Measuring Residential Energy Efficiency Improvements with DEA

Peter Grösche

RWI Essen*

September 18, 2006

Abstract

This paper measures energy efficiency improvements in the US residential sector between 1997 and 2001 and relies on the benchmarking idea of Philipsen et al. (1997, 1998). Using an an intertemporal Data Envelopment Analysis (DEA) model and US households data, we compute for each household its distance to the empirical observed technology frontier. On the basis of simple nonparametric tests, we check for significant energy efficiency improvements between the years. Our results indicate a significant improvement in energy efficiency for all considered dwelling categories: mobile homes, single-family houses, and apartments.

JEL: C6, D13, Q41,

Key words: Residential Energy Efficiency Analysis, Data Envelopment Analysis

Preliminary version

Do not quote without permission

*Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Hohenzollernstr. 1-3, 45128 Essen, Germany. Email: groesche@rwi-essen.de

1 Introduction

Residential energy efficiency is in the focus of energy policy for more than two decades and is still a source for a fruitful academic discussion. Efficiency improvements are considered as a possible option for environmental protection and against potential energy scarcities.

Despite its promising prospects, it is far from clear how to measure energy efficiency improvements accurately. There are a number of approaches and concepts. Well known examples are the 'Energy Usage to Gross Domestic Production' and the so called 'Activity-Structure-Intensity Index', see Ang (2006) for a recent review. Phylipsen et al. (1997, 1998) measures energy efficiency in the cement industry using a benchmarking approach. The industries are compared with a best practice benchmark and the magnitude of deviation from the benchmark denotes the degree of inefficiency.

This paper measures energy efficiency improvements of US households between 1997 and 2001, and relies basically on the idea of Phylipsen et al. Using *Data-Envelopment-Analysis* (DEA) – a nonparametric frontier estimation technique – we calculate a best-practice benchmark for the households. By using simple nonparametric statistical tests, we check if a significant improvement of residential energy efficiency has occured between 1997 and 2001.

The outline of the paper is as follows. Section 2 starts with a definition of energy efficiency and gives way to the DEA methodology. Section 3 provides an overview on our data set. In Section 4, we discuss our results, while section 5 concludes.

2 Measuring and testing for energy efficiency

This section consist of three parts. We start with the traditional definition of energy in Section 2.1 and show in Section 2.2 how to measure residential energy efficiency with DEA. We point out that DEA corresponds with the traditional definition of energy efficiency. While thitherto the analysis considers only one point of time, we introduce in Section 2.3 another time period and show how to test for energy efficiency *improvements* between the periods..

2.1 The standard approach

Residential energy consumption derives from the demand of energy services, such as the demand for thermal comfort. The households 'produce' those services by using a set of fuel inputs in their respective energy commodities (e.g. their heating equipment). By this means, the standard approach to measure residential energy efficiency draws on the framework of Becker's (1965) home-production function, which regards households as a 'small factory' (Becker 1965: 496). Along these lines, Wirl (1997:17) defines residential energy efficiency as "the ratio between service output and energy input":

$$\eta = \frac{\text{service output}}{\text{energy input}}. \tag{1}$$

An *improvement* of energy efficiency would result in an increase of η.[1] This ratio is easy to calculate as long as output denotes only a single service, such as 'space heated by used energy', and can be done for a single household or even for the whole residential sector. But households produce a bunch of services from energy, therefore for each service usually a separate efficiency ratio is computed.

To get finally an entire picture of residential energy efficiency one has to aggregate the various efficiency ratios among services, while paying attention to the different measuring units of the ratios.[2] For example, one can 'normalize' the ratios by dividing each η with a certain benchmark (e.g. the respective theoretical maximum), such that the respective measuring units cancel out. The resulting

[1] Often the literature refers to the reciprocal of η, the energy intensity (see e.g. Haas 1997. Schipper et al. 2001).

[2] While the efficiency of space heating will be expressed e.g. in 'heated square-foot per kWh', water heating efficiency will be probably measured in 'liters of water heated per kWh'.

relative efficiency ratio $\tilde{\eta}$ is dimensionless. Thereafter one can aggregate all $\tilde{\eta}$ with a scheme of a-priori fixed weights that may reflect the shares from total energy consumption of each service.

Such an aggregation procedure is associated with some difficulties. First, it requires the specification of the amount of energy, consumed by each service output. Unfortunately, such numbers are hardly to obtain.

Second, to analyze overall residential energy efficiency over time the relative efficiencies ratios $\tilde{\eta}$ from different time periods must refer to the same basis, thus, the same benchmarks. This might arouse some ambiguities about the choice of the benchmark. If for example the 'state-of-the-art' is used as benchmark because the 'theoretical maximum value' is unknown, it is likely that this alternative benchmark will not be constant over time.

Third, an a priori weighting scheme uses fix weights for all households in all time periods. That is, space heating accounts always for e.g. 70% of households' total energy demand. Such a pattern might be correct for a certain point of time. But since technical progress is likely to affect the commodities of the several energy services differently,[3] the real consumption pattern will change over time, and a fix weighting scheme will not reflect the efficiency trend accurately. On the other hand, adjusting the weighting scheme to actual circumstances leads to non-comparability of two points in time.

2.2 Energy efficiency and DEA

Using DEA can avoid the mentioned difficulties. This methodology neither requires a separation of households total energy consumption by service output, nor does DEA require the provision of a 'normalizing' benchmark in advance. DEA treats each household as a producing unit in a multi-input/multi-output environment and computes a so called *best-practice frontier* as benchmark from

[3]For example, an important progress in heating technology will decrease both the households total energy demand and the share for space heating from total demand.

the data at hand. DEA considers from the outset the whole process of transforming energy into demanded services and computes an overall efficiency indicator for the household.

To see how DEA fits into the definition (1) of energy efficiency let s_{jl} be the jth service ($j = 1, \ldots, J$) from household l ($l = 1, \ldots, L$), and let e_{kl} be its kth fuel input ($k = 1, \ldots, K$). Each household uses at least one input to produce at least one service. Assuming that household o wants to consume as much service as possible by using as few inputs as possible, the goal is to:

$$\max_{u\,v} \eta_o = \frac{\sum_j u_{jo} s_{jo}}{\sum_k v_{ko} e_{ko}} \qquad \text{subject to} \tag{2a}$$

$$\frac{\sum_j u_{jo} s_{jl}}{\sum_k v_{ko} e_{kl}} \leq 1 \quad l = 1, \ldots, L, \tag{2b}$$

with household-specific weights $u_{jo} \geq 0, v_{ko} \geq 0$, assigned to the outputs and inputs, respectively. Problem (2) has to be solved for each of the L households.

Basically, equation (2a) resembles the definition of energy efficiency in equation (1). The nominator consists of an aggregated service output from the evaluated household o, the denominator is its aggregated energy input. To maximize η_o, the weighting scheme (u, v) is determined such that household o appears 'as good as possible', in reference to all L households as can be seen in the L constraints (one for each household) of (2b). Here, similar efficiency ratios for all L households are computed but the optimal weighting scheme of household o is used. Note that the weights u_{jo} and v_{ko} do not differ between (2a) and (2b). If we now find for any household l a higher efficiency ratio than for household o, then household l outperforms household o with o's own optimal weighting scheme. Such a superior household can serve as benchmark for household o. The constraints in (2b) ensure further that $0 < \eta_o \leq 1$, where $\eta_o = 1$ indicates an efficient household.

By simply adding the constraint $\sum_k v_{ko} e_{ko} = 1$ we linearize Problem (2), and get a DEA program in primal or "multiplier" form (Cooper et al. 2004), which can be solved for each household by linear programming.

Dual to the above multiplier form is the "envelopment model" (Cooper et al. 2004):

$$\text{(3a)} \qquad \min \ \theta_o$$

subject to:

$$\text{(3b)} \qquad \theta_o e_{ko} \geq \sum_l \lambda_l e_{kl} \quad k = 1, \ldots, K$$

$$\text{(3c)} \qquad s_{jo} \leq \sum_l \lambda_l s_{jl} \quad j = 1, \ldots, J$$

$$\text{(3d)} \qquad \lambda_l \geq 0 \qquad l = 1, \ldots, L.$$

Program (3) "envelopes" the data points by calculating a piece-wise linear boundary of the reference technology, commonly called *frontier*. The reference technology is reflected by the right-hand side of the constraints (3b) and (3c). Household o is compared with the frontier given by a convex combination of the sample observations. θ_o and the L household-specific weights λ_l are the unknown parameter that have to be determined by solving program (3).

θ_o serves as measure of inefficiency with $0 < \theta_o \leq 1$, and can be interpreted as equiproportional detraction of all employed inputs such that household o reaches the frontier. For example, if $\theta_o = 0.8$, then household o must reduce its energy consumption e_{ko} by 20% to become efficient. Thus, $(1 - \theta_o)e_{ko}$ is the conservation potential of household o. The program (3) searches for the smallest θ_o such that household o employs at most as much inputs (3b) and produces at least as much output (3c) as the reference technology. Since the inefficiency measure affects only on the inputs e_{ko}, model (3) is usually called an input-oriented DEA model.

Hitherto, we have assumed an underlying technology with constant return to scale (CRS) and strong disposability for in- and outputs. Including the additional

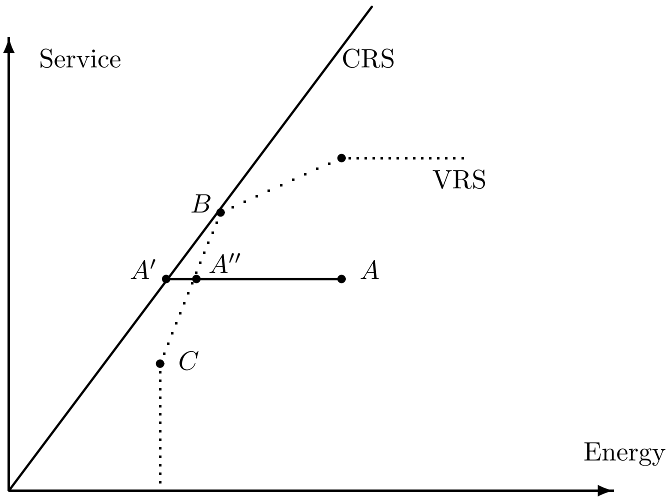


Figure 1: Input-oriented DEA

constraint (Banker, Charness, Cooper 1984)

$$\sum_l \lambda_l = 1 \tag{3e}$$

relaxes the CRS assumption and allows for increasing, constant, and decreasing returns to scale. The literature refers to the DEA problem (3a)-(3d) as (input-oriented) constant returns to scale (CRS) model. Adding (3e), we get the (input-oriented) variable returns to scale (VRS) model.

Figure 1 illustrates this distinction for a one-input/one-output case. It shows a CRS technology space as convex cone as well as a VRS technology space depicted by the dotted frontier. Point B lies on the efficient frontier of the CRS technology. B is also efficient under VRS technology, while Point C is only efficient under a VRS technology. Point A is inefficient in both cases.

Under a CRS technology, a proportional scaling of B still belongs to the frontier and such a scaling ($\lambda_B B$) can serve as benchmark. The CRS-DEA program (3a)-(3d) would compute for A a permissible energy consumption of $\theta_A^{\text{CRS}} A$ and would project point A onto the frontier of the convex cone CRS:

$$A' = \theta_A^{\text{CRS}} A = \lambda_B B. \tag{4}$$

Point A can reduce its energy consumption by $(1 - \theta_A^{CRS})$ percent until it would reach the frontier of the CRS technology. The projection A' is a scaling (a convex combination) of B with $0 < \lambda_b < 1$.

Assuming a VRS technology, constraint (3e) would force the parameter vector λ to add up to unity. In Figure 1 an input-oriented VRS model like (3a)-(3e) would evaluate point A against the piece of frontier spanned by B and C with increasing returns to scale. Thus, the efficient projection is

$$A'' = \theta_A^{\text{VRS}} A = \lambda_B B + \lambda_C C, \tag{5}$$

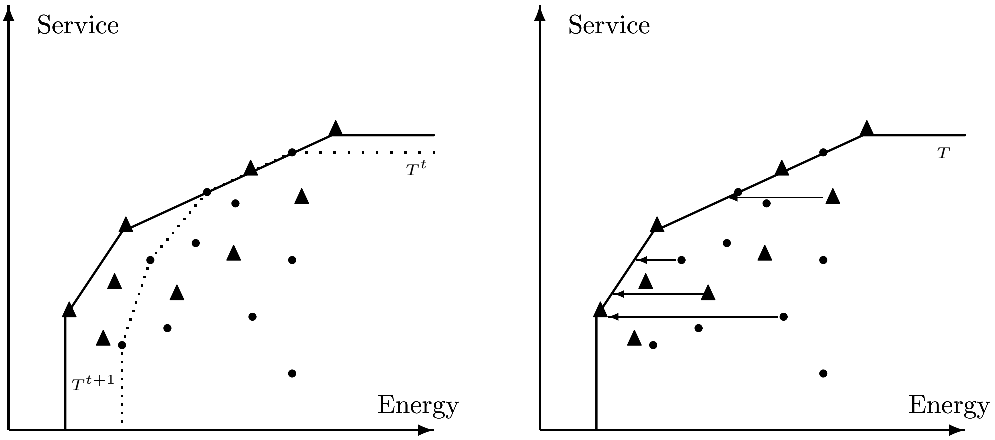
with the convex combination $\lambda_B B + \lambda_C C$ as benchmark and $\lambda_B + \lambda_C = 1$.

2.3 Measuring for improvements in energy efficiency

Assume that we have two samples of households from consecutive time periods t and $t+1$. Define S^t as the set of L households belonging to period t, each having at t a certain production plan $(e, s) \in T^t$, while T^t describing the technology available at t. In Figure 2(a) the set S^t is illustrated by the circles and the technology frontier of T^t is reflected by the dotted line. Let S^{t+1} be the succeeding set in time of S^t that contains M households with production plan $(e, s) \in T^{t+1}$. Figure 2(a) shows S^{t+1} as triangles with the solid line as the technology frontier of T^{t+1}.

An improvement of energy efficiency from time t to $t+1$ has occurred in Figure 2(a) since the frontier of the technology has shifted outwards and enhances the technology space. Thus, what was possible for households of S^{t+1} was not necessarily possible for households of S^t. However, some regions of the technology space might not be affected from such a progress. Then T^t and T^{t+1} overlap in these regions and the frontiers superimpose partially as depicted in Figure 2(a).

The sets S^t and S^{t+1} base on two cross sectional samples, from which we construct a pooled set $S = S^t \bigcup S^{t+1}$ with $L + M$ members and the technology $T = T^t \bigcup T^{t+1}$. Following Tulkens and Vanden Eackhaut (1995), we call S the *intertemporal production set*.



(a) Consecutive Production Sets

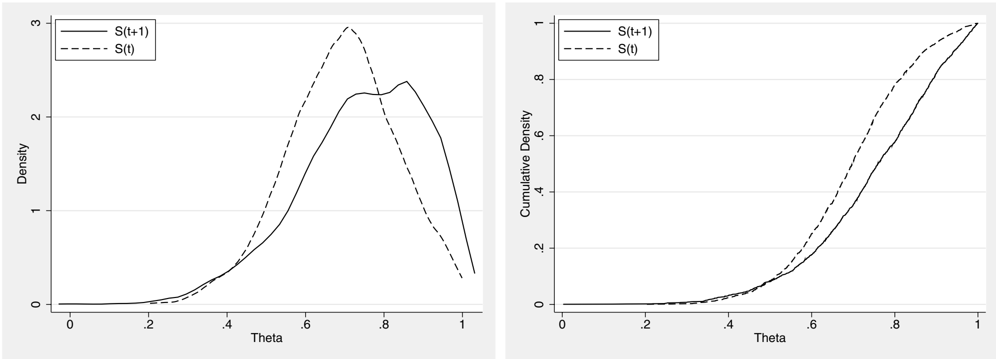
(b) Intertemporal Production Set

A • indicates an observation from S^t, ▲ an observation from S^{t+1}. In Figure 2(a) the technology T^t (below the dotted line) is a subset of T^{t+1}. Figure 2(b): Measuring for each household the distance to the frontier of the intertemporal production set indicates if an improvement in energy efficiency has occurred.

Figure 2: Production Sets

To test for an improvement in residential energy efficiency, we can measure for each household its distance to the frontier of T, as depicted in Figure 2(b). If an improvement has occurred, we expect in general a larger distance for households coming from the subset S^t, compared to households coming from S^{t+1}. This is simply because parts of the technology T were not available at time t. On the other hand, if no improvement at all has occurred and both technologies are identical, it should not matter from which subset a household has arisen from.

The distance to the frontier of T will be estimated by solving a DEA program like (3). As a result, we obtain for each household an estimate θ_l and can test if the distribution of θ differs between the two subsets S^t and S^{t+1}. Since $0 < \theta_l \leq 1$ and efficiency increases with θ_l, the more efficient subset should exhibit relatively more mass near 1 and less mass as θ decreases. In other words: The cumulative density of the more efficient subset should be quite flat in the lower regions of θ but becomes steep as θ converges to 1. The reverse is true for the inferior subset.



(a) Empirical Densities (b) Cumulative Density

The distributions for both variables were generated by 2000 random draws using a normal distribution (μ, σ^2) right-truncated at 1. Var.1 $\sim tN(0.8, 0.04)$; Var.2 $\sim tN(0.7, 0.02)$.

Figure 3: Comparing distributions of inefficiency scores

Thus, the cumulative density of the more efficient subset should lie well below its counterpart of the other subset.

Consider Figure 3 for which we have created distributions of θ for two artificial sets S^t and S^{t+1}. In figure 3(a), we see their empirical densities, figure 3(b) shows their empirical cumulative densities. S^{t+1} has more mass as θ approaches 1 but less mass as θ departs. Accordingly, the cumulative density of S^{t+1} lies well below the cumulative density of S^t. Thus, there is a substantial difference between the two cumulative densities, indicating an energy efficiency improvement.

To test if this improvement is significant we employ simple nonparametric tests, which are in line with the nonparametric nature of the DEA methodology. The *Kolmogorov-Smirnov* (KS) *Two-Sample Test* focuses on the maximal vertical deviation between the two cumulative densities. The *Robust-Rank-Order Test* draws inference from a rank vector. Both tests are discussed in more detail in the appendix.

3 The data

We use data coming from the US *Residential Energy Consumption Survey* (RECS), conducted regulary by the US Energy Information Administration (EIA).[4] For our purpose we use the surveys of 1997 (S^t) and 2001 (S^{t+1}) to check for energy efficiency improvements. Thus, $S = 1997 \bigcup 2001$. Each survey contains micro data for household and dwelling characteristics, energy consumption and expenditures, the endowment with several appliances and to some extent also behavioral consumption patterns.

While both survey come along with together more than 10,000 households, we had to drop a couple of observations. First of all, we drop households using coal, wood, district heating and renewable energies (heat pumps, solar heaters), since the amounts of energy consumed from these sources are not reported. Further, we drop all households which were disconnected from energy supply, e.g. due to an unpaid utility bill, because their energy consumption does not reflect their energy demand.[5] We remain with 7,533 households in total from which 3,969 come from the 1997 survey, and 3,564 households from 2001.

The surveys categorize residents into three dwelling types: mobile homes, single-family houses, and apartments. Our analysis is conducted for each of the three dwelling types separately because the dwelling categories are not comparable. For instance, since the heat loss of a dwelling is correlated with the ratio of outer surface to dwelling size, the heat loss of an apartment is smaller than the loss of a single-family house. Benchmarking a house against an apartment would hence be incorrect, even if both exhibit the same dwelling size.

Table 1 summarizes the employed input and output data by dwelling type. As the only input we use the household's total energy consumption, expressed in British Thermal Units (BTU). Similar to Schipper et *al.* (2001), we decompose

[4] The data are available online at http://www.eia.doe.gov/emeu/recs/contents.html.

[5] The household has to suffer a loss of comfort which is not in line with the idea of energy efficiency.

	Mobile Homes	Single-Family Houses	Apartments
Observations	# 516	# 5124	# 1893
-with basin	# 23	# 598	# 72
-with homebusiness	# 27	# 355	# 65

	Unit	Mean	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.
Total Energy	Mill. BTU	83	10	321	118	6	584	62	3	525
Heatnorm	square feet	901	0	4545	1986	0	15804	890	0	8685
Coolnorm	square feet	552	0	4663	1096	0	17497	555	0	7200
Persons	number	2.5	1	9	2.7	1	15	2.2	1	8
TVs	number	1.9	0	7	2.3	0	13	1.6	0	6
VCRs	number	1.3	0	11	1.5	0	11	1.1	0	11
Computers	number	0.3	0	8	0.6	0	7	0.4	0	9
Fridges/Freezers	number	1.4	1	4	1.7	0	5	1.1	0	4

Table 1: Data Summary

total service production into the main end uses space heating and cooling, water heating, cooking, and electric appliances.

Using the information about the heated and cooled space, measured in square feet respectively, we calculate the following "normals" to account for climate conditions that influence the households' energy consumption for space heating and cooling:

$$\text{heatnorm} = \frac{\text{heated space} \times HDD}{\text{1997 US. average } HDD} \tag{6}$$

$$\text{coolnorm} = \frac{\text{cooled space} \times CDD}{\text{1997 US. average } CDD}, \tag{7}$$

where HDD and CDD denotes heating and cooling degree days, respectively. HDDs are calculated as difference between 65° F indoor temperature and the daily average outdoor temperature below 65° F, summed among all days of a year. CDDs are calculated in a respective manner. A household which has to bridge a large gap between indoor and outdoor temperature gets in this way a grade up for each unit of (heated or cooled) space. For the sake of working with familiar numbers, a normalization with US average HDDs and CDDs takes place. We include the variables “heatnorm” and “coolnorm” as proxy for the service demand of space heating and cooling.

The number of persons in the household are used as a surrogate for hot water and cooking demands.

We incorporate the following electric appliances: TV-sets, videos and DVDs, computers, fridges and freezers. The number of the respective equipment serves as approximization of the associated service demand.

We further include two binary variables: basin (d_1) and home business (d_2). The first one indicates if a household has to heat a swimming-pool, a hot tub or a large aquarium. The second one takes into account that some households run a business activity at home and may therefore use more energy as a ‘normal’ resident. The binary variables are defined as:

$$d_{cl} = \begin{cases} 0 & \text{, if household } l \text{ is not equipped with device } c, \\ -1 & \text{, if household } l \text{ is equipped with device } c, \end{cases} \tag{8}$$

and we insert in our DEA model (3) for each device c the additional constraint

$$d_{co} \geq \sum_l \lambda_l d_{cl}. \tag{9}$$

These constraints restrict the set of reference observations to those households, which operate in the same or a more difficult environment as the household under evaluation. For example, if household o owns a swimming-pool, it does not make much sense to benchmark o against households without a pool, since o produces more service (heating the pool) for which it demands energy. Therefore, constraint (9) selects only households as possible benchmarks which likewise own a pool. On the other hand, better equipped households may serve as benchmark since they operate under more difficult conditions in terms of energy consumption. For more details, see the appendix.

For each household we solve the following input-oriented VRS-DEA program:

(10a) $$\min \ \theta_o$$

subject to:

(10b) $$\theta_o e_o \geq \sum_{l \in H} \lambda_l e_l$$

(10c) $$s_{jo} \leq \sum_{l \in H} \lambda_l s_{jl} \quad j = 1, \ldots, 7$$

(10d) $$d_{co} \geq \sum_{l \in H} \lambda_l d_{cl} \quad c = 1, 2$$

(10e) $$1 = \sum_{l \in H} \lambda_l$$

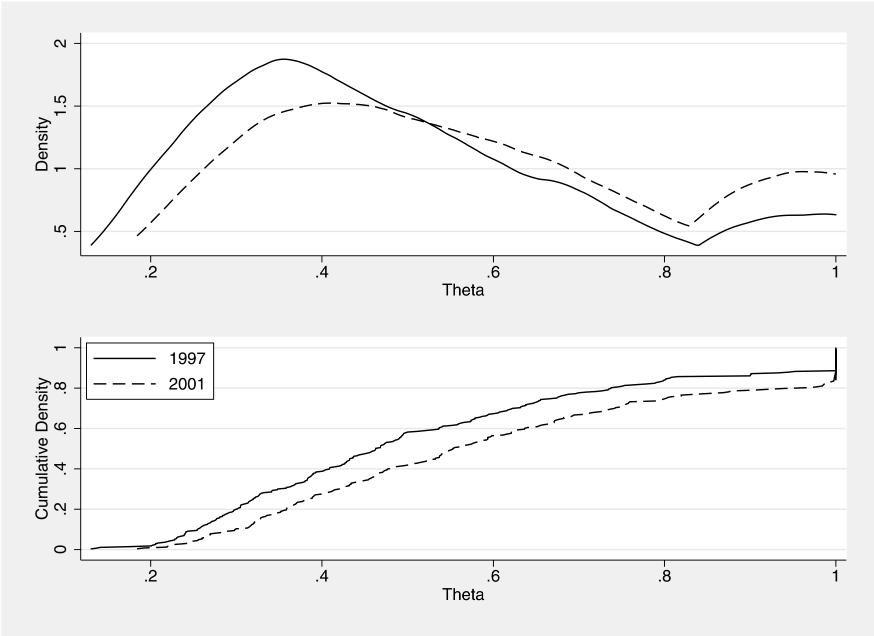
(10f) $$0 \leq \lambda_l,$$

where H denotes a subset that contains all observations of the same dwelling category as the household currently under evaluation. Since we employ the aggregated energy consumption e_o of household o as the only input, the expression $(1 - \theta_o)e_o$ can directly be interpreted as its energy conservation potential. We obtain for each dwelling type a distribution of θ which contains observations from 1997 and 2001. By using the Kolmogorov-Smirnov test and the Robust-Rank-Order test we can check, if significant improvements in energy efficiency between the years has occurred.

4 Results

In this section we will discuss the results for each dwelling type separately. We start with mobile homes and will then proceed with single-family houses and apartments.

The mean efficiency score θ for mobile homes in each year gives a first hint that some efficiency improvement between 1997 and 2001 has occured. The average



Number of observation: 516; from 1997: 273; from 2001: 243.

Figure 4: Results Mobile Homes

θ of 1997 amounts to 0.524 whereas the average θ of 2001 is 0.605. The mean energy conservation potential $(1 - \theta)$ decreases from 47.6% in 1997 to 39.5% in 2001.

Figure 4 shows the empirical densities and cumulative densities for the two years. The 2001 density has comparably less mass in lower regions of θ, but more mass in higher regions. Remember that a larger θ indicates higher efficiency. The cumulative density of 2001 lies well below the cumulative density of 1997, figuring out that some progress in energy efficiency has taken place.

The Kolmogorov-Smirnov (KS) test computes a maximal distance of $D = 0.1668$ between the two cumulative densities, which is significantly different from zero (p-value: 0.001). Likewise, the Robust-Rank-Order (RRO) test computes a test statistic of $\tilde{U} = 3.7392$. Since $\tilde{U}$ is approximately standard normal distributed, the observed $\tilde{U}$ is well apart from any popular critical value. Thus, both tests indicate a statistical significant improvement in energy efficiency for mobile homes between 1997 and 2001.

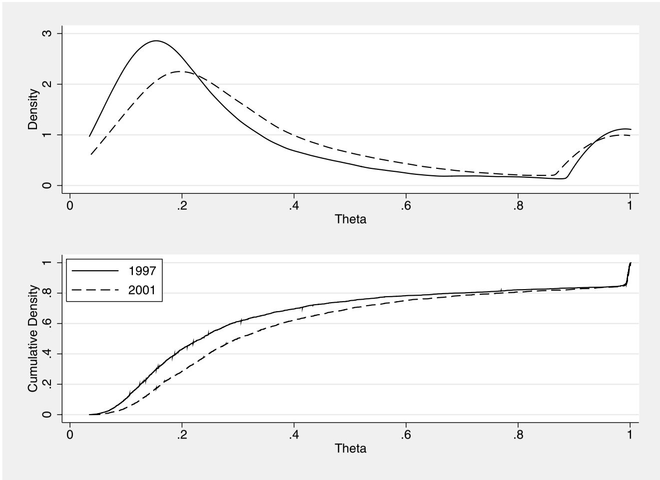
Figure 5 shows the results for single-family houses and for apartments. Figure 5(a) indicates an improvement in energy efficiency for single-family houses. The average θ for 1997 is 0.374, the average θ for 2001 is 0.425. The distribution seems to shift to the right between the years, into more efficient regions of θ. Accordingly, the 2001 cumulative density lies well below its 1997 counterpart. The KS test computes a maximal distance of $D = 0.1497$ ($p = 0$), the RRO test calculates a $\tilde{U} = 10.0724$. Thus, both tests indicate a significant improvement in energy efficiency from 1997 to 2001 for single-family houses.

Figure 5(b) shows the results for apartments. Likewise, there seems to be an improvement of energy efficiency in apartments between 1997 and 2001. The average θ increases from 0.263 in 1997 to 0.351 in 2001. Thus, while about 73.7% of consumed energy in apartments could have been saved in 1997, this conservation potential decreases until 2001 to 64.9%. The improvement is significant since the KS test computes a $D = 0.1694$ (p-value=0) and the RRO test computes an $\tilde{U} = 8.4398$.

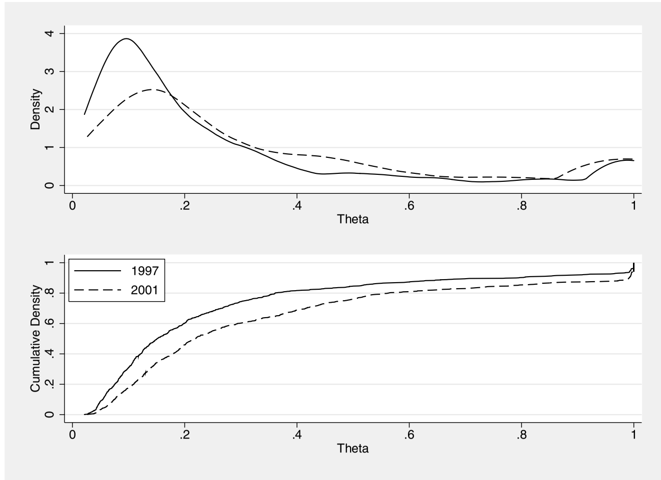
All distributions for the three dwelling categories exhibit something like a two peak pattern for both years. While the empirical densities have approximatively normal shape in the range of $0.1 \leq \theta \leq 0.8$ they exhibit another peak around $\theta = 0.9$. The two-peak pattern is very pronounced for single-family houses and apartments (Figure 5) but less extreme for mobile homes.

This let us presume that there might be a systematic difference between these two peaks, maybe even an omitted important variable. That variable should be somehow related to to space heating and cooling, since these services usually account for the major share of households energy consumption. Our first guess was that our climate correction via HDDs and CDDs was insufficient and remaining regional climate influences cause the special pattern of our results.[6] Other can-

[6]Remember that we include heating and cooling degree days to capture climate conditions.



(a) Single-Family Houses



(b) Apartments

For Single-Family Houses: Number of observation: 5121; from 1997: 2776; from 2001: 2345; For Apartments: Number of observation: 1893; from 1997: 917; from 2001: 976.

Figure 5: Results for Single-Family Houses and Apartments

didates are the type and the age of the heating and cooling system, differences concerning the used fuel or the buildings's construction year. Unfortunately, we lack information on the construction material of the building and its insulation standard.

If the sought variable sorts households in either of the two peaks, there must be a strong correlation between θ and this variable. We use Spearman rank correlation (Siegel and Castellan 1988) for remaining climate influences[7] and ordinary χ^2 tests for all other considered candidates to check for any relevant relationship that can explain the special pattern of our results.

Unfortunately, we found no sufficient explanation for our results among the considered candidates. There seems to be some remaining climate influences not captured by our climate correction since the Spearman correlation is significant. But the coefficient is rather weak such that it seems implausible that the remaining climate influences are responsible for our special results pattern. The type and age of the heating equipment cannot contribute anything to explain the results for single-family houses and apartments, the choice of the fuel and the age of the building cannot explain the two peaks in our results for apartments.

Finally, our results may be caused by "superinsulated" low-energy or even passiv houses which therefore serve as benchmark and constitute the right peak of our efficiency distributions. Ordinary, such buildings are not comparable to "normal" houses and should therefore constitute a separate dwelling category. But since we lack information on the building's insulation standard, there is no way to check this hypotheses.

Although we used the standard approach for climate correction it might be possible, that the relationship is not linear and we therefore will find some remaining influences.

[7]We aggregated HDD and CDD into a variable "climate-days" for those households that have either heating or cooling equipment, or both. If e.g. a household does not have air conditioning, then it cannot consume energy for cooling and thus CDD are not took into consideration. If the household neither owns heating or cooling equipment, the variable "climate-days"=0.

5 Summary

In this paper we measure energy efficiency improvements of US residents between 1997 and 2001. We treat a household as producing unit that transforms energy into demanded energy services. Using intertemporal Data-Envelopment-Analysis we compute for each household an overall inefficiency measure, expressed as distance to the best-practice technology frontier of the respective dwelling category.

We found for all three considered dwelling categories – mobile homes, single-family houses, and apartments – an improvement of energy efficiency, that is, the energy conservation potential decreased between 1997 and 2001. Simple non-parametric tests confirm that this trend is statistically significant.

Our analysis draws on the comparison of time specific inefficiency distributions instead of considering average numbers over time, like the energy intensity. However, our computed inefficiency distributions have a quite idiosyncratic two-peak shape. Although we intensive checked if dwelling characteristics may have caused this shape, we still miss the explanation. This lets us presume that the underlying reason is a determinant not included in our data set, for instance the insulation standard of the building. It is desirable to apply the proposed methodology to an other data set where such information is included.

Acknowledgments

The author is indebted to Manuel Frondel and Colin Vance who contributed very helpful comments on this paper.

References

ANG, B.W. (2006), Monitoring changes in economy-wide energy efficiency: From Energy-GDP ratio to composite efficiency index, *Energy Policy*, Vol.34:574-582.

BANKER, R.D., A. CHARNES AND W.W. COOPER (1984), Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis, *Management Science*, Vol.30:1078-1092.

BANKER, R.D. AND R.C. MOREY (1986), The use of categorical variables in Data Envelopment Analysis, *Management Science*, Vol.32:1613-1627.

BECKER, G.S. (1965), A theory of the allocation of time, *The Economic Journal*, Vol.75:493-517.

COOPER, W.W., L.M. SEIFORD AND J. ZHU (2004), Data Envelopment Analysis: History, models and interpretations, in: *Handbook on Data Envelopment Analysis*, pp. 1-39, Kluwer Academic Publisher, Boston.

HAAS, R. (1997), Energy efficiency indicators in the residential sector: What do we know and what has to be ensured, *Energy Policy*, Vol.25:789-802.

PHYLIPSEN, G.J.M., K. BLOK AND E. WORRELL (1997), International comparisons of energy efficiency-methodologies for the manufacturing industry, *Energy Policy*, Vol.25:715-725.

PHYLIPSEN, G.J.M., K. BLOK AND E. WORRELL (1998), *Handbook on international comparisons of energy efficiency in the manufacturing industry*, Department of Science, Technology and Society, Utrecht University, The Netherlands.

SCHIPPER, L., F. UNANDER, S. MURTISHAW AND M. TING (2001), Indicators of energy use and carbon emissions: Explaining the energy economy link, *Annual Review of Energy and the Environment*, Vol.26:49-81.

SIEGEL, S. AND N.J. CASTELLAN (1988), *Nonparametric statistics for the behavioral sciences*, McGraw-Hill, New York.

TULKENS, H. AND P. VANDEN EECKAUT (1995), Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects, *European Journal of Operational Research*, Vol.80:474-499.

WIRL, F. (1997), *The economics of conservation programs*, Kluwer Academic Publishers, Dordrecht.

Appendix

A Nonparametric tests

The *Kolmogorov-Smirnov* (KS) *Two-Sample Test* examines if two independent samples might come from the same population.[8] It focuses on the maximal vertical deviation between the two empirical cumulative density distributions. As a one-tail test, it tests H_0: "No difference in the distributions" against H_1: "One distribution is stochastically larger", where 'stochastically larger' means in our context 'more efficient', since efficiency increases with θ. Basically, the KS test computes along the range of θ the difference between the two cumulative density functions:

$$D = \max\{F_{S^t}(\theta) - F_{S^{t+1}}(\theta)\} \quad (11)$$

and tests if this difference is "large enough" to reject H_0. The KS test is standard in most statistical software packages and can therefore be easily adopted. If no such option is available, Siegel and Castellan (1988) report tables for critical values.

Another method of comparing difference between the two distributions is to rank all households by their θ in ascending order. Thus, the largest θ leads to the highest rank. Now, if the set S^{t+1} is more efficient as S^t, the bulk of S^{t+1} should be assigned with higher rank numbers as the bulk of S^t. On the other hand, if no difference can be found between the sets, the observations of both sets should be distributed randomly across the rank vector. The *Robust-Rank-Order*

[8] For more than two samples one can use the Jonckheere test for ordered alternatives, described in e.g. Siegel and Castellan (1988).

Test (Siegel and Castellan 1988) inspects if the assigned ranks for one set are significantly larger. For both sets, this test counts for each household how many observations *of the other set* have a smaller rank, viz. are less efficient. For example, if a household l belongs to S^t and has rank 15 we count for the ranks 1 to 14 how many households belong to S^{t+1}. If this count is 6, then $U_l(S^t) = 6$.

After doing this count for all members in both sets, the respective means and variabilities can be calculated:

$$U(S^{t+1}) = \sum_{l=1}^{M} U_l(S^{t+1})/M \tag{12a}$$

$$U(S^t) = \sum_{l=1}^{L} U_l(S^t)/L \tag{12b}$$

$$V_{S^{t+1}} = \sum_{l=1}^{M} [U_l(S^{t+1}) - U(S^{t+1})]^2 \tag{12c}$$

$$V_{S^t} = \sum_{l=1}^{L} [U_l(S^t) - U(S^t)]^2 \tag{12d}$$

The test statistics:

$$\tilde{U} = \frac{L\,U(S^t) - M\,U(S^{t+1})}{2\sqrt{V_{S^t} + V_{S^{t+1}} + U(S^t)U(S^{t+1})}} \tag{13}$$

approaches the unit normal distribution for large samples. For small samples ($L \leq M \leq 12$) tabulated values are available in Siegel and Castellan (1988).

One important occurence with DEA are ties across groups in the rank assignment. Again, let household l belong to S^t while household m belongs to S^{t+1}, but $\theta_l = \theta_m$, thus both households have the same rank and are tied. Usually a DEA program detects several entities to be efficient and assigns therefore $\theta = 1$ for some households. Siegel and Castellan (1988:143) describe how to deal with such ties: Observations of the other group but with the same rank are judged only with 1/2. This means for the above household l in S^t: Count all households in S^{t+1} with lower rank + 1/2 times the households in S^{t+1} with the same rank.

B DEA with categorical variables

To the best of our knowledge, Banker and Morey (1986) were the first who developed a procedure of nested reference technologies to include categorical variables into DEA. Their suggestion requires that an order of the categories has a meaningful interpretation. The procedure restricts the set of observations serving as reference points only to those observations that operate in the same or a more unfavorable environment as the household currently under evaluation.

Banker and Morey (1986) concentrate on categorical inputs, but we have to deal with a binary output variable. Thus, we take up the original idea and alter it slightly such that the nesting solution is suitable for our problem.

Assume we have an output dummy variable d_{cl} which indicates that household l is equipped with device c, e.g. a heated swimming-pool. Thus, the household 'produces' a certain service with this device, viz. thermal comfort in his pool. The question is how to control for such a binary variable in DEA?

Since the calculated benchmark of a DEA analysis often consists of a convex combination of various households with different values for a certain output, a convex combination of the output values must be feasible and at best even an meaningful interpretation. Both is not applicable for a convex combination of a dummy variable. Thus, we cannot include d_{cl} as ordinary output variable.

The idea of Banker and Morey (1986) relies basically on benchmark selection. Compare two households: all other things equal, household A is equipped with a pool whereas B is not. Usually we would expect that household A has a larger energy demand since he has to heat the pool and operates under a more difficult environment for energy conservation than B. Benchmarking A against B is meaningless since A produces more service. On the other hand, benchmarking B against A makes indeed sense. If household B uses more energy than A, then B is of course inefficient: it produces less services while using more energy.

Generally, we can benchmark a less equipped household against a better equipped one but the reverse is not reasonable. Therefore we have to restrict the reference set for better equipped households.

To formalize this we define d_{cl} as:

$$d_{cl} = \begin{cases} 0, & \text{if } l \text{ is not equipped with device } c, \\ -1, & \text{if } l \text{ is equipped with device } c, \end{cases} \tag{14}$$

and introduce for each device the additional constraint

$$d_{co} \geq \sum_{l} \lambda_l d_{cl} \tag{15}$$

into the DEA program (3a)-(3e). If household o is equipped with device c, then constraint (15) holds with equality and narrows the benchmark selection set to households which are likewise endowed with device c. The right hand side of (15) must add up to -1, which can only be achieved by households for which $d_{cl} = -1$, since $d_{cl} \in \{-1, 0\}$ and $\lambda_l \geq 0$ and $\sum_l \lambda_l = 1$.

If instead $d_{co} = 0$, then constraint (15) is not binding. The right hand side must add up to 0 which can be achieved by households with either $d_{cl} = 0$ or $d_{cl} = -1$. Thus, a household without a pool can also be benchmarked against households with a pool.