. The analysis will assume that there exists a single energy-efficient lighting measure. Conceptually, this assumption is manifested within a larger framework where lighting fixtures are uniform across customers in terms of their physical makeup: The detailed set of numbers which are found in the forthcoming framework are consistent with the situation where all fixtures begin with four standard 40-watt lamps and two standard core & coil ballasts, and end with two energy-efficient "T8" lamps and a single electronic ballast; the single lighting measure can then be defined as the fixture retrofit itself. (In addition, if the number of fixtures per square foot is constant across customers, it follows that the lighting measure yields a constant energy-demand savings per square foot.) In reality such a retrofit entails a capital cost of approximately $44 per fixture, and the customer incentive payment that is available under SDG&E's Commercial Lighting Retrofit program is approximately one-fourth of this sum. The associated energy-demand impact would be expected to equal approximately two-thirds of the fixture's original energy demand, which changes as a result of the retrofit by approximately 123 watts, from approximately 184 watts to 61 watts.

Figure 2



The numbers that are given in the previous paragraph allow for the calculation of the lighting measure's financial payback period. Payback (measured in years) equals the customer's initial capital expense divided by the dollar value of annual energy savings. The customer's initial capital expense is $44 less the 25% incentive payment if the customer adopts the measure under the program. Annual energy savings per fixture depend on the cost of energy (set at 9¢ per kWh), annual hours-of-operation for lighting for customer j during iteration i (h_{ij}), and the energy demand savings ((184-61)/1,000 = .123 kilowatts):

$$\text{PAYBACK}_{ij} = \frac{(\text{Initial Capital Cost per Fixture})}{(\$/\text{kWh})(h_{ij})(\text{Energy Demand Savings})} = \frac{(1-c)(\$44)}{(\$.09)(h_{ij})(.123\text{ kW})}$$

Under the customer incentive, c = 25%; c = 0 if the customer adopts the measure outside of the program.

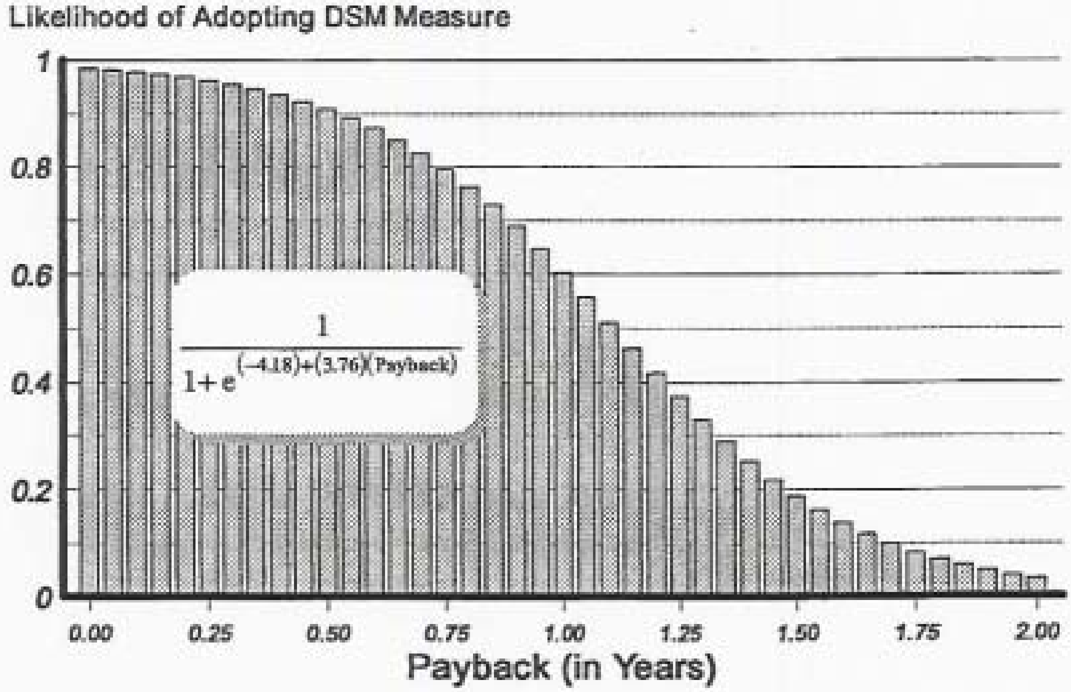
Mathematical Specification of the Investment Function

As will be discussed later, it will be assumed that each customer can consider installing the measure within an exogenously determined area of the customer's facility. This area will be referred to as the "square footage of affected lighted space" and will range from 10-100% of the total square footage of the facility. The probability q that the customer goes on to adopt the single lighting measure (to be installed across the affected lighted space) is assumed to be a simple logit-function of payback:

$$(3)\quad q(h_{ij},c)=\left\{1+\exp\left[(-4.18)+(3.76)(\mathrm{PAYBACK}_{ij})\right]\right\}^{-1}=\left\{1+\exp\left[(-4.18)+(3.76)\left(\frac{(1-c)(\$44)}{(\$.09)(h_{ij})(.123\,\mathrm{kW})}\right)\right]\right\}^{-1}$$

Equation (3), which is graphed as a function of payback in Figure 3, is the *Investment Decision Function* that governs measure adoption.

Figure 3



For any iteration within the study, average hours-of-operation were approximately $\overline{h}_i = 3,000$; a quick check would verify,

$$\text{(4)} \quad \begin{aligned} &\text{PAYBACK}_{ij}\Big|_{c=0\%,\,h_{ij}=3,000} = 1\tfrac{1}{3}\text{ years}, \quad \text{PAYBACK}_{ij}\Big|_{c=25\%,\,h_{ij}=3,000} = 1\text{ year} \\ &q(3,000,0\%) = 0.3, \quad q(3,000,25\%) = 0.6 \end{aligned}$$

The role of the Investment Decision Function will now be discussed.

Generating Adopters and Non-Adopters Within the Participant and Nonparticipant Groups

The Participant Group and the Nonparticipant Group: Preliminary Considerations

Although the issue of self-selection bias is, in practice, a serious one, it is not an issue that is central to this study. In other words, given the regulatory issues that generated this study, it is not useful to simulate the self-selection phenomenon in the study as a means of evaluating the measurement protocols. It is more useful to "give the protocols a chance" in the study, by constructing the simulated environment in such a way as to free the protocols framework from the self-selection issue. In this way, the other fundamental properties of the protocols' approach can be compared to those of the Discrete Choice Analysis approach.

To accomplish this, the simulation study assumes that for each iteration there are 400 customers (see Figure 2), 200 of which are unaware of the DSM program (and the associated customer incentive payments) and 200 of which are aware of the DSM program; the 200 customers that are aware of the program will be referred to as "participants" (they are participants in the sense that they face the "incentivized" price for the DSM measure, since they are aware of the program) and the 200 customers that are unaware of the program will be referred to as "nonparticipants." This allows the protocols' method for calculating net impact to function, since nonparticipants will not "select themselves" into the program (since they are unaware of its existence), and there will be a group of nonparticipants that will adopt the DSM measure. Details of this portion of the study will now be considered.

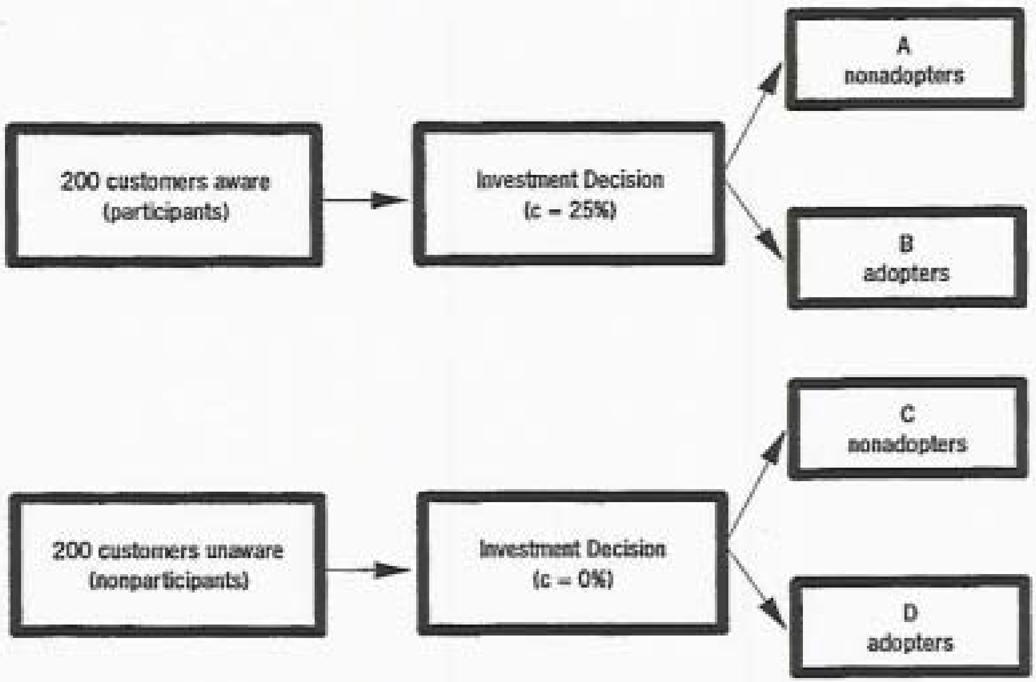
The Participant Group and the Nonparticipant Group: Details

The SAS random-number generator was used extensively within the study. This random number generator generates a uniformly distributed random variable that ranges between zero and one. With the proper linear transformation, the random number generator can be used to generate any uniformly distributed random variable, with an arbitrarily specified mean and range. We adopt the following notation:

Notation: Let "$u(\mu \pm \delta)$" represent the realized value of a uniformly distributed random variable which has mean μ, and lower and upper bounds of μ-δ and μ+δ, respectively, on its probability distribution.

Given that the framework contains customers that are aware and unaware, a major step in the study is to establish--through the Investment Decision Function--the adopters within each group. The breakout of customers is depicted in Figure 4:

Figure 4



It follows from previous definitions that,

$$\text{prob}\{u(.5\pm.5) \le q(h_{ij},c)\} = q(h_{ij},c)$$

Then during iteration i, the scheme for determining adopters is:

$$\{u(.5\pm.5)_{ij} \le q(h_{ij},0\%)\} \quad \text{implies} \quad \{\text{The Unaware Customer j is an Adopter During Iteration i}\}$$

$$\{u(.5\pm.5)_{ij} \le q(h_{ij},25\%)\} \quad \text{implies} \quad \{\text{The Aware Customer j is an Adopter During Iteration i}\}$$

It follows from equation (4) that for each iteration approximately 120 of the 200 customers that are aware of the program will become adopters (60% of 200), while approximately 60 of the 200 customer that are unaware will be adopters (30% of 200).

The Investment Decision Function is obviously at the heart of the net-to-gross issue. However, at this point we leave the Investment Decision Function until the associated estimation issues are addressed later in the report.

Consumption by End-Use and Conventional Conditional Demand Analysis

Introduction

Three end-uses were established for the study, two of which were indoor lighting and air conditioning. The third end-use is "cooking," although the basic structure of this component of the model could be associated with a variety of different end-uses. The lighting and air conditioning components of the model were developed because of their importance in reality, and it is the lighting component of the model that is the DSM-related element in the study.[2] The cooking end-use simply enhances the model in line with the fundamental structure of models of this sort that are estimated from actual data. The details of the model now follow.

Building Characteristics and Lighting

A customer's building is the primary means by which a single customer contributes to the study. In turn, buildings are primarily characterized by their square footage. The study is consistent with an analysis of medium office buildings of approximately 25,000 square feet. The previously mentioned SAS random number generator was used to determine building square footage, during iteration i, customer j:

(5) $$\mathrm{SQFT}_{ij} = u(25{,}000 \pm 6{,}250)_{ij}$$

Therefore, for any iteration, buildings are expected to range from 18,750 to 31,250 square feet.[3]

For all customers, lighting-related electricity usage--before any lighting retrofit--is assumed to be constant at 2 watts per square foot. Since an assumption has already been made that corresponding watts per fixture is 184, it follows that fixtures are placed at one per 92 square-foot area. Annual hours-of-operation for lighting are assigned as according to,

(6) $$h_{ij} = u(3{,}000 \pm 500)_{ij}$$

Base energy consumption for lighting (energy consumption for lighting before any lighting retrofit) during month t (in kWh) is then,

(7) $$\mathrm{LIGHT}_{ijt}^{\mathrm{base}} = (2\ \mathrm{watts})\left[\left(\frac{h_{ij}}{12\ \mathrm{months}}\right)\left(\mathrm{SQFT}_{ij}\right)(1/1{,}000)\right]$$

---

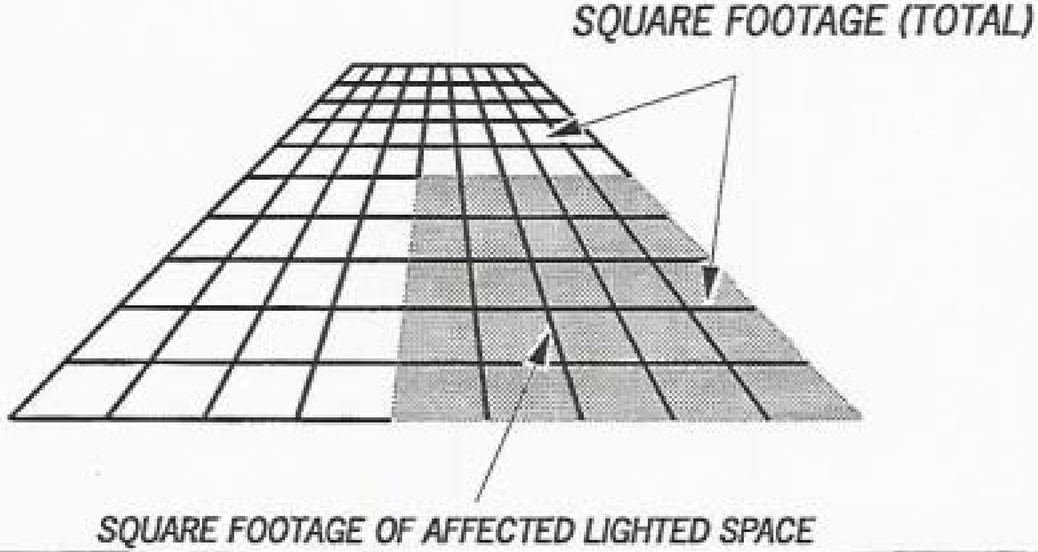
[2]Heating and ventilation will be ignored in the model.

[3]It should be emphasized here that the SAS random number generator is simply used at this point to establish, within the iteration, the distribution of buildings across square footage; the results of this process do not generate anything that for analytical purposes constitutes a *bona fide* random variable.

As mentioned, each customer can consider installing the measure within an exogenously determined area of the customer's facility.[4] This area will be referred to as the "square footage of affected lighted space" and will range from 10-100% of the total square footage of the facility (see Figure 5):

$$\mathrm{SQFT}_{ij}^{A} = u(.55 \pm .45)\left(\mathrm{SQFT}_{ij}\right)$$

Figure 5
The Square Footage of Affected Lighted Space is 10-100% of the Total Square Footage for the Facility



As already established, the energy demand savings that is associated with the lighting retrofit is a fraction (123 watts)/(184 watts) of the original demand of 2 watts per square foot, so that the energy consumption for lighting is,

$$\mathrm{LIGHT}_{ijt} = \mathrm{LIGHT}_{ijt}^{base} - \Delta\mathrm{LIGHT}_{ijt} \tag{8}$$

[4]The square footage of affected lighted space is exogenous in the sense that the proportion of total square footage where fixtures can be retrofited is simply assigned at random within the model, rather than determined by some other key element of the analysis. It is the decision whether or not to retrofit the square footage of affected lighted space that is endogenous to the model, through the Investment Decision Function.

where the change to base lighting is the change that is associated with the square footage of affected lighted space:

$$(9) \qquad \Delta LIGHT_{ijt} = \left(\frac{123}{184}\right)(2\text{ watts})\left[\left(d_{ijt}\right)\left(\frac{h_{ij}}{12\text{ months}}\right)\left(SQFT_{ij}^{A}\right)(1/1{,}000)\right]$$

The variable d_{ijt} is an econometric *dummy* variable that is zero prior to any retrofit and one at the point of retrofit and thereafter. For each customer, during each iteration, 36 months of consumption data (which includes lighting-related consumption) were generated within the model. For those customers who adopted (in line with the Investment Decision Function), the retrofit date was assigned to customers from months 13 to 24, with customers basically uniformly distributed across these twelve months.

Equations (7), (8), and (9) constitute a complete specification of monthly energy consumption for the end-use that is lighting. Specifically, the square-bracketed factors that can be found within equations (7) and (9) will act as regressors in the complete regression model that will be presented in later sections.

Air Conditioning

The air conditioning (AC) component of the model will now be developed. In line with the econometric approach that is generally taken the AC component will be based primarily on weather variables and building characteristics. This approach is motivated by the fact that building *heat gain* through external building surfaces--which depends primarily on weather variables and building characteristics--must be at the heart of the AC model.[5]

A regression model that allows for heat gain through building surfaces--by *conduction* through surfaces, and by *solar gain* through windows--will have in it one or more components that are of the following general form (for building j during month t):

$$(10) \qquad \text{AC Regression Component} = \beta\times(\text{WEATHER VARIABLE})_{jt}\times(\text{SURFACE AREA})_{j}$$

For the simulation study, calibrating a set of AC regression components that are of this general form is a task that is conceptually straightforward: the product of the coefficient β and the weather variable needs to have the correct units of measurement, so that when this product is multiplied by the appropriate surface area variable the result is a sensible value for monthly AC energy consumption. While this could be accomplished in a variety of ways, certain elements of the energy engineering literature are apposite: *ASHRAE* (1989, pp. 26.32-26.62) contains the ***CLTD/CLF Calculation Procedure*** (CLTD for "Cooling Load Temperature Difference", and CLF for "Cooling Load Factor") for estimating hourly AC load in buildings, based on the physical characteristics of the building in question as well as its geographic location. Although the procedure is less accurate than other more sophisticated estimation procedures that can be found in the same body of literature, it is designed for ease of calculation and reasonable accuracy. More important for our purposes, the ***CLTD/CLF Calculation Procedure*** expresses hourly

[5]Heat gain from internal sources (lighting, people, and office equipment) will also be considered.

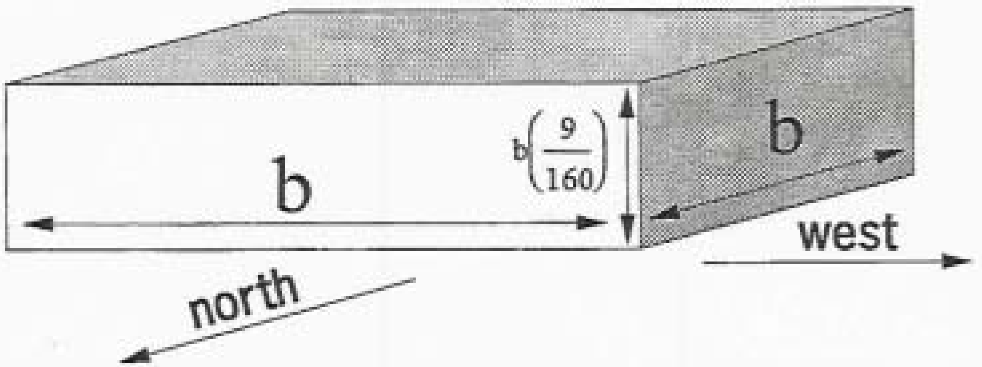
AC load in "regression-like" terms that are comparable to the monthly expression in equation (10). Shortly (after a discussion of the supposed physical makeup of the study's simulated buildings), the ***CLTD/CLF Calculation Procedure*** will be exploited as a means of calibrating the AC component of the model.

As already stated, at the heart of the AC issue is the heat gain that is associated with conduction through building surfaces (roofs, walls, and windows) and with solar gain through windows. This leads to the consideration of building dimensions, building orientation, and the physical makeup of buildings. So that the study as a whole would remain manageable, the study's simulated buildings were single-story, square buildings of width and depth "b" feet (see Figure 6), so that in light of equation (5) the building dimension for customer j is (suppressing the iteration subscript i at this point),

$$b_j = \sqrt{SQFT_j}$$

The ratio of the building height to width (or depth) was set at 9/160; this leads to reasonable results for building height in light of the variation in square footage across buildings that has already been established in equation (5). The buildings have an exact north/south/east/west orientation as in Figure 6, and windows comprise one-third of the surface area of walls.

Figure 6
Building Proportions and Orientation



Note that several implications concerning wall (non-window) area and orientation, window area and orientation, and roof area follow from these assumptions (continuing to suppress the iteration subscript i):

$$AREA_j^{window,r} = \left(\frac{1}{3}\right)\left(\frac{9}{160}\right)SQFT_j, \quad r = \text{north, south, east, west}$$

(11) $$AREA_j^{wall,r} = \left(\frac{2}{3}\right)\left(\frac{9}{160}\right)SQFT_j, \quad r = \text{north, south, east, west}$$

$$AREA_j^{roof} = SQFT_j$$

As mentioned, the *ASHRAE* standards have been used to calibrate the AC portion of the model. In particular, CLTD is a temperature variable (in units of °F) for which standards have been developed, standards that allow for the estimation of the hourly heat gain that is associated with conduction through exterior surfaces. These *ASHRAE* standards are available by geographic location, surface-type (roof, wall, and window), surface orientation (north, south, east, west), month, and hour. Table 1 contains the CLTD standards for roofs that were used in the study ($CLTD_{t\tau}^{roof}$). Two additional sets of four tables (tables that are comparable to Table 1, for walls and windows, each of which has four possible orientations) are available from the same source. The CLTD standards are designed to be used in conjunction with a consociate set of standardized "heat transfer coefficients" (defined by surface type), so that the final calculation for heat gain has the proper units. For solar gain, the "CLF" portion of ***CLTD/CLF Calculation Procedure*** pertains to standards that are similarly available--again, by geographic location, orientation, month, and hour.

The conduction (CLTD) and solar (CLF) gain portions of heat gain will now be modeled. In accordance with Table 1 let,

$$CLTD_{t\tau}^{s,r} = \text{ASHRAE CLTD Standard (in °F) for:} \quad \begin{array}{l} \{\text{surface } s = \text{roof, wall, window}\} \\ \{\text{orientation } r = \text{north, south, east, west}\} \\ \{\text{month } t = \text{January, February, ... December}\} \\ \{\text{hour } \tau = 1, \ldots 24\} \end{array}$$

Let U^s (in units of Btu/h·ft² °F) denote the standardized heat gain coefficient for surface s (assumed to be uniform across buildings, for surface s). Accordingly, the standard for conduction heat gain (in Btu/h) for month t at hour τ through these three surfaces is,

$$U^{roof}\left(CLTD_{t\tau}^{roof}\right)\left(AREA_j^{roof}\right)+\sum_{s=\substack{window,\\wall}}\left\{\left(U^s\right)\left(\sum_{r=\substack{north,south,\\east,west}}\left(CLTD_{t\tau}^{s,r}\right)\left(AREA_j^{s,r}\right)\right)\right\}$$

At this point a simplifying assumption is made. If $CLTD_{t\tau}^{s,r}$ for windows and walls is roughly proportional to $CLTD_{t\tau}^{roof}$, it follows from (11) that conduction heat gain for month t at hour τ is,

$$HEATGAIN_{jt\tau}^{conduction}=\left(U^{conduction}\right)\left(CLTD_{t\tau}^{roof}\right)\left(SQFT_j\right),$$

where the parameter $U^{conduction}$ (the "composite heat transfer coefficient") is a complicated function of the factors of proportionality and every building parameter that has been defined to this point.

For solar gain through windows, the *ASHRAE* literature can be similarly exploited, although for compactness the description of this effort will be parsimonious.[6] Standards have been established which allow for the construction of a solar gain variable $X_{t\tau}^r$ as well as a final expression for solar heat gain for the building (in units of Btu/h-ft²):

$$HEATGAIN_{jt\tau}^{solar}=(SC)\left(\sum_{r=\substack{north,south,\\east,west}}\left(X_{t\tau}^r\right)\left(AREA_j^{window,r}\right)\right)$$

The standardized parameter SC is a unit-free factor of proportionality ("shading coefficient") that depends on the physical properties of windows (assumed uniform across buildings, in the study). If $X_{t\tau}^r$ is similarly assumed to be roughly proportional to $CLTD_{t\tau}^{roof}$, it follows from (11) that solar heat gain for month t at hour τ is,

$$HEATGAIN_{jt\tau}^{solar}=\left(U^{solar}\right)\left(CLTD_{t\tau}^{roof}\right)\left(SQFT_j\right)$$

where the parameter U^{solar} is a function of the factor of proportionality and building parameters.

The final expression for heat gain through surfaces is then,

$$\begin{aligned}HEATGAIN_{jt\tau} &= HEATGAIN_{jt\tau}^{conduction}+HEATGAIN_{jt\tau}^{solar}\\ &=\left(U^{conduction}+U^{solar}\right)\left(CLTD_{t\tau}^{roof}\right)\left(SQFT_j\right)=(U)\left(CLTD_{t\tau}^{roof}\right)\left(SQFT_j\right)\end{aligned}$$

[6]The standards and the associated calculation for solar heat gain are readily available in ASHRAE (1989), Chapter 26. A brief summary is available on p. 26.33.

Table 1

ASHRAE Standards for Cooling Load Temperature Difference (°F) for Flat Roofs $(CLTD_{t\tau}^{roof})$

Hour (τ) → Month (t) ↓	1	2	3	4	5	6	7	8	9	10	11	12	
Jan	19	16	14	11	8	6	3	1	0	0	1	3	
Feb	24	21	19	16	13	11	8	6	5	5	6	8	
Mar	29	26	24	21	18	16	13	11	10	10	11	13	
Apr	33	30	28	25	22	20	17	15	14	14	15	17	
May	35	32	30	27	24	22	19	17	16	16	17	19	
Jun	36	33	31	28	25	23	20	18	17	17	18	20	
Jul	35	32	30	27	24	22	19	17	16	16	17	19	
Aug	33	30	28	25	22	20	17	15	14	14	15	17	
Sep	29	26	24	21	18	16	13	11	10	10	11	13	
Oct	24	21	19	16	13	11	8	6	5	5	6	8	
Nov	19	16	14	11	8	6	3	1	0	0	1	3	
Dec	17	14	12	9	6	4	1	-1	-2	-2	-1	1	

Hour (τ) → Month (t) ↓	13	14	15	16	17	18	19	20	21	22	23	24	$(CLTD_t^{roof}) = (30)\sum_{\tau=1}^{24}(CLTD_{t\tau}^{roof})$
Jan	6	10	15	19	23	26	28	29	29	27	25	22	10,230
Feb	11	15	20	24	28	31	33	34	34	32	30	27	13,830
Mar	16	20	25	29	33	36	38	39	39	37	35	32	17,430
Apr	20	24	29	33	37	40	42	43	43	41	39	36	20,310
May	22	26	31	35	39	42	44	45	45	43	41	38	21,750
Jun	23	27	32	36	40	43	45	46	46	44	42	39	22,470
Jul	22	26	31	35	39	42	44	45	45	43	41	38	21,750
Aug	20	24	29	33	37	40	42	43	43	41	39	36	20,310
Sep	16	20	25	29	33	36	38	39	39	37	35	32	17,430
Oct	11	15	20	24	28	31	33	34	34	32	30	27	13,830
Nov	6	10	15	19	23	26	28	29	29	27	25	22	10,230
Dec	4	8	13	17	21	24	26	27	27	25	23	20	8,790

Within the category "With Suspended Ceiling," roof type #7 was selected from *ASHRAE* (1989) Table 29 (p. 26.34). The values in Table 29 were modified for latitude based on *ASHRAE* (1989) Table 32 (p. 26.37), for North latitude 32 deg.

Based on this, the final expression for Btu's per month (assuming thirty days per month) is,

$$(12) \qquad (30)\sum_{\tau=1}^{24}\text{HEATGAIN}_{jt\tau}=(\text{U})(\text{CLTD}_t^{\text{roof}})(\text{SQFT}_j)$$

where,

$$(\text{CLTD}_t^{\text{roof}})=(30)\sum_{\tau=1}^{24}(\text{CLTD}_{t\tau}^{\text{roof}})$$

Table 1 contains the final $\text{CLTD}_t^{\text{roof}}$ values that have been constructed for the study. In turn, based on reasonable assumptions for building characteristics, the composite heat transfer coefficient was calculated:

$$\text{U} = .197 \quad \text{Btu}/\text{SQFT}_j\ {}^\circ\text{F}$$

As a result, equation (12) can be constructed in its entirety.

However, the final step in constructing the regression component is to specify AC efficiency so that monthly Btu's are translated into kWh for the month. Nine watt-hours per Btu was selected as the overall AC efficiency factor, so that the monthly regression component for AC consumption (for heat gain through surfaces) is (except for one forthcoming modification),

$$\left(\frac{.197}{9}\right)\left[(\text{CLTD}_t^{\text{roof}})(\text{SQFT}_j)/1{,}000\right]$$

While the expression in square brackets might be a suitable regressor for the simulation study, only a single set of twelve monthly values for $\text{CLTD}_t^{\text{roof}}$ is available as standards (as is evidenced in Table 1). As a result, the use of this set of standards within the regressor entails the implicit assumption that weather for any given month-type (say, April) is constant across the years in the simulation study. However, a more well-rounded model would allow weather to change from year to year. As a result, in the study, for iteration i, $\text{CLTD}_t^{\text{roof}}$ was allowed to vary across years by ±20% of the standard, so that the final form for the AC regression component for heat gain through external building surfaces has the form (reintroducing the iteration subscript i),

$$(13) \qquad \text{AC}_{ijt}^{\text{external}}=\left(\frac{.197}{9}\right)\left[\left(\overline{\text{CLTD}_{it}^{\text{roof}}}\right)(\text{SQFT}_{ij})/1{,}000\right].$$

where,

$$(14) \qquad \overline{\text{CLTD}_{it}^{\text{roof}}}=(\text{CLTD}_{it}^{\text{roof}})(u(1\pm .2)_{it})$$

Note that equation (13) is of the general form expressed in equation (10), so that the goal of constructing a regression component of this form has been satisfied.

Heat gain from internal sources is also a consideration. AC consumption related to lighting is simply proportional to the lighting component of the model

(15) $$AC_{ijt}^{light} = \left(\frac{3.412}{9}\right)\left[LIGHT_{ijt}^{base} - \Delta LIGHT_{ijt}\right],$$

where 3.412 is the appropriate Btu scaling coefficient. Looking ahead, it is clear that within the overall regression equation, the coefficient $(3.412/9)$ will simply add to the regression coefficients that will be associated with the regressors in equations (7) and (9).

For heat gain from building occupants, a standard of 450 Btu/h per occupant was adopted so that for eight-hour workdays during a thirty-day month the associated cooling consumption (in kWh) is,

$$AC_{ijt}^{occupants} = \left(\frac{1}{9}\right)\frac{(8\times 30)(450\ \text{BTu/h})}{1{,}000}\left[\left(OCCUPANTS_{ij}\right)\right]$$

In the study, the number of occupants was determined according to,

$$OCCUPANTS_{ij} = \left(u\left(\frac{2}{192}\pm\frac{1}{192}\right)_{ij}\right)\left(SQFT_{ij}\right).$$

This implies a density of one employee per area eight to fourteen feet-square, or an occupant on average every 100 square-feet.

For heat gain from office equipment, a standard of 1,800 Btu/h per unit of equipment was adopted (in line with standards for heat gain from personal computers) so that for eight-hour workdays during a thirty-day month, the associated cooling consumption (in kWh) is,

$$AC_{ijt}^{equipment} = \left(\frac{1}{9}\right)\frac{(8\times 30)(1{,}800\ \text{BTu/h})}{1{,}000}\left[\left(EQUIP_{ij}\right)\right]$$

On average, the number of units of equipment was set to one-half unit per occupant:

$$EQUIP_{ij} = \left(u(.5\pm .5)_{ij}\right)\left(OCCUPANTS_{ij}\right).$$

The final expression for AC consumption is then,

(16) $$AC_{ijt} = AC_{ijt}^{external} + AC_{ijt}^{light} + AC_{ijt}^{occupants} + AC_{ijt}^{equipment}$$

The exact structure of this expression will be considered as the model develops.

Cooking

A third end-use was added to the model: "cooking." While the specific nature of the model's third end-use has been specified ("cooking"), it is unnecessary to specify the end-use as such; any one of a number of labels (e.g., "outdoor lighting," "refrigeration") could be placed on this end-use. The cooking component of monthly consumption was set at 2,000 kWh (for those customers that had cooking facilities), and, for any iteration, on average 60% of the customers were determined to have cooking facilities. The exact structure of the cooking component of the regression model is then,

(17) $\quad \mathrm{COOK}_{ijt} = (2{,}000)\left[d_{ij}^{cook}\right],$

where d_{ij}^{cook} is an econometric dummy variable that takes on the value one if customer j has cooking facilities during iteration i, and zero otherwise.

The Addition of Random Disturbances and Conventional Conditional Demand Analysis

These three end-uses--lighting, air conditioning, and cooking--make up total monthly energy consumption for the simulation study. To fully formulate the energy consumption issue as a regression problem, we now view equations (8), (16), and (17) as the *expected values* for consumption for the three respective end-uses; in this case actual monthly consumption kWh_{ijt} will equal their sum plus a random disturbance term ε_{ijt}:

$$\mathrm{kWh}_{ijt} = \mathrm{LIGHT}_{ijt} + \mathrm{AC}_{ijt} + \mathrm{COOK}_{ijt} + \varepsilon_{ijt}$$

The following regression equation results from making substitutions from various definitions that are associated with equations (8), (16), and (17) (regressors are in square brackets):

$$\begin{aligned}
\mathrm{kWh}_{ijt} = {} & \beta_1\left[\left(\frac{h_{ij}}{12\ \text{months}}\right)\left(\frac{\mathrm{SQFT}_{ij}}{1{,}000}\right)\right] + \beta_2\left[\mathrm{OCCUPANTS}_{ij}\right] + \beta_3\left[\mathrm{EQUIP}_{ij}\right] + \beta_4\left[d_{ij}^{cook}\right] \\
(18) \qquad & + \beta_5\left[\left(\overline{\mathrm{CLTD}}_{it}^{roof}\right)\left(\frac{\mathrm{SQFT}_{ij}}{1{,}000}\right)\right] \\
& + \beta_6\left[\left(d_{ijt}\right)\left(\frac{h_{ij}}{12\ \text{months}}\right)\left(\frac{\mathrm{SQFT}_{ij}^{A}}{1{,}000}\right)\right] + \varepsilon_{ijt}
\end{aligned}$$

where,

$$\beta_1 = (2\ \text{watts})\left(1+\frac{3.412}{9}\right),\quad \beta_2 = \left(\frac{1}{9}\right)\frac{(8\times 30)(450\ \mathrm{BTu/h})}{1{,}000},\quad \beta_3 = \left(\frac{1}{9}\right)\frac{(8\times 30)(1{,}800\ \mathrm{BTu/h})}{1{,}000},$$

$$\beta_4 = 2{,}000, \qquad \beta_5 = \left(\frac{.197}{9}\right), \qquad \beta_6 = -(2\ \text{watts})\left(\frac{123}{184}\right)\left(1+\frac{3.412}{9}\right) = -1.84$$

Equation (18) is the final regression equation for the simulation study. This regression equation is used directly for the C-CDA portion of the study, and it is this equation that will be manipulated to arrive at the S-CDA framework.

Note that the lighting coefficients β_1 and β_6 contain adjustments for lighting/cooling interactions (3.412/9 is the cooling-related watts, per watt of lighting); this adjustment is necessary for estimation purposes since the lighting regressors in equation (8) are exactly proportional to those that are implicit in the AC equation (15); note also that β_1 and β_6 are in units of watts. Most important, note that β_1 is the lighting Base Usage coefficient around which the Appendix Z issue revolves, and β_6 is the gross energy impact of the lighting measure (1.84 watts per square foot of affected lighted space). Of course, β_1 will be estimated using only C-CDA, and β_6 will be estimated using both the C-CDA and S-CDA estimators.

> Key Point: The gross impact of the measure is 1.84 watts per square foot

A key element of the analysis will be to specify an error-variance model for the disturbance term ε_{ijt} in equation (18). The error-variance model will simply be,

$$\text{(19)} \qquad E\left(\varepsilon_{ijt}^2\right) = \sigma_{ij}^2$$

The actual values for the disturbances is determined by a zero-mean uniform probability distribution:

$$\varepsilon_{ijt} = k_{ij}\left(\overline{kWh_{ij}}\right)\left(\sqrt{3}\right)\left(u(0 \pm .5)_{ijt}\right)$$

where $\overline{kWh_{ij}}$ is the average for the three-year monthly series that can be generated using equation (18) (ignoring ε_{ijt}), and $k_{ij} = u(.2 \pm .1)_{ij}$ is a scaling factor for the disturbances. The implication of this overall structure for the disturbance term is that the standard deviation of the disturbance term can range (for customer j during iteration i) from 10-30% of the average value for expected monthly consumption. As a result, while the disturbances are homoscedastic for customer j during iteration i, the variance of the disturbances is unrestricted across customers and iterations. The implications for estimation are discussed at a later point (see the "**Basic Accuracy**" section).

As a check on the reasonableness of the model, some simulated data should be examined. The following values were adopted as a basis for generating a single three-year monthly energy consumption series:

$$h_{ij} = 3{,}000 \qquad SQFT_{ij} = 25{,}000 \qquad SQFT_{ij}^A = (.90)SQFT_{ij}$$

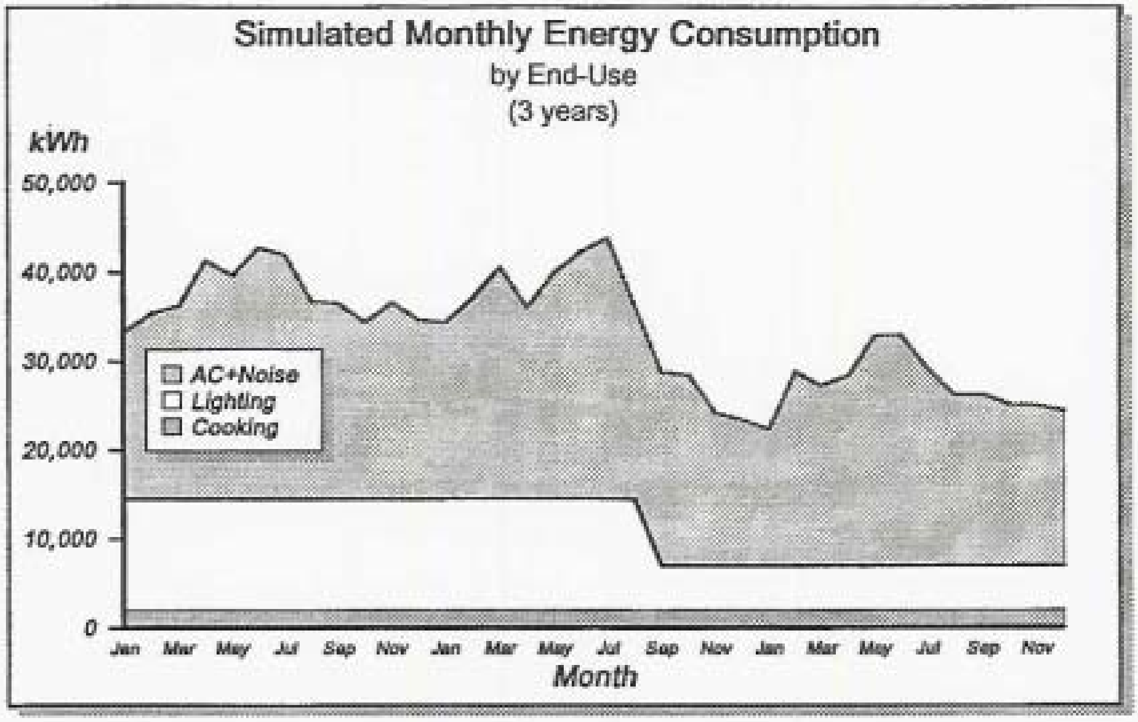
$$OCCUPANTS_{ij} = 260 \qquad EQUIP_{ij} = 130 \qquad d_{ij}^{cook} = 1$$

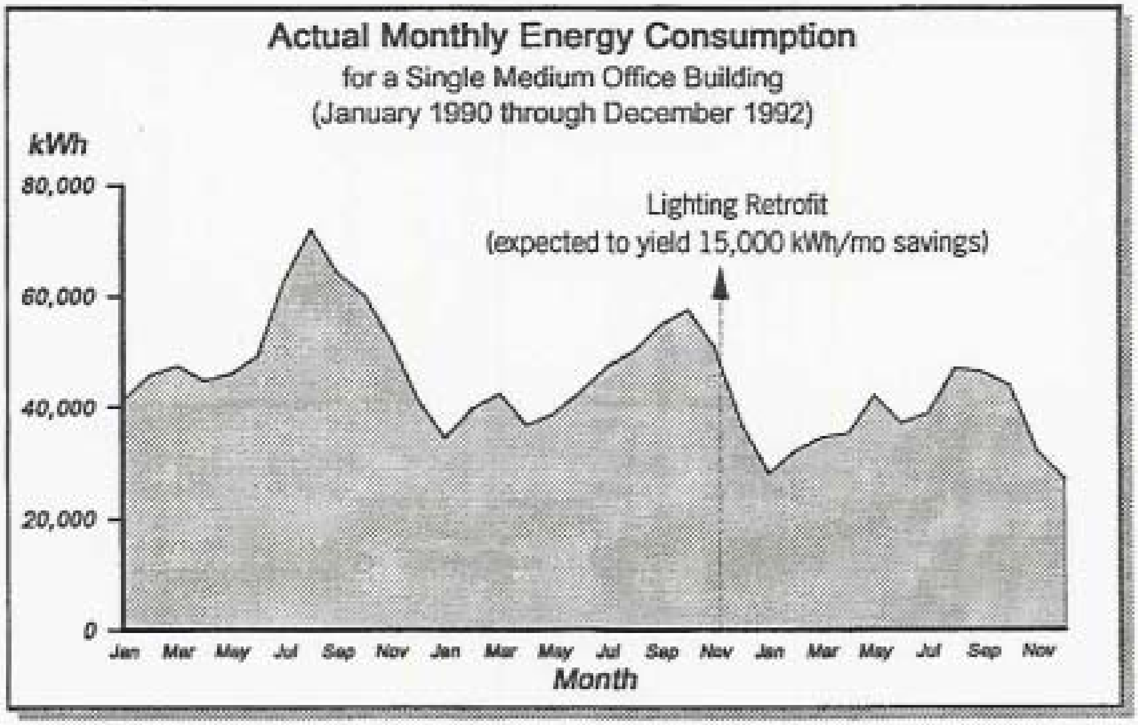
The weather variable $\overline{CLTD_{it}}^{roof}$ was evaluated using Table 1 and the procedure that is associated with equation (14), and the dummy variable d_{ijt} was set at one after August of the second year. The following choices determined the disturbance terms:

$$k_{ij} = 0.15$$

$$\overline{kWh_{ij}} = 20{,}000$$

Figure 7


Simulated Monthly Energy Consumption
by End-Use
(3 years)
kWh
50,000
40,000
30,000
20,000
10,000
0
AC+Noise
Lighting
Cooking
Jan Mar May Jul Sep Nov Jan Mar May Jul Sep Nov Jan Mar May Jul Sep Nov
Month


Actual Monthly Energy Consumption
for a Single Medium Office Building
(January 1990 through December 1992)
kWh
80,000
60,000
40,000
20,000
0
Lighting Retrofit
(expected to yield 15,000 kWh/mo savings)
Jan Mar May Jul Sep Nov Jan Mar May Jul Sep Nov Jan Mar May Jul Sep Nov
Month

The result is the simulated consumption data found in the top portion of Figure 7. As a check on reasonableness, these results are compared in Figure 7 to actual monthly consumption data for a particular medium-sized office building within the SDG&E service territory.

Simplified Conditional Demand Analysis

The General Framework[7]

At this point, the C-CDA framework is generalized so that a general S-CDA framework can be derived. (During this effort the iteration subscript i will be suppressed.) C-CDA is based on a regression model of energy consumption for customer j at time t as it depends on a set of K regressors f_k that are functions of a vector of weather variables and customer characteristics $\bar{X}_{jt}$:

$$(20) \qquad kWh_{jt} = \sum_{k=1}^{K} \beta_k \{ f_k(\bar{X}_{jt}) \} + \varepsilon_{jt}.$$

The K regressors f_k represent the end-use components of energy consumption and their relevant interactions. Specifically, C-CDA is based on the complete and explicit mathematical specification of the K functions f_k, and the completion of the data-collection effort that is associated with $\bar{X}_{jt}$.

To arrive at a S-CDA framework, we note that generally at least some (say, K^*) of the regressors <u>*which are not central to the analysis*</u> have factors g_k that are independent of time:[8]

$$(21) \qquad f_k(\bar{X}_{jt}) = \{ g_k(\bar{w}_{jk}) \} \{ h_k(\bar{x}_{jk}, \bar{z}_{jkt}) \}, \quad \bar{X}_{jt} = \{ \bar{w}_{jk}, \bar{x}_{jk}, \bar{z}_{jkt} \}, \quad k = 1, \ldots K^*.$$

Defining,

$$(22) \qquad \beta_{k,j} = \beta_k \{ g_k(\bar{w}_{jk}) \},$$

the C-CDA equation (20) can be rewritten as,

$$kWh_{jt} = \sum_{k=1}^{K^*} \beta_{k,j} \{ h_k(\bar{x}_{jk}, \bar{z}_{jt}) \} + \sum_{k=K^*+1}^{K} \beta_k \{ f_k(\bar{X}_{jt}) \} + \varepsilon_{jt}.$$

Define Ω as the set of values for k (within the K^* terms above) for which the corresponding regression component is entirely independent of time, so that $h_k = 1$ for each k in Ω. This yields the final S-CDA regression model:

[7] For the most part, this section is taken from SDG&E report MIAP-92-P50-S01-R320, *Commercial/Industrial Energy Efficiency Incentives: Lighting Retrofit, Estimation of Gross Energy-Demand Impacts*, June 1993. This report gives the results of a highly successful application of S-CDA to data that are associated with the lighting retrofit program.

[8] The vector $\bar{x}_{jk}$ is included as a possible argument to the function h_k to emphasize the point that some of the arguments to the function h_k may be independent of time.

$$\begin{aligned} kWh_{jt} &= \sum_{k \in \Omega} \beta_{k,j} + \sum_{k \in \Omega} \beta_{k,j} \left\{ h_k\left(\bar{x}_{jk}, \bar{z}_{jt}\right) \right\} + \sum_{k=K^*+1}^{K} \beta_k \left\{ f_k\left(\bar{X}_{jt}\right) \right\} + \varepsilon_{jt} \\ &= \beta_{0,j} + \sum_{k \in \Omega} \beta_{k,j} \left\{ h_k\left(\bar{x}_{jk}, \bar{z}_{jt}\right) \right\} + \sum_{k=K^*+1}^{K} \beta_k \left\{ f_k\left(\bar{X}_{jt}\right) \right\} + \varepsilon_{jt} \end{aligned} \tag{23}$$

Note the important features of equation (23), in contrast to equation (20). In limiting the number of regressors to K, equation (20) requires the explicit mathematical specification of the functions g_k, as well as the data elements of $\bar{w}_{jk}$. However, while equation (23) involves a large number of econometric *dummy* variables, it is independent of both the functions g_k and the data that are associated with the vector $\bar{w}_{jk}$. (While computer memory issues may need to be resolved, modern software tools such as *SAS* PROC GLM are readily available to meet the task of constructing the large number of econometric *dummy* variables that are associated with equation (23).) In addition, the above formulation points to additional flexibility that is available. A more flexible alternative to the specifications that are represented in equations (20)-(22) is,

$$kWh_{jt} = \sum_{k=1}^{K} \beta_k \left\{ f_{k,j}\left(\bar{X}_{jt}\right) \right\} + \varepsilon_{jt}, \quad f_{k,j}\left(\bar{X}_{jt}\right) = \left\{ g_{k,j}\left(\bar{w}_{jk}\right) \right\} \left\{ h_k\left(\bar{x}_{jt}, \bar{z}_{jt}\right) \right\}, \quad \beta_{k,j} = \beta_k \left\{ g_{k,j}\left(\bar{w}_{jk}\right) \right\}$$

This set of three equations allows for the failure, across customers, of the K^* mathematical specifications g_k that must hold when C-CDA is applied. Yet these equations also yield equation (23), pointing to the added freedom of specification that is associated with S-CDA.

These facts elucidate the fundamental value of S-CDA in comparison to C-CDA. In cases where only the $K - K^*$ regression coefficients β_k of equation (23) are of genuine interest, the remaining regression coefficients of equation (23) can be estimated unconstrained across customers, without regard to the specification of the functions g_k and without the data collection efforts that are associated with the vector of customer characteristics $\bar{w}_{jk}$. In general, the application of S-CDA will minimize the estimation errors that result when the regression components $g_{k,j}\left(\bar{w}_{jk}\right)$ are erroneously constructed. Specifically, S-CDA will minimize the impact in the case where the functions g_k (or $g_{k,j}$) are mathematically misspecified, and *errors-in-variables bias* in the case where the data that are associated with $\bar{w}_i$ are recorded with errors.

The framework represented by equations (21)-(23) addresses a more general econometric issue. Demand-side measurement activity is often characterized by an almost automatic tendency toward C-CDA. The assumption in these cases is that it is possible (and even necessary) to estimate the K parameters of equation (20) and, equivalently, to explicitly model and estimate all the end-uses and their interactions. Certainly this tendency does not constitute careful econometric modeling. On the other hand, the framework (21)-(23) accounts for the fact that in using cross-section/time-series data, careful consideration should always be made as to which elements of the regression are explicitly modeled and which are not. These sorts of modeling considerations are emphasized in basic econometric texts (see Judge, *et al.*, 1985b, Chapter 13), yet they are frequently ignored within many current demand-side measurement efforts.

The Application of the Simplified Conditional Demand Analysis Framework

Certainly equation (18) is of the general form given in equation (20) for K = 6, and (continuing to suppress the iteration subscript i),

$$\bar{X}_{jt} = \left\{ h_j \quad SQFT_j \quad d_{jt} \quad SQFT_j^A \quad \overline{CLTD}_t^{roof} \quad OCCUPANTS_j \quad EQUIP_j \quad d_j^{cook} \right\}$$

The associated functions f_k can be easily constructed. In line with equation (21), for $K^* = 5$ let,

$$\begin{pmatrix}
f_1(\bar{X}_{jt}) = \{g_1(\bar{w}_{j1})\}\{h_1(\bar{x}_{j1}, \bar{z}_{j1t})\} & = \left\{\beta_1\left[\left(\frac{h_j}{12\text{ months}}\right)\left(\frac{SQFT_j}{1{,}000}\right)\right]\right\}\{1\} & = \beta_{1,j}\{1\} \\
f_2(\bar{X}_{jt}) = \{g_2(\bar{w}_{j2})\}\{h_2(\bar{x}_{j2}, \bar{z}_{j2t})\} & = \{\beta_2[OCCUPANTS_j]\}\{1\} & = \beta_{2,j}\{1\} \\
f_3(\bar{X}_{jt}) = \{g_3(\bar{w}_{j3})\}\{h_3(\bar{x}_{j3}, \bar{z}_{j3t})\} & = \{\beta_3[EQUIP_j]\}\{1\} & = \beta_{3,j}\{1\} \\
f_4(\bar{X}_{jt}) = \{g_4(\bar{w}_{j4})\}\{h_4(\bar{x}_{j4}, \bar{z}_{j4t})\} & = \left\{\beta_4\left[d_{ij}^{cook}\right]\right\}\{1\} & = \beta_{4,j}\{1\} \\
f_5(\bar{X}_{jt}) = \{g_5(\bar{w}_{j5})\}\{h_5(\bar{x}_{j5}, \bar{z}_{j5t})\} & = \left\{\beta_5\left[\left(\frac{SQFT_j}{1{,}000}\right)\right]\right\}\left\{\overline{CLTD}_t^{roof}\right\} & = \beta_{5,j}\left\{\overline{CLTD}_t^{roof}\right\}
\end{pmatrix}$$

Recognizing that $\Omega = \{k = 1,2,3,4\}$, and that $K - K^* = 1$ (the single coefficient β_6 is of prime interest), the final S-CDA equation (23) is (re-introducing the iteration subscript i),

$$\text{(24)} \qquad kWh_{ijt} = \beta_{0,ij} + \beta_{5,ij}\left\{\overline{CLTD}_{it}^{roof}\right\} + \beta_6\left[\left(d_{ijt}\right)\left(\frac{h_{ij}}{12\text{ months}}\right)\left(\frac{SQFT_{ij}^A}{1{,}000}\right)\right] + \varepsilon_{ijt}$$

where,

$$\beta_{0,ij} = \left\{\beta_1\left[\left(\frac{h_{ij}}{12 \text{ months}}\right)\left(\frac{SQFT_{ij}}{1{,}000}\right)\right]\right\} + \left\{\beta_2\left[OCCUPANTS_{ij}\right]\right\}$$
$$+\left\{\beta_3\left[EQUIP_{ij}\right]\right\} + \left\{\beta_4\left[d_{ij}^{cook}\right]\right\} = \beta_{1,ij} + \beta_{2,ij} + \beta_{3,ij} + \beta_{4,ij}$$

Note the properties of equation (24), which is the final S-CDA regression equation for the simulation study. First, the customer-specific intercept $\beta_{0,ij}$ subsumes the lighting Base Usage coefficient β_1, so that this parameter is not identified (in the econometric sense) when S-CDA is applied. In addition, the key lighting-savings parameter β_6 is constrained across customers while the remaining regression coefficients--the intercepts $\beta_{0,ij}$ and the weather coefficients $\beta_{5,ij}$--are to be estimated unconstrained across customers. Therefore, equation (24) does not depend on the data elements,

$$(25) \qquad \left\{SQFT_{ij} \quad OCCUPANTS_j \quad EQUIP_j \quad d_j^{cook}\right\}$$

which constitute (for any k) the elements of $\tilde{w}_{jk}$ of equations (21) and (22). This fact points to the significant value of the S-CDA framework in that these data elements--which are required for the implementation of C-CDA--are very difficult and expensive to collect in that they require detailed on-site collection of data, and they are typically not part of the relevant DSM program database. (Moreover, the only reason for collecting these data would be for purposes of estimating the unneeded coefficients of equation (18), or perhaps for some gain in efficiency that comes from exploiting the exact structure of equation (18). However, this last argument requires the assumption that the exact form of the regression equation is known, whereas as already stated the S-CDA framework allows for a great deal of flexibility in this area; the possible gain in efficiency is an issue that will be addressed as the study results are reported in the "**Results**" portion of this report.) The data elements that are present in equation (24) are much more easily obtained and may well be included in the program database.

It is the prime purpose of this study to analyze the properties of least-squares regression estimators as they are applied to the two alternative specifications (18) and (24), in light of the fact that it is the goal of each estimator to estimate only the single lighting-savings parameter β_6.

Results

Final Structure of the Simulation Study

Figure 8 shows the final structure of the simulation study, based on the groups A-D of Figure 4. The top portion of Figure 8 describes the part of the study that supports the Appendix Z filing. With respect to the net-to-gross issue, the data that come from groups A-D (including hours-of-operation) are used for discrete choice model estimation; the hours-of-operation for the participant group (A and B jointly) are then used to estimate the net-to-gross ratio by estimating participants' behavior in the absence of the customer incentive payment. With respect to the gross savings issue, only the energy consumption data from group B (participants who adopt) enter into the S-CDA model; groups C and D are kept out of the model since the Appendix Z filing specifies this position, and group A (nonadopters) can add nothing to the S-CDA model due to the fact that the only parameter that is estimated across customers is the parameter β_6 of equation (24), which only applies in the case of adoption.

The bottom portion of Figure 8 shows the part of the study that is associated with the protocols. The participant and nonparticipant groups are used in generating separate estimates of gross savings, providing the basis for an estimate of the net impact.[9]

Figure 8 specifically points to the fact that the participant groups A and B will be used in comparing the S-CDA model of Appendix Z and the C-CDA model of the protocols. As a result, for any iteration, with 200 customers in each of the two groups (the participant group and the nonparticipant group), the S-CDA model will be estimated using approximately 120 adopters (based on the 60% figure of equation (4)), while the C-CDA model will be estimated with the 200 participants, with approximately 120 of them also being adopters.

Details of the simulation study will now be provided. The results concerning the estimation of the gross impact will be considered first due to the relative importance of this issue and to the fact that the net impact as calculated under the protocols depends on the estimates across the participant and nonparticipant groups of the gross energy impact. Estimation results for the net-to-gross and net impact issues will then be presented.

The Basic Accuracy of the C-CDA and S-CDA Estimators of Gross Impact

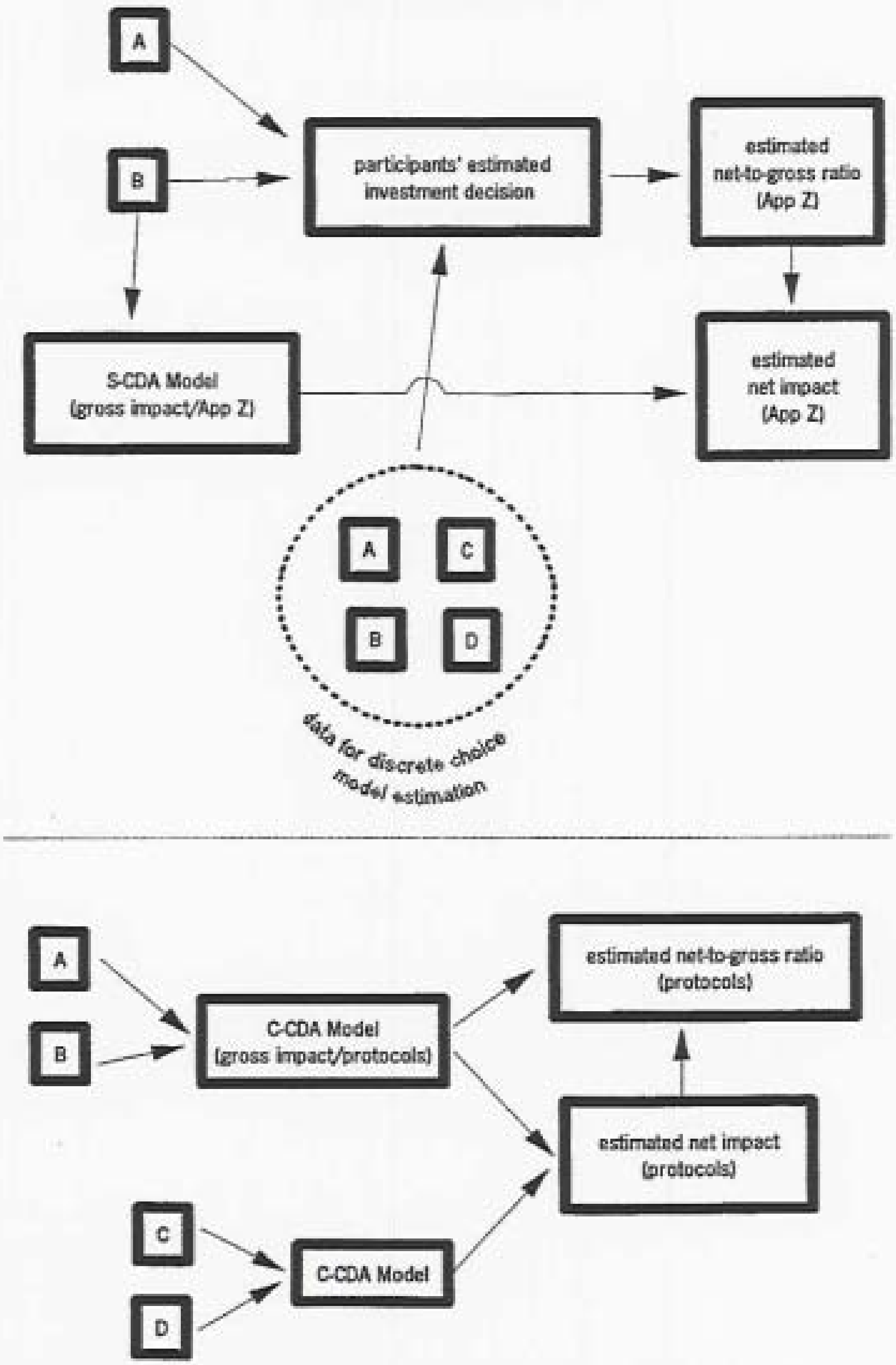
As already mentioned, Figure 8 points to the fact that the participant groups A and B will be used in estimating the S-CDA model of Appendix Z and the C-CDA model of the protocols (although since group A cannot possibly add information to the S-CDA model, the estimation of the S-CDA model will be based entirely on simulated data from group B). This allows for the two competing models to be on equal footing for comparison purposes.

Since every data element of the regression equations (18) and (24) has been clearly established, the estimation of the key lighting-savings parameter β_6, for each of the equations (18) and (24) (and for any number of iterations), is a straightforward task. The only complicating factor in applying *least-squares* techniques are the implications for efficient estimation that are associated with the error-variance model of equation (19). The task is to estimate the customer-specific error-variance σ_{ij}^2 (assumed, in the study, to be unknown) which can form the basis for an application of *weighted least-squares*. This can be accomplished by obtaining adequate estimates of

[9] This is in accordance with the protocols' Table C-4.

Figure 8

Structure of the Simulation Study



the disturbance terms ε_{ijt} that are common to equations (18) and (24). While for this purpose either of the equations (18) and (24) could be estimated using *ordinary least-squares*, a third regression equation was defined:

$$kWh_{ijt} = \beta_{0,ij} + \beta_{5,ij}\left\{\overline{CLTD_{it}^{roof}}\right\} + \beta_{6,ij}\left[\left(d_{ijt}\right)\right] + \varepsilon_{ijt}$$

This regression equation can be viewed as the S-CDA equation (24) in its completely unconstrained form. In light of the error-variance model of equation (19), the *ordinary least-squares* estimator will be unbiased and efficient if applied (by customer) to this regression equation, and the resulting mean squared error constitutes a consistent estimator for the customer-specific error-variance σ_{ij}^2; the inverse of this estimate serves as the basis for the application of *weighted least-squares* to equation (18), the C-CDA equation. This two-phase least-squares technique was the fundamental technique that was adopted within the study.

Moreover, given the coding conventions of the *SAS* procedure PROC GLM, the estimation of the S-CDA equation (24) is nearly as simple. After defining a customer indicator variable "CUST," the single line of programming code that defines the regression equation (24) within PROC GLM is,

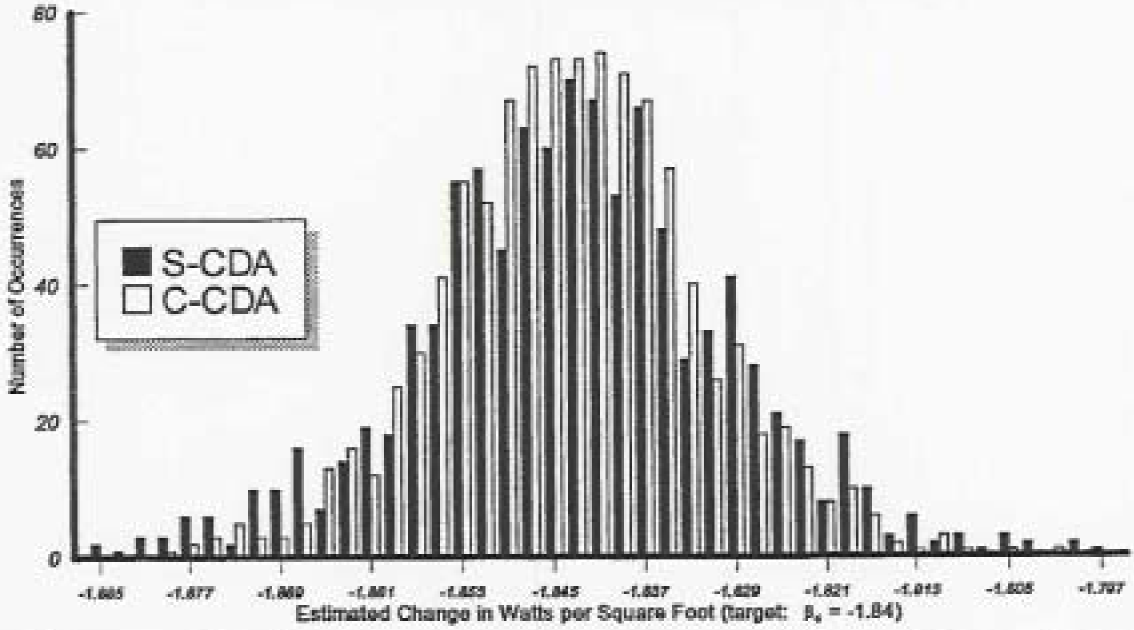
```
MODEL KWH = CUST  CUST*CLTD  D*HOURS*SQFTAFF ;
```

where the variable names in the above piece of code have an obvious relationship to the variables in equation (24) (e.g., HOURS = $h_{ij}/12{,}000$). The "CUST" and "CUST*" pieces of the code tell the program to estimate a separate intercept $\beta_{0,ij}$ for each customer and a separate weather coefficient $\beta_{5,ij}$ for each customer, respectively. However, since the third term in this expression ("D*HOURS*SQFTAFF") does <u>not</u> contain an element "CUST*," the coefficient on this term (β_6) is <u>constrained across customers</u>. (The weighting scheme that has been described in the previous paragraph was used here as well.)

Figure 9 gives some of the most important results of the simulation study. The *weighted least-squares* estimates of the lighting savings parameter β_6 (the gross impact of the single lighting measure), for the C-CDA equation (18) and the S-CDA equation (24), are given for 1,000 simulation computer runs. Figure 9 shows that both estimators are obviously (not surprisingly) unbiased, since each distribution has a mean estimate for β_6 of -1.84 watts. More important, Figure 9 shows, given the distribution of the estimated values for β_6, that the two estimators for β_6 are extremely close in terms of accuracy; the two distributions are virtually congruent, although the tails of the S-CDA distribution are slightly more pronounced (in fact, the standard deviation of the C-CDA distribution (0.011) is 80% of the standard deviation of the S-CDA distribution (0.014)). This is a remarkable result in light of the fact that the S-CDA estimator does not depend on the significant data elements (25): square footage of the building, the number of building occupants, equipment inventories, and cooking data. In other words, the added accuracy that would, in practice, be associated with the data elements (25) is minimal. Moreover, this small degree of added accuracy is predicated on the data-collection accuracy of the data elements (25) (e.g., Can the square footage of a building be consistently and accurately measured?), and the exact mathematical specification of the of the C-CDA equation (18). It is the purpose of the next section to describe the basic properties of the two competing estimators under conditions of data-collection errors and modeling misspecification.

Figure 9

The Distribution of Results (1,000 iterations) Over the Estimated Change in Energy Usage



The Properties of the C-CDA and S-CDA Estimators Under More Realistic Conditions

Of great concern is the performance of the C-CDA and S-CDA estimators under more realistic conditions. As already mentioned, the S-CDA framework is theoretically more robust in cases where data are collected with errors, and when the regression model in equation (18) is misspecified by the analyst. It is the purpose of this section to present the simulation results that support this.

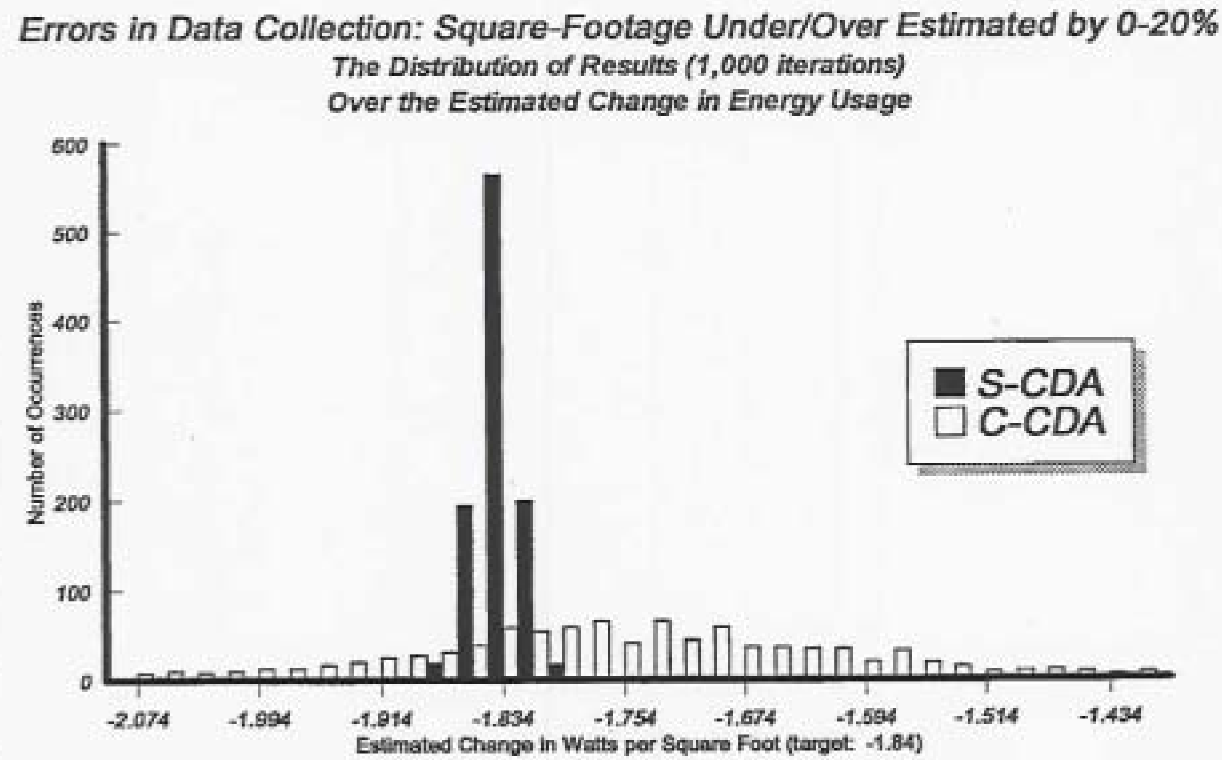
Errors-in-Variables Bias

It is well known that when regression models contain regressors for which data are collected with errors, the resulting *errors-in-variables bias* (see Judge, 1985a, pp. 532-535) will impact the estimation process, and that parameter estimates will tend to be biased toward zero. Figure 10 shows the results for the C-CDA and S-CDA models in a simulated case of *errors-in-variables bias*. These results are based on the case where the regressor,

$$\mathrm{SQFT}_{ij}^{*} = \left(1+u(0\pm0.2)\right)_{ij}\left(\mathrm{SQFT}_{ij}\right)$$

is substituted, during the estimation phase, for $SQFT_{ij}$ in equation (18). (This structure for $SQFT^{*}_{ij}$ implies that the building square footage on average will be recorded correctly, but in any one instance will be recorded with error by as much as ±20%.) Figure 10 shows the dramatic *errors-in-variables bias* that is associated with the C-CDA estimator in this case, as well as a dramatic decrease in accuracy. The resulting estimates are clearly biased toward zero away from the true value for the measure savings parameter β_6 (-1.84 watts). Moreover, since equation (24) does not depend on $SQFT_{ij}$, the variance of S-CDA model in equation (24) is unaffected by the errors (Figure 10 also shows S-CDA results for an additional 1,000 iterations). These results support the notion that the cost of estimating the individual end-use elements that are contained in equation (18) will most likely be a substantial *errors-in-variables bias* and a significant decrease in accuracy.

Figure 10



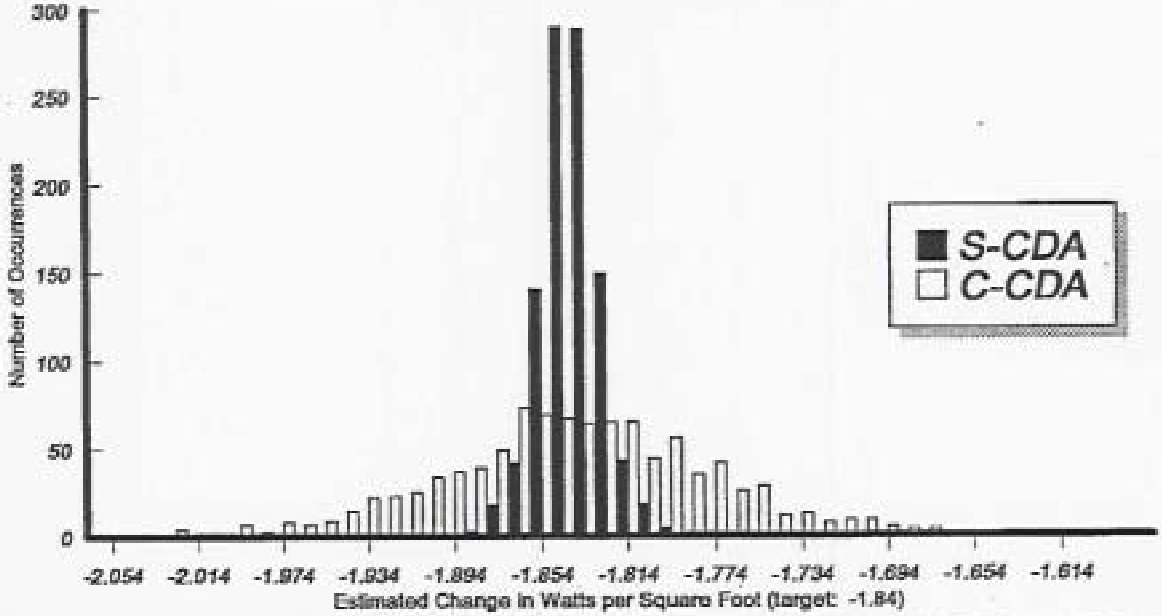
Misspecification of the Regression Equation

Problems can also occur when the regression equation is mathematically misspecified (see Pindyck & Rubinfeld, 1981). Figure 11 shows the results of omitting the "cooking" end-use (2,000 kWh per month) from the C-CDA regression equation (18) during the estimation phase. The variance of the C-CDA estimator is clearly dramatically increased (most likely due to the fact that the misspecifcation of the regression equation tends to impact the error term in the equation), although the associated bias is not conspicuous (undoubtedly due to the

Figure 11

Misspecification Errors: A Minor End-Use is Omitted from the Model

The Distribution of Results (1,000 iterations) Over the Estimated Change in Energy Usage



small consumption that is associated with this end-use). Since equation (24) does not depend on the cooking indicator d_{ij}^{cook}, the S-CDA model in equation (24) is unaffected by the misspecification (Figure 11 also shows S-CDA results for an additional 1,000 iterations). These results show that the accuracy of C-CDA is significantly impacted by *misspecifcation error*.

Analysis of Net-to-Gross/Net Impact

At this point, consider the issue of net impact. The fundamental task is to compare the protocols' approach--that of equation (2)--with the approach that is advocated in the Appendix Z filing--that of using S-CDA in conjunction with Discrete Choice Analysis. However, rather than considering the net impact alone, the net-to-gross ratio (the ratio of the net impact of the program to the gross impact) will be considered first. This will be useful in that the net-to-gross ratio is a unit free number that can be readily explained and understood. In addition, under the Appendix Z filing the net-to-gross ratio is estimated separately from the gross impact, using Discrete Choice Analysis (while this is not the case under the protocols' framework). This implies that it will be valuable to consider the properties of Discrete Choice Analysis on their own.

In order to derive the net-to-gross ratio note that within the context of the simulation study, $-\beta_6$ of equation (18) is the gross savings for the measure (1.84 watts per square foot of affected lighted space). For N participants the expected gross impact of the program (in watts) is (suppressing the iteration subscript i),

(30) $$\left(\text{Nonparticipant Group Load Impact}\right)=\sum_{j\in D}\left(\hat{\beta}_6^{\text{nonpart}}\right)\left(\text{SQFT}_j^{\text{affected}}\right)\Big/\sum_{j\in\{C\text{ or }D\}}\left(\text{SQFT}_j^{\text{affected}}\right)$$

(In line with the protocols' Table C4, the parameter β_6 is estimated separately for participants and nonparticipants, using the C-CDA equation (18) and the *least-squares* procedure that has already been described.) The first term should equal approximately 60% of 1.84 watts, while the second term should equal approximately 30% of 1.84 watts. The estimated net-to-gross ratio is then

(31) $$\left(\frac{\text{Net}}{\text{Gross}}\right)_{\text{protocols}}=1-\left\{\left[\frac{\sum_{j\in D}\left(\hat{\beta}_6^{\text{nonpart}}\right)\left(\text{SQFT}_j^{\text{affected}}\right)}{\sum_{j\in\{C\text{ or }D\}}\left(\text{SQFT}_j^{\text{affected}}\right)}\Bigg/\frac{\sum_{j\in A}\left(\hat{\beta}_6^{\text{part}}\right)\left(\text{SQFT}_j^{\text{affected}}\right)}{\sum_{j\in\{A\text{ or }B\}}\left(\text{SQFT}_j^{\text{affected}}\right)}\right]\right\},$$

which should also (as in the case of Discrete Choice Analysis) be very close to one-half $(1-(.3/.6))$.

Estimation Results: Net-to-Gross Ratio

Figure 12 gives the main results for comparing the alternative estimators for the net-to-gross ratio. (The underlying assumptions are those of perfect data collection and model specification, as in the case of Figure 9.) The means of the distributions are 0.47 and 0.48 for the Discrete Choice Analysis (with *maximum likelihood* estimation) and the protocols' methods, respectively. With respect to accuracy, the results show that under the best conditions the two estimators are comparable, with there being some evidence of a wider distribution for the protocols' method: the standard deviation of the distribution for the protocols' method (.071) is approximately 50% higher than that of the Discrete Choice Analysis method (.048). This is most likely attributable to the fact that the estimate of the net-to-gross ratio in equation (31) depends on the two C-CDA estimates of load impact $\hat{\beta}_6^{\text{nonpart}}$ and $\hat{\beta}_6^{\text{part}}$. This tends to add more variation to the estimate since the estimation technique does not directly address the issue of investment decisions; rather, it approaches the net impact issue indirectly, through the energy consumption portion of the model and the twofold application of C-CDA.

Figure 13 gives results that are related to Figure 10 (where building square footage on average was recorded correctly, but in any one instance was be recorded with error by as much as ±20%). Note that the bias that is evident in the protocols' estimator for the gross impact (see Figure 10) is not apparent in Figure 13, although the increase in variation is still conspicuous. This is not surprising since the calculation in equation (31) will cause the bias that is in both $\hat{\beta}_6^{\text{nonpart}}$ and $\hat{\beta}_6^{\text{part}}$ to cancel. In fact, the mean for the distribution in Figure 13 for the protocols' estimator is 0.48, as in Figure 12. However, in Figure 13, the standard deviation of the protocol's estimator is twice that for the Discrete Choice Analysis method, so that still follows that the *errors-in-variables* impact on the accuracy of the protocols' estimator is still very significant.

Estimation Results: Net Impact

Equations (26) and (27) provide the basis for estimating net load impact per square footage of affected lighted space under the Appendix Z filing (with β_6 estimated using S-CDA):

$$\frac{(\beta_6)\left(\sum_{j\in\{A\,or\,B\}} q(h_j, 25\%)\left(SQFT_j^{affected}\right) - \sum_{j\in\{A\,or\,B\}} q(h_j, 0\%)\left(SQFT_j^{affected}\right)\right)}{\sum_{j\in\{A\,or\,B\}}\left(SQFT_j^{affected}\right)}$$

The true value for this parameter was calculated using the actual values for the Investment Decision Function and realized values for the square footage of affected lighted space over a large number of customers in groups A and B, with the result being -0.52 watts. In turn, equations (29) and (30) provide the basis for estimating Net Load Impact (per square footage of affected lighted space) under the protocols (with β_6 estimated using C-CDA on participants and nonparticipants):

$$(32)\qquad \frac{\left(\hat{\beta}_6^{part}\right)\sum_{j\in B}\left(SQFT_j^{affected}\right)}{\sum_{j\in\{A\,or\,B\}}\left(SQFT_j^{affected}\right)} - \frac{\left(\hat{\beta}_6^{nonpart}\right)\sum_{j\in D}\left(SQFT_j^{affected}\right)}{\sum_{j\in\{C\,or\,D\}}\left(SQFT_j^{affected}\right)}$$

Earlier in this report, when the C-CDA and S-CDA <u>gross-impact</u> estimators were being compared, it was noted that the S-CDA estimator had a slightly higher standard deviation (as evidenced by Figure 9). However, Figure 12 showed that the net-to-gross estimator from the protocols has a larger variance, due to the two-fold application of C-CDA. This alerts us to the fact that either of the estimators for net impact may end up with a higher degree of variation.

The results given in Figure 14 address this issue. There it is obvious that the net-impact estimator from the protocols has a higher standard deviation (.098) relative to that of the Appendix Z estimator (0.067), by 50%. This implies a greater degree of accuracy on the part of the Appendix Z estimator for the Net Load Impact.

Summary of Results

One of the main issues in SDG&E's Appendix Z filing has been addressed in Figure 9. There it is clear that when site-specific data are collected accurately and when the mathematics of the C-CDA regression are correctly specified (which implies that the S-CDA regression equation is correctly specified), the two estimators for the gross impact of the measure--C-CDA and S-CDA--are comparable in terms of accuracy. However, Figures 10 and 11 show that when site-specific data are collected with error or when the mathematics of the C-CDA regression are misspecified, the S-CDA estimator for the gross impact of the DSM measure greatly outperforms the C-CDA estimator in terms of accuracy. This is an important although predictable result, since the S-CDA estimator is not heavily dependent on those data which are extremely expensive to collect and which are very difficult to collect with accuracy.

On the issue of net impact, Figure 12 points out that the Discrete Choice Analysis approach to estimating the net-to-gross ratio is somewhat more accurate than the protocols' approach, due to the fact that Discrete Choice Analysis approaches the issue of net impact directly, while the protocols' approach is an indirect one where the C-CDA estimator is applied twice during the net impact calculation. In addition, Figure 13 indicates that the relative inaccuracy of the protocols' approach is increased further when site-specific data are collected with error.

Most important, Figure 14 shows that when S-CDA and Discrete Choice Analysis--the two techniques that are advanced in the Appendix Z filing--are used jointly, the final result is a more accurate estimate of Net Load Impact.

Figure 12

The Distribution of Results (1,000 iterations) Over the Estimated Net-to-Gross Ratio

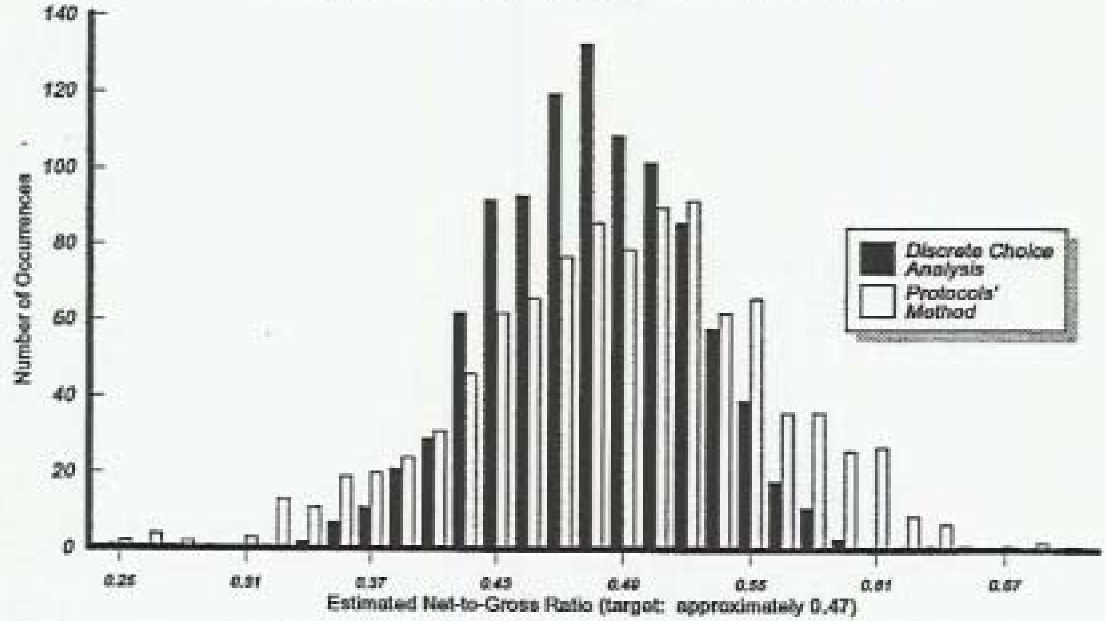


Figure 13

Errors in Data Collection: Square-Footage Under/Over Estimated by 0-20%

The Distribution of Results (1,000 iterations)
Over the Estimated Net-to-Gross Ratio

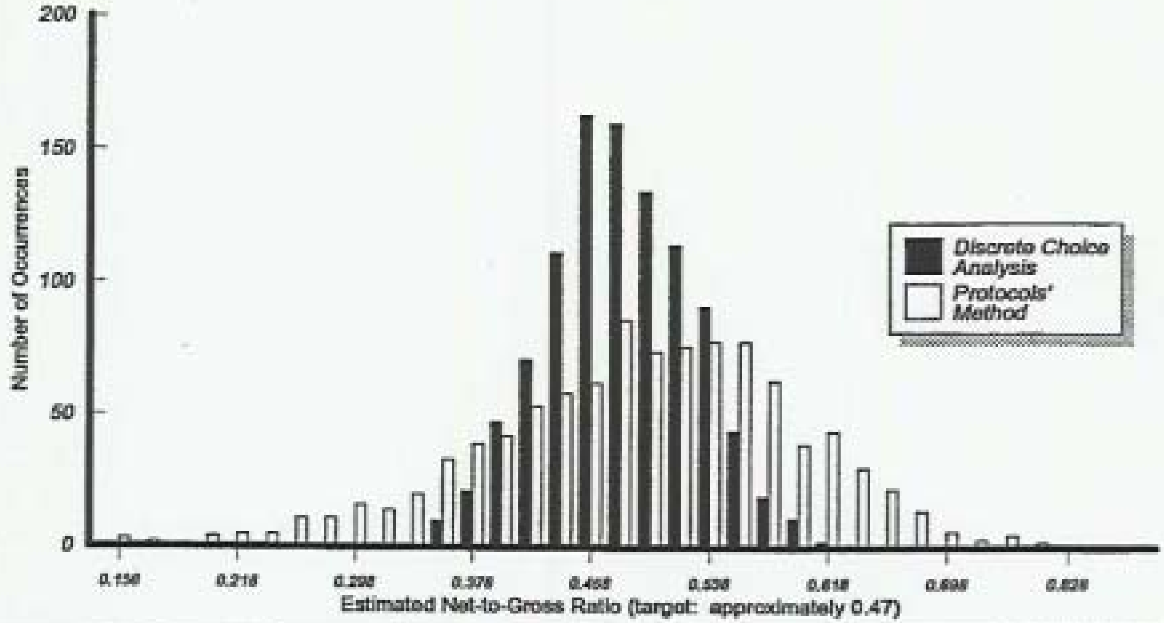
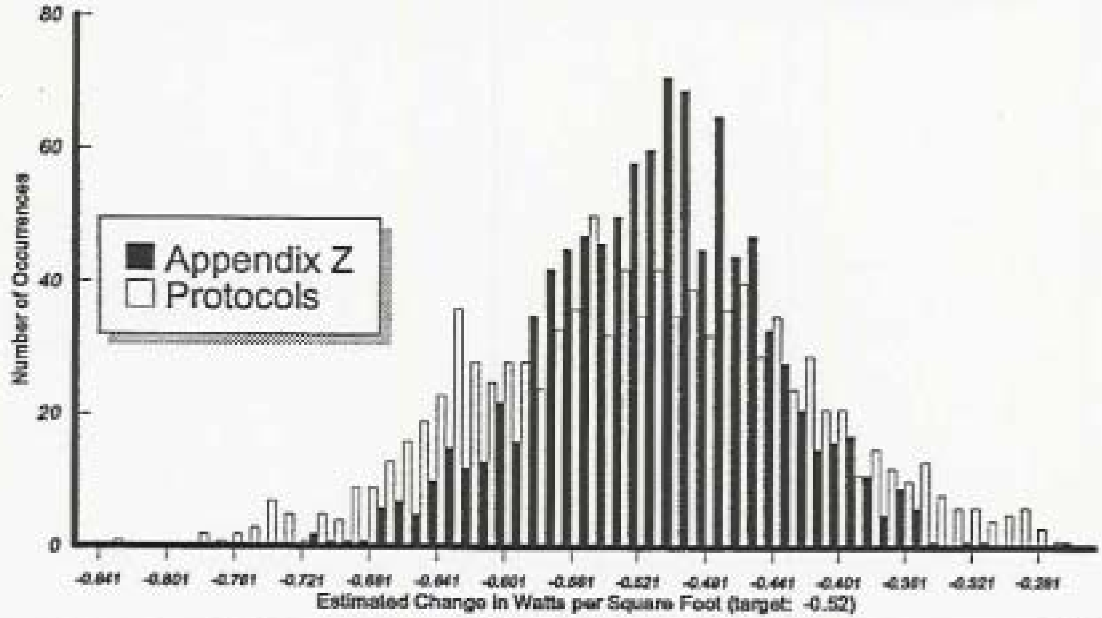


Figure 14

The Distribution of Results (1,000 iterations)
Over the Estimated Net Change in Energy Usage



References

Aigner, Dennis J. and Peter Schonfeld, "Experimental Design for Direct Metering of Residential Electricity End-Uses," Econometric Decision Making: Games, Econometrics and Optimisation. Contributions in Honour of Jacques H. Dreze, Edited by J.J. Gabszewicz, *et al.*, North-Holland, 1990.

American Society of Heating, Refrigerating and Air Conditioning Engineers, 1989 ASHRAE Handbook Fundamental I-P Edition, Atlanta, GA, 1989.

Bartels, Robert and Denzig Fiebig, "Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads," Energy Journal, 11(4):79-97, 1990.

Caves, D.W., K.E. Train, *et al.*, "A Bayesian Approach to Combining Conditional Demand and Engineering Models of Electricity Usage," Review of Economics and Statistics 69: 438-448, 1987.

Engle, R.F., C.W.J. Granger and R. Ramanathan, "Regional Load Curve Models, Volumes 2,4," EPRI EA-1672, Final Report, Project RP-1008, 1981.

Fiebig, D., R.A. Bartels and D.J. Aigner, "A Random Coefficient Approach to the Estimation of Residential End-Use Load Profiles," Journal of Econometrics, 50:297-327, 1991.

Hendricks, W. and R. Koenker, "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity," Journal of the American Statistical Association, March 1992

Hendry, David F., "Monte Carlo Experimentation in Econometrics," Handbook of Econometrics, Chpt. 16, Griliches and Intrilligator, North-Holland, 1984.

Hill, D., "The Time-of-Day Demand for Electricity by End-Use: An Analysis of Wisconsin Data," Analysis of Residential Response to Time-of-Day Prices, Report EA-2380, EPRI, Palo Alto, CA, 1982.

Judge, G.G., R. Carter Hill, *et al.*, Introduction to the Theory and Practice of Econometrics, New York, NY: John Wiley & Sons, 1985a.

Judge, G.G., W. E. Griffiths, *et al.*, The Theory and Practice of Econometrics, Second Edition, New York, NY: John Wiley & Sons, 1985b.

Lawrence, A.G. and M. Parti, "Survey of Conditional Energy Demand Models for Estimating Residential Unit Energy Consumption Coefficients," Report EA-3410, EPRI, Palo Alto, CA, 1984.

Maddala, G.S., Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press, 1983.

Parti, Michael and C. Parti, "The Total and Appliance Specific Conditional Demand for Electricity in the Household Sector," The Bell Journal of Economics, 2(1), Spring 1980.

Pindyck, R. S. and D. L. Rubinfeld, Econometric Models & Economic Forecasts, Second Edition, New York, NY: McGraw-Hill, 1981.

Regional Economic Research, "Commercial End-Use Data Development Handbook: COMMEND Market Profiles and Parameters, Volume 2: COMMEND Data and Parameter Development Techniques," EM-5703, EPRI, Palo Alto, CA, 1988.

Regional Economic Research, "Commercial Sector Regional Data--1986 NBECS Market Profiles," Energy Information Administration, 1991.

Schiffman, Dean A., Athena Besa, *et al.*, Commercial/Industrial Energy Efficiency Incentives: Lighting Retrofit Estimation of Gross Energy-Demand Impacts, San Diego Gas & Electric, MIAP-92-P50-S01-R320, June 1993.

Sebold, F.D. and K.M. Parris, "Residential End-Use Energy Consumption: A Survey of Conditional Demand Estimates" Report CU-9487, Research Project 2547-1, EPRI, Palo Alto, CA, Oct. 1989.

SAS Code for Simulation Study[10]

```

%LET N_RUN= 100;
DATA PART;
 DO XRUN=1 TO &N_RUN;
 N=200; * N = NUMBER OF CUSTOMERS IN EACH GROUP;
  DO P_NP=1 TO 2;
   DO NCUST=1 TO N;
   CUST_CNT+1;
   IF CUST_CNT=1 THEN GROUP+1;
   IF CUST_CNT=INT(N/100)+1 THEN CUST_CNT=0;
   CUST=(P_NP-1)*N+NCUST;
   HOURS=ROUND(2500+1000*UNIFORM(0));
    IF P_NP=1 THEN AWARE_YN='YES';
    IF P_NP=2 THEN AWARE_YN='NO';
   * IF AWARE_YN='NO' THEN WT_CATM=1;
   * ELSE WT_CATM=0;
    WT_CATM=1;
    IF AWARE_YN='YES' THEN SUBSIDY=.2493;
    ELSE SUBSIDY=0;
    PAYBACK=(1-SUBSIDY)*44/(.09*HOURS*123/1000);
    PAYBACK=ROUND(PAYBACK*100)/100;
    G1=-4.1773;
    G2= 3.7617;
    LOGIT=1/(1+EXP(G1+G2*PAYBACK));
    IF UNIFORM(0)<LOGIT THEN ADOPT_YN='YES'; ELSE
 ADOPT_YN='NO';
   SQFT=ROUND(25000*(0.75+0.5*UNIFORM(0)));
   SQFT_AFF=ROUND((0.10+0.90*UNIFORM(0))*SQFT);
   SQFT_DUM=SQFT*(1+(2/5)*(UNIFORM(0)-0.5));
   SQFT_APX=SQFT_AFF; * LINE ADDED 06/22/93 ;
   IF ADOPT_YN='NO' THEN SQFT_AFF=0;
   SAVINGS=2.0*SQFT_AFF*123/184;
   IF UNIFORM(0)<0.6 THEN COOK_YN=1; ELSE
 COOK_YN=0;
    IF COOK_YN=1 AND UNIFORM(0)<0.8 THEN
 COOK_YND=1;
    ELSE COOK_YND=0;
  OUTPUT;
   END;
  END;
 END;

 PROC SORT DATA=PART;
 BY XRUN AWARE_YN ADOPT_YN;

 PROC MEANS NOPRINT DATA=PART;
 VAR SQFT_APX;
BY XRUN AWARE_YN ADOPT_YN;
OUTPUT OUT=SQFT1 SUM=SQFT1 ;
PROC MEANS NOPRINT DATA=PART;
VAR SQFT_APX;
BY XRUN AWARE_YN ;
OUTPUT OUT=SQFT2 SUM=SQFT2 ;

DATA COMPLETE;
 DO XRUN=1 TO &N_RUN;
  DO NP=1 TO 4;
  SQFT1=0;
   IF NP=1 THEN DO;
   AWARE_YN='NO '; ADOPT_YN='NO '; END;
   IF NP=2 THEN DO;
   AWARE_YN='NO '; ADOPT_YN='YES'; END;
   IF NP=3 THEN DO;
   AWARE_YN='YES'; ADOPT_YN='NO '; END;
   IF NP=4 THEN DO;
   AWARE_YN='YES'; ADOPT_YN='YES'; END;
  OUTPUT;
  END;
  END;
  DROP NP;

  DATA SQFT1;
 MERGE COMPLETE SQFT1;
 BY XRUN AWARE_YN ADOPT_YN;

 DATA WSQFT;
 MERGE SQFT1 SQFT2 (KEEP=XRUN AWARE_YN SQFT2);
 BY XRUN AWARE_YN;
 IF AWARE_YN='NO' AND ADOPT_YN='NO' THEN
DELETE;
 IF AWARE_YN='YES' AND ADOPT_YN='NO' THEN
DELETE;
 SQFT3=LAG(SQFT1); SQFT4=LAG(SQFT2);
 IF AWARE_YN='NO' AND ADOPT_YN='YES' THEN
DELETE;
 WSQFT=(SQFT3/SQFT4)/(SQFT1/SQFT2);
 KEEP XRUN WSQFT;

 PROC DATASETS;
 DELETE SQFT1 SQFT2 COMPLETE;

 PROC FREQ DATA=PART;
 BY XRUN;
 TABLES AWARE_YN*ADOPT_YN / OUT=COUNT
NOPRINT;

 PROC TRANSPOSE PREFIX=N DATA=COUNT
OUT=COUNT;
 VAR COUNT;
 BY XRUN AWARE_YN;

```

[10] This code may contain portions that are obsolete.

```
DATA COUNT;
SET COUNT;
N_YES=N2; N_NO=N1;
DROP _LABEL_ _NAME_ N2 N1;

DATA SAVSQFT;
SET PART;
KEEP XRUN AWARE_YN ADOPT_YN CUST SQFT_AFF
HOURS COOK_YN COOK_YND;

PROC SORT DATA=SAVSQFT;
BY XRUN AWARE_YN CUST;

PROC SORT DATA = PART;
BY XRUN;

PROC CATMOD ;
RESPONSE LOGIT/OUTEST=CMPARMS;
DIRECT PAYBACK;
MODEL ADOPT_YN=PAYBACK/ NOPARM NOITER
NOPROFILE             NODESIGN NOGLS
ML EPSILON=.000005;
BY XRUN;
WEIGHT WT_CATM;
TITLE2 'PROC CATMOD ON DATA=PART';

DATA CMPARMS;
SET CMPARMS;
IF _TYPE_='PARMS';


DATA PART_NG;
MERGE PART CMPARMS;
BY XRUN;
IF AWARE_YN='YES';
PBCK_NOS=44/(.09*HOURS*123/1000);
PBCK_NOS=ROUND(PBCK_NOS*100)/100;
* LOGIT FOR PARTICIPANTS WITH SUBSISDY
EVALUATED AT ZERO (ESTIMATED);
LGT_NOS=1/(1+EXP(B1+B2*PBCK_NOS));
* LOGIT FOR PARTICIPANTS WITH SUBSISDY
EVALUATED AT ZERO (ACTUAL);
LGT_NOSA=1/(1+EXP(G1+G2*PBCK_NOS));
PBCK_NOX=(1-SUBSIDY)*44/(0.09*HOURS*123/1000);
PBCK_NOX=ROUND(PBCK_NOX*100)/100;
LGT_NOX=1/(1+EXP(B1+B2*PBCK_NOX));
DROP _TYPE_ _NAME_ N_RUN N P_NP;

PROC SUMMARY DATA=PART_NG;
VAR LGT_NOS LGT_NOSA LGT_NOX;
BY XRUN;
* WEIGHT SAVINGS;
WEIGHT SQFT_AFF;
OUTPUT OUT=NET_G MEAN=;

DATA NET_G;
SET NET_G;
NG_REL=(1-LGT_NOS)/(1-LGT_NOSA);
* NG_APPZ=(1-LGT_NOS)/(1-LGT_NOSA);
NG_APPZ= 1-(LGT_NOS/LGT_NOX);
DROP _TYPE_ _FREQ_;

DATA PART;
SET PART;
N_OCCUP=ROUND(((1/(3*64))+UNIFORM(0)*2/(3*64))*SQF
T);
N_EQUIP=ROUND(UNIFORM(0)*N_OCCUP);
INSTALLD=ROUND(13+11*UNIFORM(0));
DMENDUSE=0;
IF UNIFORM(0) LE .4 THEN DMENDUSE=N_OCCUP;
* IF AWARE_YN='NO' OR ADOPT_YN='YES';
 DO YEAR=1 TO 3;
 DO MONTH=1 TO 12;
  M_COUNT=(YEAR-1)*12+MONTH;
    IF MONTH= 1 THEN
CDH=10230*(0.8+0.4*UNIFORM(0));
    IF MONTH= 2 THEN
CDH=13830*(0.8+0.4*UNIFORM(0));
    IF MONTH= 3 THEN
CDH=17430*(0.8+0.4*UNIFORM(0));
    IF MONTH= 4 THEN
CDH=20310*(0.8+0.4*UNIFORM(0));
    IF MONTH= 5 THEN
CDH=21750*(0.8+0.4*UNIFORM(0));
    IF MONTH= 6 THEN
CDH=22470*(0.8+0.4*UNIFORM(0));
    IF MONTH= 7 THEN
CDH=21750*(0.8+0.4*UNIFORM(0));
    IF MONTH= 8 THEN
CDH=20310*(0.8+0.4*UNIFORM(0));
    IF MONTH= 9 THEN
CDH=17430*(0.8+0.4*UNIFORM(0));
    IF MONTH=10 THEN
CDH=13830*(0.8+0.4*UNIFORM(0));
    IF MONTH=11 THEN
CDH=10230*(0.8+0.4*UNIFORM(0));
    IF MONTH=12 THEN CDH=
8790*(0.8+0.4*UNIFORM(0));

 TAU_IT=0;IF INSTALLD LE M_COUNT THEN TAU_IT=1;
 IF ADOPT_YN='NO' THEN TAU_IT=0  ;
*AC REGRESSORS;
 XAC_1=SQFT*CDH/1000;
 XAC_1_D=SQFT_DUM*CDH/1000;
 XAC_2=SQFT*(HOURS/12)/1000;
 XAC_4=N_OCCUP*8*30/1000;
 XAC_5=N_EQUIP*8*30/1000;
 XAC_6=SQFT_AFF*(HOURS/12)*TAU_IT/1000;
```

```
 BAC_1=0.197/9;
 BAC_2=3.412*(2/9);
 BAC_4=450/9;
 BAC_5=1800/9;
 BAC_6=-(3.412)*(123/184)*2/9;
EXP_Y=
 BAC_1*XAC_1+
 BAC_2*XAC_2+
 BAC_4*XAC_4+
 BAC_5*XAC_5+
 BAC_6*XAC_6;
*LIGHTING REGRESSORS;
 XLT_1=SQFT*(HOURS/12)/1000;
 XLT_1_D=SQFT_DUM*(HOURS/12)/1000;
 XLT_2=SQFT_AFF*(HOURS/12)*TAU_IT/1000;
 BLT_1=2;
 BLT_2=-(123/184)*2;
EXP_Y=EXP_Y+BLT_1*XLT_1+BLT_2*XLT_2;
BCOOK=2000;
EXP_Y=EXP_Y+BCOOK*COOK_YN;

OUTPUT;
END; END;

PROC SORT DATA=PART OUT=PART;
BY XRUN AWARE_YN CUST;

PROC SUMMARY DATA=PART;
VAR EXP_Y;
BY XRUN AWARE_YN CUST;
OUTPUT OUT=GET_MEAN MEAN=AVG_KWHM;

DATA GET_MEAN;
SET GET_MEAN;
SE_SCALE=(1-UNIFORM(0))*0.1+UNIFORM(0)*0.30;
GAMMA=1;
SE_SCALE=GAMMA*SE_SCALE;

DATA PART;
MERGE PART GET_MEAN;
BY XRUN AWARE_YN CUST;
KWH_M=EXP_Y+AVG_KWHM*SE_SCALE*(3**0.5)*(UNIFO
RM(0)-0.5);
DROP
N_RUN N P_NP SUBSIDY PAYBACK G1 G2 LOGIT
BAC_1 BAC_4-BAC_6 BLT_1 BLT_2 _TYPE_ _FREQ_;

PROC SORT DATA=PART OUT=PART;
BY XRUN CUST AWARE_YN ADOPT_YN;
*****;
  PROC REG DATA=PART OUTEST=PARAMS NOPRINT;
  BY XRUN CUST AWARE_YN ADOPT_YN;
  MODEL KWH_M=CDH XLT_2;
  OUTPUT OUT=PART R=EHAT_ITR;
  TITLE2 'PROC REG ON DATA=PART';
*****;
DATA PART;SET PART;
EHAT_2=EHAT_ITR**2;
DROP EHAT_ITR;
*****;
  PROC SUMMARY DATA=PART;
  VAR EHAT_2;
  BY XRUN CUST AWARE_YN ADOPT_YN;
  OUTPUT OUT=SIGMA_2 MEAN=SIG2_RC;

  DATA SIGMA_2;
  SET SIGMA_2;
  IF SIG2_RC GT .5 THEN WTFACTOR=1/SIG2_RC; ELSE
WTFACTOR=.;
  KEEP XRUN CUST AWARE_YN ADOPT_YN WTFACTOR;



*****;
  DATA PARAMS;
  SET PARAMS;
  KEEP XRUN AWARE_YN ADOPT_YN CUST XLT_2;
  RENAME XLT_2=BXLT_2;
*****;
  PROC SORT DATA=SAVSQFT OUT=SAVSQFT;
  BY XRUN CUST AWARE_YN ADOPT_YN;
*****;
  DATA PARAMS;
  MERGE PARAMS SAVSQFT;
  BY XRUN CUST AWARE_YN ADOPT_YN;
  IF ADOPT_YN='NO' THEN BXLT_2=0;
*****;
  PROC SORT DATA=PARAMS OUT=PARAMS;
  BY XRUN AWARE_YN ADOPT_YN;
*****;
  PROC SUMMARY DATA=PARAMS;
  VAR BXLT_2;
  BY XRUN AWARE_YN ADOPT_YN;
  OUTPUT OUT=RES3_20 MEAN=SAV_SQFT;
  WEIGHT SQFT_AFF;
*****;
  DATA RES3_20;
  SET RES3_20;
  IF ADOPT_YN='NO' THEN SAV_SQFT=0;
  KEEP XRUN AWARE_YN ADOPT_YN SAV_SQFT;
*****;
  DATA PART;
  MERGE PART SIGMA_2;
  BY XRUN CUST AWARE_YN ADOPT_YN;
  PROC SORT DATA=PART OUT=PART2;
  BY XRUN AWARE_YN;

PROC REG DATA=PART2 OUTEST=ONE NOPRINT;
```

```
MODEL KWH_M = XAC_1 XAC_4 XAC_5 XLT_1 XLT_2
/NOINT;
* MODEL KWH_M = XAC_1 XAC_4 XAC_5 XLT_1 XLT_2
COOK_YN     /NOINT;
* MODEL KWH_M = XAC_1 XAC_4 XAC_5 XLT_1 XLT_2
COOK_YND    /NOINT;
* MODEL KWH_M = XAC_1_D XAC_4 XAC_5 XLT_1_D
XLT_2 COOK_YN  /NOINT;
WEIGHT WTFACTOR;
BY XRUN AWARE_YN;
TITLE2 'PROC REG ON DATA=PART2';
DATA DELTA;
MERGE ONE COUNT;
BY XRUN AWARE_YN;
IF AWARE_YN='YES' THEN DELTA=XLT_2;
ELSE DELTA=(N_YES/(N_NO+N_YES))*XLT_2;

IF AWARE_YN='YES' THEN XLT_2_P =XLT_2;
IF AWARE_YN='NO'  THEN XLT_2_NP=XLT_2;

DATA DELTA;
MERGE DELTA WSQFT;
BY XRUN;

DATA A;
SET DELTA;
IF XLT_2_P NE .;
KEEP XRUN XLT_2_P;
DATA B;
SET DELTA;
KEEP XRUN XLT_2_NP WSQFT;
IF XLT_2_NP NE .;

DATA A; MERGE A B; BY XRUN;
NG_PROTO=1-(XLT_2_NP/XLT_2_P)*WSQFT ;

PROC TRANSPOSE DATA=DELTA PREFIX=DELTA
OUT=DELTA;
BY XRUN;
VAR DELTA;

DATA DELTA;
SET DELTA;
DLTA_AWN=DELTA1;
DLTA_AWY=DELTA2;
NTG_PROT=1-(DLTA_AWN/DLTA_AWY);
DROP DELTA1 DELTA2 _NAME_;

DATA NET_G;
MERGE NET_G DELTA A;
BY XRUN;
NG_REL_P=NTG_PROT/(1-LGT_NOSA);

DATA PART2;
SET PART2;
IF AWARE_YN='YES' AND ADOPT_YN='YES';
HOURS=HOURS/(12*1000);
LAST_REG=SQFT_AFF*HOURS*TAU_IT;

PROC GLM DATA=PART2 NOPRINT;
CLASS CUST;
MODEL KWH_M=
CUST
CUST*CDH
SQFT_AFF*HOURS*TAU_IT
/NOINT SOLUTION;
BY XRUN;
WEIGHT WTFACTOR;
OUTPUT OUT=PARTTWO P=KWH_MHAT;
TITLE2 'PROC GLM ON DATA=PART2';

PROC SORT DATA=PARTTWO;
BY XRUN CUST;

DATA AUX; *THIS MINI DATA SET SELECTS THE FIRST
CUSTOMER OF EACH XRUN;
SET PARTTWO;
BY XRUN ;
IF FIRST.XRUN;
KEEP XRUN CUST;

DATA PARTTWO;
MERGE PARTTWO AUX (IN=A);
BY XRUN CUST;
IF A;

PROC REG DATA=PARTTWO NOPRINT
OUTEST=RESULTS;
MODEL KWH_MHAT=CDH LAST_REG;
BY XRUN ;
* WEIGHT WTFACTOR;
TITLE2 'PROC REG ON PREDICTED VALUES FROM PROC
GLM';

DATA RESULTS;
MERGE RESULTS NET_G;
BY XRUN;
KEEP
NG_APPZ LAST_REG XLT_2_P NG_PROTO;
```

Addendum

Purpose of the Addendum

The purpose of the August addendum to the July 1993 report is threefold. First, some minor corrections and enhancements have been made to the text and equations of the July 1993 report.[11] Second, after interested parties had read the July 1993 version of this report, requests were made to make additional simulation runs. Third, the S-CDA model was applied to the nonparticipants of the simulation study, in order to demonstrate that the S-CDA model could be used to construct--using the framework of the protocols--an estimate of Net Load Impact, provided that nonparticipants adopt the DSM measure in significant quantities.

With respect to the second purpose of the Addendum (as just described above), the main issue that led to requests for additional simulation runs is that of the impact of changes in energy consumption other than the change that is associated with the DSM measure. This issue has to do with whether the C-CDA and S-CDA frameworks can be appropriately adapted when, for example, a customer who adopts a DSM lighting measure also adopts (perhaps outside the DSM programs of the utility) a cooking or space-cooling measure as well. While it is a fact (as the mathematics surrounding equation (23) implies) that the S-CDA framework--like the C-CDA framework--can readily be adapted given these sorts of changes, SDG&E has agreed to make the additional simulation runs so that this fact would be manifest. Among the overall simulation results that will be presented, two sets of results (one for cooking and one for space cooling) pertain to adapting the C-CDA and S-CDA frameworks to the sorts of changes that have just been described, assuming the collection of appropriate site-specific data.[12] Two additional sets of results are presented which compare the properties of the C-CDA and S-CDA estimators in the case where the changes that have just been described are not accounted for in the modeling and data-collection process.

Summary of Results

As already stated, this Addendum is designed to address the issue of changes in energy consumption other than those that are associated with the DSM lighting measure that is the focus of the main body of results. When other changes in energy consumption occur, they may be correctly accounted for in the modeling and data-collection process, or they may be wrongly omitted from consideration. The fundamental issue is the impact on the C-CDA and S-CDA estimators of the gross impact of the DSM lighting measure in each of these instances.

Based on the evidence that will follow, it can be concluded that the C-CDA and S-CDA estimators perform comparably in the case where the changes in energy consumption are correctly accounted for in the

[11] These changes will not be enumerated here. The only notable changes were corrections to the section of the July 1993 report entitled *Estimation Results: Net Impact*, wherein the conceptual errors were orginally made concerning the net load impact for the program.

[12] SDG&E wishes to emphasize that its position on S-CDA does not contain any element which limits the collection of useful site-specific data, especially those data which pertain to changes in energy consumption apart from DSM program changes. See the portion of this report entitled *Prescriptive Elements of the Protocols: The "Base Usage" Issue* (unchanged since the July 1993 version of this report) for a description of the Company's position on this issue.

modeling and data-collection process, with the C-CDA estimator showing a slightly greater degree of accuracy (as was the case in Figure 9, where the two estimators were compared in the absence of other changes in energy consumption). In addition, the C-CDA and S-CDA estimators are affected in duplicate fashion, when the changes in energy consumption are wrongly omitted from consideration. In summary, the C-CDA and S-CDA estimators perform comparably with respect to the issue of other changes in energy consumption.

Furthermore, the study results show that the S-CDA model can indeed be applied to nonparticipants (provided that they adopt the measure in significant numbers), so that an estimate of Net Load Impact can be derived in line with the protocols.

Changes in Consumption for Other End-Uses: C-CDA versus S-CDA

As already mentioned, the general S-CDA equation (23) is certainly capable of accommodating changes in energy consumption other than those that are associated with DSM programs. This will be demonstrated using two specific cases:

> Case 1: Among those customers who have energy consumption for cooking, 20% will have (at some point in time) a change in cooking consumption from 2,000 kWh per month to 1,200 kWh per month.
>
> Case 2: 15% of all customers will have (at some point in time) a change in space-cooling consumption. The change will be a 20% reduction in cooling-related consumption (a 20% gain in efficiency).

As will be shown, while the C-CDA model can easily be modified to handle Cases 1 and 2, the task is nearly as simple for the S-CDA model.

A Change in Cooking Consumption (Case 1)

Modifying the C-CDA Regression Equation

Recalling that the basic model is a 36-month model, and designating the change in cooking consumption as occurring (for the 20% of cooking customers who have changes) during one of those months, the revised C-CDA equation is,

$$\mathrm{kWh}_{ijt} = \beta_1\left[\left(\frac{h_{ij}}{12\text{ months}}\right)\left(\frac{\mathrm{SQFT}_{ij}}{1{,}000}\right)\right] + \beta_2\left[\mathrm{OCCUPANTS}_{ij}\right] + \beta_3\left[\mathrm{EQUIP}_{ij}\right]$$
$$+\beta_4\left[d_{ij}^{cook}\right] + \Delta\beta_4\left[\left(d_{ij}^{cook}\right)\left(x_{ijt}^{cook}\right)\right] + \beta_5\left[\left(\overline{\mathrm{CLTD}}_{it}^{roof}\right)\left(\frac{\mathrm{SQFT}_{ij}}{1{,}000}\right)\right] + \beta_6\left[\left(d_{ijt}\right)\left(\frac{h_{ij}}{12\text{ months}}\right)\left(\frac{\mathrm{SQFT}_{ij}^{A}}{1{,}000}\right)\right] + \varepsilon_{ijt}$$

where, $\beta_4 = 2,000$ kWh, $\Delta\beta_4 = 1,200 - 2,000 = -800$ kWh, and,

$$\left\{\begin{array}{ll} x_{ijt}^{cook} = 1 & \text{if cooking is an end-use } (d_{ij}^{cook} = 1) \text{ } \underline{\text{and}} \text{ there has been a } \underline{\text{change}} \text{ in cooking} \\ x_{ijt}^{cook} = 0 & \text{otherwise} \end{array}\right\}$$

The above definition implies that x_{ijt}^{cook} is zero: (a) throughout time, for those customers who do not have cooking, (b) throughout time, for those customers who have cooking but where no change has occurred, and (c) for the time period before the cooking-usage change, for those customers with cooking who have had a change.

__Modifying the S-CDA Regression Equation__

For Case 1, the appropriate S-CDA regression equation is,

$$\text{(33)}\quad kWh_{ijt} = \beta_{0,ij} + \beta_{5,ij}\left[\overline{CLTD_{it}^{roof}}\right] + \Delta\beta_{4,ij}\left[\left(x_{ijt}^{cook}\right)\right] + \beta_6\left[\left(d_{ijt}\right)\left(\frac{h_{ij}}{12 \text{ months}}\right)\left(\frac{SQFT_{ij}^{A}}{1,000}\right)\right] + \varepsilon_{ijt}$$

where,

$$\text{(34)}\quad \Delta\beta_{4,ij} = \Delta\beta_4\left[\left(d_{ij}^{cook}\right)\right]$$

This specification allows the cooking coefficient $\Delta\beta_{4,ij}$ to be estimated for each customer j (which allows for more flexibility, across customers, in the parameter $\Delta\beta_4$), so that the specification does not depend on the data element d_{ij}^{cook}. Of course, during estimation the constraint $\Delta\beta_{4,ij} = 0$ must be imposed (the associated regressor must be omitted) for those j where $x_{ijt}^{cook} = 0$ for all t (where there is no cooking, or when cooking does not change).

__Results for Case 1 (Cooking Changes) When There Are No Specification or Data Errors__

Figure 15 shows that the S-CDA estimator is comparable, in terms of accuracy, to the C-CDA estimator in that the two distributions are virtually congruent.[13] This implies that data on the <u>timing of the change</u> leads to

[13] Although the distributions are, by inspection, virtually congruent, the standard deviation of the S-CDA distribution is 0.101 (due to the fact that a small fraction of the results were fairly far from the mean), while the standard deviation of the distribution for the C-CDA estimator is 0.011. Although the standard deviation for the S-CDA estimator is much higher, it should be noted that 90% of the results for the S-CDA estimator were within ±2 C-CDA standard deviations of the C-CDA mean, and 95% of the results for the S-CDA estimator were within ±2.5 C-CDA standard deviations of the C-CDA mean. This implies that the two distributions are virtually congruent, except for 5% (or less) of the S-CDA results, and that, in this case, the standard deviation is not a good indicator of the true breadth of the S-CDA distribution. In fact, removing the two largest outliers from the S-CDA results (-4.735 and -3.00 watts) causes the standard deviation for the S-CDA distribution to drop from 0.101 to 0.024.

virtually the same level of accuracy, compared to the case where a formal mathematical model is built around the exact structure of the change. This certainly is a very useful result.

Along these lines, in viewing the results of Figure 15, it is important to note that equation (33) is actually a conservative version of the S-CDA regression equation, in that the potential accuracy of the S-CDA estimator is understated (most likely by only a small amount). Equation (33) does not impose $\Delta\beta_4$ across customers, in spite of the fact that the data-collection efforts that are associated with x_{ijt}^{cook} might well allow this. That is, equation (34) could be substituted into equation (33), in which case $\Delta\beta_4$ could be imposed as a regression coefficient that is constant across customers; the corresponding regression variable would be $\left(x_{ijt}^{cook}\right)\left(d_{ij}\right)$. Exploiting this structure would add some accuracy to the S-CDA framework.

A Change in Space Cooling Consumption (Case 2):

Modifying the C-CDA Regression Equation

Designating the change in space cooling consumption as occurring (for the 15% of customers who have changes) during one of the 36 months of the model, the revised C-CDA equation is,

$$\begin{aligned} kWh_{ijt} = {} & \beta_1\left[\left(\frac{h_{ij}}{12\text{ months}}\right)\left(\frac{SQFT_{ij}}{1{,}000}\right)\right]+\beta_2\left[OCCUPANTS_{ij}\right]+\beta_3\left[EQUIP_{ij}\right]+\beta_4\left[d_{ij}^{cook}\right] \\ & +\beta_5\left[\left(\overline{CLTD}_{it}^{roof}\right)\left(\frac{SQFT_{ij}}{1{,}000}\right)\right]+\Delta\beta 5\left[\left(x_{ijt}^{cool}\right)\left(\overline{CLTD}_{it}^{roof}\right)\left(\frac{SQFT_{ij}}{1{,}000}\right)\right]+\beta_6\left[\left(d_{ijt}\right)\left(\frac{h_{ij}}{12\text{ months}}\right)\left(\frac{SQFT_{ij}^{A}}{1{,}000}\right)\right]+\varepsilon_{ijt} \end{aligned}$$

where,

$$\beta_5=\frac{.197}{9},\quad \Delta\beta_5=(.8)\left(\frac{.197}{9}\right)-\left(\frac{.197}{9}\right)=(-.2)\left(\frac{.197}{9}\right).$$

and,

$$\left\{\begin{array}{ll} x_{ijt}^{cool}=1 & \text{if there has been a }\underline{\text{change}}\text{ in cooling} \\ x_{ijt}^{cool}=0 & \text{otherwise} \end{array}\right\}$$

The above definition implies that x_{ijt}^{cool} is zero throughout time for those customers who have not had a change in cooling, and during the time period before the cooling-usage change for those customers who have had a change.

Modifying the S-CDA Regression Equation

For Case 2, the appropriate S-CDA regression equation is,

$$kWh_{ijt} = \beta_{0,ij} + \beta_{5,ij}\left[\overline{CLTD}_{it}^{roof}\right] + \Delta\beta_{5,ij}\left[\left(x_{ijt}^{cool}\right)\left(\overline{CLTD}_{it}^{roof}\right)\right] + \beta_6\left[\left(d_{ijt}\right)\left(\frac{h_{ij}}{12 \text{ months}}\right)\left(\frac{SQFT_{ij}^{A}}{1,000}\right)\right] + \varepsilon_{ijt}$$

where,

(35) $$\Delta\beta_{5,ij} = \Delta\beta_5\left[\frac{SQFT_{ij}}{1,000}\right]$$

Of course, during estimation the constraint $\Delta\beta_{5,ij} = 0$ must be imposed (the associated regressor must be omitted) for those j where $x_{ijt}^{cool} = 0$ for all t. (This specification allows the cooling coefficient $\Delta\beta_{5,ij}$ to be estimated for each customer j, which allows for more flexibility, across customers, in the parameter $\Delta\beta_5$.)

Results for Case 2 (Space-Cooling Changes) When There Are No Specification or Data Errors

Figure 16 shows that, in terms of accuracy, the S-CDA estimator for the gross impact of the measure is again comparable to that of the C-CDA estimator in that the two distributions are virtually congruent (the standard deviation of the distribution for the S-CDA estimator is 0.013, while that of the distribution for the C-CDA estimator is 0.011). (Note once again that the full structure of equation (35) is not exploited in the specification of the S-CDA regression equation, since collecting data on x_{ijt}^{cool} may also allow for the collection of square footage data, in which case the coefficient $\Delta\beta_5$ could be constrained across customers. This would lead to some increase in the accuracy of the S-CDA estimator, most likely by a small amount.)

C-CDA and S-CDA When Changes in Other End-Uses are Not Accounted For

Figures 17 and 18 show the relative accuracy of the S-CDA and C-CDA estimators when the non-lighting changes in energy consumption (previously described as Cases 1 and 2) are not accounted for in the modeling and data-collection process. (Specifically, the premise is that, with respect to model estimation, the C-CDA and S-CDA regression equations contain no elements related to the changes, and they are estimated according to equations (18) and (24) as found in the body of this report.) Figure 17 contains results for Case 1, which involves a change in cooking consumption. Although each of the estimators appears to involve a small bias (the means of the distributions are -1.856 and -1.851 for S-CDA and C-CDA, respectively), the distributions seem to have widened comparably (the standard deviations of the distributions are 0.021 and 0.019 for S-CDA and C-CDA, respectively).

relative to those of Figure 9 . The conclusion would seem to be that the two alternative estimators suffer equally when a misspecification of this sort occurs.

Similarly, Figure 18 contains results for Case 2, which involves a change in space-cooling consumption. Each of the estimators again appears to involve a small bias (the means of the distributions are -1.875 and -1.866, for S-CDA and C-CDA, respectively), and the distributions seem to have again widened comparably (the standard deviations of the distributions are 0.029 and 0.033 for S-CDA and C-CDA, respectively). Once again, the evidence points to the conclusion that the two alternative estimators suffer equally when a misspecification of this sort occurs.

Applying S-CDA Within the Net Load Impact Calculation of the Protocols

The S-CDA model can certainly be employed in the Net Load Impact equation (32), in place of the C-CDA model, and results similar to those of Figure 14 can be constructed. Figure 19 gives results in the case where both the C-CDA and S-CDA estimators are applied to the Net Load Impact equation (32). The distributions of the estimated Net Load Impact are virtually identical (each with a mean of -0.52 and a standard deviation of 0.0.98), implying that in the S-CDA model can be used effectively in the case where the protocols version of the Net Load Impact calculation is utilized.

Figure 15

Results with Other Change in Operation: Cooking Usage Changes

The Distribution of Results (1,000 iterations)
Over the Estimated Change in Energy Usage

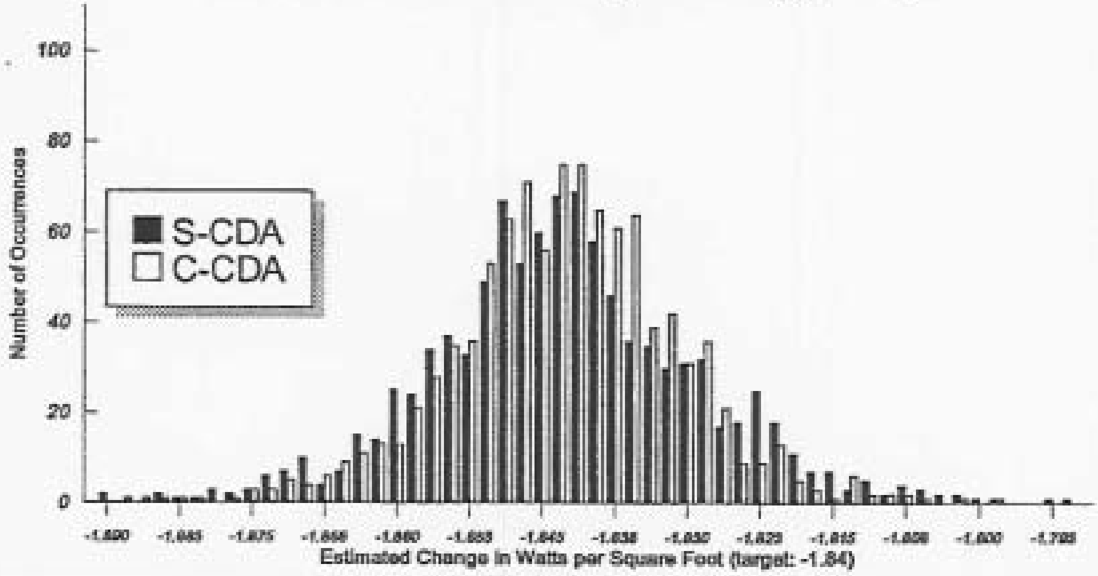


Figure 16

Results with Other Change in Operation: Cooling Usage Changes

The Distribution of Results (1,000 iterations)
Over the Estimated Change in Energy Usage

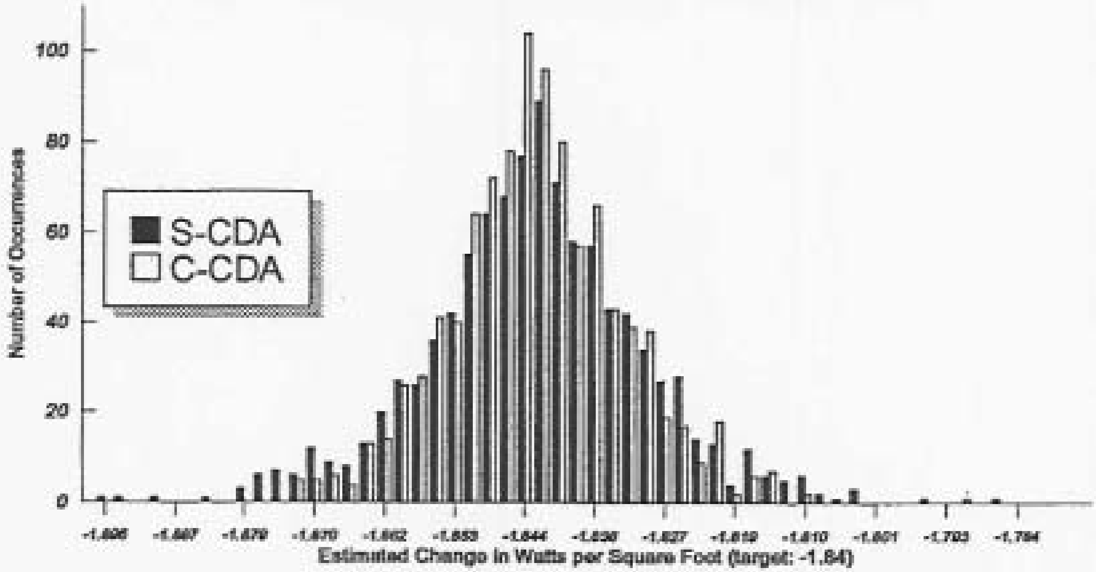


Figure 17

Results with Other Change in Operation: Change in Cooking Not Accounted For

The Distribution of Results (1,000 iterations)

Over the Estimated Change in Energy Usage

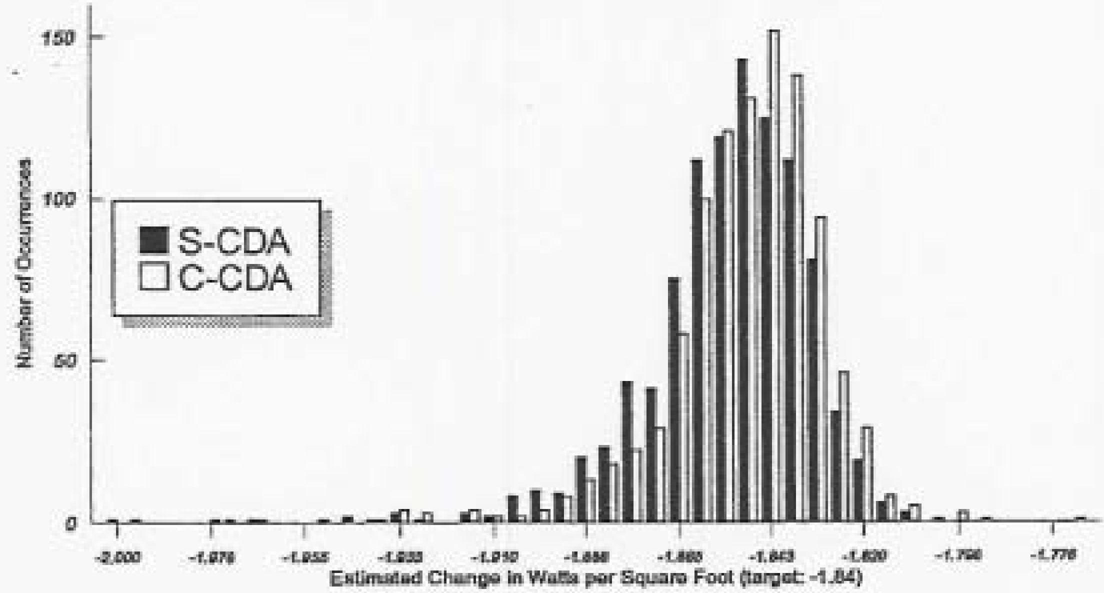


Figure 18

Results with Other Change in Operation: Change in Cooling Not Accounted For

The Distribution of Results (1,000 iterations)

Over the Estimated Change in Energy Usage

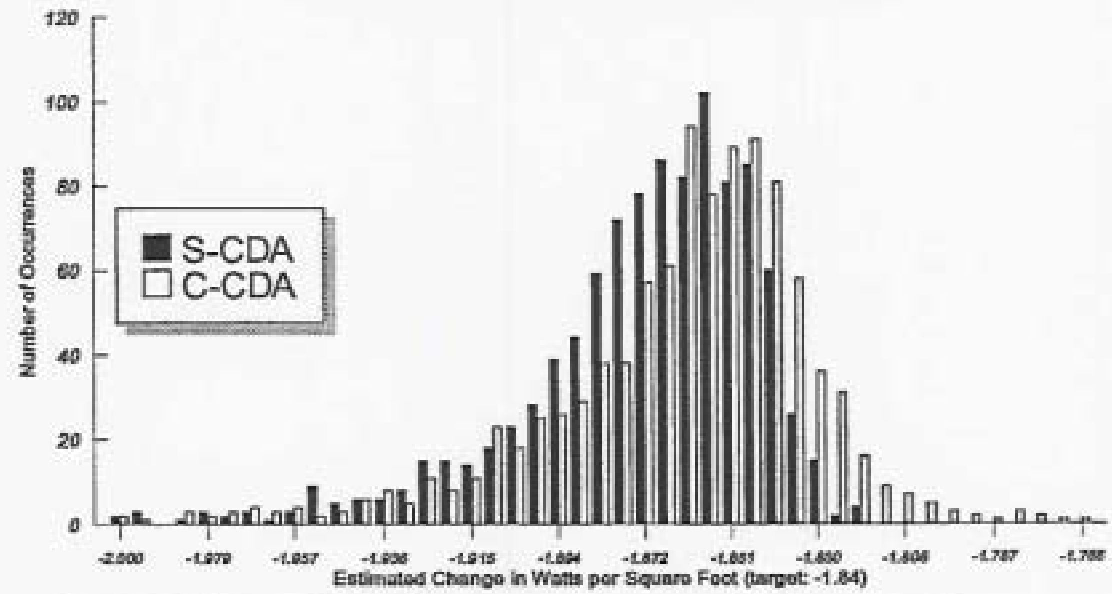


Figure 19

The Distribution of Results (1,000 iterations) Over the Estimated Net Change in Energy Usage as Estimated Under the Protocols