Pergamon

Energy 25 (2000) 339–354

ENERGY

www.elsevier.com/locate/energy

Accuracy of home energy rating systems

Jeff Ross Stein, Alan Meier *

Lawrence Berkeley National Laboratory, Building 90, Room 2000, Berkeley, CA 94720, USA

Received 11 January 1999

Abstract

We compared Home Energy Rating Systems (HERS) ratings and actual utility billing data for about 500 houses in four states. We found that HERS can, on average, predict annual energy cost accurately. However, on an individual house basis the agreement between predicted energy cost and actual energy cost was often poor, especially for older houses. Discrepancies between predicted and actual energy use have important implications for the true cost-effectiveness of HERS-recommended improvements. There was no clear relationship between rating score and actual energy cost. Given these results, HERS providers need to give consumers more information about accuracy and how to interpret ratings. They should also take greater advantage of the thousands of ratings that have been conducted in order to assess and improve HERS' ability to predict energy use.

1. Introduction

Approximately 20% of all the energy consumed in the United States is consumed by the residential sector. Much of this energy can now be saved cost-effectively 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 [1].

A HERS is a computer-simulation-based method for assessing a home's predicted energy use under standard conditions and its potential for improvement. There are dozens of private and state-government-affiliated organizations that provide HERS ratings. Raters are typically not

* Corresponding author. Fax: +1-510-486-4673.
E-mail address: akmeier@lbl.gov (A. Meier).

PII: S0360-5442(99)00072-9

employed by the HERS provider but rather they are usually independent energy consultants, mechanical contractors, etc. Most HERS providers buy or lease their simulation software from one of two private software firms. Japan, Germany and Canada are all considering adopting HERS similar to the US model.

HERS typically generate three types of output:

- *Rating score* — rating scores are usually on a scale to 0 to 100 points or one to five stars. The score is based on a comparison between the rated house and a reference house that meets a desired energy code or standard but is tailored to the same dimensions, equipment type and fuel mix as the rated house. Rating scores have many direct applications, such as allowing the buyer of a high-scoring home to qualify for a larger mortgage.
- *Energy use/cost predictions* — HERS make energy use and energy cost predictions for specific end-uses like heating and hot water and for the whole house. Unlike scores, which are *relative* to a reference house (i.e., a house's own potential given its shape, fuel mix, etc.), predictions are *absolute* measures and can be used to compare houses in the same way that miles-per-gallon (MPG) ratings are used to compare cars.
- *Recommendations* — HERS produce a list of recommended improvements that are calculated to be cost-effective on a life-cycle basis for a specific house, such as adding attic insulation or replacing old heating or cooling equipment. HERS-recommended improvements can be financed in an Energy Improvement Mortgage [2].

1.1. Problem statement

HERS and related energy-efficiency financing products have been available in the USA since the late 1970s; nevertheless, they have only penetrated a small share of the potential market. While predictive ability of rating systems is generally not considered by HERS experts to be one of the most significant barriers right now to widespread use of HERS, all agree that predictive ability is important for credibility and success in the long run and that research must be conducted to assess and improve predictive ability. To date, however, very little research has been done on the subject and almost no data have been made publicly available. Therefore, we sought to assess HERS predictive ability by comparing home energy ratings with actual utility billing data.

1.2. Sources of difference

There are many potential sources of difference that can account for HERS discrepancies. The potential sources of difference can be divided into *system differences* (i.e., discrepancies that can be addressed by improving the HERS) and *natural uncertainty* (i.e., differences beyond the control of the rating system). These sources are summarized in Table 1 and described below.

1.2.1. Rater inspection

Lack of adequate training or experience or lack of diligence in conducting inspections could cause a rater to overlook 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.

Table 1
Factors affecting accuracy of HERS output

Potential sources of discrepancy	HERS output type		
	Rating score	Energy cost prediction	Recommended improvements
System differences			
Rater inspection	×	×	×
Default values (permanent features)	×	×	×
Default values (occupant behavior)	×	×	×
Simulation algorithms	×	×	×
Energy prices		×	×
Default values (recommended features)			×
Equipment remaining life			×
Discount rate			×
Natural uncertainty			
Actual occupant behavior		×	×
Actual weather		×	

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.

1.2.2. 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 differences introduced by such physical assumptions is to collect more actual data. For example, a HERS provider in Kansas 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.

1.2.3. Default values (occupant behavior)

Since ratings are designed to rate the house and not the occupants, a HERS must rely on default values for the number of occupants, thermostat settings, hot water usage, appliance usage, etc. In addition to assumptions of how the permanent space conditioning and water heating equipment is used, HERS must also estimate all ‘other’ energy uses in order to estimate total energy use and energy cost. Thus, assumptions are made explicitly or implicitly about the existence and energy use of non-permanent features such as space heaters, portable air conditioners, refrigerators, TVs, non-hardwired lights, etc. Incorrect assumptions about typical behavior can result in large average differences between a HERS prediction and a specific house.

1.2.4. Simulation algorithms

The software must be able to simulate accurately 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 predictive ability.

1.2.5. Energy prices

A HERS must use correct current utility rates and must also forecast, explicitly or implicitly, the cost of energy over the payback period of the recommended equipment.

1.2.6. Default values (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.

1.2.7. Remaining equipment life

Many HERS recommendations focus on replacing equipment nearing its life expectancy and soon requiring replacement. The estimated energy saving 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.

1.2.8. 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.

1.2.9. Actual occupant behavior

Research has shown that actual occupant behavior is probably the single most significant determinant of actual energy use. Sonderegger [3] 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 [4] 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 difference, but inherent variation, or natural uncertainty, in occupant behavior will always affect the precision of individual ratings. In other words, even if there are no HERS system differences, the predicted energy use or energy cost can be off by 50% or more due to occupant behavior.

1.2.10. 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.

2. Previous work

There have been numerous comments about the general uncertainties associated with estimating energy use and savings [5], especially with respect to simulation models [6]. However, there exist very little published data on HERS' predictive ability.

2.1. HERS Council

The HERS Council compared billing data with ratings for five houses rated with the Virginia Home Energy Rating Organization (VA HERO) software [7–9]. The report does not indicate how this sample of houses was selected for analysis or the house age, rating date or rater(s). Table 2 summarizes the HERS Council's findings. HERS' predictive ability 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.

2.2. 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", including DOE 2.1E-W54 [10]. 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. 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 can simulate other physical features such as heating,

Table 2
HERS Council accuracy data

	House A	House B	House C	House D	House E
Location (state)	VA	VA	VA	VA	VA
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 ($)	1543		1310	582	793
Actual electricity cost ($)	1384		1233	665	547
Electricity cost discrepancy (%)	10		6	−12	45
Gas cost estimate ($)		347[a]	946[a]		
Actual gas cost ($)		723		521	
Gas cost discrepancy (%)		−52		82	
Total cost discrepancy (%)	10	−52	6	29	45

[a] Gas estimate is for heating and domestic hot water (DHW) only. It is unclear if gas is used for other end-uses.

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 discrepancy. 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, such as incorrect occupant behavior assumptions, non-typical occupant behavior and rater inspection errors.

Nisson tested two programs with the HERS BESTEST: Rem/Design and HOT2000 [6]. Rem/Design, a HERS tool developed by Architectural Energy Corporation of Boulder, Colorado, is used by several HERS providers in the USA. HOT2000 was developed by the Canadian government and is used extensively in Canada and the USA 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 with actual billing data for about 50 homes in Canada [6,11,12]. Both showed agreement within 10% between estimated energy use and actual metered energy use. However, these programs differ fundamentally from most US HERS software in that they incorporate detailed data on actual occupant behavior, making the ratings more accurate but not ‘occupant blind’.

2.3. BEERS

Researchers at the Florida Solar Energy Center (FSEC) analyzed the predictive ability of the Florida Building Energy-Efficiency Rating System (BEERS), a HERS developed by FSEC and used widely in Florida [13]. Four-hundred-and-twenty-three new Florida houses (built between 1991 and 1993) were rated using BEERS. 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 pool heater energy, and refrigerator energy. 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.

Unfortunately, Florida Power and Light (FPL), the sponsor of the research project, did not allow FSEC to make the full report or the data publicly available. FPL did allow Philip Fairey of FSEC to give a qualitative description of the results at a recent conference [13]. 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. Perhaps the most valuable finding of the FSEC analysis is the fact that it was possible to improve significantly the predictive ability 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.

To summarize the previous work, the HERS Council report gives a picture of HERS' predictive ability in the field but it represents a small sample size of unknown origin, i.e., it is unclear if the ratings are representative of normal ratings. The FSEC study covers a large sample size and it is particularly valuable because it involves end-use metering, but the ratings are not representative and the results are not fully available. The HERS BESTEST represents a valuable tool for testing the simulation portion of a HERS, but not for assessing the overall accuracy of actual ratings. Thus, to put our research in context, our research represents the only effort that has been undertaken to analyze and publish data on the accuracy of actual HERS ratings; i.e., to compare rating output with actual energy use data for a significant sample of homes rated under normal circumstances.

3. Data

Although tens of thousands of houses have been rated in the last several years, few HERS providers were willing and able to supply us with data. The first data set we received and the one we examined most extensively, 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. The other three rating systems for which we received ratings and utility data were Energy Rated Homes of Colorado, Home Energy Ratings of Ohio, and Midwest Energy, a utility company and HERS provider in Kansas (see Table 3). These HERS all used different rating software and had slightly different ratings procedures (e.g., the CHEERS ratings did not include blower door testing, while the other ratings did). All of the HERS providers assured us that the samples were representative in terms of the types of house they rate and the expected accuracy of ratings. In most cases, the HERS providers also provided us with utility billing data for the year preceding the rating date. CHEERS provided us with three years of utility data. We collected actual weather data (from local NOAA weather stations) for the study period for all of the locations and found that the actual weather was close enough to the long-term average climate that weather normalization was not necessary.

4. Results

While utility billing data can only be used to measure the predictive ability of whole house energy rating (or electricity and gas predictions) directly, we were able to infer some results regarding HERS' ability to achieve the objectives of each of the three types of HERS output: scores, energy predictions and recommendations.

Table 3
Summary of case study results

	CHEERS (all homes)	CHEERS (new only)	Kansas	HERO — Ohio	ERHC — Colorado[a]
Sample size	185	30	16	14	276
Average year built	1959	1990–1994	1995	N/A	1969
Blower door test?	no	no	yes	yes	yes
Average HDD/yr for 1984–1995	2791	2791	4954	5371	6254
HDD in study yr (s)	3% below avg.	3% below avg.	N/A	5% above avg.	3% above avg.
Average actual energy cost ($)	1154	1327	1462	1697	135 MBtu
Average predicted energy cost ($)	1585	1026	1531	1402	120 MBtu
Average energy cost discrepancy[b] (%)	51	−8	−7	−14	−3
Standard deviation in differences (%)	80	44	15	20	35

[a] Energy cost data were not available for Colorado, so difference and standard deviations refer to site energy use.
[b] Average energy cost difference refers to the average of the individual differences and is not equivalent to the difference of the average.

4.1. Score

None of the HERS showed a clear relationship between rating score and total energy use or energy cost (see Fig. 1). This is an expected result because a rating measures only that house's potential for efficiency improvement; ratings are not designed to compare different houses. Nevertheless, consumers and even some HERS-related housing programs expect and assume that houses with higher rating scores will have lower energy costs. Our results show that houses with higher scores did not tend to use any less energy than houses with lower scores. The dashed line in Fig. 1 shows the relationship that one would expect (and that CHEERS predicts): as score increases, energy cost decreases. The solid line shows that, in reality, average energy cost is constant at about $1000/yr regardless of the score. One possible explanation is the 'take-back 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 convert some of the expected savings into higher levels of service.

4.2. Energy predictions

Three of the four HERS — Kansas, Ohio and Colorado — were quite accurate, on average, at predicting the actual energy cost or energy use (see Table 3). For example, on average, the Colorado system underpredicted the actual energy use by 3% (see Fig. 2). The fourth system, CHEERS, overpredicted the actual energy cost by about 50%, but for newer houses CHEERS

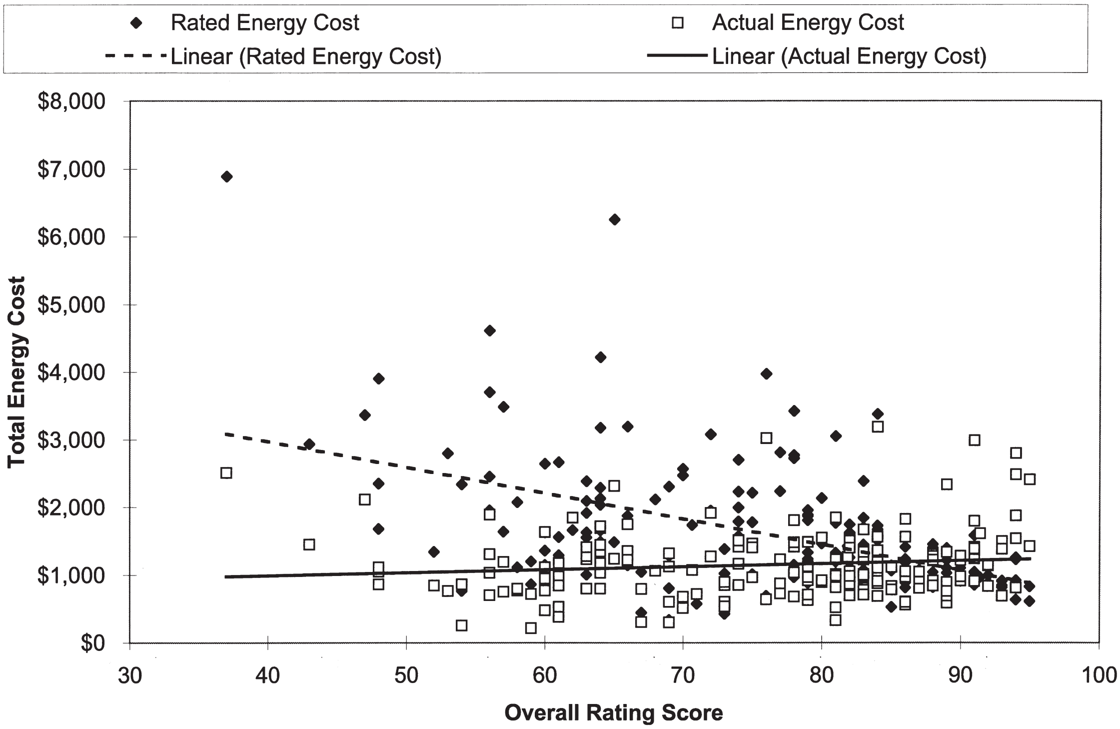

Fig. 1. CHEERS: actual energy use versus rating score.

was more accurate (underpredicting by 8%, on average) (see Fig. 3). However, while the average differences were quite low in most cases, the individual differences were often quite high. For example, the standard deviation for the CHEERS predictions was 80%, meaning that about 1/3 of the houses were overpredicted by more than 130% or underpredicted by more than about 30%. While much of the individual difference is attributable to actual occupant behavior, the magnitude of some of the differences (and the consistent tendency in the case of CHEERS to overpredict energy use) clearly implies that it is possible to improve both the average and individual differences by addressing system differences.

Comparing directly the predictive ability of the rating systems based on these case studies is risky because each sample of homes and each HERS is unique. Differences in the samples and the rating systems include the following:

- *Average house age* — the Kansas homes were almost all new, while the CHEERS houses were significantly older than the other groups. Old houses are harder to rate than new ones because it is difficult to assess the presence and extent of retrofits (like wall insulation).
- *Diagnostic testing* — only the CHEERS ratings 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 improvements in ratings systems were made in the

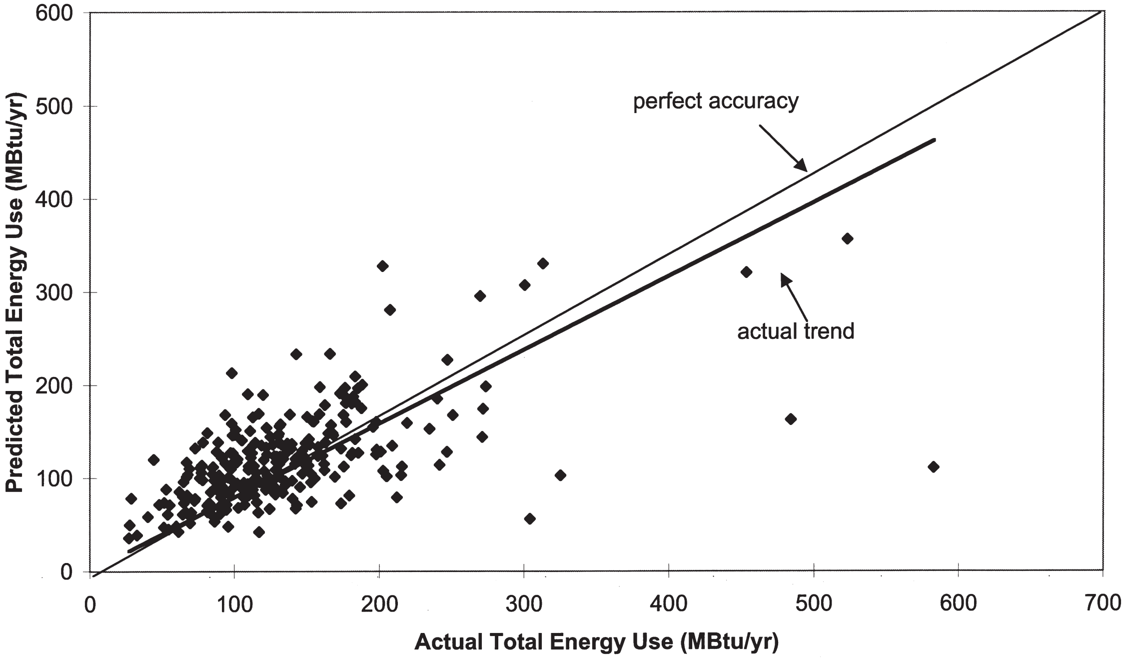

Fig. 2. Colorado: predicted versus actual energy use.

intervening years, including new simulation engines and a user-friendly Windows-based interface. These improvements can improve the predictive ability and consistency of ratings.

- *Climate* — one of the most significant differences between the sample groups is the severity of weather. The California climates have much milder winters than the other locations, and the Colorado location clearly has the most severe winter climate. As Pettersen [4] and others have found, and as our results appear to substantiate, it is harder to predict energy use in mild climates than in more severe climates. Intuitively, this makes sense because, for example, the percentage difference in heating energy between a thermostat setting of 70°F (21°C) and 75°F (24°C) is much larger when the outdoor temperature is 60°F (16°C) than when the outdoor temperature is 30°F (−1°C). (Table 3 shows the average heating degree days in base 65°F for each location.) Furthermore, space conditioning (the focus of HERS) is a smaller percentage of total energy use in mild climate. This all means that it is harder to have accurate ratings for both energy ‘misers’ and energy ‘hogs’ in mild climates than it is in more extreme climates. This difficulty can be seen by plotting regression lines of the average difference versus the rating score for each of the HERS (see Fig. 4). CHEERS, for example, appears to be well calibrated for high-efficiency houses but to be considerably off target for lower-scoring houses. Colorado is well calibrated for medium-efficiency houses and is not too far off, on average, for either the very high- or very low-efficiency houses. The slope of this regression line may be an inevitable function of climate severity. Indeed, CHEERS has the mildest climate and the steepest regression line, while Colorado has the most severe climate and the flattest regression line.

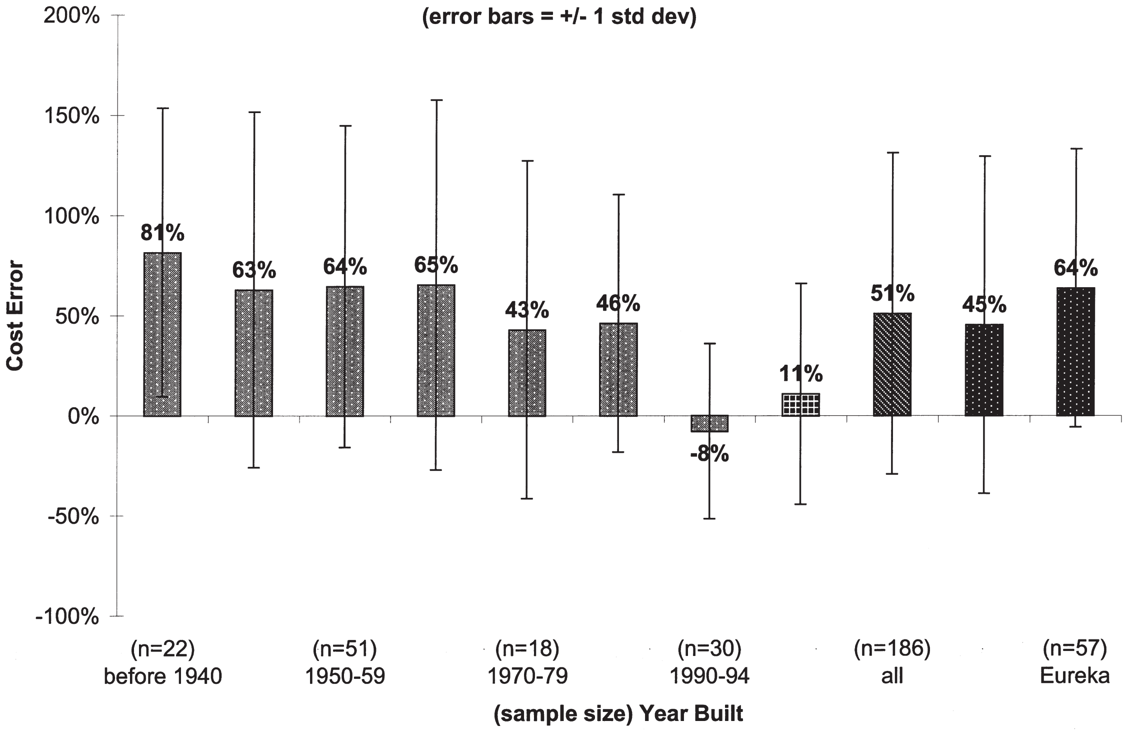

Fig. 3. CHEERS: cost difference versus age group and climate zone.

4.3. Quality of recommendations

In order to gauge the predictive ability of recommendations we compared the actual energy use of the CHEERS homes to the total energy savings that CHEERS predicted the occupants would receive if they implemented all the rating recommendations (see Fig. 5). 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. In general, it is likely that a significant fraction of HERS recommendations will not be cost-effective because there does not have to be any safety factor built into a recommendation. A recommendation is generally considered cost-effective if it is predicted to have a net present value of greater than zero, i.e., a positive expected cash flow (monthly financing cost<monthly savings).

4.4. Implications of limited predictive ability

4.4.1. Consumer risk

The consumer risks that he will buy the wrong house as a result of inaccurate scores or cost estimates or that he 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

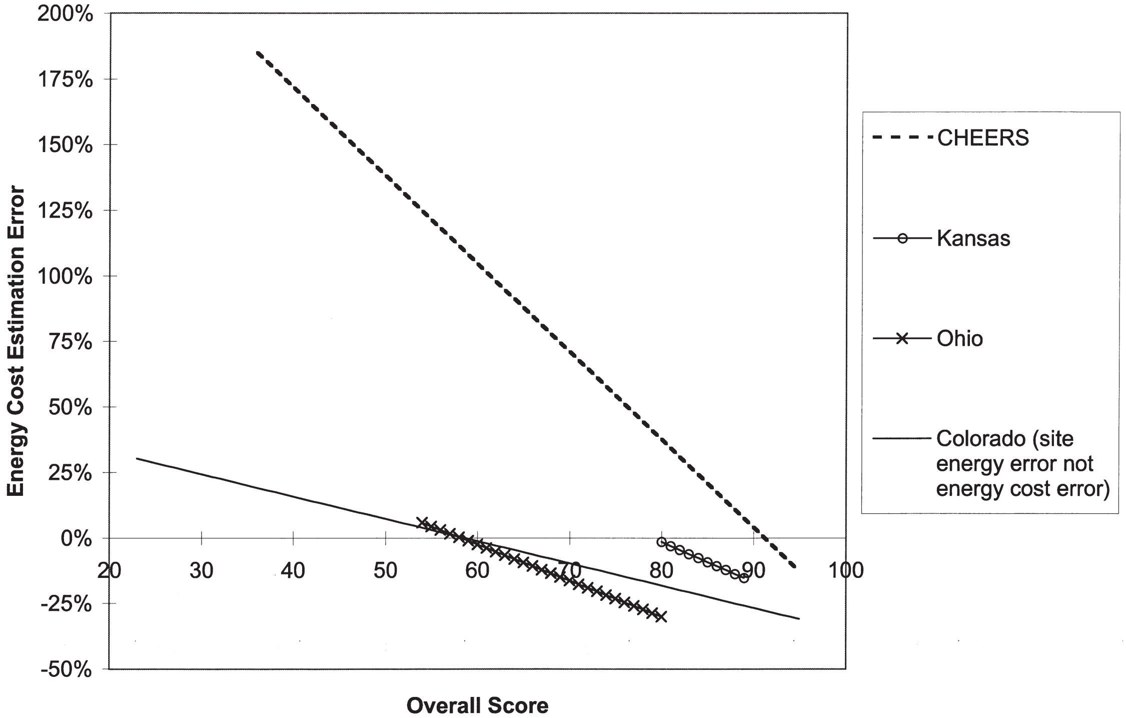

Fig. 4. Linear regression line of energy cost difference versus rating score.

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.

The risk of making an uneconomical investment depends on the accuracy of recommendations for that specific house and 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 consumer's monthly costs (energy costs+mortgage payment) to go up instead of down, as expected. However, as Horowitz [5] and others have noted, the consumer is generally not at any greater risk of defaulting on the mortgage.

On the other hand, many recommended improvements also provide intangible benefits such as increased comfort, reduced noise, greater security and better esthetics. Moreover, HERS ratings usually give the customer detailed economic information about each of the recommendations, such as simple payback period. Thus a skeptical or risk-averse consumer has the option to pick

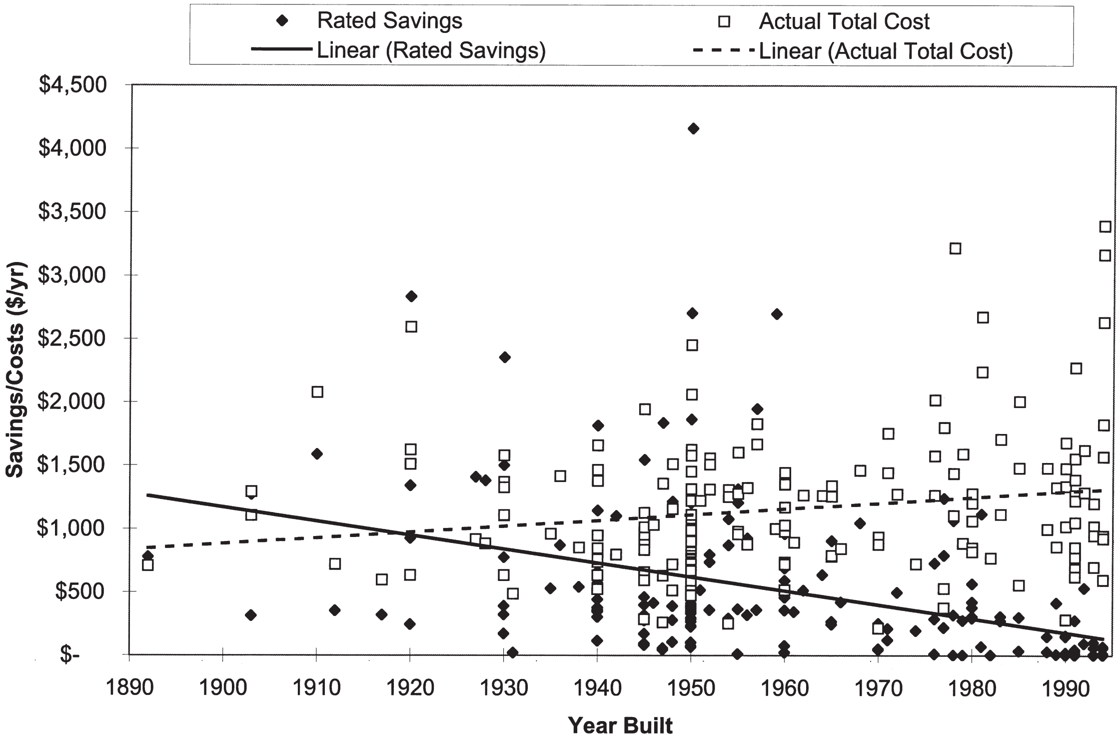

Fig. 5. CHEERS: estimated potential savings versus actual energy cost.

and choose recommendations with more attractive economics and thereby mitigate the risk of a bad investment. Furthermore, HERS recommendations are often not very sensitive to the accuracy of the rating. For example, suppose a HERS rating calculated that installing a hot water tank wrap in a particular house had a payback of one year. Even if the rating overpredicted hot water use by 300%, the tank wrap still would have a payback of about three years, which is a very healthy 37% return on investment.

4.4.2. Lender risk

Interestingly, there is almost no risk to energy mortgage lenders due to the accuracy of HERS. According to Horowitz [5], energy mortgages do not 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 of NREL, 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. According to Horowitz, 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.

4.4.3. 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 misled 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.

5. Conclusions

Hundreds of thousands of homes have been rated by HERS across the country in the last several years. The ratings and utility bills for these houses represent a relatively easy and low-cost opportunity to improve the predictive ability of rating systems. Utility data can be used not only to validate predictive ability but also to calibrate rating systems and to help identify and correct specific system differences.

5.1. Calibration

HERS programs could adjust the ratings to reflect observed discrepancies between the ratings and utility bills. This calibration corresponds to changes of slope in the results shown in Fig. 4. The calibration would adjust the results so as to flatten the regression lines and minimize the differences in energy costs over the whole range of scores. Note that this procedure improves the average but does not affect the variance.

5.2. Discrepancy correlations and correction

Analysis of billing data can be taken a step further by looking for correlations between success in predictive ability and house characteristics. For example, we found that CHEERS overpredicted gas use more in Eureka (a relatively cold climate) than in Fresno (a relatively hot climate), and overpredicted electricity use more in Fresno. The more heating or cooling required, the greater the overstatement. This suggests that CHEERS is using incorrect heating and cooling setpoints or infiltration rates, or conduction rates, etc. These differences still do not pinpoint the problem, but they help. The CHEERS data also showed that some raters more successfully predicted energy use than other raters, which emphasizes the need for rater training, oversight, retraining and the need to minimize rater judgment calls in the rating procedures.

While utility billing data can be a valuable and inexpensive way to improve HERS predictive ability, it does not give the whole picture. Other types of research are also needed to document and improve predictive ability. For example, the HERS BESTEST is a valuable tool for testing the simulation properties of HERS [10,14,15]. As the FSEC study showed, submetering of particular end-uses can be critical for improving the predictive ability of end-use energy predictions.

5.3. Disclaimers

In addition to conducting research to improve HERS predictive ability, HERS providers also need to give consumers more information about the accuracy of ratings and how to interpret ratings. For example, rating scores are calculated in a rather complicated manner and HERS ratings generally do not explain how scores are calculated or how they should be interpreted. Our results indicate that consumers may be misled if they think rating scores can be used to compare houses in the same way that miles-per-gallon ratings can be used to compare cars. Furthermore, ratings that calculate energy costs or life-cycle savings to four significant digits are likely to give consumers a false sense of confidence (e.g., “Upgrading the cooling system to SEER 12.0 will save $2166 on a lifecycle basis”). A range of savings that encompasses much of the uncertainty in the calculation is probably more appropriate (e.g., “Upgrading the cooling system to SEER 12.0 is likely save $1000–$3000 on a lifecycle basis under typical conditions”).

The absence of disclaimers and the lack of publicly available data demonstrating predictive ability are probably impeding the acceptance and growth of HERS amongst certain consumers, lenders and other groups nationwide. Furthermore, a lack of credibility may eventually undermine HERS. For these reasons, HERS organizations and HERS providers must continue to document and improve HERS tools and procedures.

Acknowledgements

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

References

[1] Home energy ratings: a primer for weatherization programs. Residential Energy Services Network under contract to the National Association of State Energy Officials (NASEO), 1996.

[2] Luboff J. Making energy mortgages work. Home Energy Mag 1995;(May/June).
[3] Sonderegger RC. Energy and Buildings 1977/78;1:313.
[4] Pettersen TD. Energy and Buildings 1994;21:209.
[5] Horowitz M. Innovative financing of home efficiency improvements. In: Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings, Vol. 2. American Council for an Energy-Efficient Economy, 1996:97.
[6] Nisson JDN. The accuracy of building energy simulation software? Energy Design Update 1996;(April).
[7] Guidelines for uniformity: voluntary procedures for home energy ratings, version I. Home Energy Rating System Council (HERSC), 1995.
[8] Guidelines for uniformity: voluntary procedures for home energy ratings, Version II. Home Energy Rating System Council (HERSC), 1996.
[9] HERS projected energy use v. actual utility bill data. Home Energy Rating System Council (HERSC), 1996.
[10] Judkoff R, Neymark J. Home energy rating system building energy simulation test (HERS BESTEST): volume I and II, Tier 1 and Tier 2 tests user's manual, NREL/TP-472-7332. National Renewable Energy Laboratory, 1995.
[11] Efficiency housing database development: field energy audit of existing dwellings — phase I report for the Alberta homes. Energy Technology Branch/CANMET, Department of Natural Resources (NRCan), 1995.
[12] AUDIT 2000 training and reference manual. Energy Technology Branch/CANMET, Department of Natural Resources (NRCan), 1995.
[13] Fairey P. BEERS: how accurate? Presented at HERSC-FREE '97, 1997.
[14] Judkoff R, Neymark J. J Solar Energy Eng 1995;117:7.
[15] Judkoff R, Neymark J. International Energy Agency building energy simulation test (BESTEST) and diagnostic method, NREL/TP-472-6231. National Renewable Energy Laboratory, 1995.