LBNL Report No. 40394

ACCURACY OF HOME ENERGY RATING SYSTEMS

Jeff Ross Stein

Building Energy Analysis Group, Bldg 90-2000
Environmental Energy Technology Division
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

June 1997

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technology, State and Community Programs of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Abstract

Actual residential energy bills were compared to estimated energy use and energy costs for four Home Energy Rating Systems (HERS) including the CHEERS Version I rating system. CHEERS tended to overestimate the actual energy cost by approximately 50%, with large variation in the accuracy of individual ratings. In addition to normal variations in occupant behavior, possible sources of error include inadequate rater training and incorrect assumptions about average occupant preferences for thermostat setpoints. However, for houses built between 1990 and 1994 CHEERS' average error was relatively small. The other three HERS all exhibited relatively small average errors in the estimated energy cost or energy use. Contrary to expectations, none of the HERS showed any clear relationship between the rating score and actual energy cost. Furthermore, all of the case studies tended to support the hypothesis that occupants of more energy efficient houses "takeback" some of their energy savings by using more energy services. This makes it difficult to use standard occupant behavior assumptions for all house types and still expect average error to be small for all types of houses. The case studies also demonstrated that it is more difficult to accurately predict energy use in mild climates than in more severe climates. While accuracy does not appear to be a major concern of the HERS industry or lenders who participate in HERS programs, it is still important, especially for consumers who make investments based on HERS ratings. Improvements in accuracy can be made with additional research and possibly through some fundamental modifications in the rating systems.

Acknowledgments

This report is written in the third person because there were many people who contributed to it, either by supplying data, or feedback, or in a variety of other ways. I am very grateful to these people. They include:

Ed Vine (LBNL)
Bruce Wilcox (Berkeley Solar Group)
Charles Segerstrom (PG&E)
Jeanne Byrne (Home Energy Magazine)
Ron Judkoff (NREL)
Lois Arena, Steve Andrews, and Jay Luboff (Energy Rated Homes of Colorado)
Brian Dreiling (Midwest Energy)
Ruth Kubinski and Sally Lukens (Home Energy Ratings of Ohio)
Cynthia Gardstein (HERS Council)
Doug Swartz (Ft. Collins Power & Light)
Brian Pon (LBNL)
Judy Roberson (LBNL)
Gene Rochlin (ERG)
Stephen Handorf (independent consultant)
Philip Fairey (Florida Solar Energy Center)
Phil Carlton (CHEERS)
Robert Clear (LBNL)
Vadim Belotserkovsky (Natural Resources Canada).
Richard Faesy (Energy Rated Homes of Vermont)
Michael Holt (Architectural Energy Corporation)
Barbara Farhar (NREL)
Rebecca Vories (Infinite Energy)

Robert Scott of CHEERS deserves special thanks for supplying the data that serve as the core of this research. As opposed to many HERS providers, Robert was eager to supply data and to work with me. Finally, I would particularly like to thank my primary advisor, Alan Meier of LBNL, for his extensive guidance, insight, and encouragement.

Table of Contents

Tables

Figures

I. Executive Summary

Approximately 20% of all the energy consumed in the United States is consumed by the residential sector. Much of this energy can now be cost-effectively saved by constructing new houses to be more energy efficient and by retrofitting existing houses with more efficient equipment. Unfortunately, most of the opportunities to save energy, natural resources, and money in houses are not captured because of market barriers such as lack of information and lack of financing. Home Energy Rating Systems (HERS) and related financial products, like Energy Improvement Mortgages (EIMs), have the potential to facilitate identification and financing of a tremendous number of such opportunities. A recent study by the Indiana State Energy Office estimated that penetration by HERS and EIMs into just 20% of the nation's annual housing market would result in:

- 1.6 million home energy ratings.
- $8 billion of cost-effective energy improvements installed and financed through the mortgage loan process.
- 100 billion Btu's in energy savings.
- 5 million fewer tons of CO_2 emissions (NASEO 1996).

A HERS is a computer simulation-based method for assessing a home's estimated energy use under standard conditions (similar to a miles-per-gallon rating for a car) and its potential for improvement. A rating usually requires a detailed home inspection by a trained rater. HERS typically generate three types of output: (1) a rating score (e.g. 0 to 100 points); (2) energy use and energy cost estimates for specific end-uses like heating and hot water and for the whole house; and (3) a list of recommended improvements that are calculated to be cost-effective.

Rating a house is difficult because every house is different and there are many potential sources of error such as rater mistakes, imprecise simulation algorithms, and incorrect assumptions about physical features like air infiltration rates. Furthermore, ratings are designed to rate the house and not the occupants so standard assumptions are made for all occupant-related inputs such as the number of people, number of appliances, and thermostat settings. Thus a rating that is accurate for the "typical" family could still be highly inaccurate for any particular family. (Accuracy can be roughly defined as the degree to which rating estimates correspond to actual home energy use, energy cost, and potential savings.)

While accuracy of rating systems is currently not considered by HERS experts to be the most important barrier to widespread use of HERS, all agree that accuracy is important for long-term credibility and success and that research is needed to assess and improve accuracy. To date, however, very little research has been done on the subject and almost no data has been made publicly available. Thus the goals of this research have been:

- to collect and analyze accuracy data,
- to examine the role of accuracy in the success of HERS, and
- to explore ways in which accuracy can be improved.

The principal methodology was to compare estimated energy use and energy cost from ratings with actual energy use and cost from utility bills. We sought data from HERS organizations on houses that have recently been rated and for which utility billing data had already been collected or could easily be collected. Although tens of thousands of houses have been rated in the last several years, few HERS organizations were willing and able to supply actual ratings.

The first data set we received and the one we most extensively examined, was from the California Home Energy Efficiency Rating System (CHEERS). CHEERS supplied us with approximately 200 ratings--about 1/3 from Eureka (a relatively cold California climate) and 2/3 from Fresno (a relatively hot California climate). The houses were rated in 1994 using CHEERS Rate Tool Version I, which has since been replaced by the entirely new Version II tool. CHEERS also helped us to obtain three years of monthly gas and electric billing data for the rated homes from Pacific Gas and Electric.

The average energy cost estimation error for the CHEERS sample was about 50%. In other words, CHEERS tended to overestimate the actual energy cost by about 50%. The standard deviation was 80%, meaning that about 1/3 of the houses were overestimated by more than 130% or underestimated by more than about 30% (see Table 1). While some of the estimation error is attributable to occupant behavior, the magnitude of some of the errors and the consistent tendency to overestimate energy use clearly implies that it is possible to improve both the average error and the variance by addressing other sources of error.

While utility billing data for rated homes cannot pinpoint specific sources of error in ratings, they can yield valuable clues for improving HERS. For example, the fact that CHEERS overestimated gas use more in Eureka than in Fresno, and the fact that CHEERS overestimated electricity use more in Fresno, tends to indicate that CHEERS may be using incorrect heating and cooling setpoints or infiltration rates, or conduction rates, etc. The more heating or cooling required, the greater the overstatement. Another important trend found in the CHEERS data is that some raters tended to produce more accurate ratings than other raters, which emphasizes the need for rater training, oversight, retraining, and the need to minimize rater judgment calls in the rating procedures. We also found that the average cost estimation error and standard deviation decreased as house age decreased. In fact, for houses built between 1990 and 1994, CHEERS underestimated energy cost by 8% on average.

Table 1. Summary of Case Study Results

	CHEERS (all homes)	CHEERS (new only)	Midwest-Kansas	HERO-Ohio	ERHC-Colorado *
sample size	185	30	16	14	276
avg. yr. built	1959	'90-'94	1995	N/A	1969
blower door test?	no	no	yes	yes	yes
avg. energy cost error	**51%**	**-8%**	**-7%**	**-14%**	**-3%**
std dev in errors	**80%**	**44%**	**15%**	**20%**	**35%**
avg. actual energy cost	$1,154	$1,327	$1,462	$1,697	
std dev in actual cost	46%	48%	24%	41%	51%
avg. HDD/yr. for '84-'95	2791	2791	4954	5371	6254
HDD in study yr.(s)	3% below avg.	3% below avg.	N/A	5% above avg.	3% above avg.
* Error and standard deviations for Colorado are for site energy use not energy cost because actual cost data were not available					

Other case study data that we received showed a smaller average estimation error and smaller standard deviation. For example, for a sample of 276 houses rated by Energy Rated Homes of Colorado certified raters, the average energy use estimate was only 3% lower than the average actual energy use (see Table 1). However, directly comparing the accuracy of the rating systems based on these case studies is almost like comparing apples and oranges because each sample of homes and each HERS is unique. Differences in the samples and rating systems include the following:

- **Severity of climate** - The two California climates have much milder winters than the other locations, and the Colorado location clearly has the most severe winter climate. As Petersen (1994) and others have found, and as our results appear to substantiate, it is harder to estimate energy use in mild climates than in more severe climates. Intuitively, this also makes sense because, for example, the percentage difference in heating energy between a thermostat setting of 70°F and 75°F is much larger when the outdoor temperature is 60°F than when the outdoor temperature is 30°F. (Table 1 shows the average heating degree days for each location.)

- **Average age of the housing sample** - The Kansas homes, for example, were almost all new ones, while the CHEERS houses were significantly older than the other groups. Old houses tend to be harder to estimate than new ones.
- **Occupant-related input** - Unlike the other HERS, Home Energy Ratings of Ohio (HERO) ratings allow the rater to input occupant-specific characteristics such as the number of actual occupants.
- **Diagnostic testing** - The CHEERS ratings were the only ones that did not include blower door testing. Thus they probably cost less to conduct but are also less precise.
- **Rating dates** - The CHEERS ratings were all performed in 1994 compared with 1996 for most of the other ratings. Significant progress in ratings systems was made in that period. For example, the CHEERS ratings were conducted using a rather user-unfriendly DOS-based program that has since been replaced with a more user-friendly Windows version that is likely to reduce rater errors.

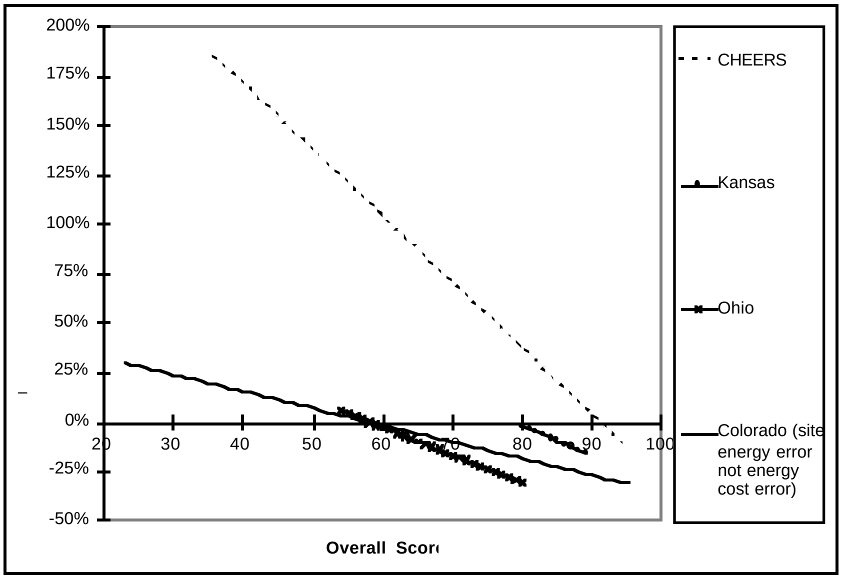
One of our most surprising discoveries was that none of the HERS we examined showed any clear relationship between rating score and total energy use or energy cost. Technically, rating scores only measure a house's individual potential for energy improvement and therefore should not be used to compare different houses. However, many consumers and HERS-related housing programs expect and assume that houses with higher rating scores will have lower energy costs. Yet even when compared to houses of similar size, ones with higher scores did not tend to use any less energy than homes with lower scores. One possible explanation is the "takeback effect" which says that the higher scoring houses are indeed more efficient and would use less energy if they were operated in the same manner as lower scoring houses but they are not operated in the same manner. Occupants of more efficient, higher scoring houses are likely to be more affluent, to have more appliances, and to choose more comfortable heating and cooling setpoints. Thus they "takeback" some of the expected savings in higher levels of service.

The takeback effect is supported by another phenomenon common to all the HERS studied: a clearly downward sloping trendline, or linear regression line, when estimation error is plotted against overall score (see Figure 1). For example, CHEERS tended to significantly overestimate energy cost for houses with very low scores. Overestimation decreases as the score increases to the point where CHEERS tends to underestimate energy cost for houses with very high scores. While the CHEERS trendline appears to cross the x-axis (0% estimation error) at a score of around 90, the Kansas trendline crosses at a score of about 80, and the Colorado and Ohio rating systems both have their lowest average error at a score of approximately 60. Thus it appear that it is possible to adjust the point where this trendline intersects the x-axis but not to flatten out the downward slope because HERS assume the same standard occupant behavior for all house types, when in fact, behavior varies according to the energy efficiency of the house.

Furthermore, in addition to being a function of socio-economic behavior dynamics, the particular slope of this trendline may be an inevitable function of climate severity. The percentage difference in energy use between the energy "misers" and the energy "hogs" in a mild climate is greater than the difference in a severe climate. Indeed, California has the mildest climate and the steepest trendline, while Colorado has the most severe climate and the flattest trendline.

The trendline can be translated vertically (i.e. the average error can be brought close to zero) by calibrating ratings with utility billing data for a statistically representative sample of houses. Calibration means adjusting one or more of a number of standard assumptions such as heating setpoint, cooling setpoint, home operating profile, internal gains, infiltration rates, and hot water usage.

Figure 1. Linear Regression of Energy Cost Error vs. Rating Score



Making accurate recommendations for cost-effective improvements is the most difficult objective of a HERS; it is also the most difficult objective to validate. One way for us to gauge the accuracy of recommendations was to compare the actual energy use of the CHEERS homes to the predicted total energy savings if the occupants implemented all the recommendations. Since many of the ratings predicted that it was possible to save over 50%, and in some cases over 100%, of current consumption, it is likely that at least some of the recommendations would not be cost effective.

The HERS marketplace (homeowners, banks, HERS providers, builders, etc.) does not seem to demand a high degree of accuracy. HERS experts do not rank accuracy high on the list of keys to HERS success, and do not report many questions or complaints about accuracy from consumers. In response to this perception, HERS providers choose not to discuss accuracy in marketing literature or on the rating forms themselves. Indeed, it appears that lending institutions that participate in Energy Improvement Mortgages and other HERS-related energy efficiency financing products are not exposed to any significant risk related to HERS accuracy because the agreement between expected and actual energy bills does not appear to affect the probability that a homeowner will default on a mortgage (Horowitz 1996).

Homeowners, however, are bearing at least some risk related to HERS accuracy. One risk is that a homeowner will incorrectly conclude that one home is more energy efficient than another based on inaccurate HERS scores. Another risk is of making an uneconomical investment based on an inaccurate recommendation. This risk is real and may be significant. It is ironic to note, however, that a homeowner can make an uneconomical investment and never know it. There is another risk that both the homeowner and the lender face which is probably more significant than the risk from accuracy: the risk that the homeowner (or the bank, in the case of foreclosure) will not be able to resell the house for a price that recovers both the original sale price plus the cost of the energy improvements (i.e. the risk that the market will not value energy efficiency features at their true economic value).

This is not to say that HERS accuracy is unimportant. Lack of data regarding accuracy may be impeding the growth and acceptance of HERS amongst certain consumers, lenders, and other groups nationwide. Furthermore, a lack of accuracy may eventually impact some HERS and cause irreparable credibility problems, which could spread to all HERS. For these reasons, HERS organizations and HERS providers continue to strive to improve accuracy.

One way to improve accuracy is to collect more utility billing data on some of the thousands of homes that have been rated with HERS in the last several years. A great deal of very detailed data now exist and must simply be gathered. As this research project has demonstrated, this sort of analysis can be fairly inexpensive to perform and can yield valuable information for calibrating rating systems and for identifying sources of error. Other forms of research such as submetering of energy end-uses, software-to-software comparisons, and pre/post-retrofit analyses are also needed.

Based on the wealth of knowledge that can be gained from comparing ratings and utility data, HERS providers may want to consider some fundamental modifications to HERS. For example, it may be possible to greatly improve HERS accuracy by incorporating a few key pieces of information about the current or prospective occupants into a rating. Occupant-specific input could be as simple as the number of occupants or could include other characteristics such as preferred temperature settings, appliances owned, annual weeks of vacation, hours at home, etc. Another modification that has the potential to greatly reduce the cost of performing ratings without compromising accuracy is to switch from the current simulation-based system to a prescriptive rating system. Detailed statistical analysis of ratings and utility bills could yield a regression equation that allows accurate prediction of energy use and energy costs based on a much smaller number of variables than are now collected for ratings. Such a system has the ability to account for the “takeback effect” and could generate scores and recommendations as well.

II. Introduction

A. Residential Market Barriers

The environmental impact of energy production and use is well documented. Residential energy use is a major component of total energy use. Much of the energy used in houses can now be cost-effectively saved by constructing them more efficiently or by retrofitting existing houses with more efficient equipment. Unfortunately, many of the opportunities to save energy and money in houses are not captured. There are a number of well documented market failures that explain this phenomena including lack of information and bias towards up-front capital costs rather than life-cycle costs. The net loss to society from these missed opportunities in terms of wasted money and wasted natural resources is huge (NASEO 1996).

Since the 1970’s energy efficiency proponents have developed a number of regulatory policies (e.g. Demand Side Management; building codes) and voluntary programs that have addressed these market barriers. However, regulatory mechanisms are increasingly endangered in the current atmosphere of deregulation and therefore there is a greater need for voluntary, market-based mechanisms. One such mechanism is Home Energy Rating Systems and the associated energy-efficiency financing products.

B. HERS Concept

A Home Energy Rating System (HERS) is a system for collecting and processing the data needed to produce a home energy rating, which is a standard measurement of a home’s energy efficiency and its potential for improvement.

1. Input Data

The first step in a HERS rating is for a certified rater to collect the necessary data on the house to be rated. Usually a site visit is required but it is possible to rate a house based on the engineering drawings. The required pieces of information vary depending on the particular HERS but generally include the following:

- utility providers and rate schedules
- house age, location (climate zone), and orientation (front azimuth)

- house dimensions (e.g. wall lengths and heights, overhang sizes, etc.)
- thermal envelope composition characteristics (e.g., wall, window, roof, and basement types, insulation levels, window shades, etc.)
- size, efficiency, age, and fuel type of heating system, permanent air conditioning system, and water heater
- inventory of lamp type and wattages in permanent lighting fixtures
- existence and fuel type of other appliances (clothes dryer, cooking range, pool, hot tub, etc.)

Often it is not possible for a rater to gather certain pieces of information so he/she must make assumptions either based on a formula or based on experience. For example, in older houses the furnace size and efficiency is often not clearly marked on the furnace. The rater may make an educated guess or consult a default table that lists average values based on house age.

a) Diagnostics

Rather than use standard assumptions for certain house parameters, it is sometimes possible to use testing equipment to empirically determine parameter values. For example, the accuracy of a rating can be improved by conducting air leakage tests of a house's envelope and it's duct system using blower door testing equipment. Diagnostic testing equipment can also be used to ascertain the efficiency of a furnace. Diagnostic testing can improve the accuracy of ratings but it increases the cost of ratings by increasing the amount of time and equipment required to conduct a rating.

2. Simulation Tool

Input data are fed into a computer program. There are hundreds of different software programs for simulating building energy use. Programs have been developed for many types of buildings (commercial, residential, etc.) and for many different uses (building design, HVAC equipment sizing, code compliance, etc.). HERS software is a small subset of this universe of building software. Most HERS simulation tools build a model of the house and simulate energy use for one year using standard assumptions for the required information that is not gathered by the rater. Assumptions are always made for the following:

- typical weather for that climate zone (based on climate records)
- standard occupant behavior (thermostat setpoints, hot water usage, personal appliance usage, etc.)

Even if actual weather data are available for a particular year or if occupant data are available, this information cannot be incorporated into the simulation because HERS are designed to rate the house irrespective of any particular occupants or the weather in a particular year. Most HERS also model and simulate a "reference" house that has the same dimensions, orientation, basement type, and fuel type as the rated house but is constructed to a particular building code such as CABO MEC (Council of American Building Officials Model Energy Code) or California's Title 24 (an energy code for new construction in California).

3. Output

The HERS software produces three types of output:

a) Rating Score

A score is a measure of a house's efficiency relative to its reference house. Scoring is usually on a zero-to-100-point scale or a zero-to-five-star scale (the higher the score, the more efficient the house). Scores are usually computed for a house's overall energy use and for several end-uses including heating, cooling, and water heating. Typically, if a house receives a score of about 80, then it is performing as well as if it were built to the building code of the reference house. However, different houses can be built to code and have very different energy use on a whole house or square foot basis due to differences in shape, orientation, basement type, shading, etc. Therefore, unlike a mile-per-gallon rating or an appliance Energy Guide Label, a HERS score is technically not designed for comparison between houses.

There is currently considerable controversy within the HERS community over how to weight different fuels in the scoring process. Some people argue that a BTU is a BTU and they should all be weighted equally, regardless of the source (this is referred to as “site energy” scoring). Others point out that a BTU of electricity usually costs more than three times as much as a BTU from natural gas. They prefer a “source energy” scoring system that includes the efficiency of the power plant and distribution losses in the utility wires, thereby more closely reflecting the true cost to the consumer.

b) Energy and Cost Estimate

A HERS gives an estimate of annual energy use and energy cost for a house as it currently exists, assuming typical occupant behavior. Estimates are also broken down into end-uses. End-use estimates are more important than overall estimates because they are more dependent on the permanent features of the house which HERS are designed to characterize and improve. Overall estimates include end-use estimates but are also largely determined by the “other” end-uses like cooking, refrigeration, and electronics, which HERS are not designed to measure. Thus standard assumptions are made for all “other” energy uses based on house size. Unlike the rating score, the energy use and energy cost estimates are designed to be used to compare different houses. The Score and Estimation portion of a HERS for a particular house might look as follows:

Table 2. Sample HERS Scores and Estimates

	Score	Estimated Annual	Annual Energy Use	
		Energy Cost ($)	Electricity (kWh)	Gas (kBtu)
OVERALL	76	1162	6489	63807
Heating	87	219	131	37519
Cooling	97	16	124	0
Hot Water	64	119	0	22096
All Other Energy Uses		809	6233	4191

c) Recommendations

A HERS will recommend energy efficiency improvements that are cost-effective on a life-cycle basis for a particular house. There are different methods for estimating cost effectiveness. According to a common method a measure is cost effective if its annual energy cost savings are greater than the annual cost of the item if the *incremental* capital cost is financed over the lifetime of the item at a certain interest rate. Incremental capital cost implies that for each recommendation to replace a piece of equipment, that equipment is already in need of replacement. The incremental cost, for example, is the difference between (a) the cost of replacing an old furnace with a new one of the same efficiency and (b) the cost of replacing the old furnace with a higher efficiency furnace.

The Recommendations section of a HERS usually shows the estimated incremental installation cost of each item and one or more economic measurements of cost-effectiveness such as internal rate of return or simple payback period. The Recommendations section of a HERS might look as follows:

Table 3. Sample HERS Recommendations

Interest Rate: 8.5%				
Recommendations List	**Useful Life (yrs)**	**Annual Savings ($)**	**Lifecycle Savings ($)**	**Estimated Cost ($)**
Upgrade attic insulation from R-7 to R-30	30	143	1533	450
Upgrade floor insulation from 0 to R-19	30	32	344	450
Upgrade wall insulation from 0 to R-13	30	534	5738	1150
Upgrade hot water equipment efficiency	15	68	567	425
Replace single glaze, metal windows with double pane, low-e, vinyl windows	30	177	1906	4100
Upgrade heating efficiency to AFUE 0.80	30	174	1870	1400
TOTAL		**1128**	**11957**	**7975**
Simple Pay Back Period for Package of Energy Improvements is 7 years				

C. HERS Providers

There are dozens of private and state and local governmental organizations that provide HERS. A common type of HERS provider is a private non-profit organization that may be a member of a national HERS organization. Private HERS providers often receive funding support from utilities and government agencies. Most HERS, however, are expected to eventually become financially self-supporting based on fees for processing ratings and other services. Raters are typically not employed by the HERS provider but rather they are usually independent energy consultants, mechanical contractors, etc. Raters charge customers for ratings (ratings are often subsidized or financed) and pay a fee to the HERS provider for processing the ratings and for training and certification. Most HERS providers buy or lease their HERS software from one of two private software firms. Many providers have developed their own software. It is common for HERS providers to switch to a different software provider or update to a newer software version as new and better products come onto the market.

D. HERS Applications

1. Information

The simplest use of a HERS rating is to provide information to homebuyers and others. In the same way that a car shopper may want to know a car's fuel-economy rating, a HERS rating provides a homebuyer with valuable information on a home's estimated energy use and potential for improvement. Many people believe that the availability of HERS information will increase the value of a house and encourage home builders and owners to improve energy efficiency.

2. Labeling/Certification

A number of labeling and certification programs have been developed that rely on a HERS to demonstrate that a particular home qualifies for that particular label. Often the label has name recognition in the marketplace and a house with the label could therefore be worth more than a similar house without the label. The label or certificate can also qualify the owner or builder for other incentives, including publicity that can help market a house. For example, a home in Colorado can qualify for the E-Star Certificate if it receives 4 or more stars on a HERS rating conducted by an Energy Rated Homes of Colorado certified home energy rater (RESNET Notes, 9/26/96). Curiously, most labeling programs rely on the overall score to qualify a house, rather than estimated energy costs, even though scores are not designed to be used to compare different houses.

3. New Construction Incentives

Many utilities and state and local government agencies offer various financial incentives to home builders who meet certain efficiency standards (Vine 1995). The State of Michigan, for example, recently awarded grants to five builders who will build five star energy rated homes (RESNET Notes, 2/27/97). Communities in the Chicago area that are experiencing explosive growth are considering using HERS ratings and a modified permit fee schedule to allow rebates of high rated efficiency homes (RESNET Notes, 8/16/96).

4. Energy Efficiency Financing

There are at least two types of Energy Mortgages (EMs).

a) Energy Efficient Mortgages

An Energy Efficient Mortgage (EEM) allows the buyer of a highly efficient house to qualify with a lower income. One type of EEM is a "two percent stretch" of the debt to income qualifying ratio. Typically a homeowner's monthly PITI payment (principal, interest, taxes, and insurance) cannot exceed 28% of his/her monthly income. However, if the house meets a certain threshold on the HERS scoring system, then the homeowner's PITI payment can be as much as 30% of his/her monthly income. The rationale is that the homeowner will be paying less for energy and will have more money left for mortgage payments.

b) Energy Improvement Mortgages

An Energy Improvement Mortgage (EIM) allows a homebuyer to incorporate the cost of cost-effective energy retrofits into a home loan. For example, if a HERS rating recommended $5,000 of cost-effective energy retrofits, then the lender can add this amount to a normal mortgage, making it an Energy Improvement Mortgage. Again, the rationale is that monthly energy savings will be larger than the increase in monthly mortgage payments.

Local mortgage banks in many parts of the country have been able to offer EEMs and EIMs because government or quasi-governmental agencies in the national secondary mortgage market purchase, guarantee, or insure EIMs and EEMs. The Veterans Administration (VA) and Federal Housing Administration (FHA) provide loan insurance or guarantees to primary market lenders who offer EMs that meet certain criteria. Fannie Mae and Freddie Mac offer to purchase EMs that meet their criteria. Now private banks that purchase "jumbo" mortgages on the secondary market are also beginning to offer EMs. (Jumbos are large mortgages that exceed the mortgage size limits of the governmental and quasi-governmental players.)

c) Home Improvement Loans

Utilities often form partnerships with local lenders to finance cost-effective home improvements, such as adding insulation or replacing heating equipment. These are unsecured loans, often at below-market rates, and can range from $1,000 to $25,000. Loan payments can even be included on a customer's utility bill.

E. Problem Statement

HERS and HERS-related financing products have only captured a tiny fraction of the potential market. In order to increase the penetration rate, HERS need to address a number of challenges and questions. One such question is: Do they work? More specifically, will a house really use less energy than another house if it has a higher HERS score? Will a house really cost as much to operate as the rating says? And will recommended improvements really be cost effective? Clearly a rating will not always yield the exact right answer to these questions, but how close will the rating come to the answer? Buyers, lenders, builders and others who make investments based on HERS output need to know what degree of risk they are being exposed to and what the chances are that the rating is steering them in the wrong direction. Validation data are needed for at least two reasons: (1) to demonstrate to consumers, funding agencies and others that HERS are technically sound, and (2) to gather data necessary to improve the accuracy of HERS.

Unfortunately, almost no information to address the question of HERS accuracy has been made publicly available. Recognizing the need (for data) to evaluate and improve rating tool accuracy, the HERS Council Guidelines call for all HERS providers to maintain a database of HERS input and output data, utility bill releases and information on retrofits for at least 10% of homes rated annually, or 500 homes annually, which ever is less (HERS Council Guidelines Version 2.0)[1].

In a report entitled "HERS Projected Energy Use versus Actual Utility Bill Analysis," the HERS Council concludes that:

- It is difficult to obtain meaningful data in order to connect predicted energy use with actual consumption determined from utility bills.
- Collecting validation data and developing appropriate data analysis procedures is important for improving HERS accuracy (HERS Council 12/1996).

The National Renewable Energy Laboratory, which has been involved in developing procedures for assessing HERS accuracy, also recognizes the need for "further development of empirical validation methods appropriate for testing HERS software" (NREL, 1995).

Thus the primary objectives of this research were:

- to collect HERS input and output data and utility bill data for homes that have been rated with HERS.
- to develop appropriate data analysis procedures for interpreting validation data in order to describe and improve HERS accuracy.
- to examine the role that accuracy plays in the future success of HERS and HERS-related energy efficiency financing programs.

III. Accuracy

A. Defining Accuracy

Accuracy can be roughly defined as how closely HERS output corresponds to actual data. Certainly no one expects HERS to be perfectly accurate because there is an almost infinite number of variables that affect energy use and HERS must make simplifying assumptions. Sources of inaccuracy or error can be divided into two types: natural uncertainty and system error. Natural uncertainty is due to the uncertainty in occupant behavior and weather, i.e. the variables for which HERS are not responsible. System error includes rater mistakes, and inaccuracy in assumptions about physical features such as infiltration, i.e. the variables for which HERS are responsible. While there are ways of estimating accuracy, it can be very difficult to determine if the source of error is natural uncertainty or system error. However, both types of error are relevant--system error affect all types of HERS output and natural uncertainty affects accuracy of energy estimates and recommendations.

---

[1] To increase the credibility of HERS with national lenders, Congress directed the Department of Energy to develop voluntary uniform guidelines for rating systems (Federal Register 1995). DOE, in turn, contracted with the national HERS Council, which includes representatives from the HERS providers, utilities, electric and gas industry groups, builders, primary and secondary lenders, appraisers, equipment manufacturers, realtors, and consumers. The HERS Council Guidelines, which have not been finalized or adopted by DOE, include a list of minimum rated features, on-site inspection procedures, standard occupant behavior assumptions and quality assurance prescriptions.

Accuracy has different meanings for each of the three types of HERS output and for how those pieces of information are used. Methods for estimating accuracy will also be different for each type of output.

1. Rating Score Accuracy

Strictly speaking, a scoring system is accurate if a house built exactly to code scores an 80, a house using no energy scores a 100, the most inefficient possible house scores a zero, and all other houses are placed appropriately along that continuum. Since a score is relative only to the potential performance of a particular house, a house's score can be unrelated to its energy cost relative to other houses and still be perfectly accurate. For example, if House A, a large, rectangular, single-story house, is built exactly to code it would score exactly an 80 and might use $800 per year in energy. House B is a small, square, two-story house that uses inefficient single pane windows. Since it is below code it scores a 75, but since it is smaller and its shape is inherently more efficient than that of House A it only uses $600 per year in energy. Thus, strictly speaking, it is not appropriate to judge score accuracy by comparing scores and actual energy use of different houses.

However, since a score still should have some approximate relationship to actual energy usage and since score is commonly used as a surrogate for energy costs it is useful to judge a scoring system by comparing score to actual energy cost. Thus, generally speaking, a scoring system is accurate if houses with high scores use less energy than houses with low scores (particularly when comparing houses of similar floor area). The clarity of this trend can be considered a measure of score accuracy. Since scores are primarily used by consumers for comparison, a HERS scoring system need only be *relatively* accurate not *absolutely* accurate, i.e. the actual numerical scores are not important as long as the houses are ranked in the correct order.

2. Energy Use (Cost) Estimation Accuracy

Accuracy of energy use or energy cost estimates is more straightforward. Error is defined as the percentage that the rated amount is above the actual amount. The basic formula is the same for all types of error: total cost error, total energy use error, gas use error, electricity use error, etc.

$$error = \frac{rated}{actual} - 1$$

A positive error means the rating is overestimating actual use and a negative error means that the rating is underestimating actual use. For example, if a house is rated or estimated to use $1,500 per year in energy and actual uses only $1,000, then the energy cost error is 50%, i.e. the rating overestimates energy use by 50%. Each end-use estimate (heating, cooling, etc.) made by a HERS will have an accuracy associated with it. End-use accuracy is more important that overall energy use or energy cost accuracy because the end-use estimates are used to determine the cost-effective recommendations. However, usually it is not possible to directly measure end-use accuracy because end-uses are rarely sub-metered.

The estimation accuracy of a HERS can be judged by looking at the bias and precision of a sample of ratings. Bias is measured by averaging the individual errors. Precision can be judged by the standard deviation of the errors.

Certainly energy and cost estimates should be relatively accurate, in order to use them to compare different houses, but in some applications there is also a need for estimates to be absolutely accurate. For example, the accuracy of recommendations depends on the absolute accuracy of end-use energy estimates.

3. Recommendation Accuracy

A particular recommendation is accurate if it turns out to be cost effective, i.e. it pays for itself in savings in a reasonable period of time. To truly judge whether specific recommendations turned out to be cost-effective would require detailed end-use monitoring before, and for a considerable period after, installation of recommended features. The accuracy of a package of recommendations could be roughly judged by comparing the actual energy use before and after retrofit with the rating's estimated change in energy. Recommendation accuracy can also be roughly judged by comparing accuracy of energy use or cost estimates of existing houses with and without recommended features. For example, if a rating system accurately estimates houses with inefficient furnaces and also accurately estimates houses with efficient furnaces then there is a good chance that it can accurately recommend new furnaces.

B. Sources of Error

There are a number of sources of error for each of the three types of HERS output. These sources are described in detail below. Table 4 shows which sources of error apply to each type of HERS output.

Table 4. Factors Affecting Accuracy of HERS Output

	HERS Output Type		
Potential Sources of Error	Rating Score	Energy Cost Estimation	Recommended Improvements
Rater Inspection	X	X	X
Default Values (permanent features)	X	X	X
Simulation Algorithms	X	X	X
Occupant Behavior		X	X
Weather		X	
Energy Prices		X	X
Default Values (recommended features)			X
Equipment Remaining Life			X
Discount Rate			X

1. Rater Inspection

Lack of adequate training or experience or lack of diligence in conducting inspections could cause a rater to miss information about a house or make other mistakes. If the rater is not well trained in using the software, mistakes could also be made entering data into the computer. Rater training and quality control are very important. Some raters are also equipment contractors who use ratings as a sales tool. They may be tempted to use assumptions that make retrofits look more attractive.

2. Default Values (permanent features)

Since it is impossible to determine all the properties of a house, it is necessary to make assumptions relating to the building envelope, the distribution system, and equipment efficiencies such as air infiltration rate, duct leakage rate, furnace efficiency, etc. One way to reduce errors introduced by such physical assumptions is to collect more actual data. For example, Brian Dreiling of Midwest Energy (a HERS provider in Kansas), has found that infiltration rates measured by blower door tests can be 50% greater or less than default infiltration rates (personal communication 2/97). Thus there is a tradeoff between rating accuracy and the cost of collecting more field data.

3. Simulation Algorithms

The software must be able to accurately simulate the heating and cooling loads and the performance of the house based on the input and assumptions. Thus, the physics of the software program affect accuracy.

4. Occupant Behavior

Occupant behavior is the primary source of variability in individual ratings. In addition, incorrect assumptions about typical behavior can result in large average errors. Assumptions must be made regarding number of occupants, thermostat settings, hot water usage, appliance usage, etc. Some HERS estimate the number of occupants based on floor area, others use the number of bedrooms. Research has shown that occupant behavior can have a very significant effect on actual energy use. Sonderegger (1977/78) examined the variation in winter natural gas consumption of 205 townhouses in Princeton, New Jersey. Physical features--such as the position of the townhouse (end or non-end unit), the number of bedrooms, and the amount of insulated glass--accounted for 54% of the variation in energy use. Differences in occupant behavior were associated with much of the remaining 46% variation. In a similar study, Pettersen (1994) concluded that if the inhabitants' behavior is unknown, it is impossible to predict the total energy consumption more accurately than ± 15-20%. In mild winter climates heating energy use cannot be estimated more closely than within an interval of ±35-40% of average consumption if the behavior of the inhabitants is unknown. However, in cold climates the heating consumption will be within an interval of ± 20-25%. Thus, selecting average behavior assumptions can reduce HERS average error, but inherent variation, or natural uncertainty, in occupant behavior will always affect the precision of individual ratings.

A related assumption concerns the use of wood burning fireplaces for space heating. According to California Energy Commission Survey of Occupancy Patterns and Energy Consumption in New California Houses (1984-88) 8% of the house occupants reported wood as the main heating fuel and 24% reported using wood for heating. However, if a fossil fuel heating system (gas, electric, propane, etc.) exists in a house, then HERS assume that this is the only means of heating. Thus wood use by particular occupants can affect accuracy of ratings for those cases.

In addition to affecting the end-uses that HERS are most concerned with (heating, cooling, and hot water), occupant behavior determines the all "other" energy uses that HERS must also estimate in order to estimate total energy use and energy cost. Assumptions are made explicitly or implicitly about the existence and energy use of nonpermanent features such as space heaters, portable air conditioners, refrigerators, TVs, nonhardwired lights, etc.. Incorrect assumptions about these energy uses or natural variation in them can affect overall energy use and cost estimates.

a) Takeback Effect

The takeback effect is a theory that says occupants of energy efficient houses chose a higher level of energy service than occupants of inefficient houses, thereby "taking back" some of the expected savings. For example, if a house is inefficient and is very expensive to heat, the occupants might settle for a lower temperature in the winter in order to save money. In contrast, occupants of a more efficient house might opt for warmer temperatures in the winter. In other words, the savings have been achieved but they do not show up on utility bills because they have been consumed in the form of greater service. Economists have referred to this phenomenon as a "rational economic model of behavior." While it may be rational, it makes it very difficult for a HERS to estimate average behavior when the average is dependent on house efficiency. On the other hand, after reviewing 15 studies on takeback effect, Nadel (1993) concluded that little if any takeback is likely for residential space heating, water heating, or lighting.

5. Weather

Since HERS assume typical weather, abnormally cold or hot weather in a particular year can cause a house to use more or less energy than estimated.

6. Energy Prices

A HERS must use correct current energy prices and must also forecast, explicitly or implicitly, the cost of energy over the payback period of the recommended equipment. Using correct current prices is important for estimating current energy costs, while price forecasts affect the accuracy of recommendations. Until now, finding current energy prices has been relatively straightforward since most houses are served by regulated monopoly utilities. As retail competition reaches the residential market, this job will become much more challenging. There will be hundreds of new utility providers with complex rate structures and consequently there is likely to be significant variability in prices and uncertainty in long term prices.

7. Physical Assumptions (recommended features)

Assumptions are made about the energy performance of the recommended features. Assumptions must also be made about the marginal cost of more efficient equipment.

8. Remaining Equipment Life

A significant assumption for many recommended features is that the existing equipment is near the end of its useful life and needs to be replaced anyway. The estimated energy savings of the recommended equipment is therefore only compared to the difference between the price of a new unit with the same efficiency as the old unit and the price of the recommended unit. However, if the existing equipment is likely to last several more years then it has residual value and estimated savings should be compared to the marginal cost of the new equipment plus some residual value or early retirement penalty on the old equipment.

9. Discount Rate

In order to make accurate recommendations, the opportunity cost of capital for financing improvements must be assumed (e.g., for an Energy Improvement Mortgage the discount rate would be the mortgage interest rate). If a different discount rate is available to the consumer then the list of cost-effective recommendations could be different.

C. Previous Work on HERS Accuracy

While there is very little published data on HERS accuracy, there is some anecdotal evidence that there are accuracy concerns with building energy simulation in general and with HERS in particular. Horowitz (1996) admits “it is no secret in the energy services community that engineering estimates of energy savings are imprecise.” According to Charles Segerstrom of PG&E, HERS tools overestimate energy use and he has heard suggestions from HERS experts to simply reduce all estimates by 1/3 to 2/3 (personal communication 2/96). Other HERS experts we spoke with expressed similar suspicions (personal communication with Doug Swartz 3/97). According to Ned Nisson, “even the most sophisticated software may over- or under-predict actual energy use by as much as 60%. . . Although few people like to admit it, even the best hourly simulation programs like DOE-2.1 are not necessarily ‘accurate.’ This rather well-kept secret has been known for some time” (Nisson 4/96)

1. HERS Council - VA HERO

In a recent report the HERS Council compared billing data to ratings for five houses rated with the Virginia Home Energy Rating Organization (VA HERO) software (HERS Council, 12/96). The report does not indicate how this sample of houses was selected for analysis or the house age, rating date, or rater(s). Table 5 summarizes the HERS Council’s findings. It is interesting to note that accuracy was generally better for total energy cost estimates than for gas or electricity cost estimates. This suggests that it is possible that total cost estimates could be accurate but recommendations based on specific end-use estimates could be inaccurate.

Table 5. VA HERO Accuracy Data (source: HERS Council 12/96)

	House A	House B	House C	House D	House E
location (state)		PA	PA	PA	PA
total score	59	69	72	69	55
fuel type	all electric	gas/electric	all electric	gas/electric	oil/electric
blower door test?	yes	yes	yes	yes	yes
electricity cost estimate ($)	1,543		1,310	582	793
actual electricity cost ($)	1,384		1,233	665	547
elec. cost error (%)	**10%**		**6%**	**-12%**	**45%**
gas cost estimate ($)		347*		946*	
actual gas cost ($)		723		521	
gas error (%)		**-52%**		**82%**	
total cost error (%)	**10%**	**-52%**	**6%**	**29%**	**45%**
* Gas estimate is for heating and DHW only. It is unclear if gas is used for other end-uses.					

2. BESTEST

In order to assess HERS accuracy the National Renewable Energy Laboratory (NREL) has developed the HERS Building Energy Simulation Test (HERS BESTEST), which is a procedure for comparing HERS output to the output of "several of the best public-domain, state-of-the-art building energy simulation programs available in the United States" (NREL, 1995). These programs currently are BLAST 3.0 Level 215, DOE2.1E-W54, and SERIRES/SUNCODE 5.7. Very simple houses in two locations (Colorado Springs, CO and Las Vegas, NV) are simulated with these "state-of-the-art" programs using standard HERS assumptions for occupancy and physical properties. The same houses are simulated with the HERS tool being tested. In order to pass the test, the HERS estimate for heating and cooling load (in MBtu/y) must be within a certain percentage of the estimate from the "state-of-the-art" programs for each of these hypothetical houses. Of course, BESTEST makes the underlying assumption that the "state-of-the-art" programs are accurate.

NREL has performed the BESTEST on at least five HERS programs. However, the tests were done on a nondisclosure basis. The results have not been made publicly available and the HERS programs have not been identified (personal communication with Ron Judkoff, 11/95, 11/96). We have learned from CHEERS that CHEERS was one of the HERS tested and that its performance was similar to that of the other four HERS tested.

Currently, the BESTEST is designed only to assess a HERS's ability to properly simulate heat transfer through the thermal envelope of conventional homes. NREL is planning to expand the BESTEST to be able to assess how accurately a HERS is simulating other physical features such as heating, ventilating and air conditioning (HVAC) equipment. However, even in future incarnations, BESTEST is basically only testing the "physics and math" of the HERS program and not other potential sources of error. Furthermore, BESTEST does not evaluate the ability to model very complex or unusual buildings. Thus even if a HERS passes the BESTEST there are other potential sources of error that could make it inaccurate. Potential sources of error not accounted for in BESTEST include: incorrect occupant behavior assumptions, nontypical occupant behavior, and rater inspection errors.

Ned Nisson tested two programs with the HERS BESTEST: Rem/Design and HOT2000 (Energy Design Update, 4/1996). Rem/Design, a HERS tool developed by Architectural Energy Corporation of Boulder Colorado, is used by several HERS providers in the US. HOT2000 was developed by the Canadian government and is used extensively in Canada and the US but not for HERS applications. Nisson found that HOT2000 and Rem/Design both agreed well with the best simulation programs. Except for a few near misses, both programs fell within the BESTEST allowable range for 19 heating cases in Colorado Springs and 10 cooling cases in Las Vegas.

Ratings from HOT2000 and a related program, AUDIT2000, have been compared to actual billing data for about 50 homes in Canada (Energy Design Update 4/96; NRCAN 4/95). Both showed agreement within 10% between estimated energy use and actual metered energy use. However, these programs differ fundamentally from US HERS software in that they incorporate occupant specific behavior, making the ratings more accurate but not "occupant blind."

3. FSEC - BEERS

Researchers at the Florida Solar Energy Center (FSEC) have recently used data from Florida Power and Light's $6 million New Home Research Project to analyze the accuracy of the Florida Building Energy-Efficiency Rating System (BEERS), a HERS developed by FSEC and used widely in Florida (Fairey, 1997). As part of the project 423 new Florida houses (built between 1991 and 1993) were rated using BEERS. These ratings are not necessarily representative of normal BEERS ratings because, as part of a research project, they were subject to a higher degree of quality control than normal. For example, the ratings were not conducted by typical BEERS raters but by specially trained FSEC contractors. The ratings were also the highest of the three classes of BEERS ratings, Class 3, which includes envelope (blower door) and duct leakage diagnostics.

In addition to utility bill data on whole house consumption, specific energy end-uses were monitored in many of the 423 homes including heating energy, hot water energy, cooling energy, pool pump and heater energy, and refrigerator energy. Since BEERS generates energy use estimates for each of these end-uses and others, it was possible to draw conclusions about the accuracy of each of these end-use estimates.

On a whole-house basis, BEERS tended to slightly overestimate actual total energy use. However, some of the end-uses were overestimated, on average, while others were underestimated, on average. For example, heating and hot water energy use were overestimated while pool energy was underestimated. The fact that the total energy use estimate was quite accurate while some of the end-use estimates were not very accurate has important implications for the accuracy of specific recommendations that are based on the end-use estimates.

Another interesting finding of the analysis was the fact that the variance in individual error was larger for heating than for cooling. Given that Florida is a cooling-dominated climate, with few very cold days, this result suggests that it is harder to estimate heating energy use in mild winters than to estimate cooling in climates with extreme summer weather.

Perhaps the most valuable finding of the FSEC analysis is the fact that it was possible to significantly improve the accuracy of the rating tool based on the data collected. Statistical analysis was used to derive clearer relationships between the input data and end-use consumption. For example, a better correlation coefficient was developed for the relationship between hot water energy use and the number of bedrooms.

IV. Methodology

A. Data Collection

Given the amount of time it takes to rate a single house (1-3 hours), the best way to compare ratings to utility data is to collect data on houses that have already been rated and then seek utility data. We sought data on rated homes for which utility data had already been collected or could easily be collected. Appendix A contains a partial list of the organizations we contacted seeking ratings and utility data. These included HERS providers, HERS software companies, national HERS organizations, utilities, and government agencies.

Although hundreds of thousands of houses have been rated by various HERS over the last several years, we found it very difficult to get ratings and utility data for several reasons, including:

- Some HERS software developers and HERS providers have collected validation data but are unwilling to make their data publicly available.
- HERS providers have validation data or at least some bill releases but do not have the staff resources to assemble data from a significant number of ratings.
- Utility bill releases have not been collected from the house occupants so billing data cannot be collected due for privacy reasons. (However, in some cases we were able to collect utility data without bill releases from HERS providers and utilities on a nondisclosure basis, where the name and address fields had been erased from the records.)
- Utilities are reluctant to reveal customer data.
- Many houses rated by HERS are newly constructed houses that do not have a billing history.
- Many rated houses have different utility providers for electricity and gas. Many houses also use non-utility fuels such as wood, oil, or propane making it very difficult to collect consumption data.

Where data were available we sought a representative sample of ratings, i.e. typical houses rated by typical raters. Ideally, it is best if the provider of the data selects the sample randomly and does not know the accuracy of the ratings beforehand. This was not always the case. In some cases the sample was provided by HERS providers who already knew the accuracy of the ratings and thus could have introduced a sampling bias. Given the large variability in houses and ratings, we sought as many houses as possible but had to settle for what the data providers were willing to supply.

At a minimum, we sought the estimated annual energy cost and at least one year's worth of electricity and gas consumption and cost data. Where possible, we also collected characteristics of the house (age, size, style, etc.) and additional rating data (rater, score, end-use estimates, recommendations, etc.). HERS providers typically do not keep track of recommended improvements. Therefore, our case studies focused on the ratings and actual energy use and not on actual energy savings.

B. Data Analysis

Before comparing the ratings to the utility bills, it was often necessary to convert the billing data to a usable 12 month format. For example, some houses had only 9 or 10 months of data. In most of these cases we simply extrapolated the average over 12 months. In a few cases, extremely low consumption data in one or more months indicated that the house was empty or the data were incorrect. In some of these cases, we extrapolated from months that had reasonable consumption. In many cases, we decided there were too few months of valid data, and these houses were removed from the analysis. While none of the utility data were weather normalized, we did attempt to estimate the potential error introduced into the results due to differences between typical and actual weather by estimating the relationship between heating degree days and monthly and annual energy consumption.

The first level of analysis was to compare the estimated energy cost with the actual energy cost using graphing and statistical techniques (such as mean, standard deviation, and paired t-test) to determine if the means of the rated costs and the actual costs were statistically different. Annual cost and fuel use (electricity, gas) estimates were analyzed in this way. In one case, some end-use monitoring was conducted allowing for analysis of end-use estimates. We were not able to collect before and after retrofit utility data to compare estimated savings with actual savings.

The second level of analysis was to look for correlations between estimation errors and house characteristics. A number of graphical and statistical techniques were used. For example, errors were compared for various subsets of the samples based on house location, size, age, equipment type, etc. Multiple linear regression analysis was used to determine which known house variables were most significant in predicting estimation error.

C. Assessing the Significance of Accuracy

In addition to literature review, the primary method for investigating the significance of accuracy was through informal discussions with stakeholders. Discussions were held with most of the potential data sources listed in Appendix A as well as with national laboratory researchers and with lenders offering EEMs.

V. CHEERS Case Study

The California Home Energy Efficiency Rating System (CHEERS) was founded in October 1990 as a private nonprofit HERS provider. CHEERS is supported by a coalition of six major California utilities, financial industry associations, building and housing industry associations, and consumer and environmental advocacy groups. CHEERS has a network of about 100 independent certified raters in California, who have conducted several thousand ratings for a variety of purposes, including energy efficiency mortgages.

CHEERS has recently adopted a new windows-based software tool (Version 2.0) that uses an entirely different simulation engine. The ratings analyzed in this study were all conducted using the original CHEERS software (Version 1.0) which employs the CALRES calculation engine and has a DOS-based user interface. Therefore, all information presented and conclusions drawn here refer only to the original CHEERS system (Version I) and not the current CHEERS system (Version II).

CHEERS gives a score (0 to 100) for the following end-uses: heating, cooling, hot water, lighting, and overall efficiency. Scoring is based on source energy, e.g. 10,239 Btu per kWh of electricity. A CHEERS rating also gives estimates for energy use (kWh of electricity and/or kBtu of fuel) and annual cost ($) for the following end-uses: heating, cooling, hot water, lighting, and other (cooking, pool heater, etc.), and total.

CHEERS ratings are conducted by certified raters that must pass a week-long training course. CHEERS ratings typically do not include blower door testing or other forms of diagnostic testing, nor is testing required by the current HERS Council Guidelines. (Blower door testing is encouraged by both CHEERS and the HERS Council Guidelines.)

It is worth noting that in some ways, CHEERS is handicapped by its climate. All CHEERS ratings to date have been in California, which has a relatively mild climate compared to much of the country. Since heating and cooling loads are smaller than other places, there will be fewer cost-effective opportunities to save heating and cooling energy. Moreover, heating and cooling are a smaller percentage of total energy costs and non-permanent appliances are therefore a bigger percentage. Also, energy use is more sensitive to small changes in thermostat settings and other occupant behavior. Thus it is more difficult to accurately estimate total energy costs.

A. Data

After initial discussions, we requested from CHEERS a sample of approximately 200 rated houses--half from the Fresno area and half from the Eureka area. These areas were chosen because they are diverse climates and within the Pacific Gas and Electric (PG&E) service territory[2]. CHEERS queried their database and came up with about 150 houses in Fresno and 70 in Eureka. The house files were sent by CHEERS to PG&E who provided monthly gas and electricity consumption and cost data for most of these houses from

---

[2] Fresno, which is in the Central Valley approximately 150 miles southeast of the San Francisco Bay Area, has hot summers and mild winters. Eureka, on the Pacific Coast approximately 200 miles north of the San Francisco Bay Area, has mild summers and colder winters than Fresno. According to the California Energy Commission, Fresno averages 2228 heating degree days per year and 1997 cooling degree days per year, while Eureka averages 4085 heating degree days.

January 1993 through December 1995. Thus, the CHEERS sample was chosen randomly, and CHEERS did not know the accuracy before selecting the sample or passing the data to us.

CHEERS provided us with most, but not all, of the input and output data for the houses. Input data included:

- floor area, house volume, and number of storeys
- year built
- heating and cooling equipment fuel and efficiency; distribution system; duct location (conditioned or unconditioned space); duct insulation
- presence of setback thermostat
- presence of utility fuel (natural gas) fireplace (CHEERS does not distinguish between non-utility fuel fireplace and no fireplace)
- hot water equipment type, efficiency, size, tank wrap, and distribution type
- pool or spa pump and heater fuel and size; pool and spa covers
- presence of window labels
- type and wattage of all light bulbs in permanent light fixtures
- clothes dryer and cooking range fuel type
- rater id number and date of rating

The input data provided did <u>not</u> include:

- dimensions of walls, windows, etc.
- window types
- insulation values for walls, ceilings, etc.

The output data included:

- scores for heating, cooling, hot water, lighting, and overall for each house "as is"
- annual energy use and cost estimates for heating, cooling, hot water, lighting, other, and overall for each house "as is"
- scores for heating, cooling, hot water, lighting, and overall for each house "as recommended"
- annual energy use and cost estimates for heating, cooling, hot water, lighting, other, and overall for each house "as recommended"

The output data did <u>not</u> include:

- list of recommended features

The gas and electric billing data from PG&E were found to have a number of problems, therefore a data filtering process was devised to remove, or correct for, bad data (see Appendix B). After processing the utility bill data for the 193 houses, we ended up with gas and electric data for 186 houses, electric only data for 5 houses, and gas only data for 1 house. Of the remaining records, less than 1% of the data points were considered incorrect and altered.

B. Analysis

1. Estimation Accuracy

While the data available cannot be used to directly validate each of the end-use estimates, it can be used to validate the estimates for gas, electricity and total energy and cost.

a) Energy Cost

Figure 2 shows that for the entire sample there was a clear bias towards overestimating the total actual energy cost. The average actual energy cost was $1,154/year, while the average rated energy cost was

$1,585/year. Thus the error of the average was 37%. (Error is defined as the percentage that the rated amount is above the actual amount).

Figure 2. CHEERS: Rated vs Actual Total Energy Cost

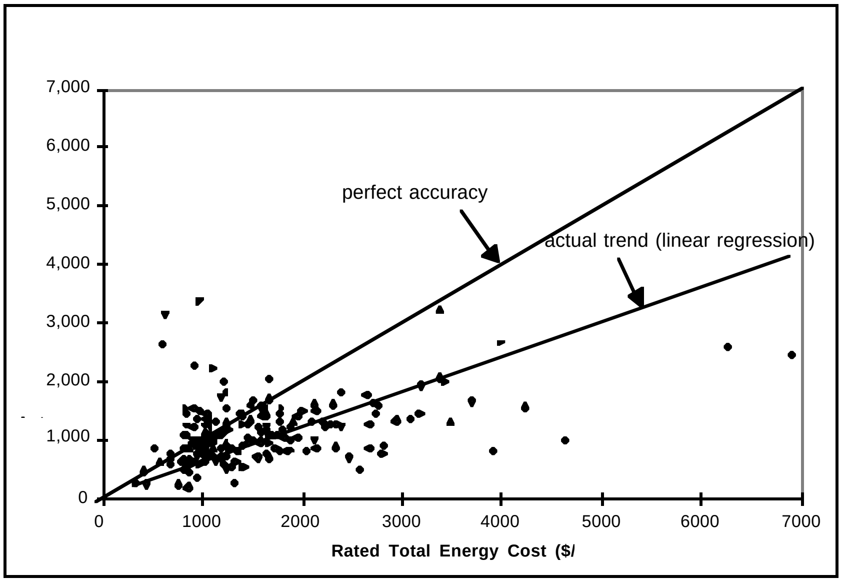


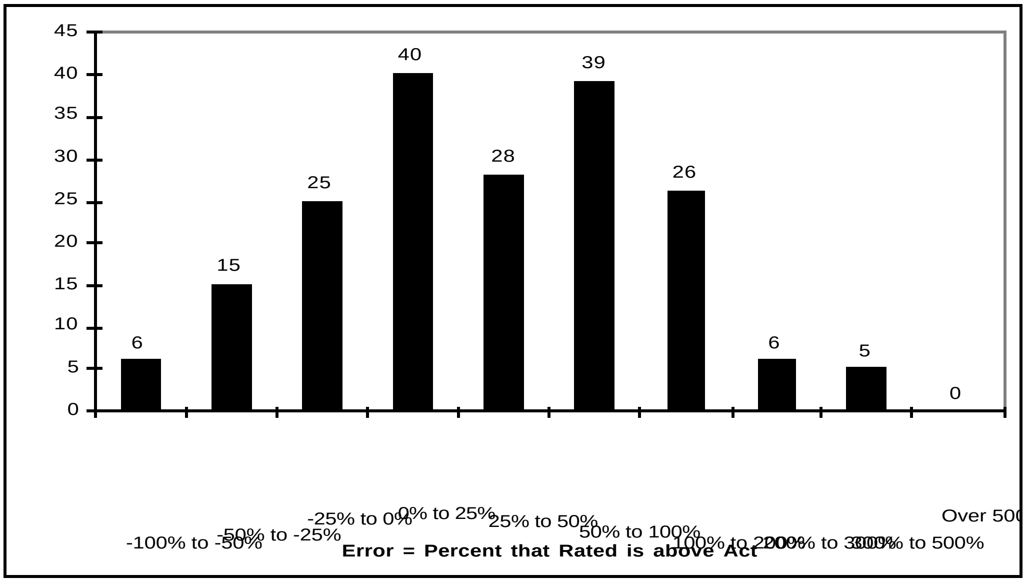
Figure 3 shows the wide distribution of the individual errors. For example, for five houses CHEERS overestimated the total energy cost by 300% to 500%. The average error was 51%. This may seem inconsistent with the 37% error of the average but they actually refer to different things mathematically (error of the average is the ratio of the average amounts and the average error is the average of the ratios). In this case, the average error, 51%, is probably the most useful term because this is the expected error for any individual rating. Thus all references to average error refer to the average of the individual error not the error of the averages.

$$Avg\ Error = \frac{\sum_{n=1}^{N}\left(\frac{rated\ \ amount}{actual\ \ amount}-1\right)}{N} \neq \frac{\sum rated\ \ amounts}{\sum actual\ \ amounts}$$

$$where\ N = number\ of\ ratings$$

Statistical analysis of the total cost estimate was performed using a paired T-test. This analysis confirmed that the means of the rated costs and the actual costs were statistically different.

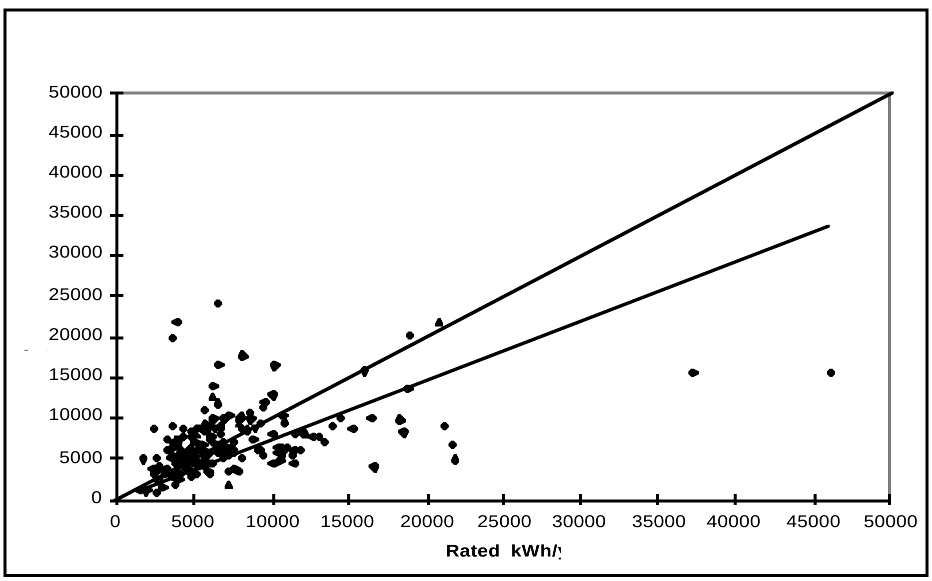
Figure 3. CHEERS: Histogram of Energy Cost Error



b) Electricity Use Estimate

Figure 4 also shows that on an individual level the rated electricity use is often substantially different from the actual amount. However, there is less of an overall bias towards overstating the electricity use than total energy cost. The average error is 18%. A paired t-Test showed that one cannot conclude with 95% confidence that the differences are significant.

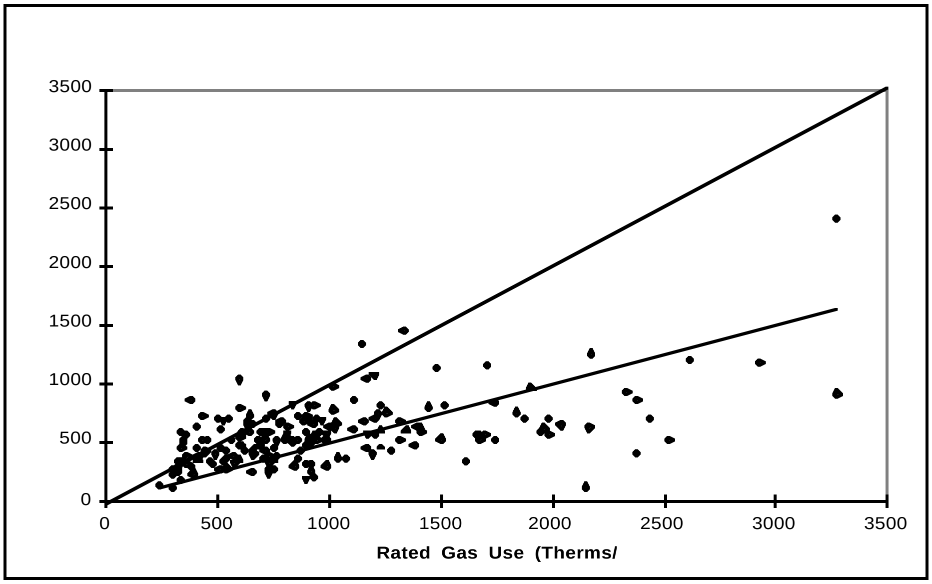
Figure 4. CHEERS: Rated vs. Actual Electricity Use



c) Gas Use Estimate

Figure 5 shows a clear tendency to overestimate the actual gas use. The average of the errors is 86% and the difference between rated gas use and actual gas use were significantly different.

Figure 5. CHEERS: Rated vs Actual Gas Use



d) Utility Rate Schedule Accuracy

Since both the CHEERS and PG&E data include energy use and energy cost data, we have the opportunity to verify that the energy prices used by CHEERS are reasonably correct. Figure 6 and Figure 7 compare the rated price of electricity and gas to the actual prices paid according to the utility bills (keep in mind the x and y scales when examining these figures). Although it appears that CHEERS is not always using the correct price, the average prices are quite close. The average rated electricity price is $0.120/kWh compared to $0.117 for the average actual price (3% average error). The average rated gas price is $0.64/therm, while the average actual gas price is $0.57/therm (12% average error). The differences seen here between actual and rated average unit price could reflect the fact that utility tariffs do not have a fixed unit price. Typically a tariff will include a fixed cost plus a certain price per kWh or therm for the first X number of kWh or therms and then a different unit price for any amount above X (i.e. unit price is a function of the number of units consumed). Therefore, CHEERS is probably using the correct tariff structure but since the estimated usage is different from the actual usage, the unit price will be off.

Figure 6. CHEERS: Rated vs. Actual Electricity Price

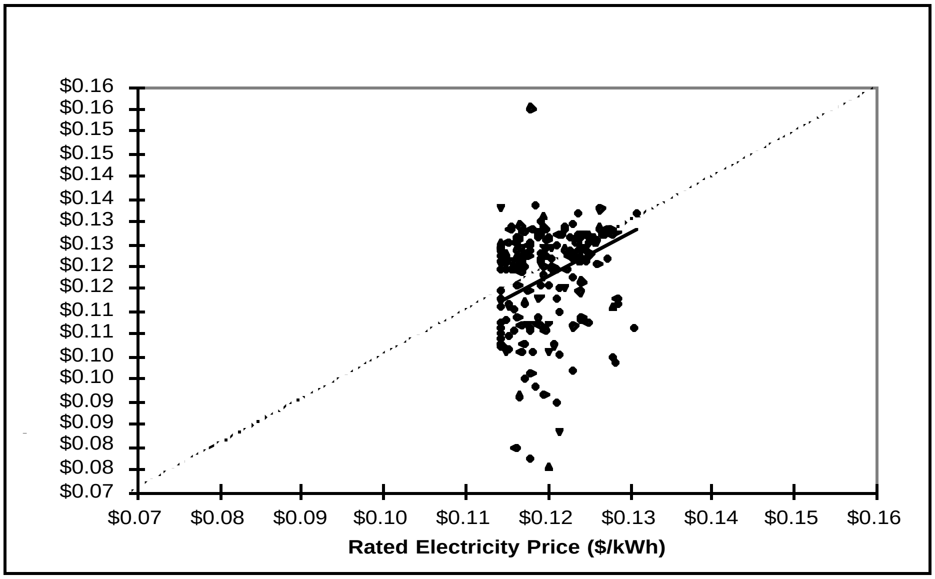
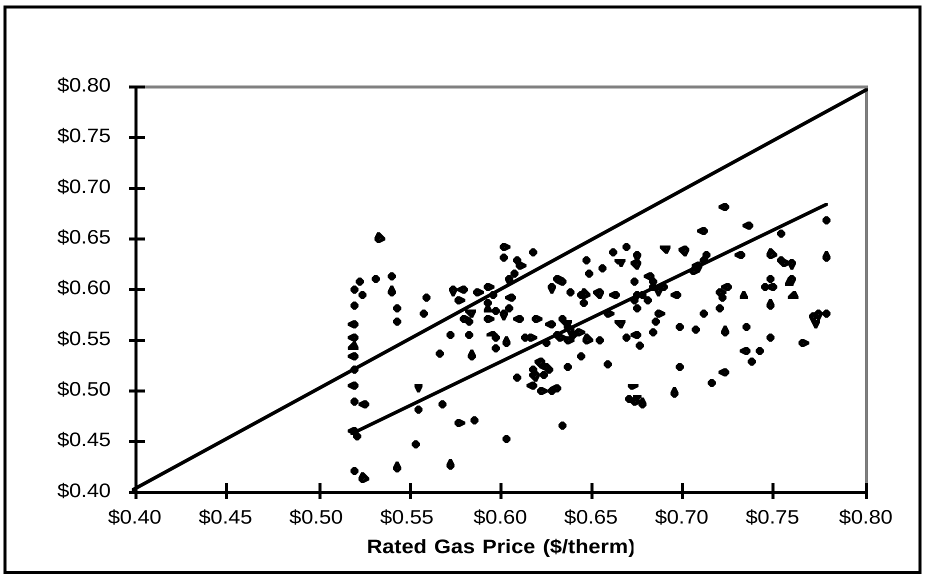


Figure 7. CHEERS: Rated vs. Actual Natural Gas Price



e) Uncertainty

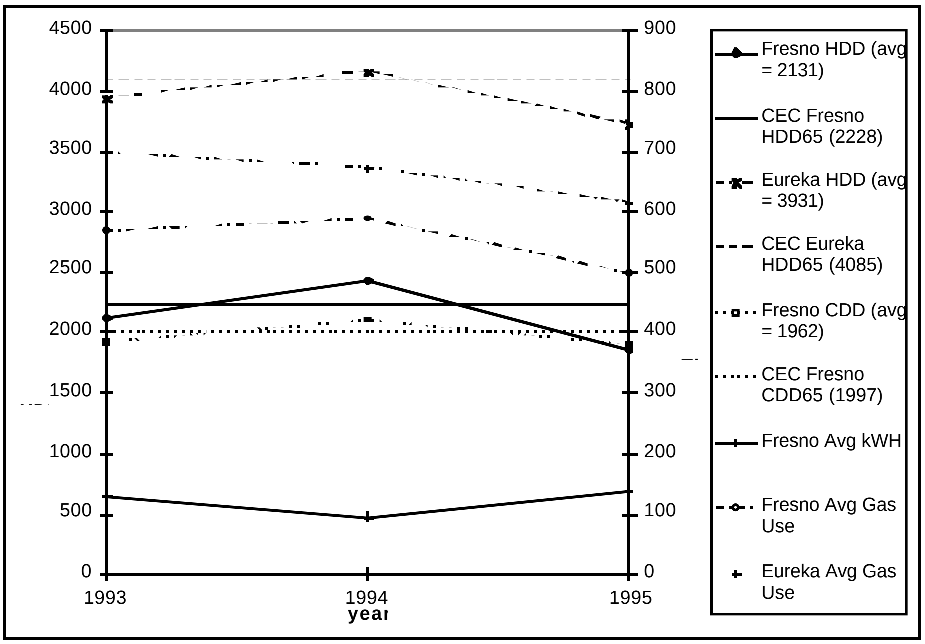
There are a number of factors that must be accounted for in assessing the accuracy of the data. By addressing these factors, we are able to determine if the results are reasonably accurate.

(1) Weather Normalization

Like all HERS, CHEERS uses default weather conditions based on long-term averages. If a particular winter were much colder than the default values, then CHEERS might underestimate actual heating energy and the next year the opposite might be true. If the heating degree days (HDD) or cooling degree days (CDD) for a particular year are significantly different from the default values, it is appropriate to weather normalize the energy use data or the rating results by scaling one or the other by an appropriate factor.

To determine the need for weather normalization, we first divided the houses into the two climates: Eureka (CEC Climate Zone 1) and Fresno (CEC Climate Zone 13). Figure 8 compares the CHEERS default values for HDD and CDD (from the CEC) with actual HDDs and CDDs as measured by the National Climate Data Center at the Eureka and Fresno weather stations. For both locations, the actual HDDs were slightly below normal in 1993, slightly above normal in 1994 and below in 1995. Similarly, 1993 was a slightly cooler summer in Fresno than usual, 1994 slightly hotter, and 1995 slightly cooler than usual. (Eureka CDDs are not plotted because none of the homes in Eureka have built-in air conditioning.) The weather during the monitoring period (1993-1994) is not significantly different from the long term average weather assumed by CHEERS. Therefore differences between actual weather and assumed weather is not a significant source of error in these ratings.

Figure 8. CHEERS: Annual HDD and CDD (actual and CHEERS default) and Actual Energy Use



It is interesting to note that there is no clear correlation between energy use and climate data. For example, average gas use in Eureka decreased from 1993 to 1994 even though 1994 winter months were colder but then gas use decrease from 1994 to 1995, as expected. The trend in average annual electricity use in Fresno is exactly opposite of what one would expect from the climate data. More energy is used in the milder summers of 1993 and 1995 than in the hotter summer of 1994.

On the other hand, Figure 9 and Figure 10 show a clear relationship between monthly HDDs and gas use in Eureka. Based on the trendline in Figure 10 it is possible to determine that a correction factor of about 4% could be used to weather normalize CHEERS gas estimates or actual gas use data for Eureka. For example, we could increase the actual gas use data by 4% or decrease the CHEERS gas estimates by 4%. However, this is clearly well within the uncertainty range for this analysis. Thus weather normalization of Eureka gas data is not used.

Figure 9. CHEERS: Heating Degree Days and Average Gas Use for Eureka

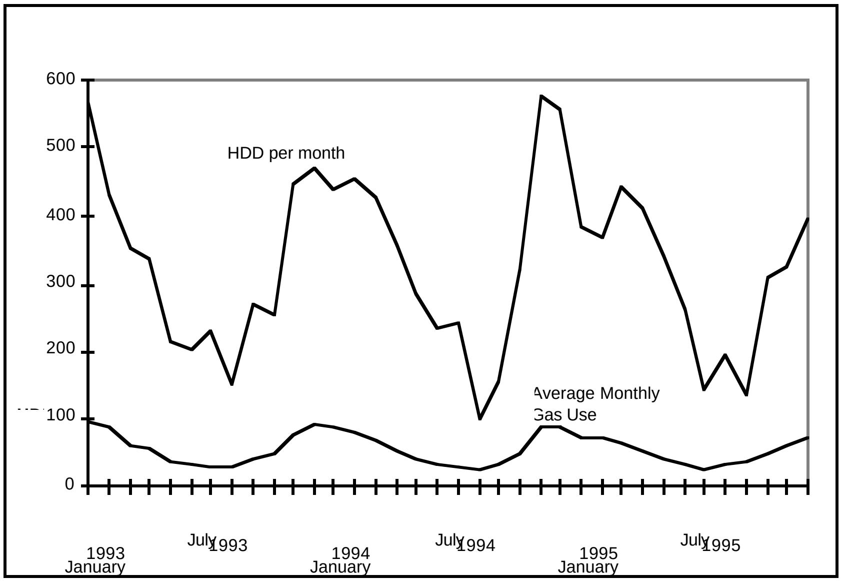
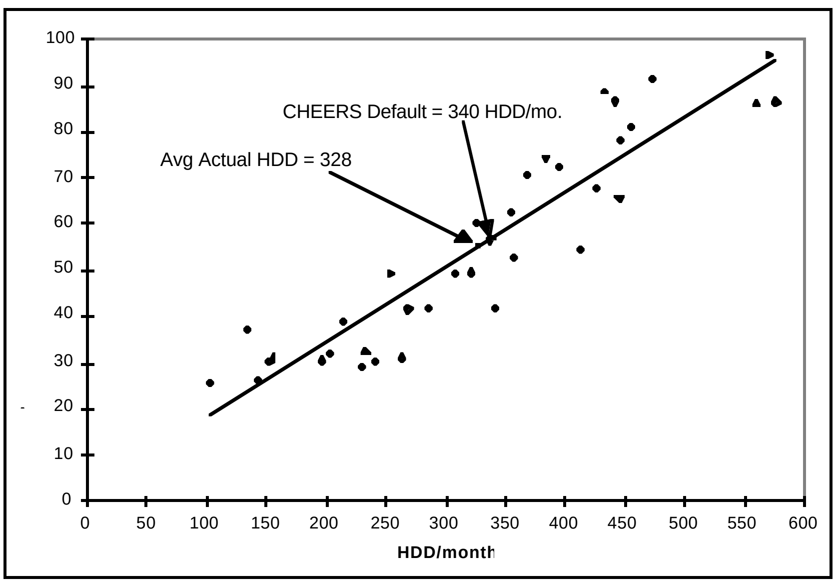


Figure 10. CHEERS: Gas Use/month vs HDD/month in Eureka



The difference between the actual HDDs and CDDs and the CHEERS defaults for Fresno is even smaller than for Eureka. Furthermore, the relationship between HDDs and gas and between CDDs and electricity is less pronounced than the HDD-gas relationship in Eureka. Thus there is also no need to weather normalize the Fresno data.

(2) Post-Rating Retrofits

Most of the houses were rated in mid-1994; some were rated in late 1993. If retrofits were undertaken as a result of the ratings (or as a result of obsolescence, etc.) after the ratings, then it would not be appropriate to compare the CHEERS estimates of current energy use with energy use for periods after the rating. (A CHEERS representative did not believe that any of these rating were used for Energy Mortgages.)

To get an idea from the data of whether retrofits were undertaken after the ratings were performed we plotted electricity consumption in 1993 vs. consumption in 1995. Figure 11 shows large variation in energy use on an individual basis but essentially no change on an average basis. Figure 12, however shows a slight decrease in average natural gas use from 1993 to 1995. Some of this decrease could be attributable to the milder winter period in 1995, but it could also mean that energy retrofits were undertaken after the ratings. It may also reflect the inevitable replacement of old, inefficient furnaces and water heaters with new, more efficient models, either before or after the ratings. Furnaces and water heaters are typically replaced about every 10-20 years (conversation with Alan Meier). However, given the small size of the change in gas use, the error introduced into our results by comparing ratings to pre- and post-rating energy data is probably small compared to the overall uncertainty in the results.

Figure 11. CHEERS: 1993 vs 1995 Electricity Use

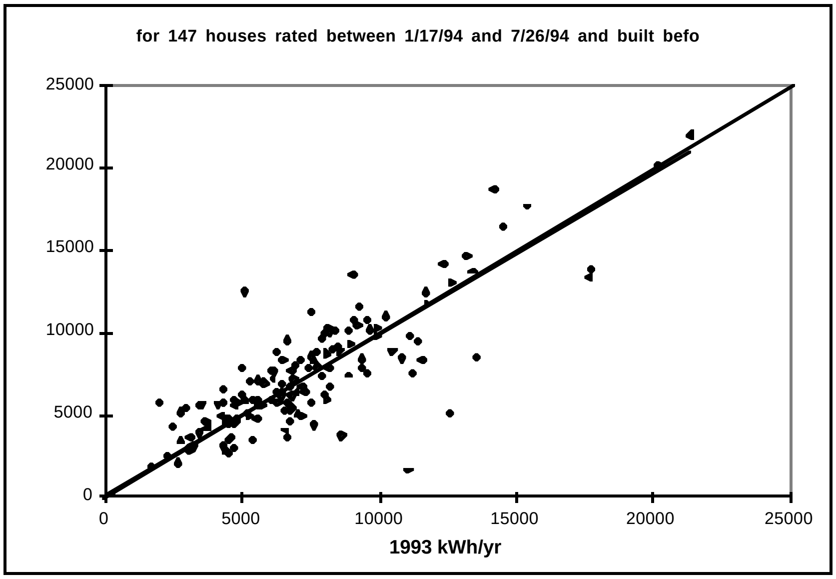
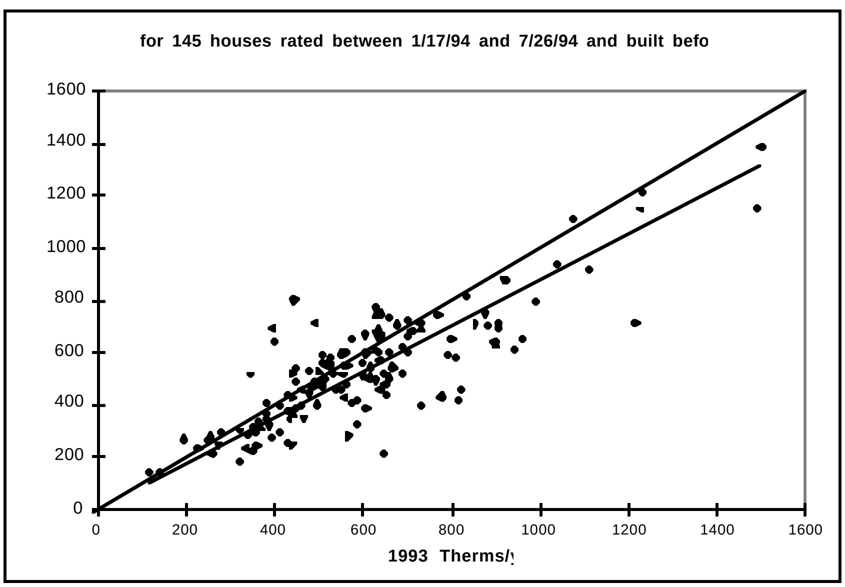


Figure 12. CHEERS: 1993 vs 1995 Gas Use



(3) Problems with Utility Data

There are at least two potential sources of uncertainty regarding the utility data: (1) some houses may be only partially occupied, and (2) the utility data may simply be wrong.

CHEERS, like all HERS, makes standard assumptions about occupant behavior, including vacations. It is not appropriate to compare ratings with energy use of unoccupied houses. If, for example, a family had a second home and only occupied the rated home eight months out of the year, neither the family, nor the HERS provider, would expect the rating to be accurate. Another possible source of error would be if the occupants of a rated house moved out, and there was an extended vacant period before new occupants arrived.

Possible sources of bad utility data include: errors in meter reading and errors in data handling before we received the data. Some filtering of the utility data was performed (as described in Appendix B) to remove clearly bad data and house-years with long periods of clear nonoccupancy. The remaining data has been extensively scrutinized. While there are some suspicious records, the vast majority of the data are within a reasonable range of expected values and follow believable monthly and season patterns. The average gas use in Eureka (shown in Figure 9) is an example of a realistic level of consumption and seasonal pattern.

2. Accuracy Correlations

In order to help isolate the sources of error in the ratings we attempted to determine what types of houses CHEERS is better at estimating, i.e. what types of houses had lower average errors and standard deviations. Two basic methods were used for finding such correlations. The first method is to subdivide the sample population according to different house characteristics and see if certain subsets are more accurate than others. The second method is multiple linear regression analysis. Using the first method, a number of variables were isolated. The ones that exhibited the strongest effect on accuracy are discussed below.

a) Climate Zone

Overestimation of gas use is greater in Eureka (Climate Zone 1) than in Fresno (Zone 13) (See Figure 13). Figure 14 shows that electricity use overestimation is greater in Fresno. Statistical analysis shows that there is no general bias in electricity estimation for Eureka. However, there is statistical significance to the overestimation of gas in Eureka and gas and electricity in Fresno.

Figure 13. CHEERS: Gas Use Error by Climate Zone

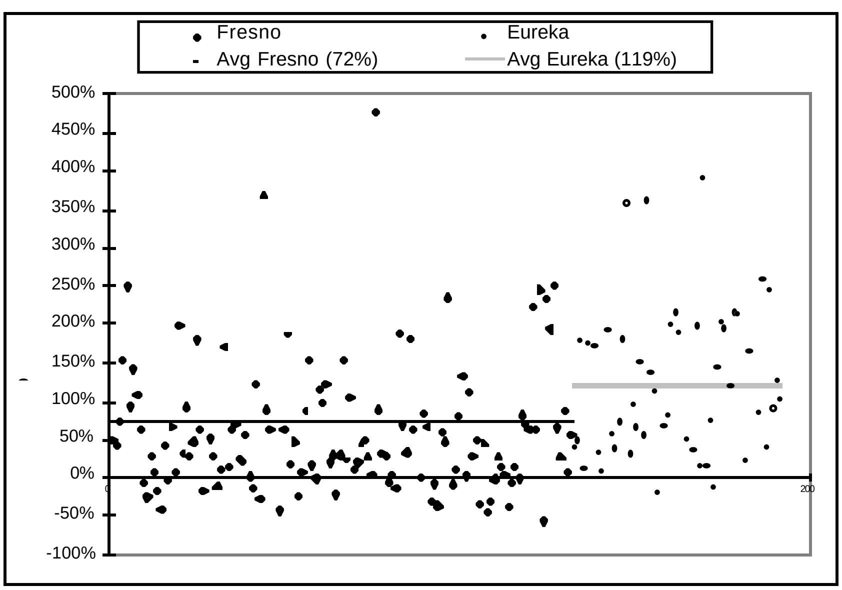


Figure 14. CHEERS: Electricity Error vs Climate Zone and Air Conditioning Type

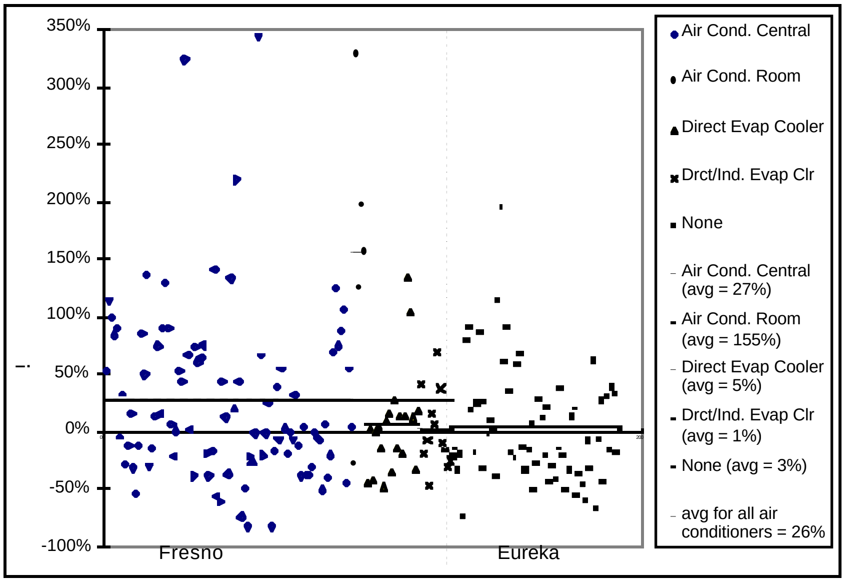


Figure 15 shows the type of heating, cooling, and hot water equipment in the test houses in each climate zone. Figure 16 shows the CHEERS estimated breakdown of electricity and gas use by end-use in each climate zone. From these figures it is clear that Eureka homes require more heating than Fresno homes and that Fresno homes require more cooling than Eureka homes, which do not have air conditioning.

Figure 15. CHEERS: Equipment Type

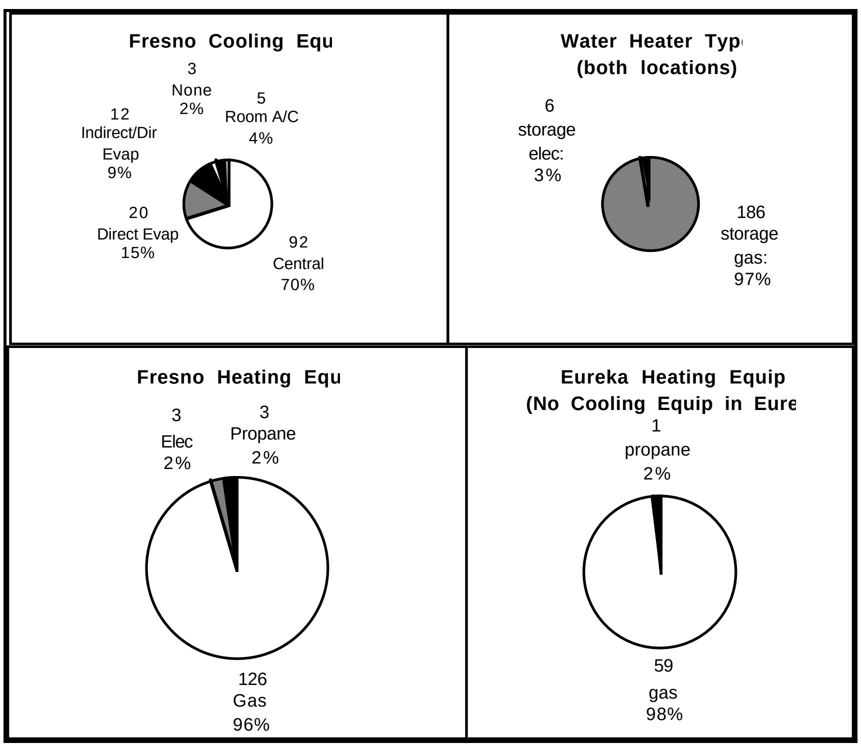
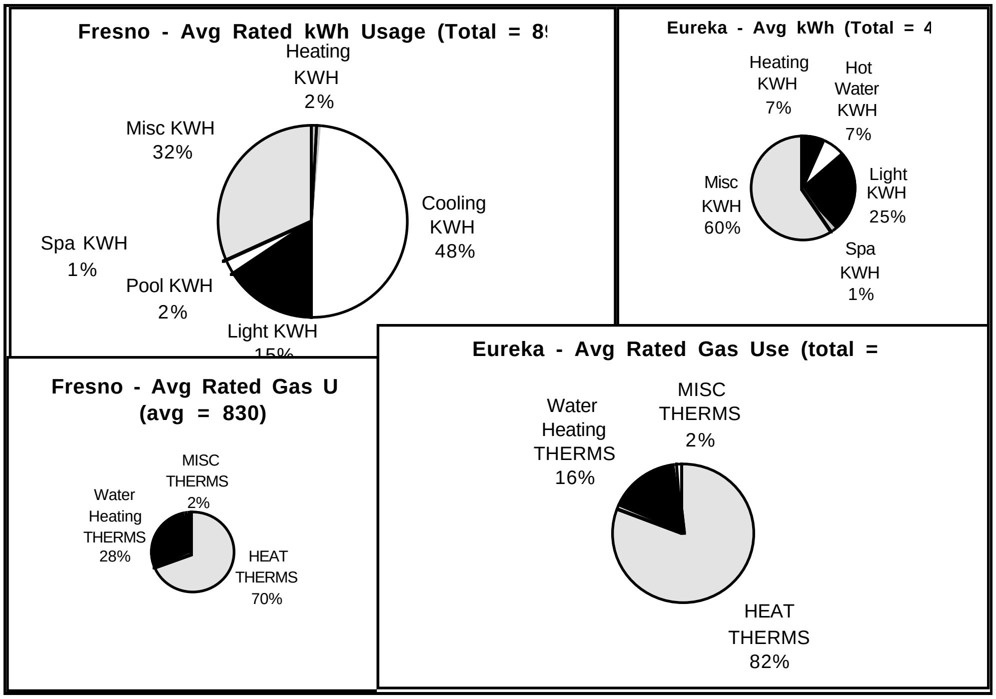


Figure 16. CHEERS: Average End-Use Estimates



Thus it appears that CHEERS accurately estimates miscellaneous electricity use but overestimates heating and cooling energy use. The more heating or cooling required, the greater the overstatement (See Table 6). Table 6 also shows that it is more difficult to estimate heating energy use in the milder winter climate of Fresno (standard deviation = 159%) than in the colder winter climate of Eureka (standard deviation = 93%).

Table 6. Summary of CHEERS Results

Location	Avg. Year Built	Avg. Area (ft^2)	Avg. Elec. (kWh/y)	Avg. Gas (therm/y)	Avg. Cost ($/y)	Elec. Error	Elec. Std Dev	Gas Error	Gas Std Dev	Cost Error	Cost Std Dev
Fresno	1965	1524	8019	541	1247	25%	74%	72%	159%	45%	84%
Eureka	1947	1340	5005	655	945	4%	50%	119%	93%	64%	66%
Both	1959	1466	7072	576	1154	18	68	86	144	51%	80%

b) House Age

The accuracy of gas use estimation and total energy cost estimation improves the newer the house is. In fact, for houses built between 1990 and 1994, the average CHEERS rating underestimated total energy use by 8%. And for all homes built since 1977 (inception of Title 24), the average rating overestimated total energy use by only 11%. Again large variation in accuracy is seen for these groups (See Figure 17 and Figure 18). For example, for 52 homes built since 1977, the standard deviation in the total cost error is 55%, meaning that approximately one third of the errors are either above 66% or below -44%.

Figure 17. CHEERS: Cost Error vs Age Group and Climate Zone

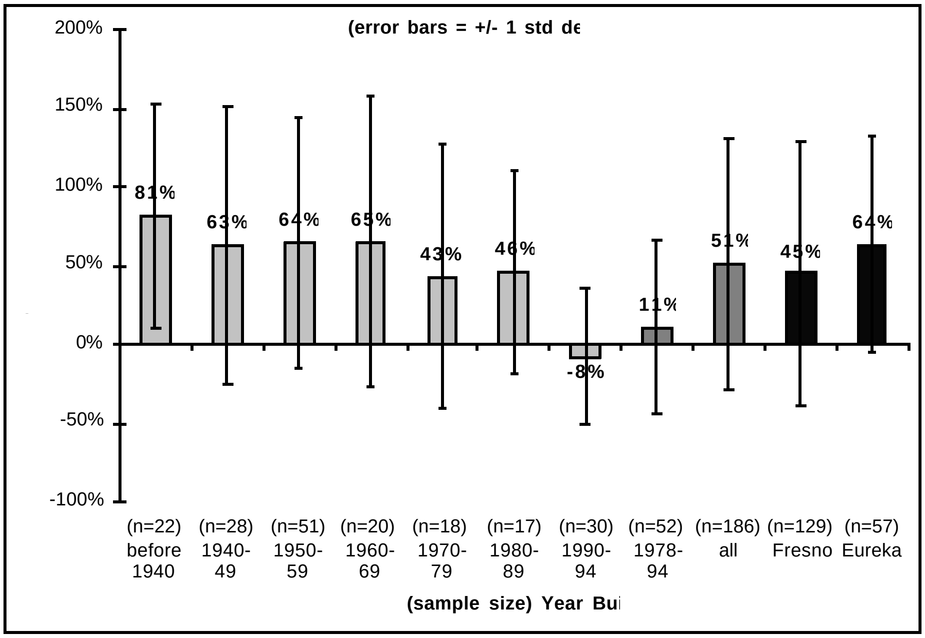
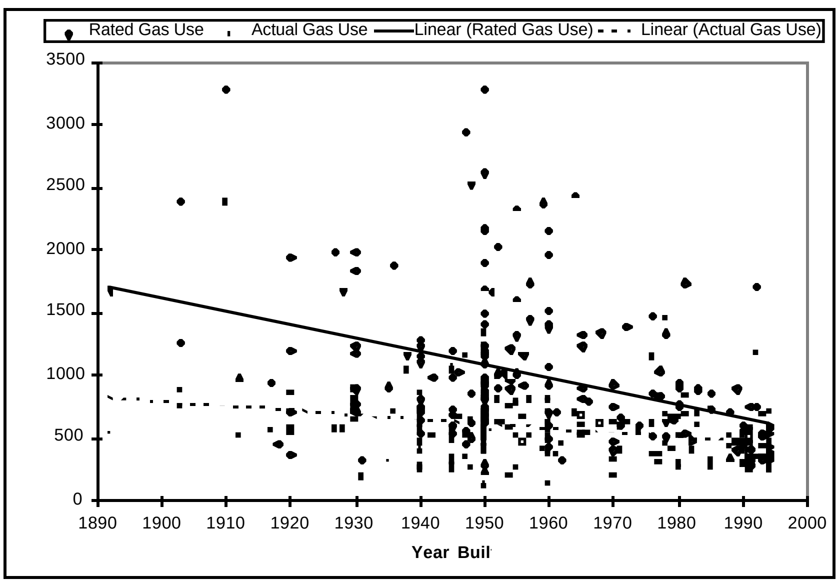


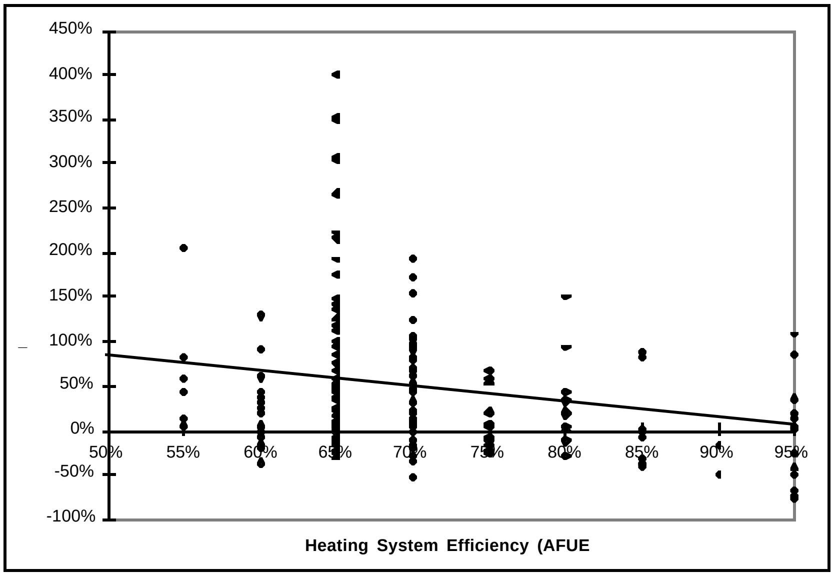
Figure 18. CHEERS: Rated and Actual Gas Use vs House Age



c) Energy Efficient Features

One of the strongest correlations found was between energy cost accuracy and the presence of certain energy efficient features. For example, the average energy cost error for homes with simple thermostats was 65% (sample size = 118 houses), while the average energy cost error for houses with the more advanced night setback thermostats was only 22% (65 houses). Similarly, the average energy cost error for houses with older, “non-labeled” windows was 73% (109 houses), while the average energy cost error for houses with “labeled” windows was only 20% (77 houses)[3]. Furthermore, Figure 19 shows a clear relationship between accuracy of gas estimation and furnace efficiency. For high efficiency furnaces, CHEERS is quite accurate on average, but as furnace efficiency decreases, overestimation increases. Since the presence of these energy efficient features correlates closely with age, the correlation between accuracy and energy efficient features supports the hypothesis that newer, more efficient houses can be rated more accurately.

Figure 19. CHEERS: Energy Cost Error vs Heating System Efficiency



Since we do not have the complete CHEERS file on each house we could not compare accuracy to some of the other energy efficient features such as wall or attic insulation. The overall house score is intended as a measure of a house’s energy efficiency. Correlations between score and accuracy are discussed in a later section of this report.

d) Rater

The CHEERS ratings were conducted by seven different raters. As Figure 20 shows, some raters were clearly more accurate than others. This appears to be strong evidence that the skill of the rater plays an important role in accuracy. However, as with other correlations discussed above, correlation does not prove causality. Some raters could perform better than others not necessarily because they are better raters but because they happen to have easier houses to rate. For example, rater 305 had a lower average error and lower standard deviation than rater 304 but 305’s houses were also much newer on average (1980 versus 1956). On the other hand, rater 303 did considerably worse than rater 304 even though 303’s houses were newer. Given the large role that the rater plays in the rating process and the strength of the correlation shown in Figure 20, it seems clear that the rater can be a significant source of error.

[3] A window label means that type of window has undergone testing by ANSI, a nationally recognized testing institute, and is considered to have better thermal performance than non-labeled windows, which are usually older.

Figure 20. CHEERS: Energy Cost Error for 7 Raters

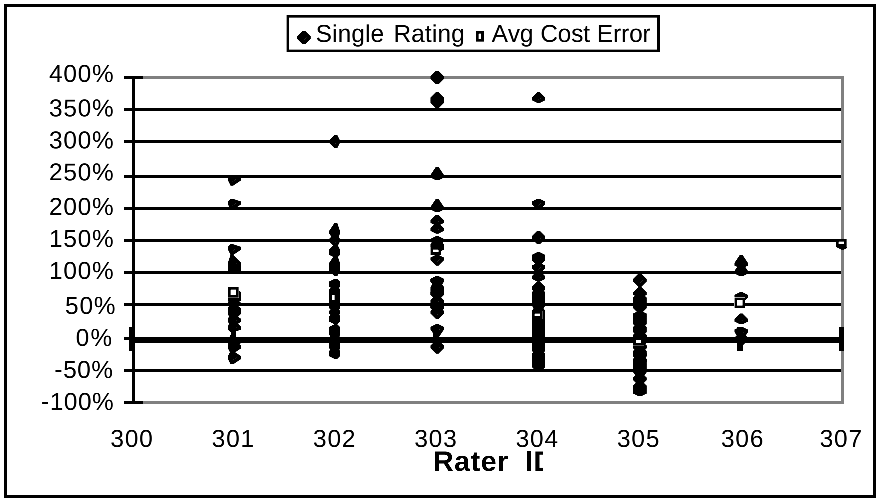


Table 7. Summary of CHEERS Results by Rater

Rater ID Number	301	302	303	304	305	306	307
No. Ratings	19	38	25	61	35	6	1
Zone	1	1	13	13	13	13	13
Avg Yr Built	1936	1953	1963	1956	1980	1977	1956
Avg Cost Error	69%	61%	132%	36%	-6%	54%	144%
Std dev in Cost Error	72%	65%	111%	65%	42%	49%	

e) Outlier Analysis

Detailed analysis was undertaken of the houses for which CHEERS appears to have done a particularly poor job of estimating total energy cost. It is possible that these outliers have unusual features that CHEERS has difficulty dealing with. There were five houses with cost error < -50% (i.e. estimated energy cost was less than half the actual cost) and ten houses with cost error > 200% (estimated energy cost was at least two times the actual cost). Unfortunately, the CHEERS input data does not indicate anything particularly unusual about either of these groups of houses (see Table 8). Both groups had gas heating and hot water. None had pools, although one of the ten overestimation houses had an electric spa.

The underestimation outliers were all newer houses in Fresno. All had high efficiency equipment (e.g., duct insulation and return duct in conditioned space) and correspondingly high overall scores. Yet all used unusually high amounts of electricity, while gas use was close to the estimated amount (see Table 8). This suggests that the error could be from air conditioning or plug load use.

The overestimation outliers were all much older (average year built = 1948) and had lower efficiency equipment (e.g. unlabeled windows, no setback thermostat). While the estimated electricity and gas amounts were perhaps unreasonably high (e.g., over 20,000 kWh/yr. for a 1,150 square foot house), some of the actual electricity and gas use amounts were suspiciously low (e.g. 835 kWh/yr., which is only about $7/month for electricity).

One possible explanation for the large errors is that the overestimated houses were un-occupied or under-occupied and that the underestimated houses had unusually high numbers of occupants. Another possibility is that the utility data were corrupted in some way but still looked reasonable enough to have passed through the data cleansing filter described earlier. It is also possible that these houses showed behavior variations and the takeback effect. The newer houses might just happen to be occupied by "energy hogs," and the older ones occupied by "energy misers."

Table 8. CHEERS: Outliers

Underestimation Outliers																Rated	Actual	Rated	Actual
Record_	Cost Err	Elec Err	Gas Erro	Year B	Dryer T	Range T	Location	Windows	Heating Effic	Cooling Equip	Cooling Effi	Return Duct L	Duct Insu	Controls	kWh	kWh	Therms	Therms
83	-80%	-83%	-5%	1994	Gas	Gas	Fresno	Labeled	AFUE_0.95	Air Cond. Cer	SEER_16.0	Conditioned	R-6.3	setback	3831	21926	359	378
99	-77%	-82%	60%	1994	Gas	Gas	Fresno	Labeled	AFUE_0.95	Air Cond. Cer	SEER_16.0	Conditioned	R-6.3	setback	3490	19937	382	239
81	-72%	-73%	32%	1994	Electri	Electric	Fresno	Labeled	AFUE_0.95	Air Cond. Cer	SEER_16.0	Conditioned	R-6.3	setback	6451	24103	379	288
123	-60%	-50%	-37%	1991	Electri	Electric	Fresno	Labeled	AFUE_0.90	Air Cond. Cer	SEER_12.0	Conditioned	R-6.3	setback	6191	12457	353	563
67	-51%	-60%	-22%	1981	Electri	Electric	Fresno	Labeled	AFUE_0.70	Air Cond. Cer	SEER_08.0	Conditioned	R-4.2	no_setba	6559	16511	548	699
Average	-68%	-70%	6%	1991	Electri	Electric	Fresno	Labeled	AFUE_0.90	Air Cond. Cer	SEER_12.0	Conditioned	R-6.3	setback	5304	18987	404	433
Overestimation Outliers															Rated	Actual	Rated	Actual
Record_	Cost Err	Elec Err	Gas Erro	Year B	Dryer T	Range T	Location	Windows	Heating Effic	Cooling Equip	Cooling Effi	Return Duct L	Duct Insu	Controls	kWh	kWh	Therms	Therms
136	203%	156%	194%	1928	Gas	Gas	Fresno	NotLabeled	AFUE_0.65	Air Cond. Roo	SEER_06.0	Crawlspace	R-4.2	no_setba	11374	4439	1674	570
133	207%	125%	246%	1927	Electri	Gas	Fresno	NotLabeled	AFUE_0.65	Air Cond. Roo	SEER_10.0	Crawlspace	R-1.0	no_setba	10491	4664	1984	574
182	208%	58%	360%	1954	None	Gas	Eureka	NotLabeled	AFUE_0.55	None				no_setba	1871	1183	893	194
201	243%	36%	388%	1948	Electri	Electric	Eureka	NotLabeled	AFUE_0.65	None		Crawlspace	R-2.1	no_setba	4621	3385	2524	517
47	252%	323%	88%	1942	Electri	Electric	Fresno	NotLabeled	AFUE_0.65	Air Cond. Cer	SEER_06.0	Attic	R-2.1	no_setba	16630	3928	991	526
175	302%	194%	356%	1970	None	Gas	Eureka	NotLabeled	AFUE_0.65	None				no_setba	2458	835	926	203
78	363%	219%	476%	1959	Electri	Electric	Fresno	NotLabeled	AFUE_0.65	Air Cond. Cer	SEER_06.0	Attic	None	no_setba	21605	6774	2377	413
89	367%	344%	182%	1950	Electri	Electric	Fresno	NotLabeled	AFUE_0.65	Air Cond. Cer	SEER_06.0	Attic	R-1.0	no_setba	21776	4909	1496	531
21	367%	328%	198%	1945	Electri	Electric	Fresno	NotLabeled	AFUE_0.65	Air Cond. Roo	SEER_06.0			no_setba	7189	1682	725	243
91	400%	104%	1608%	1960	None	Electric	Fresno	NotLabeled	AFUE_0.65	Direct Evap Cooler		Attic	None	no_setba	7596	3721	2152	126
Average	291%	189%	410%	1948			Fresno	NotLabeled	AFUE_0.65		SEER_06.0			no_setba	10561	3552	1574	390

f) Multiple Linear Regression

Multiple linear regression analysis was used to determine which variables were most significant in predicting the cost estimation error. Regressions were run on various combinations of nine variables that were thought to be most significant. Several of these were numerical variables (house age, house area, heating efficiency, and overall rating score), and several were converted to dummy (0/1) variables (rater, climate zone, window labels, setback thermostat, and air conditioning). For example, the air-conditioning variable was set to 1 if the house had permanent air conditioning and 0 if it did not. Six dummy variables represented the seven raters.

The results of the regression runs were not very conclusive. The R-squares were very low for all of these runs (less than 35%), implying that no combination of these variables was particularly good at predicting the cost error. Table 9 is a sample output from one of these runs. The small P-value (P-value < 0.05) for some of the raters and for age and heating efficiency implies that these variables were significant in predicting cost error but the presence of air conditioning may not be significant.

Table 9. CHEERS: Sample Regression Output

SUMMARY OUTPUT				
Regression Statistics				
Multiple R	0.572			
R Square	0.327			
Adjusted R Square	0.290			
Standard Error	0.670			
Observations	184			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0	#N/A	#N/A	#N/A
Rater1	1.527	0.570	2.680	0.008
Rater2	1.573	0.539	2.919	0.004
Rater3	2.095	0.537	3.903	0.000
Rater4	1.126	0.510	2.207	0.029
Rater5	1.099	0.592	1.856	0.065
Rater6	1.716	0.673	2.548	0.012
Age	0.007	0.003	2.407	0.017
Heatg Effic	-1.865	0.699	-2.669	0.008
AC (1 = yes)	0.295	0.161	1.831	0.069

The general result of the regression was that:

- Other variables not included in the regression analysis play a large role in determining cost error. Occupant behavior is probably the most significant of such variables.
- Of the variables analyzed, the rater seems to be one of the most significant in predicting the cost error.
- The heating system efficiency of the house is significant. The heating system efficiency is generally reflective of the house age or overall efficiency. Thus we can say that house age/efficiency is significant.

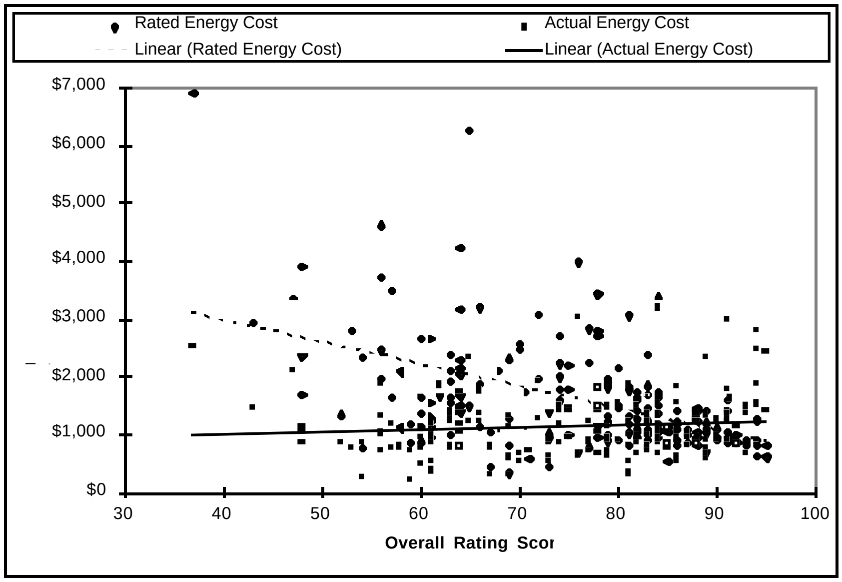
3. Score Accuracy

Unlike estimation accuracy, the actual electricity and gas use data cannot be used to directly judge score accuracy. Technically, a score is a way of judging the efficiency of a house relative to the efficiency potential for that house. Scores, therefore, are not intended to be used for comparing different houses. However, since scores are often used, formally or informally, to compare energy costs of different houses, it is worthwhile to compare scores for different houses to energy cost for different houses.

While CHEERS provides several end-use scores (heating, cooling, hot water, lighting and overall), we only have energy use data by fuel type and thus the only score that we can compare directly to energy use is the overall energy score. It is important to realize, however, that overall score is a weighted average of the end-use scores. Since there is no end-use score or even data collected on most occupant-specific features such as the efficiency of the clothes dryer, the overall score does not account for all the energy uses in a house.

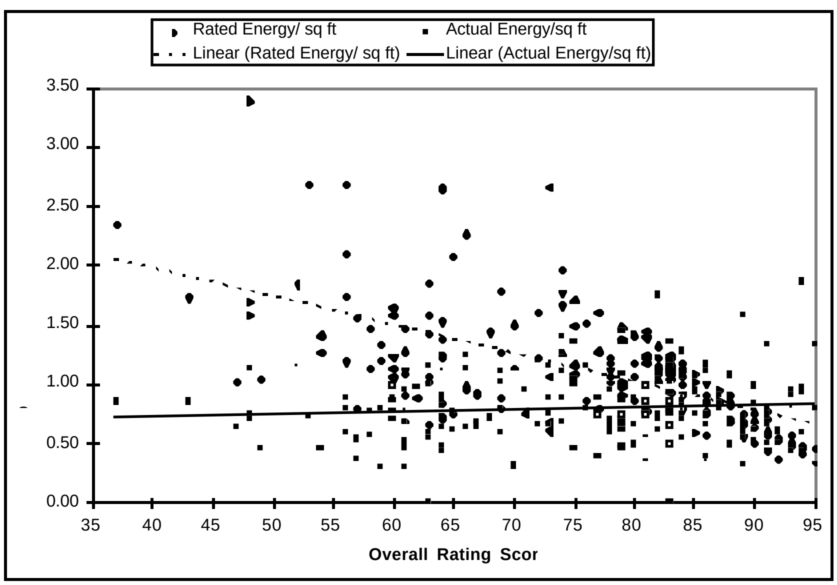
Figure 21 shows that homes with low overall rating scores are predicted to use more energy than homes with high overall scores (as one would expect), although there is significant variance. For example, several homes rated around 70 are predicted to use less than $1000/yr. while several other houses with similar scores are predicted to use over $3000/yr. Figure 21 also shows that there is almost no relationship between overall score and actual energy cost. Simple linear regression analysis also showed that there is no statistical relationship between score and actual cost. (Energy cost appears to remain fairly constant at around $1,000/yr., regardless of score.) Thus overall score should not be used to compare different houses.

Figure 21. CHEERS: Rated and Actual Energy Cost vs Rating Score



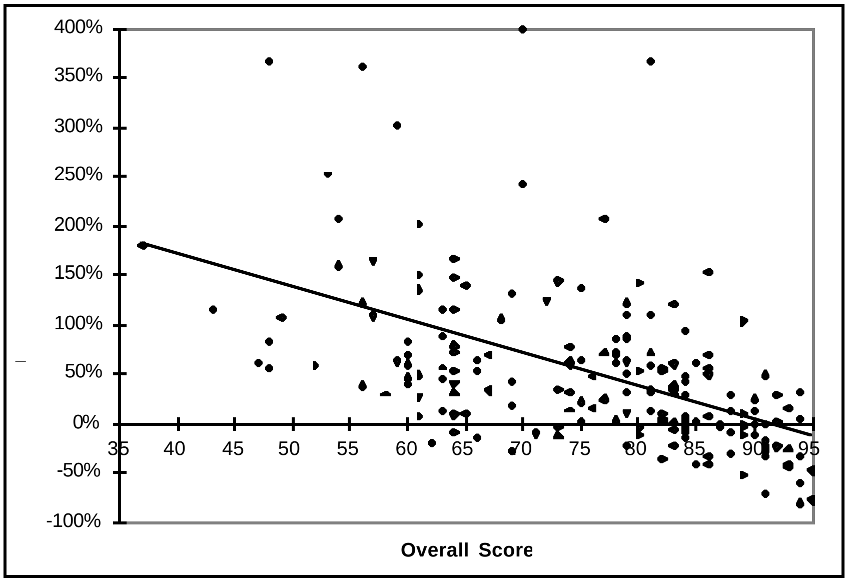
Can CHEERS ratings be used to compare energy costs for houses of the same size? Unfortunately, Figure 22 shows that if we normalize according to house size, overall score still does not correlate well with energy cost. This figure also shows that energy costs per square foot remain fairly constant regardless of overall score.

Figure 22. CHEERS: Energy Cost Per Square Foot vs Rating Score



It is worth noting that cost estimation accuracy improves as score increases. This can be seen more clearly in Figure 23. Not only does average error decrease as score increases but the standard deviation decreases as well. This result is consistent with the earlier correlation found between age and accuracy since newer houses generally have higher scores.

Figure 23. CHEERS: Energy Cost Error vs Rating Score

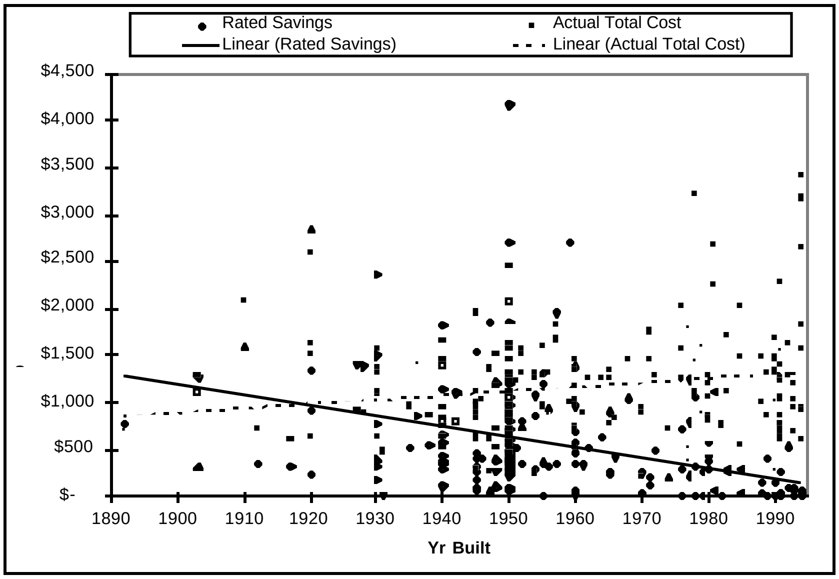


4. Recommendations Accuracy

Like score accuracy, utility data cannot be used to directly judge the accuracy of recommendations. To truly judge whether specific recommendations turned out to be cost-effective would require detailed end-use monitoring before, and for a considerable period after, installation of recommended features. However, it is useful to compare the CHEERS estimates for potential savings to the actual energy cost to get a ballpark idea of how realistic the savings figures are. While we did not receive from CHEERS the lists of recommended improvements for each house, we did receive the estimated annual energy cost if all recommended improvements were installed. The estimated potential savings is the difference between estimated current energy cost and the estimated energy cost after retrofit installation.

Figure 24 shows that the estimated potential savings decreases as house age decreases to the point where new houses are not estimated to have any potential savings. Actual energy use, however, increases slightly as house age decreases. For several of the older houses CHEERS is estimating that more energy can be saved than is actually being currently used. 13% of the houses are estimated to be able to save more energy than is currently being consumed. Clearly this is not possible. Most likely the overestimated potential savings results from overestimated energy use. Even 50% savings would be difficult to achieve considering that CHEERS recommendations do not cover energy use by non-permanent features such as refrigerators or non-permanent lights. 35% of the houses are estimated to be able to save more than 50% of the amount currently consumed. Thus it would appear that a significant percentage of recommendation packages must not be cost effective. This does not mean that each of the recommendations in a package of recommendations is not cost-effective. Conversely, given the higher than estimated energy use of many newer homes, there may be some cost-effective recommendations for newer houses that are not being recommended.

Figure 24. CHEERS: Estimated Potential Savings vs Actual Energy Cost



On the other hand, if people with older, less efficient houses are receiving a lower level of service than they would like, then it is possible that recommendations that would appear to be uneconomical can still be worthwhile if they also improve the level of service. For example, if a homeowner would like an indoor temperature of 68 degrees but the current furnace can only maintain a temperature of 63, then a new furnace that can achieve 68 might not deliver the dollar savings to pay for the incremental cost but the increased level of service makes it a good investment. At the other end of the spectrum, CHEERS may not be missing any possible recommendations for newer houses because the high energy consumption is in areas beyond CHEERS' purview, such as aquariums, portable heaters or air conditioners, etc.

C. Conclusions

On average, CHEERS tends to overestimate the total annual energy cost by approximately 50%. The variation in estimation accuracy is quite high, with a small number of the houses being overestimated by over 300%. Clearly occupant behavior contributes to the large variance. Insufficient data collection may also contribute to variance. For example, collecting actual infiltration data through blower door testing, rather than using default values, could reduce the variance.

Although it is not possible to isolate the exact sources of error in energy cost estimates, it does <u>not</u> appear that a significant source of error is the fuel price schedule used by CHEERS.

There is a clear correlation between accuracy of fuel use estimates and climate zone. Gas use overestimation is larger in Eureka, which has a larger heating load, than Fresno. On the other hand, electricity use overestimation is greater in Fresno, which has a larger cooling load. In fact, electricity use estimates are quite accurate in Eureka on average, where none of the sample houses had built-in air conditioning, i.e. no cooling load. Thus it appears that CHEERS accurately estimates miscellaneous electricity use but overstates heating and cooling energy use. The more heating or cooling required, the greater the overstatement. This could be caused by the use of unrealistic heating and cooling setpoints or assuming infiltration rates that are too high.

The weather during the monitoring period (1993-1994) is not significantly different from the long term average weather assumed by CHEERS. Therefore differences between actual weather and assumed weather is not a significant source of error in these ratings.

There is a clear correlation between house age and cost estimation error. New houses tend to be more accurate. Both the mean error and the standard deviation are smaller for newer houses than for older ones. For example, for 52 houses built since 1977, the mean cost error was 11%, which is statistically indistinguishable from 0% error given the level of uncertainty in this analysis. On the other hand, the mean cost error was 81% for 22 houses built before 1922. There are a number of possible explanations for the relationship between age and cost error:

- People may operate older, less efficient houses differently than people operate newer ones. Because it costs more to heat and cool an inefficient house and because they are often owned by poorer people than new house owners, occupants of older houses may settle for a lower level of service or comfort.
- Older houses are harder to rate, because equipment labels and documentation are less likely to be available. Older houses are also more likely to have had improvements that the owner and rater are not aware of. Also, small mistakes in rating older houses can have a larger effect on accuracy than mistakes of the same size in rating newer houses. For example, if a rater misjudges the insulation level in an older home and records R-1 instead of R-2 (off by 50%) it will have a greater effect on accuracy than if a rater misjudges the insulation level on a new home and recorded R-11 instead of R-13 (off by 15%).
- CHEERS default values for insulation, equipment efficiencies, etc. may be more inaccurate as homes get older and equipment types get more inefficient. For example, poor quality windows may not be as bad as previously thought, or old furnaces may be more or less efficient than expected.

The rater seems to play a significant role in the estimation accuracy of ratings. Some of the raters in this sample of ratings had much lower average errors and lower standard deviations than others. Although, multiple linear regression analysis did not provide strong conclusions, it did seem to indicate that the rater is a significant source of error. This certainly seems plausible given the amount of knowledge of building systems that a rater should have and the amount of "guesstimating" on the rater's part that goes into a rating. Variations between raters could also indicate that some raters were less proficient with the DOS-based Version I tool (The Windows-based Version II is more user-friendly).

Contrary to what one would expect, there appears to be no relationship between overall rating score and actual energy cost. In fact, average energy cost is fairly constant when compared to overall score. Again, it is possible that this phenomenon could be explained by the takeback effect. As overall efficiency increases people takeback more and more of the savings in increased services. Thus it is possible that overall score would correlate well with actual energy costs if all houses were operated in the same manner.

A significant percentage of the CHEERS recommendation packages are likely to not be cost-effective for the current occupants and for any potential new occupants. This appears to be the case because many of the ratings estimated that it is possible to save almost as much, if not more, energy than the houses are currently using. Since CHEERS recommendations are restricted to the end-uses that CHEERS rates (heating, cooling, lighting and hot water), cost-effective savings of even 50% are probably unrealistic.

VI. Other Validation Data

A. Kansas - Midwest Energy

Brian Dreiling of Midwest Energy, a utility company and HERS provider in Kansas, supplied us with ratings and utility data for 24 houses in the Dodge City Kansas area. He selected this particular sample based on rating date and is confident that it is a representative sample. All of the ratings were conducted by Dreiling between June 1995 and May 1996 using the Rem/Rate v8.0 software. Utility data is for the period of February 1996 to January 1997. Most of the houses were built since 1994 and Dreiling is confident that no improvements were made after the ratings. Most of the homes are single family detached but a few are duplexes. Blower door tests were conducted on all of the ratings.

We eliminated eight of these houses from our analysis because either there were too few months of gas or electric data or the utility data indicated extended periods of vacancy or corrupt data. Thus the results are for the remaining 16 houses.

As Figure 25 shows, there was close agreement between estimated energy cost and actual energy cost. The average cost error is -7%, with a standard deviation of 15%. This high degree of accuracy supports the theory that new houses are easier to rate than old ones. It also seems to show that blower door testing can improve accuracy. Figure 26 shows that as the score increases, the tendency of the rating system to underestimate energy cost increases. This downward sloping trendline was also seen for the CHEERS data (see Figure 23) and supports the "takeback effect" hypothesis that occupants increase their level of service as houses get more efficient.

Figure 25. Kansas: Estimated vs Actual Energy Cost

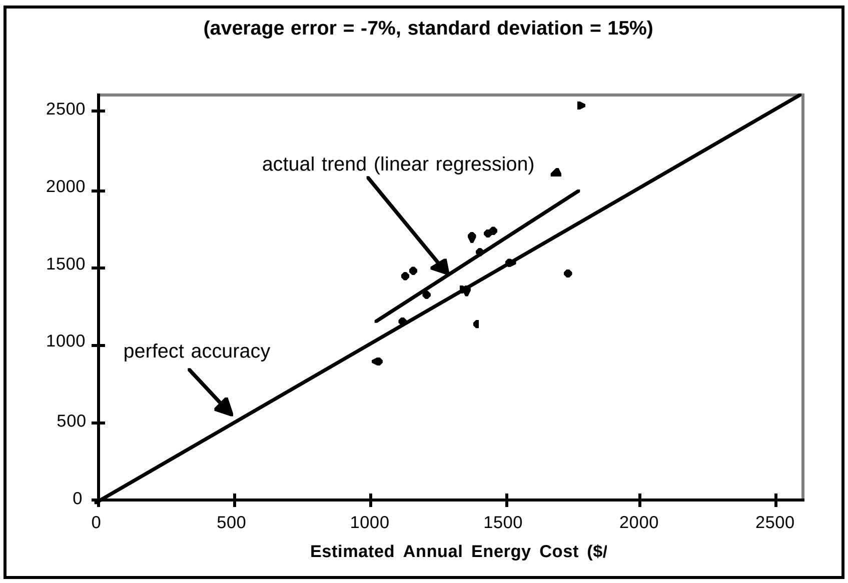
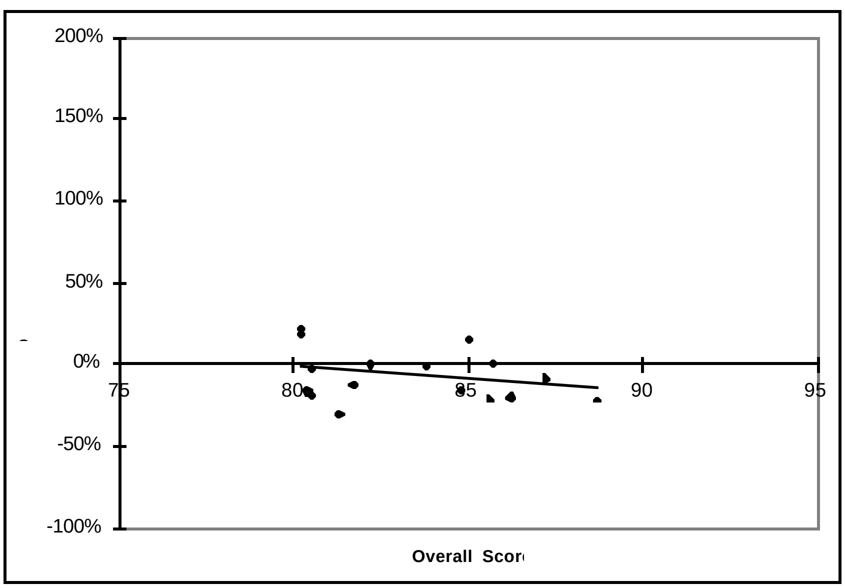


Figure 26. Kansas: Cost Error vs Rating Score



B. Ohio - HERO

Ruth Kubinski of Home Energy Ratings of Ohio Foundation, supplied us with rating and utility data for 14 houses in Ohio. This sample was selected randomly. All of the ratings were conducted in 1996 or early 1997. The utility data, which were requested from the relevant gas and electric utilities by HERO on a house by house basis, are mostly from 1995. Sally Lukens, the Executive Director of HERO, estimates that the houses are of varying ages, the houses are all single family units, and the ratings were conducted by several different raters. Blower door testing is included in all HERO ratings. Diagnostic testing is also performed to determine the actual efficiency of the heating equipment.

The software used was developed by American Electric Power and is not generally available on the market. This software is fundamentally different from other HERS software and deviates from the HERS concept described in section II in that it allows the rater to enter occupant specific data. For example, the rater can input the number of occupants. Furthermore, after collecting utility bills, the rater can adjust thermostat setpoints in the software model in order to reconcile the rating output and utility data. HERO allows for occupant specific data because the vast majority of HERO ratings are used for trouble shooting, i.e. to allow the current occupant to better understand and control energy use. Thus the more tailored the model, the more accurate it is, and the more useful it will be. Even when the rating is used by homebuyers for an Energy Efficient Mortgage, the number of future occupants is entered into the software to improve accuracy.

As Figure 27 shows, there was close agreement between estimated energy cost and actual energy cost. The average cost error is -14%, with a standard deviation of 20%. Figure 28 shows that as the overall rating score increases, the tendency of the rating system to underestimate energy cost increases. This download sloping trendline was also seen for the CHEERS and Kansas data (see Figure 23, and Figure 26) and supports the "takeback" hypothesis that occupants increase their level of service as houses get more efficient.

Figure 27. Ohio: Rated vs Actual Energy Cost

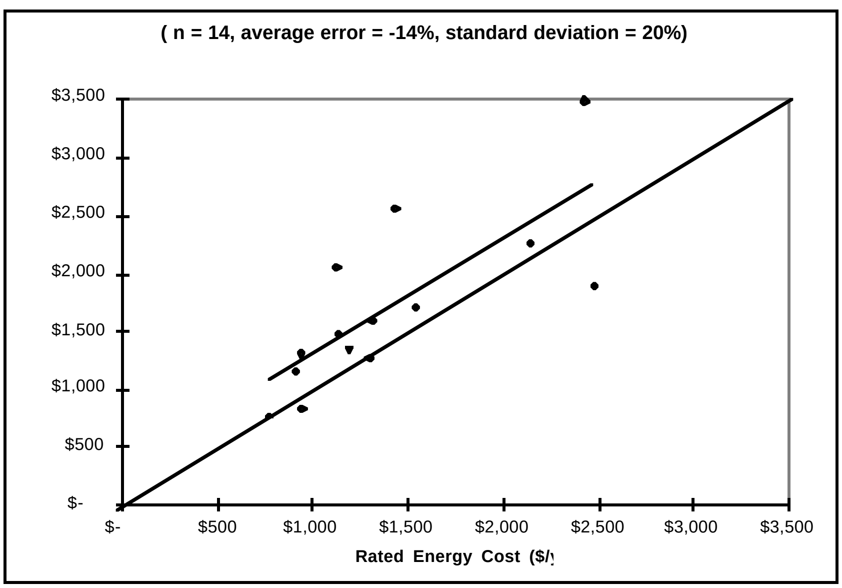
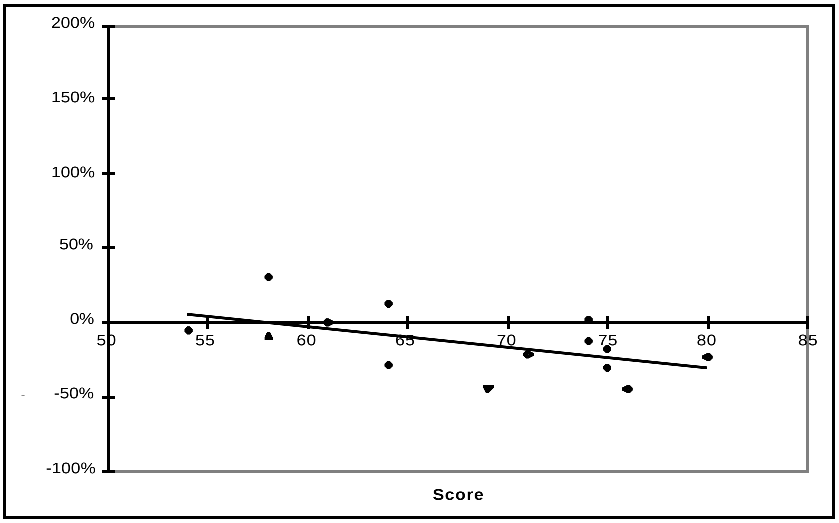


Figure 28. Ohio: Energy Cost Error vs Rating Score



C. *Energy Rated Homes of Colorado*

Lois Arena, of Energy Rated Homes of Colorado (ERHC), provided us with rating data, utility data, and some house data for 276 Colorado homes rated by several different ERHC certified raters between December 1995 and March 1997. ERHC uses the EZ Rater software developed by Evan Brown of Energy Rated Homes of America. This sample of ratings was selected randomly by Arena using the following search criteria: all ratings from around January 1996 to the present; actual bills available; houses with only gas and electric fuels; and only "as is," not new construction. Blower door testing was conducted in 275 of the 276 ratings. Utility data for ERHC ratings is collected by the rater at the time of the rating, usually from the homeowner. This procedure is different from other HERS we studied, and it gives the rater the opportunity to do some reality checking before presenting the rating to the customer or to ERHC. (We do not know if raters take advantage of this opportunity.) Raters typically collect one year of data prior to the rating date. Thus, for this sample of ratings, the utility data are primarily from 1995. We did not receive utility cost data, only kWh and ccf usage.

The houses are of all ages, with the year built ranging from 1888 to 1995 and an average year built of 1969; 91% of the houses have gas heating and 9% have electric heating; 90% have gas water heating and 10% have electric; 21% have gas clothes dryers, 76% have electric and 2% do not have dryers; 22% have gas cooking and 78% have electric cooking.

As Figure 29, Figure 30, and Figure 31 indicate, the estimated energy use is quite accurate, on average. The average error in estimated total energy use is only -3%. However, the standard deviation of 35% does imply that about one third of the ratings will be under or overestimated by more than about 35%. Electricity and natural gas use are compared by ERHC on a site energy basis (1 kWh = 3413 Btu). Figure 31 shows that the ratings tend to slightly underestimate electricity use and the larger standard deviation implies that electricity is harder to estimate than gas use (perhaps because it is more occupant dependent).

Figure 29. Colorado: Estimated vs Actual Total Energy Use

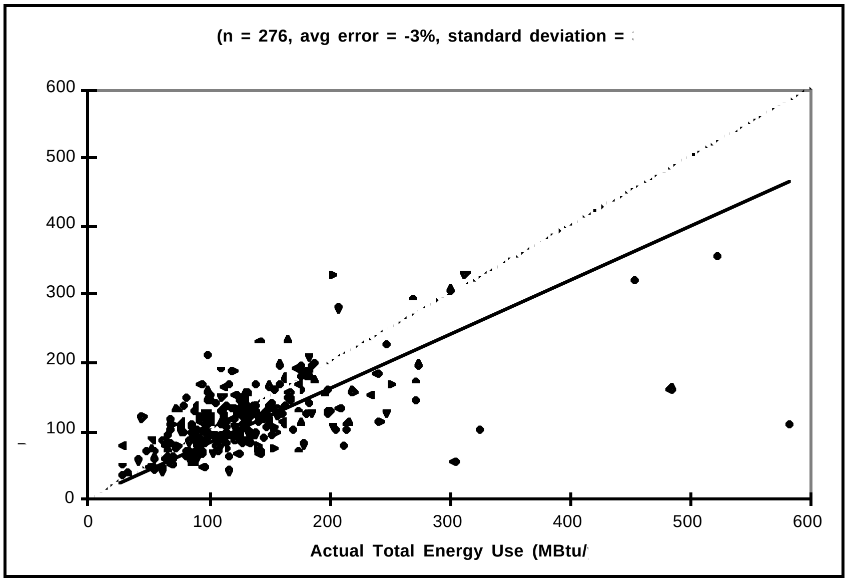


Figure 30. Colorado: Estimated vs Actual Gas Use

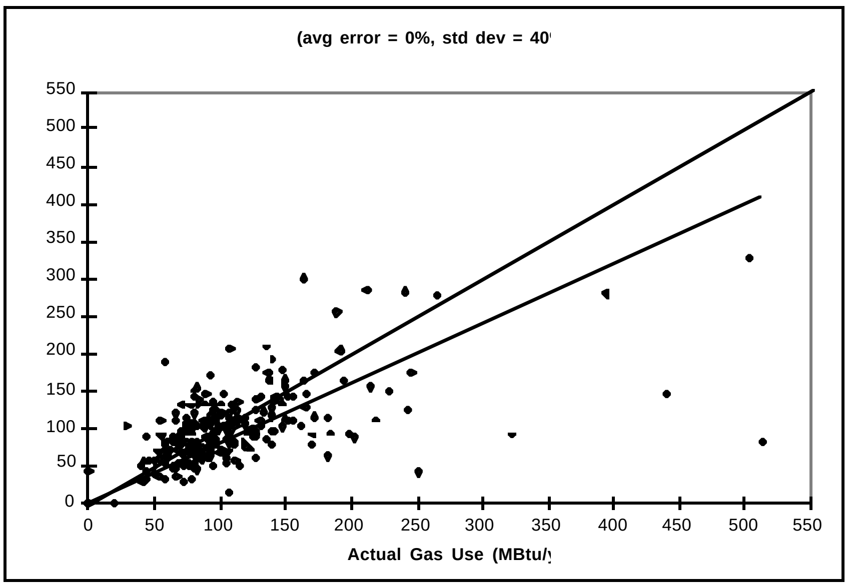
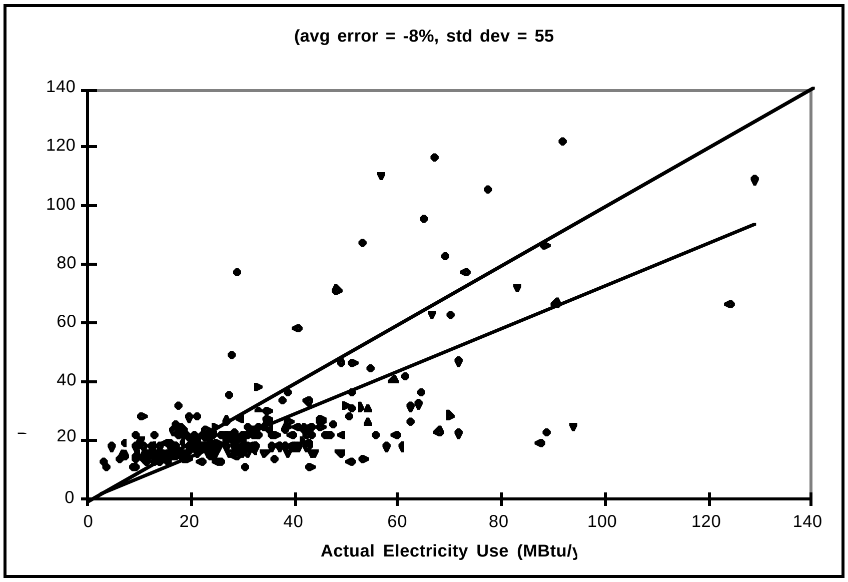


Figure 31. Colorado: Estimated vs Actual Electricity Use



Unlike the CHEERS results, there does not seem to be a clear relationship between accuracy and house age. Figure 32 shows that average error and standard deviation are roughly constant with respect to house age. Also in contrast to the CHEERS data, the Colorado ratings do not suggest that some raters are clearly more accurate than other raters. However, less active raters do sometimes tend to be less accurate on average and to have larger variances than the more active raters, which implies that rater experience improves accuracy (see Figure 33). As part of the rating, the rater records the number of occupants in the past year. These data are not used in the simulation; they are only collected to help the rater explain why predicted energy use may not match actual energy use. The assumption is that if there were fewer occupants than assumed, then the rating would overestimate energy use and vice versa (assumed number of occupants = number of bedrooms). Figure 34 tests this assumption by comparing accuracy to the agreement between the number of bedrooms and the number of occupants. Interestingly, this figure shows no relationship, i.e. even if there are more or less occupants than assumed, accuracy is unchanged. Thus none of the input factors that we analyzed (house age, rater, or assumed number of occupants) seemed to have a significant correlation to rating accuracy.

Figure 32. Colorado: House Age vs Accuracy

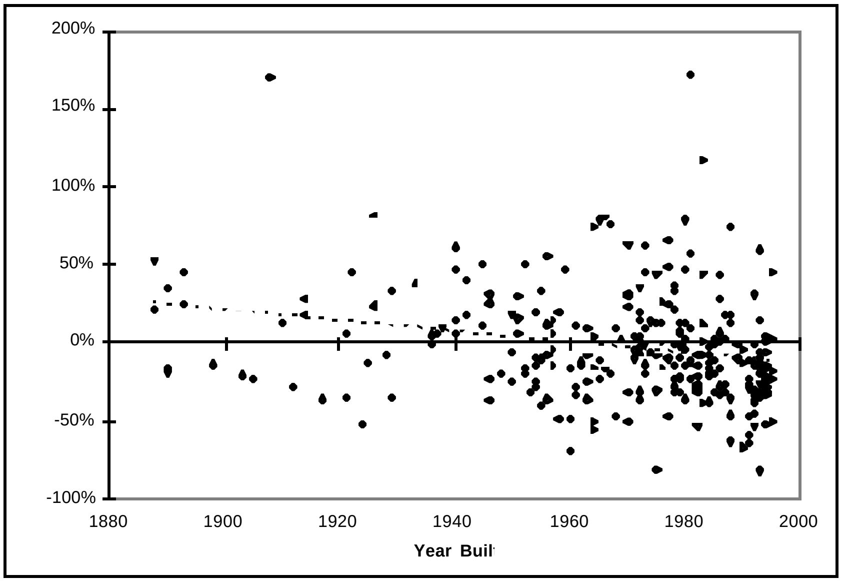


Figure 33. Colorado: Rater vs Accuracy

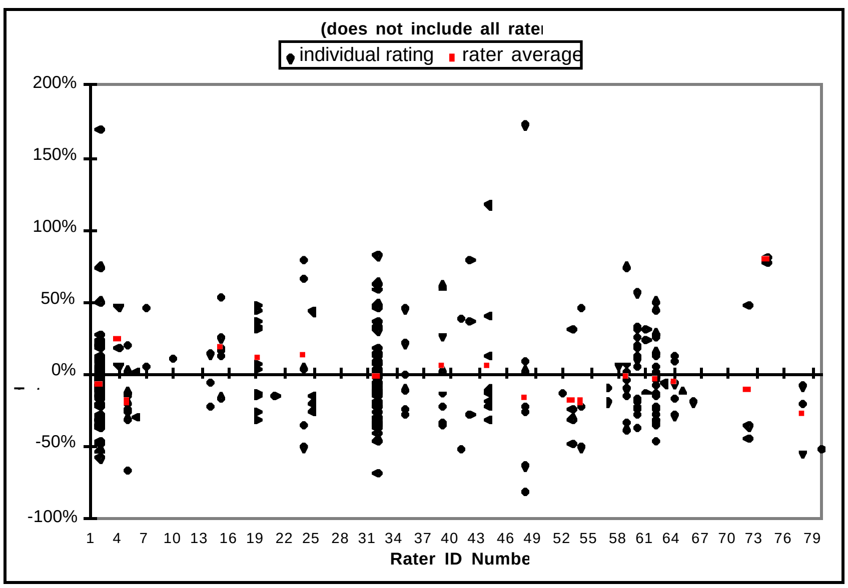
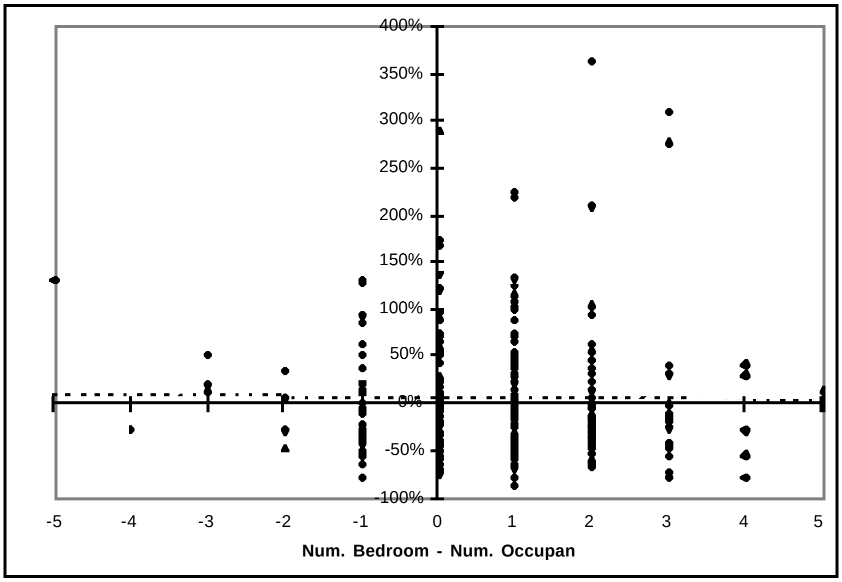


Figure 34. Colorado: Bedroom/Occupant Discrepancy vs Accuracy



Like CHEERS, there does not appear to be any clear connection between score and energy use. Figure 35 shows that as the score increases there is a slight decline in total site energy use. However, source energy use (which is closer than site energy to the consumer's energy cost) does not change with score. Figure 36 shows the same relationship as all the other HERS between score and error: as score increases error decreases. ERHC tends to overestimate energy use for low scoring houses, and to underestimate energy use for high scoring houses. A key difference between the trendline in Figure 36 and the one for the CHEERS ratings in Figure 23 is the intercept. The ERHC trendline crosses the axis (i.e. is most accurate) at a score of around 60 and the overestimates to the left of the intercept balance out the underestimates on the right. The CHEERS trendline intersects the axis at a rating score of around 90. Thus CHEERS appears to be calibrated to be most accurate for very high efficiency houses while the Colorado system is calibrated to be most accurate for medium efficiency homes.

Figure 35. Colorado: Score vs Site and Source Energy

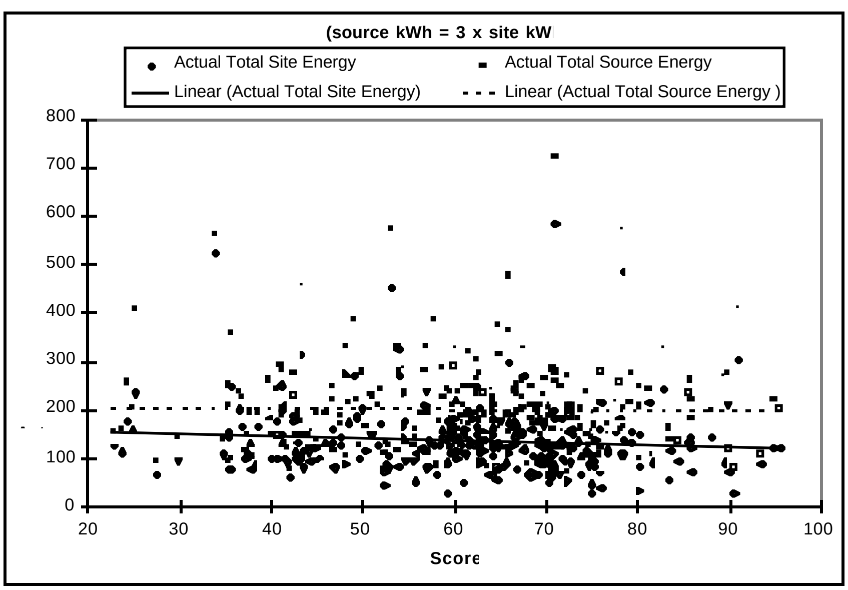
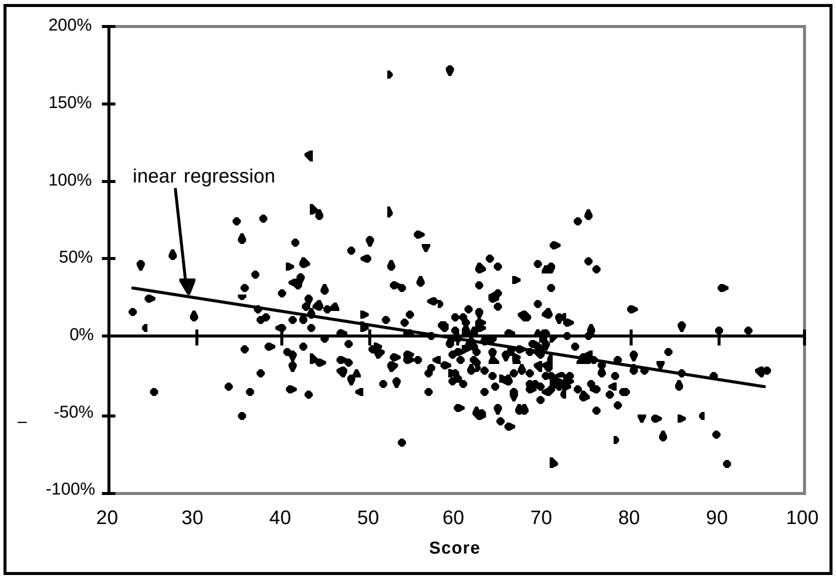


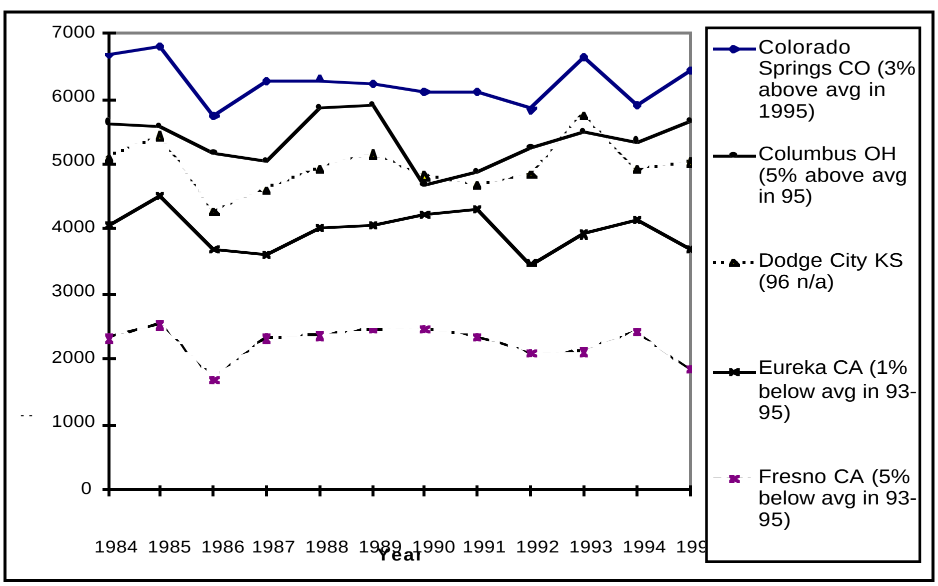
Figure 36. Colorado: Energy Use Error vs Rating Score



D. Weather in Case Study Locations

Actual weather data were collected in order to determine if the utility data were collected during unusually hot or cold years. Actual weather data is also useful for comparing the climates in the case study locations. Figure 37 shows the actual heating degree days per year from 1984 to 1995 for each of the case study locations. Since we do not know the exact locations within Ohio of the HERO homes, we have included Columbus weather data because it is roughly in the geographic center of the state. Similarly Colorado Springs, Colorado, and Dodge City, Kansas, were chosen to represent weather in those states. Figure 37 also shows that the weather was slightly colder than normal in 1995 (when utility billing data were collected) in Ohio and Colorado. Thus the actual weather might cause the ratings to slightly underestimate actual energy use. Conversely, the slightly warmer than average winters in Eureka and Fresno might cause those ratings to slightly overestimate energy use. Weather data were not available for 1996, the year of the Kansas billing data.

Figure 37. Annual Heating Degree Days for Case Study Locations



VII. How Accurate Do HERS Need to Be In Order to Be Successful?

Part of the issue here is how we define success. We could say that the automobile MPG ratings and home appliance Energy Guide Labels are a success because automakers and appliance manufacturers are required to put the stickers on their products and thus they have become “household words.” The crucial difference between these analogies and HERS is that while HERS proponents would love it if HERS ratings were required for all home sales, no one believes that this is a reasonable policy expectation. If HERS are to be successful, individuals and businesses must be willing to pay enough money for HERS ratings to support and expand the HERS infrastructure. Furthermore, HERS would not be considered successful if a large number of houses were rated but none of the HERS-dependent financial products, like Energy Improvement Mortgages, were utilized, or if no recommended improvements were implemented, or if builders were not marketing their homes with HERS labels.

A. Factors Affecting Success of HERS

Several experts who have written on the subject, and others that we have spoken with, suggest that success depends on several factors but addressing accuracy is generally not listed among them. This is not to say that they consider accuracy to be unimportant. They simply feel that HERS are sufficiently accurate and that further research and documentation would not particularly help. Factors commonly mentioned as keys to success include:

Consumer Education - Consumers must be educated about energy use and the cost-effective potential for improved comfort and reduced energy use. Most experts agree that the need for broad-based consumer marketing is the single most important factor in the success of HERS. If consumers are made aware of the existence and benefits of HERS, then they will use them. According to Vories (1991), energy ratings have failed to catch on because of a lack of visibility.

Housing Industry Education - Builders, lenders, and realtors must also be aware of HERS and convinced that recognizing and improving home efficiency can improve their business. Faesy (1988) compared the failure of a HERS in Connecticut with the success of a HERS in Vermont and concluded that the key to success was an extensive marketing, promotion, and education campaign that included education and training of real estate agents and lenders plus broad-based consumer marketing.

Better financial products - Energy-efficiency financing products must be more attractive to consumers than conventional ones and must not cause significant hassle or delay in the homebuying process (personal communication with Richard Faesy 12/95; Michael Holtz 12/95).

Credibility - Faesy (1988) and Vine et. al. (1988) point out the need for the public to perceive the HERS organization as knowledgeable, trustworthy, and helpful. However, this concern may have more to do with perceived conflicts of interest than with technical ability to perform accurate ratings. For example, consumers might be suspicious of a HERS that is sponsored by a utility company which is also motivated to sell more energy. Many see the development of national HERS guidelines by DOE as necessary for addressing credibility and conflicts of interest.

Packaging HERS with other programs - Vine et. al. (1988) found that HERS were most successful when promoted in connection with other energy efficiency programs such as free services for the correct sizing of air conditioning systems, subsidized weatherization materials, or low-cost loans.

B. Industry Perceptions of Accuracy

HERS providers generally do not believe accuracy is a concern for the consumer and thus it is not necessary to address it in marketing literature. According to Rebecca Vories, "accuracy is blown out of proportion. . . the issue is marketing" (personal communication 2/96). The only references we found to accuracy in marketing literature targeted at consumers, realtors, or weatherization program staff were general assurances that raters are "trained and experienced" and "certified", that the data is entered into a "sophisticated computer program" and that HERS "provide uniform, reliable, unbiased information" (CHEERS marketing literature; CA Assoc. of Realtors). Some of the HERS providers' output forms that the consumer receives did include a disclaimer indicating that energy cost estimates are based on average family use and that actual use will vary based on differences in lifestyle and annual weather patterns. However, some of the HERS output forms that we saw did not include any such disclaimers.

HERS providers report no complaints and no questions about accuracy from customers (personal communication with Carol Cales 2/97; Doug Swartz 3/97). According to Swartz, customers tend to focus less on the energy cost estimate and more on the rating score, which cannot be easily validated or contradicted by billing data.

While few of the HERS providers we contacted had or were willing to supply accuracy data, many were confident in the accuracy of their systems and were prepared to quantify the accuracy. Carol Cales, of Energy Rated Homes of Arkansas, recalled a study that found that cost estimates were within 10% of actual usage. Other HERS providers reported that they had performed informal comparisons between rated and actual cost data and found the ratings to be "very accurate" (personal communication with Steven Lowrie 2/97).

C. Risk Analysis

One way of assessing the degree of accuracy required is to look at the use of the rating and who is bearing risk as a result.

1. Lender Risk

In most cases, the primary lender is not bearing any risk because energy mortgages are usually resold on the secondary market to one of the government-affiliated agencies like Fannie Mae. Therefore, the secondary market lender is bearing the accuracy risk by offering EEMs and EIMs. For an EEM, the accuracy risk is that bills will not be as low as predicted and the homeowner will be stretched too thin and could default on the mortgage. The accuracy risk is similar for an EIM--the monthly savings turn out to be smaller than the increase in the monthly mortgage payments, the homeowner is short of cash and defaults. HERS providers often point to the fact that *on average* HERS can accurately predict cost savings as evidence that a bank with a portfolio of energy mortgages would be insulated from risk. However, if the variance is large and half the homes don't achieve expected savings then the bank might still be exposed to considerable risk.

It turns out, however, that there is almost no risk to the bank due to the accuracy of HERS. According to Horowitz (1996), EEMs do not do much to increase or decrease homebuyer risk of default. He contends that the only variable that research has ever shown to affect default rate is the borrower's loan-to-value ratio, not energy expenditures. For example, if the market value of a house drops below the balance remaining on the mortgage loan, then the homeowner is more likely to default. David Carey, of Fannie Mae, agrees that default rates are not affected by monthly energy savings (personal communication 2/96). According to Ron Judkoff, loan-to-income ratios have not been revised much in the last 40 years and are based on the monthly energy costs of inefficient houses from that era. However, homes are generally more efficient now, so homeowners usually can meet the mortgage even if expected energy savings do not materialize. Furthermore, even if savings do not materialize and homeowners are squeezed for cash, they are more likely to accept a lower level of comfort than to default on a mortgage. Thus there is very little risk to the lender from HERS accuracy.

Lenders, however, are very concerned about another source of risk related to HERS: the risk that the cost of retrofits will not be reflected in resale value. If a homebuyer includes the cost of recommended improvements in a mortgage and then defaults on the mortgage for an unrelated reason (e.g. he loses his job), will the bank be able to resell the house for a price that recovers its full investment (i.e. original sale price plus retrofit cost)? Thus EIMs may not increase the risk of default, but they can increase the severity of defaults (personal communication with David Carey 2/96). According to Horowitz (1996), there is strong evidence to suggest that housing markets can incorporate the value of energy efficiency into selling price if consumers are properly informed about efficiency features and energy costs. Thus, for lenders the key to success for HERS and EMs (energy mortgages) is not HERS accuracy but marketing and education so that housing markets can properly value energy efficiency. It also creates a bit of a chicken and egg problem, because consumer awareness will depend largely on the prevalence of these financial products, but lenders may be waiting for awareness to increase before promoting EMs. This is not to say that accuracy is irrelevant; if the market is to behave efficiently, it must have accurate information.

2. Consumer Risk

There are at least two types of accuracy risk that a consumer could be subject to: the risk that she will buy the wrong house as a result of inaccurate scores or cost estimates, and the risk that she will make uneconomic retrofit investments as a result of inaccurate recommendations. In the first case, it is not important for the scoring or estimates to be absolutely accurate but only relatively accurate, i.e. more efficient houses should get higher scores and lower cost estimates than less efficient houses. Since energy costs are usually quite low on a homebuyer's list of priorities, consumers are likely to use HERS scores and cost estimates in only a rough qualitative sense. Thus, unless scores or cost estimates are grossly inaccurate, consumers are probably not bearing much risk by using HERS rating to compare houses. Of course, strictly speaking, consumers should not use rating scores to compare different houses because scores are designed to rate a house's efficiency relative to code compliance, not its absolute efficiency or energy costs. Yet even if a consumer makes the common mistake of using scores to compare homes, she is probably not bearing that much risk, unless the houses use different fuel types for heating and hot water.

The risk of making an uneconomical investment depends on the accuracy of specific recommendations and depends on absolute accuracy for a particular occupant. Making recommendations is certainly the most technically difficult part of a HERS rating because of the many potential sources of error for a particular rating and for particular end-use recommendations. Even with this difficult objective, there is often no margin for error in a recommendation. An individual recommendation or package of recommendations is usually considered cost effective as long as the annual savings are greater than the extra annual mortgage cost. Furthermore, the HERS output forms that we have seen do not include any mention of the uncertainty of the calculations. Thus, the consumer is bearing the very real and perhaps considerable risk that he could end up making retrofit investments that are not cost effective. Uneconomical investment(s) could cause a consumers monthly costs (energy costs + mortgage payment) to go up instead of down, as expected. However, as Horowitz and others have noted, the consumer is generally not at any greater risk of defaulting on the mortgage. Of course, economic cost-effectiveness is only one criteria a consumer uses when making retrofit decisions. Non-economic benefits of retrofits include improved comfort, less noise, better aesthetics, and improved security.

In addition to a list of recommendations, HERS output forms typically give the consumer considerable economic information about each of the recommendations. For example the HERS form is likely to show useful life, annual savings, lifecycle savings, estimated cost, and simple payback period for each of the recommendations. A skeptical or risk averse consumer has the option to pick and choose the recommendations with more attractive economics and thereby mitigate the risk of a bad investment. For example, suppose a consumer decided only to invest in recommendations with a payback period of 3 years or less: she invests in a piece of equipment that has a useful life of 10 years and is expected to pay for itself in 3 years (which is a return on investment of 37%). Even if the savings were half of what the HERS predicted and it took 6 years to pay for itself, the investment would still yield a rate of return of 12%, which is far better than any bank savings account today.

It is interesting to note that it is entirely possible that a package of recommendations could be uneconomical and the consumer would never know and would be quite happy anyway. For example, suppose the rating for a particular house significantly overestimates the existing need for heating energy and estimates that "as is" annual energy costs are $2,000. The rating also estimates that a new furnace would save $800/yr. and would add only $600/year to the mortgage, for a net savings of $200/year and an annual energy bill of $1,200. In reality the "as is" energy cost is only $1,000 instead of $2,000. A homebuyer decides to buy the house and installs the recommended heating system. After a year the buyer is thrilled to find that his energy bill is only $600, which is half of what it was predicted to cost with the new equipment. In reality, however, the annual savings are only $400 compared to the increased mortgage cost of $600/year for a net loss of $200/year. While this may be an exaggerated example, it illustrates the point that a homebuyer who gets and EIM may never know if the investment was truly cost-effective.

3. Credibility Risk

Another form of risk is the risk that HERS will suffer serious and long term credibility problems if consumers, builders, lending institutions, funding agencies or other stakeholders conclude that HERS are significantly less accurate than they were led to believe. Other new energy efficiency technologies have suffered from this sort of backlash after encountering technical problems with initial deployments. For example, compact florescent light bulbs and solar water heaters both have had to overcome credibility problems caused by the failures of many of the earliest installations. If consumers determine that ratings are not as accurate as they expect, they may feel that they were mislead because of the lack of discussion of uncertainty in HERS literature. They may also be under the misperception that HERS are regulated by the government in the same way that other rating systems, such as automobile MPG, appliance energy labels, and food nutrition labels, are regulated.

VIII. Improving HERS Accuracy

A. Current Quality Assurance Strategies

There are a number of quality assurance procedures that are proposed in the HERS Council Guidelines. Some of these procedures are currently being used by HERS providers, but many procedures are not currently used. In addition to the HERS BESTEST discussed earlier, quality assurance focuses on two main areas:

Rater Training and Oversight - Raters must take a training course and pass written and practical exams before being certified. Once certified they are typically subject to one or more types of ongoing oversight including (1) random auditing of the raters work, (2) random re-inspection of rated homes by a QA Inspector, and (3) customer satisfaction surveys.

Calibration of Standard Assumptions - Most HERS providers start out with standard assumptions about occupant behavior. The HERS Council Guidelines, for example, specify standard assumptions to use for thermostat setpoints, internal gains, hot water use, plug loads, etc. However, after being in business for some period of time, many HERS providers try to use billing data and ratings to derive new occupant behavior assumptions that give better agreement between estimated and actual energy use. In some cases this has consisted of rather ad hoc "tweaking" based on small sample sizes, while in other cases it has been more systematic (personal communication with Doug Swartz 2/97; Lois Arena 3/97). A HERS provider may adjust one or more of the following variables: heating setpoint, cooling setpoint, thermostat setback amount and hours of setback, internal gains, gallons of hot water per bedroom, home operating profile, etc. A HERS provider may also simply apply a correction factor to fuel use or end-use estimates.

One of the potential pitfalls of this sort of calibration is that if the sample size is not sufficiently large, then the sample may not be representative of the entire population of houses in an area. Behavior that might be average for a subset may not be average for the whole group. Furthermore, even if tweaking one or more variables does improve the accuracy on average, it may not reduce the variance in accuracy. Another pitfall is that adjusting variables arbitrarily may improve accuracy of fuel use estimation or cost estimation but may hurt the accuracy of end-use estimates and specific recommendations. For example, suppose a HERS provider found that estimation accuracy could be improved by assuming less hot water use and assuming higher winter thermostat settings. The rating systems would end up recommending replacing a lot of furnaces and not replacing hot water heaters, when in reality the opposite might be more appropriate.

B. Potential Future Data Collection and Analysis

There are several types of data collection that would be useful for improving HERS accuracy.

Utility Data for Rated Homes - We have just begun to scratch the surface of the data that are potentially available for comparing ratings to utility data. Thousands of houses in dozens of states have received HERS ratings in the last several years. The data already exist, they simply must be gathered. Although it is difficult to determine the exact source of error in ratings from metered fuel use data, careful statistical analysis of data from a large, diverse sample of houses can yield valuable clues about sources of error. For example, comparing the accuracy of ratings that included blower door tests to similar ratings without blower door tests would help determine the usefulness of this added expense. Recognizing the need (for data) to evaluate and improve rating tool accuracy, the HERS Council Guidelines call for all HERS providers to maintain a database of HERS input and output data, utility bill releases and information on retrofits for at least 10% of homes rated annually, or 500 homes annually, which ever is less (HERS Council Guidelines Version 2.0).

Submetering - Long term submetering of specific energy end-uses such as heating, cooling, hot water, lighting, cooking, etc. would be extremely valuable for verifying accuracy of end-use predictions and for developing better occupant behavior assumptions and doing the sort of model calibration discussed above. Continuous monitoring of indoor and outdoor temperature can also be extremely valuable for determining actual thermostat setpoints. However, even with this type of monitoring, it may not be possible to isolate exact sources of error in particular ratings because inaccuracy in a particular end-use could still come from either occupant behavior or incorrect physical assumptions and simulation algorithms.

Submetering of Vacant Houses - Detailed submetering of a vacant house for a period of a few days can be a useful complement to the HERS BESTEST for assessing the "physics" of a HERS simulation tool. It can also be very useful for isolating rater error and determining the value of diagnostic testing and data collection versus physical assumptions. An obvious disadvantage of any type of submetering is the cost.

Pre- and Post-Retrofit Utility Data and Submetering - The accuracy of recommendations can be verified by comparing utility data before and after recommendations are installed in existing houses. One advantage of gathering this type of data is that it can be inexpensive because no additional metering equipment is needed. However, in order to isolate actual savings, it is important to account for variations in weather, new occupants, etc. Submetering can help identify savings and in some cases help tell if savings are being "taken back" in the form of increased comfort.

C. Possible HERS Modifications

While it is not entirely clear what role accuracy plays in HERS success, it is clear that HERS have not enjoyed the type of success that many in the field had hoped for or expected. There may be ways to modify the systems to improve accuracy, reduce the cost of doing ratings, and/or increase the usefulness of ratings.

1. Cost Based Scoring System

There has been considerable debate in the HERS community about what type of scoring system is both “fair” to constituencies like the electricity and gas industries and meaningful to the consumer (Nisson 7/96, CEC 1996). The current system proposed in the HERS Council Guidelines may be neither. A site energy-based system clearly favors more expensive fuels like electricity over cheaper ones like natural gas. It also has the potential to mislead consumers who expect scores to correlate to energy cost and are comparing houses that have different fuel types. An alternative scoring system would be one based on energy cost or energy cost per square foot. Since HERS already include an energy cost estimate, this alternative would basically mean doing away with the misleading score altogether and focusing on energy cost.

2. Code-Based Scoring

Another alternative to using points or stars would be to phrase the score in terms of a percentage that the house is above or below a specific energy code. For example, a house could be 10% above the CABO Model Energy Code or a furnace could be 10% below the NAECA Standard. This would avoid the common mistake of equating score with energy cost, and focus attention on what a score really is: a reference to a particular energy code. Consumers would realize that to understand their score they must understand the code. Phrasing scores in percentages above or below code is also a language that builders are familiar with from their past experience with energy codes. On the other hand, neither builders nor consumers are familiar with the nuances of the current scoring systems.

3. Hybrid Home-Occupant Rating

HERS are designed to rate the house not the occupant, so that a rating can be meaningful for any potential new occupant. However, a rating could be made more meaningful for both current and future occupants if there were some way to easily customize a rating for any potential occupant. This customization could involve various levels of sophistication depending on the amount of information gathered on the potential occupants. At the simplest level, a hybrid rating could use only the number of occupants and would not require re-simulation for each potential occupant. The estimated energy use could be expressed not as a single value but as a sliding scale as seen in this example:

Table 10. Sample Hybrid Home-Occupant Rating

	Number of People					
	1	2	3	4	5	6
Estimated Annual Energy Cost ($/yr.)	$1,000	$1,200	$1,400	$1,500	$1,600	$1,650

Dave Roberts, of Architectural Energy Corporation which authors HERS software, agrees that accuracy of predictions would improve with just the number of occupants (personal communication 2/97). Accuracy of recommendations would also improve since recommendations are based on estimated energy end-uses. Many HERS now assume that the number of occupants is equal to the number of bedrooms. Don Compton, of ENERGYPRO, L.L.C., a HERS provider, finds that HERS tend to be less accurate when the number of occupants is different from the number of bedrooms (personal communication 2/97).

A more sophisticated hybrid rating could include several occupant specific variables such as household income, house price, number of adults working at home, annual weeks of vacation, preferred temperature settings, etc. Determining which group of variables has the highest correlation to energy use would require statistical analysis of a large sample of houses for which detailed occupant data is available. While there will always be a certain randomness to occupant behavior, it is possible to reduce the uncertainty from occupant behavior. Ratings performed by the Home Energy Ratings of Ohio Foundation for existing or prospective occupants of a particular house now include the number of occupants and can include other occupant specific data such as setpoints. Ratings performed for builders or in other situations where there is no particular occupant in mind use standard occupant assumptions (personal communication with Sally Lukens 2/97).

4. Bill Reconciliation

Some HERS like ERH of Ohio and ENERGYPRO, which are primarily used for troubleshooting, hone their ratings with utility bill data. If a rating does not agree with billing data, the rater can adjust occupant related assumptions or assumptions about physical features. Steve Andrews, of ERH of Colorado, has described a procedure that could be used for reconciling billing data and rating output: Suppose a rating estimates that a house will use $500/yr. of natural gas and can save $100/yr. with a new furnace (20%). If the gas bill in the past has only been $400, then the estimated savings would be adjusted to 20% of $400 or $80 (personal communication 1/96). While this rating is now customized to the current occupant, it may also be more accurate for prospective occupants than without reconciliation. The lower-than-expected gas use could be due to occupant behavior but it could also be due to physical features that the rater was not able to account for (such as, better insulation than observed, or less infiltration than assumed).

AUDIT2000, a residential simulation program used in Canada, includes a reconciliation module. The rater can enter up to four years of monthly fuel use data into the program. The rater can then change some of the most error-sensitive parameters in the model, such as air infiltration, amount of supplemental heating and indoor temperature, and see the effect on month-by-month agreement between the model and actual energy use (NRCAN 1995b).

Utility bill reconciliation can be a valuable reality checking or quality assurance technique. Poor agreement could mean the occupants were very atypical but it could also indicate that the rater made a mistake collecting or entering data. For example, a HERS provider could institute a policy that if a rating under or overestimates actual energy use by more than 50%, then the rater must do more investigation and re-simulate. Supplemental diagnostics testing (such as blower door tests) could be triggered if the accuracy falls outside of a specified acceptable range.

Thus, utility bill reconciliation can serve a number of useful functions. Utility bills can be used to (1) improve the accuracy of a rating for the current occupant, (2) be a reality check and quality assurance technique, and (3) provide useful information about house specific data such as infiltration rate.

5. Accuracy Disclaimers

One modification that might cause confusion in the short-run, but is probably prudent in the long-run, is to give the consumer more information about the accuracy of ratings and the potential sources of error. For example, the rating forms probably should provide more explanation of the scoring system and the fact that a score is not necessarily related to energy costs and should not be used to compare houses. The consumer could also be given more information about the variability in estimating energy use or a list of some of the potential sources of error in recommendations. Some HERS providers have decided that discussing accuracy with consumers only introduces questions about accuracy that may not have existed. For this reason, they have sought not to compare ratings to utility bills, or at least not to discuss such comparisons with the consumer (personal communication with Vadim Belotserkovsky of NRCAN 3/97). The risk of alienating consumers must be weighed against potential backlash from consumers, builders and others who may feel they were misled if they find out later about the uncertainty in ratings. Furthermore, more open disclosure of accuracy information may address head-on the concerns of consumers, lenders, and others who are skeptical or distrustful of HERS.

6. Prescriptive System

A more radical modification that has the potential to greatly reduce the cost of a rating without compromising accuracy is to switch to a prescriptive rating system rather than a simulation-based system. One of the barriers that HERS face is the amount of training required to be a rater and the amount of information that a rater must collect and enter into the software. It may be possible, however, to extrapolate energy use from a smaller number of variables. Less training would be required and it would take less time to do a rating. The question is does such a set of variables exist?

Detailed statistical analysis of energy use data, house characteristics, and perhaps occupant characteristics from a very large sample of houses could yield a regression equation for predicting energy use based on variables like floor area, typical wall insulation, window type, etc. A prescriptive system could be used to produce rating scores, energy and cost estimates and even rules of thumb for cost-effective recommendations that are about as accurate as HERS simulation-based recommendations, at lower cost. For example, analysis of actual energy use data from a large sample of homes could show that for a house in a certain climate, with a certain size, insulation level, basement type, and window type, it is cost effective to replace an old furnace with a new unit. A prescriptive system would also be able to account for the “takeback effect” because, rather than using a single occupant profile, it would be able to incorporate data from the actual occupant profiles of thousands of houses with similar size, cost, efficiency and other features as any house being rated. Research may eventually show that a higher degree of accuracy can be achieved at a lower cost with a prescriptive system that includes some occupant-specific data and blower door testing.

IX. Conclusions

Accuracy is important.

HERS experts do not rank accuracy high on the list of keys to HERS success, and do not report many questions or complaints about accuracy from consumers. In response to this perception, HERS providers choose not to discuss accuracy in marketing literature or on the rating forms themselves. Indeed, it appears that lending institutions that participate in Energy Efficiency Mortgages and other HERS-related energy efficiency financing products are not exposed to any significant risk related to HERS accuracy because the agreement between expected and actual energy bills does not appear to affect the probability that a homeowner will default on a mortgage.

Homeowners, however, are bearing at least some risk related to HERS accuracy. One risk is that a homeowner will incorrectly conclude that one home is more energy efficient than another based on incorrect HERS scores. Another risk is of making an uneconomic investment based on an inaccurate recommendation. This risk is real and can be significant.

Furthermore, lack of data demonstrating accuracy is probably impeding the growth and acceptance of HERS amongst certain consumers, lenders, and other groups nationwide. Furthermore, a lack of accuracy may eventually catch up with some HERS and cause irreparable credibility problems, which could spread to all HERS

In order to achieve their objectives, HERS need to be both *relatively* accurate (e.g. scores should rank houses in the correct order) and *absolutely* accurate (i.e. so there is a high probability that recommended improvements will be cost effective for individual consumers). Furthermore, accuracy is important both on average and for individual ratings. Low average error may mean that a rating system has been well calibrated for a location and population but high individual errors can diminish the usefulness of ratings to consumers. Finally, it is most important that estimates are accurate at the end-use level (heating, cooling, and hot water) because these estimates form the basis of the cost-effective recommendations.

HERS can be highly inaccurate.

Accuracy of Energy Use/Cost Estimates

Estimated energy use can be considerably different from actual energy use both for a single house and on average for a large group of houses. This is not surprising given (1) the many potential sources of error a HERS faces in meeting its stated objectives of estimating energy use assuming standard occupant behavior, and (2) the potentially large variation in actual energy use introduced by actual occupants. However, for some HERS, the average error and some of the individual errors in energy estimates may be alarmingly high and it is probably possible to reduce both the mean and variance of such errors.

For the 193 ratings performed with CHEERS Version I software that we examined, the ratings tended to significantly overestimate both gas use and electricity use, and therefore total energy cost. For about 20% of the houses, overestimation was greater than 100%. Other HERS programs that we studied had much smaller average error and somewhat smaller variance in error, but the variance may still be cause for some concern. For example, for 276 ratings from Energy Rated Homes of Colorado, the average error in site energy use was only -3% and the standard deviation was 35%.

It is important to keep in mind, however, that direct comparison of the results from the different case studies we examined is like comparing apples and oranges because of clear differences in the sample populations. One of the most significant differences between the sample groups is the severity of weather which affects the ease of prediction. The two California climates have much milder winters than the other locations, and the Colorado location clearly has the most severe winter climate. As Petersen (1994) and others have found, and as our results appear to substantiate, it is harder to estimate energy use in mild climates than in more severe climates. Average age of the sample populations is also significant. The Kansas homes, for example, were almost all new ones, while the CHEERS houses were significantly older than the other groups. Old houses tend to be harder to estimate than new ones. The CHEERS ratings were also the only ones that did not include blower door testing. Thus they probably cost less to conduct but are also less accurate. Furthermore, the CHEERS ratings were also all performed in 1994 compared with 1996 for most of the other ratings. Significant progress in ratings systems has been made in that period. For example, the CHEERS ratings were conducted using a DOS-based program that has since been replaced with a more user-friendly Windows version.

Accuracy of Scores

While it is not stated explicitly by HERS providers, many people expect and assume that houses with higher rating scores have lower energy costs. However, none of the HERS we examined showed any clear relationship between overall rating score and actual energy use or cost. Even when compared to houses of the same size, ones with higher scores did not tend to use any less energy than ones with lower scores. The CHEERS houses with higher scores, for example, were predicted to use significantly less energy than the ones with lower scores, but in fact the higher scorers used more energy. One possible explanation is the “takeback effect” which says that the higher scoring houses are indeed more efficient and would use less energy if they were operating in the same manner as lower scoring houses but they are not operated in the same manner. Occupants of more efficient/higher scoring houses are likely to be more affluent, to have more appliances, and to choose more comfortable heating and cooling setpoints. Thus they “takeback” some of the expected savings in higher levels of service.

Accuracy of Recommendations

Making accurate recommendations is the most difficult objective of a HERS; it is also the most difficult objective to validate. One way to gauge the accuracy of recommendations that was available to us was to compare the actual energy use of the CHEERS homes to the predicted total energy savings if the occupants implemented all the recommendations. Since many of the ratings predicted that it was possible to save over 50%, and in some cases over 100%, of the current consumption, it is likely that at least some of the recommendations would not be cost-effective. The minimum requirement for cost-effectiveness is generally only a positive expected cash flow (monthly financing cost < monthly savings), i.e. there does not have to be any safety factor built into a recommendation.

HERS can be improved.

One way to improve accuracy is to collect more utility billing data on some of the thousands of homes that have been rated with HERS in the last several years. A tremendous amount of very detailed data now exists and simply must be gathered. As this research project has demonstrated, this sort of analysis can be fairly inexpensive to perform and can yield valuable information. While utility billing data for rated homes cannot pinpoint specific sources of error in ratings, they can yield valuable clues for improving HERS. For example, the fact that CHEERS overestimated gas use more in Eureka (which is colder in the winter) than in Fresno, and the fact that CHEERS overestimated electricity use more in Fresno, tends to indicate that CHEERS may be using incorrect heating and cooling setpoints or infiltration rates, or conduction rates, etc. Another important trend found in the CHEERS data is that some raters tend to produce more accurate ratings than other raters, which emphasizes the need for rater training, oversight, retraining, and the need to minimize rater judgment calls in rating procedures.

In addition to troubleshooting rating systems, utility billing data is extremely valuable for calibrating HERS for a particular climate and population. However, there is also a danger that calibration based on utility bills can lead to a false sense of accuracy because the accuracy of total energy use or total cost estimates does not necessarily indicate that all end-use estimates are also accurate. For example, end-use submetering of a large sample of houses in Florida rated with BEERS showed that total energy use estimation was fairly accurate but that some end-uses, such as heating and hot water, were overestimated, while others, such as pool energy, were underestimated. Calibrating a rating system by arbitrarily adjusting all end-uses or only certain ones may improve estimation accuracy but not necessarily end-use accuracy. End-use accuracy is critical for making accurate recommendations of cost-effective improvements. Other forms of research such as submetering of energy end-uses, software-to-software comparisons, and pre-/post-retrofit analyses are also needed.

Based on the wealth of knowledge that can be gained by examining utility bills of rated homes, HERS providers may want to consider some fundamental modifications in HERS. For example, it may be possible to greatly improve HERS accuracy by incorporating a few key pieces of information about the current or prospective occupants into a rating. Occupant-specific input could be as simple as the number of occupants or could include other characteristics such as preferred temperature settings, appliances owned, annual weeks of vacation, hours at home, etc. Another modification that has the potential to greatly reduce the cost of performing ratings without compromising accuracy is to switch to a prescriptive rating system rather than a simulation based system. Detailed statistical analysis of past ratings and utility bills could yield a regression equation that allows accurate prediction of energy costs based on a much smaller number of variables than are now collected for ratings. Such a system has the ability to account for the "takeback effect" and could be used to generate scores and recommendations as well.

Occupant behavior and climate limit the degree of accuracy that can be achieved.

Random variation in occupant behavior means there will always be some degree of uncertainty, or variance, in ratings. Furthermore, since there appears to be a relationship between a home's energy efficiency level and occupant behavior (the "takeback effect"), it appears that a rating system cannot be calibrated to have close to zero average error for all types of houses at the same time. The takeback effect is evidenced by a clearly downward sloping regression line when estimation accuracy is plotted against rating score (see Figure 1). This phenomenon was common to all the rating systems studied. It appears that it is possible to adjust the point where this trendline intersects the 0% error axis by calibrating a rating system with actual energy use data, but it is not possible to flatten out the slope of this trendline so that the average error is close to zero for both high scoring homes and for low scoring ones. Furthermore, the slope of this trendline may be an inevitable function of climate severity. The percentage difference in energy use between the energy “misers” and the energy “hogs” in a mild climate is greater than the difference in a severe climate. Indeed, California has the mildest climate and the steepest trendline, while Colorado has the most severe climate and the flattest trendline.

There is a need for disclaimers.

In order to avoid giving a false sense of accuracy, HERS providers should be more forthcoming about the technical limitations of HERS and what they can and cannot reasonably be expected to provide. For example, the rating forms probably should provide more explanation of the scoring system and the fact that a score is not necessarily related to energy costs and should not be used to compare houses. The consumer could also be given more information about the variability in estimating energy use or a list of some of the potential sources of error in recommendations. For example, instead of estimating that a house will use $1,523 of energy per year, a rating could say something to effect of, “based on the uncertainty in the rating process, we estimate with 80% confidence that this house will use between $1,300 and $1,700 of energy, if operated in a typical manner. However, variation in actual behavior and weather can result in considerably more or less energy use.”

HERS are valuable.

While this research has focused on potential technical problems and their significance and solutions, it is important not to loose sight of the benefits of Home Energy Rating Systems. HERS can play a very valuable role in transforming the market for residential energy efficiency technology by providing information and facilitating financing mechanisms. Furthermore, while some of the data gathered here have demonstrated that HERS can be highly inaccurate, much of the data also indicate that HERS can be quite accurate. The CHEERS ratings for newer homes and the Ohio, Colorado, and Kansas ratings all demonstrated that low average estimation error is possible to achieve.

It is also important to recognize that many recommended improvements have very attractive paybacks so that even if a rating is highly inaccurate for a particular customer, many of the recommendations are still likely to be cost-effective. For example, suppose a rating recommended a hot water tank blanket with a payback of 1 year. Even if the rating overestimated hot water use for that particular house and consumer by 200%, the tank blanket would still pay for itself in 3 years and the consumer would most likely come out ahead. Many recommended improvements also provide intangible benefits such as increased comfort, reduced noise, better security, and improved aesthetics. Thus a consumer may be more interested in the HERS’ ability to identify such improvements and facilitate their installation and financing, than in the actual economics of the measures. Furthermore, HERS ratings usually give the customer detailed economic information about each of the recommendations, such as simple payback period. A skeptical or risk averse consumer has the option to pick and choose recommendations with more attractive economics and thereby mitigate the risk of a bad investment.

X. Bibliography

"The Analyst as a Factor in the Prediction of Energy Consumption." *Proceeding of the 2nd International C.I.B. Symposium on Energy in the Built Environment*. Denmark. May 28-June 1, 1979.

AHFC. 1994. "Cost-Effectiveness Analysis of the Home Energy Rebate Program and the Low-Income Weatherization Program." prepared for Alaska Housing Finance Corporation, Affordable Housing and Energy Efficiency. May 9, 1994.

Alliance to Save Energy. 1996. "Energy Efficiency Financing: A Lender's Guide for Taking Advantage of the Emerging Market." Washington, D.C.: Alliance to Save Energy.

Butterfield, K. 1988. "How Reliable are Energy Audits?--Settings a Benchmark." *Home Energy Magazine*. March/April 1988.

California Association of Realtors. 1995. "Energy Efficient Mortgages: The Wave of the Future."

CEC. 1990. "Occupancy Patterns & Energy Consumption in New California Homes (1984-1988)." P400-90-009. September 1990. Sacramento, CA: California Energy Commission.

CEC. 1996. "Home Energy Rating System Technical Manual DRAFT." California Energy Commission. October 15, 1996.

CHEERS. 1991. "California Home Energy Efficiency Rating System, Inc. Research Evaluation and Design."

CHEERS. 1994. "Quality Assurance Program Summary."

CHEERS. 1995. "Rater Training Manual."

CHEERS. 1996. "CHEERS RateTool for Windows 2.0"

Collins, N. E., B. Farhar, R. Walsh. 1996. "Linking Home Energy Rating Systems with Energy-Efficiency Financing: National and State Programs." *Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings*. Vol. 2. pp. 35-46. Washington. D.C.: American Council for an Energy-Efficient Economy.

Collins, N., B. Farhar, W. Babiuch, J. Eckert. 1994. "A Plan for Evaluating Alternative Approaches to Financing Energy Improvements in Housing" NREL/TP-461-6688 (draft). August 1994. Golden, CO: National Renewable Energy Laboratory.

Faesy, A. R. 1988. "Case Study in Success: Vermont's Energy Rated Homes." *Home Energy Magazine*. Nov/Dec 1988.

Fairey, P. 1997. "BEERS: How Accurate?" presented at HERSC-FREE '97 by Philip Fairey of Florida Solar Energy Center.

Farhar, B., J. Eckert. 1993. "Energy-Efficient Mortgages and Home Energy Rating Systems: A Report on the Nation's Progress." NREL/TP-461-5478. September 1993. Golden, CO: National Renewable Energy Laboratory.

Farhar, B., N. Collins, R. W. Walsh. 1996. “Linking Home Energy Rating Systems with Energy Efficiency Financing: Progress on National and Sate Programs.” NREL/TP-460-21322. October 1996. Golden, CO: National Renewable Energy Laboratory.

Haltrecht, D. and R. Sinha. 1995. “HOT2000 Version 7 and AUDIT2000.” Ottawa, Ontario: CANMET, Department of Natural Resources.

Hammon, R., M. Modera. 1996. “Improving the Energy Efficiency of Air Distribution Systems in New California Homes.” *Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings*. Vol. 2. pp. 85-95. Washington. D.C.: American Council for an Energy-Efficient Economy.

Helke, G. A., F. Conti, B. Daniotti, R. J. Peckham. 1990. “A Detailed Comparison of Energy Audits Carried Out by Four Separate Companies on the Same Set of Buildings.” *Energy and Buildings* 14(1990) 153-164.

Hendrickson, P. 1988. “Home Energy Rating Systems: Information to Increase Efficiency.” *Home Energy Magazine*. Nov/Dec 1988.

HERSC. 1995. “Guidelines for Uniformity: Voluntary Procedures for Home Energy Ratings, Version I.” Washington, D.C.: Home Energy Rating System Council.

HERSC. 1996. “Guidelines for Uniformity: Voluntary Procedures for Home Energy Ratings, Version II.” Washington, D.C.: Home Energy Rating System Council.

HERSC. 1996. “HERS Projected Energy Use v. Actual Utility Bill Data.” Washington, D.C.: Home Energy Rating System Council.

Horowitz, M. 1996. “Innovative Financing of Home Efficiency Improvements.” *Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings*. Vol. 2. pp. 97-105. Washington. D.C.: American Council for an Energy-Efficient Economy.

Judkoff, R., J. Neymark. 1995. “A Procedure for Testing the Ability of Whole Building Energy Simulation Programs to Thermally Model the Building Fabric.” *Journal of Solar Energy Engineering*. Vol. 117. February 1995.

Judkoff, R., J. Neymark. 1995. “Home Energy Rating System Building Energy Simulation Test (HERS BESTEST): Volume I & II, Tier 1 and Tier 2 Tests User’s Manual.” NREL/TP-472-7332. November 1995. Golden, CO: National Renewable Energy Laboratory.

Judkoff, R., J. Neymark. 1995. “International Energy Agency Building Energy Simulation Test (BESTEST) and Diagnostic Method.” NREL/TP-472-6231. February 1995. Golden, CO: National Renewable Energy Laboratory.

Luboff, J. 1995 “Making Energy Mortgages Work.” *Home Energy Magazine*. May/June 1995.

Nadel, S. 1993. “The Takeback Effect: Fact or Fiction?” *Proceedings of the 1993 Energy Program Evaluation Conference*. Chicago, IL.: National Energy Program Evaluation Conference.

Nadel, S. and K. Keating. 1993. “Engineering Estimates vs. Impact Evaluation Results: How Do They Compare and Why?” *Proceedings of the 1993 Energy Program Evaluation Conference*. Chicago, IL.: National Energy Program Evaluation Conference.

NASEO. 1996. “Home Energy Ratings: A Primer for Weatherization Programs.” prepared by Residential Energy Services Network under contract to the National Association of Sate Energy Officials (NASEO).

Nisson, J.D.N. 1996. “Gas Industry Continues to Knock DOE’s Home Energy Rating System Guidelines” *Energy Design Update*. Vol. 16, No. 7. July 1996.

Nisson, J.D.N. 1996. “The Accuracy of Building Energy Simulation Software?” *Energy Design Update.* Vol. 16, No. 4. April 1996.

NRCan. 1995. “Efficiency Housing Database Development: Field Energy Audit of Existing Dwellings--Phase I Report for the Alberta Homes.” April 1995. Ottawa, Ontario: Energy Technology Branch/CANMET, Department of Natural Resources.

NRCan. 1995. “AUDIT 2000 Training and Reference Manual.” April 1995. Ottawa, Ontario: Energy Technology Branch/CANMET, Department of Natural Resources.

NREL. 1992. “Going National with HERS and EEMs: Issues and Impacts, The Collected Papers of the National Collaborative.” NREL/TP-261-4706. March 1992. Golden, CO.: National Renewable Energy Laboratory.

Pettersen, T. D. 1994. “Variation of Energy Consumption in Dwellings Due to Climate, Building and Inhabitants.” *Energy and Buildings*. No. 21. pp. 209-218.

Prindle, B. 1988. “Testing the Energy-Efficient Mortgage Program in Connecticut.” *Home Energy Magazine*. Nov/Dec 1988.

RESNET Notes. August 16, 1996. “Introducing Energy Wise Homes of Illinois, Illinois Home Energy Rating System.” National Residential Energy Services Network. http://www.natresnet.org.

RESNET Notes. Sept 27, 1996. “Energy Rated Homes of Colorado Launches E-Star Builder Program.” National Residential Energy Services Network. http://www.natresnet.org.

Sonderegger, R. C. 1977/78. “Movers and Stayers: The Resident’s Contribution to Variation across Houses in Energy Consumption for Space Heating.” in *Energy and Buildings*. pp. 313-324.

Vine, E. 1995. “Utility Residential New Construction Programs: Going Beyond the Code.” LBL-36603 UC-1322. August 1995. Berkeley, CA: Lawrence Berkeley National Laboratory.

Vine, E., B. K. Barnes, R. Ritschard. 1988. “Implementing Home Energy Rating Systems.” *Energy*. Vol. 13. No. 5. pp. 401-411.

Vories, R. 1991. “Making HERS a Household Word.” *Home Energy Magazine*. Sept/Oct 1991.

XI. Appendices

A. HERS Data Sources

Org./Contact
ERH Alaska, Barbara Collins, D.J., Steve Baden
Arkansas, Carol Cales
Rated Energy Homes Plus (CA), Ray Nelson 800-890-7929
ERH of Colorado, Lois Arena, 303-297-7395, 7325, Steve Andrews
Illinois, John Marley, 217-785-2007
Argonne Lab, Jim Cavallo, 630-252-8688
Paul Knight, own company 708-386-0345
ERH of Indiana, Mark Jansen, 317-232-8948
Iowa, Claude Papesh, 515-752-7162
Kansas Energy Star, Stan Butler, 913-296-2686
Midwest Energy, Brian Dreiling, 913-625-3437
Maine Hero, Wes Riley
Michigan HERS, John Sarver, 517-334-7234
Mississippi ERH, Linda Perry, 601-981-6699
Louisiana ERH, Ernie Singleton, 504-342-3825
New Hampshire HERO, George Hunton, 603-271-2611
David Weitz, ConServe Group, consultant to NE Utils, 603-271-2611
NE Utils, Buck Taylor, 860-665-2719
Public Service New Hampshire, Mike McQueeny, 603-634-2287
Louise Bergeron, Southern New Hampshire Services Inc (provider) 603-668-8010, Dan Gerard
HER of Ohio, Sally Lukens, 614-538-0115, Ruth
Oregon ERH, Alan Kramer, 503-986-2092
Rhode Island RISE, Debi Curry
Utah ERH, David Wilson, 801-765-0034
Vermont ERH, Richard Faesy, 802-865-3926 x 16
VT Gas Systems, Michael Russom, 802-863-4511
Arizona, Charlie Gohman
Wisconsin, Ed Carroll
HERS Council, Cynthia Gardstein
Dexter Akers, Community Concepts (HERS provider) 207-743-7716
ENERGYPRO, Don Compton, 316-264-0000
Good Cents (Southern Development, GA), Debbie Boyd, 770-821-3424
American Electric Power (util+provider), Bud Heiss, 614-223-2752
Alaska Housing Finance Corp., Phil Kaluza, 907-564-9256
National HERO, Steven Lowrie, 804-560-9134
Architectural Energy Corp (RemRate) Michael Holtz, Dave Roberts
FSEC, Phil Fairey
Volt VIEWtech, Tom Pape (SF)
Volt VIEWtech, Ed Thomas (Anaheim)
Doug Swartz, Ft. Collins Power & Light

B. CHEERS Data Filtering

The electric billing data for the CHEERS case study houses were found to have a number of problems, therefore the following filtering process was used to remove, or correct for, bad data:

1) The file from PG&E had fields for 3 years of data for 193 houses or 579 house-yrs of data

2) 2 houses with no data in any of the 3 years were removed leaving 191 houses

3) 57 house-yrs (all in 1994) were found to be corrupted (field order was scrambled) and were removed

4) 19 house-yrs had either no data or less than 300 days of data. These were removed.

At this point there were 191 houses and 497 house-yrs of data remaining. Since many house-yrs had less than 365 days of data and many of the reading periods were considerably more or less than 30 days, the monthly kWh readings were normalized to 30.42 days.

5) 4 house-yrs had less than 30 kWh/month for at least 3 months in a row. It is assumed that these houses are either unoccupied or the data is corrupted. These house-yrs were removed (0.8% of the house-yrs).

6) 40 months (out of 5916 months of data) had either no data or the reading was less than 75 kWh/mo. and was much lower than the average for that house-yr. It was assumed that there was an error in the data for these months. The values for these months were changed to the average of the adjacent months (0.7% of the data).

The natural gas billing data also had some problems. The following procedure was used to eliminate or correct for bad or missing data.

1) The file from PG&E had fields for 3 years of data for 193 houses or 579 house-yrs of data.

2) 6 houses had no data for any year, leaving 187 houses. (Houses #49, 50, 113, 149, 195, 217). Two of these houses (#49, 217) have electric DHW and propane furnaces but do have fuel ratings from CHEERS. We do not know if they have gas cooking.

3) 29 house-years had less than 300 days of data. These were removed.

4) 3 house-yrs had less than 10 therms for at least two months in Dec., Jan, and Feb. This data is considered bad and these house-years were removed (approx. 0.5% of the data). There are a number of possible explanations for why this data could be bad including improper meter reading, extended vacancy due to change in ownership, and errors in data handling before we received the data.

5) 3 house-yrs had less than 10 therms/mo. for 1 month in Dec, Jan, or Feb. These monthly values are considered erroneous and were changed to the average of the surrounding months (less than 0.1% of the data). It is important to note however, that there were many house-years with suspiciously low gas readings in months other than Dec, Jan, and Feb. For example, many houses with gas DHW had 0 or 1 therm for several months in a row in the summer period. These data were left as is.

6) 33 months of data were missing from various house-years. These months were given the average gas use of the neighboring months.

After processing the utility bill data for 193 houses, we end up with gas and electric data for 186 houses, electric only data for 5 houses, and gas only data for 1 house.