MEASURING PLANT LEVEL ENERGY EFFICIENCY AND TECHNICAL CHANGE IN THE U.S. METAL-BASED DURABLE MANUFACTURING SECTOR USING STOCHASTIC FRONTIER ANALYSIS

Gale Boyd[1] and Jonathan M. Lee[2]

[1]*Social Science Research Institute, Duke University, Box 90989, Durham NC 27708*
[2]*Department of Economics, East Carolina University, Greenville, NC 27858*

This study analyzes the electric and fuel energy efficiency for five different metal-based durable manufacturing industries in the United States over the time period 1987-2012, at the 3 digit North American Industry Classification System (NAICS) level. Using confidential plant-level data on energy use and production from the quinquennial U.S. Economic Census, a stochastic frontier regression analysis (SFA) is applied in six repeated cross sections for each five year census. The SFA control for energy prices and climate-driven energy demand (heating degree days HDD and cooling degree days CDD) due to differences in plant level locations, as well as 6-digit NAICS industry effects. Own energy price elasticities range from -.7 to -1.0, with electricity tending to have slightly higher elasticity than fuel. Mean efficiency estimates (100% = best practice level) range from a low of 33% (fuel, NAICS 334 - Computer and Electronic Products) to 86% (electricity, NAICS 332 - Fabricated Metal Products). Electric efficiency is consistently better than fuel efficiency for all NAICS. Assuming that all plants in the least efficient quartile of the efficiency distribution achieve a median level of performance, we compute the change in total energy use to be 21%. A Malmquist index is used to decompose the aggregate change in energy performance into indices of efficiency and frontier (best practice) change. Modest improvements in aggregate energy performance is mostly change in best practice, but failure to keep up with the frontier retards aggregate improvement. Given that the best practice frontier has shifted, we also find that firms entering the industry are statistically more efficient, i.e. closer to the frontier; about 0.6% for electricity and 1.7% for fuels on average.

Acknowledgements

This paper was funded by the U.S. Energy Information Administration (EIA), Office of Energy Consumption and Efficiency Analysis under contract number DE-EI0000564, in conjunction with and while working through the Leidos contract with EIA. It was prepared while the authors were Special Sworn Status researchers working under CES project #1213 at the Triangle Research Data Center, a member of the Federal Statistical Research Data Center Network (FSRDC). All results presented here have been reviewed by the Census Bureau and do not reveal any confidential information. The results do not reflect the views of the Census Bureau or EIA. The authors would like to thank Kelly Perl and Paul Otis (EIA), the conference participants at the 2016 FSRDC Annual Conference, College Station TX and the 2016 International Symposium on Energy Challenges and Mechanics, Inverness Scotland for their helpful comments on the earlier drafts.

1. Introduction

Analysis of energy demand commonly disaggregates the three major sectors (transportation, residential & commercial buildings, and industry) based on the underlying notion that energy consumption is a derived demand for energy services and the energy services for these three sectors of the economy are all very different. In industrialized economies, these three sectors each historically have consumed roughly a third of total energy. In 2015 in the U.S., transportation, residential & commercial buildings, and industry consumed 28%, 40%, and 32%, respectively.[1] Industry is arguably the most diverse in the range of energy services needed to produce the variety of products in the economy. Industrial processes range from the conversion of raw materials into intermediate products to the assembly and fabrication of final capital and consumer goods, with corresponding diversity in energy services. This diversity makes the industrial sector the most complicated of the three regarding modeling energy demand and efficiency.

The notion that energy demand suffers from the "energy paradox" or the "energy-efficiency gap" is pervasive in the energy demand literature. (Gerarden, Newell et al. 2015) define the "energy paradox" as '*the apparent reality that some energy-efficiency technologies that would pay off for adopters are nevertheless not adopted. This basic definition relates to the issue of private optimality.*' They define the "energy-efficiency gap" as the *'apparent reality that some energy-efficiency technologies that would be socially efficient are not adopted. This broader concept relates to social optimality.'* This paper uses the term "energy efficiency gap," in the narrower private notion that (Gerarden, Newell et al. 2015) term the energy paradox. An exhaustive review of the literature on the energy efficiency gap is beyond the scope of this paper; some examples include (Howarth and Andersson 1993, Jaffe and Stavins 1994, Huntington 1995, Allcott and Greenstone 2012, Boyd and Zhang 2013, Boyd and Curtis 2014, Boyd 2016). This literature can loosely be separated into two threads; quantification and identification. The second thread of this literature, reviewed by (Gerarden, Newell et al. 2015) focuses on identifying potential sources of the gap, assessing market barriers, etc. Much of the quantification literature is based on the engineering economics perspective. (Worrell, Ramesohl et al. 2004) review modeling industrial energy demand from an engineering economics perspective. The conservation supply curve (CSC) is popular representation of the engineering approach to quantification and is typified by (Rosenfeld, Atkinson et al. 1993). Engineering economics is not the only approach to the CSC. (Huntington 1995) discusses the intellectual connections between the energy efficiency gap literature and the literature on measuring production efficiency. (Blumstein and Stoft 1995, Stoft 1995) make the connections between production economics and the CSC more explicit by showing how the CSC can be constructed from the production function when inefficiency is present. This paper is a quantification of energy efficiency, but not based on an engineering economics approach. It draws on economic production theory and statistical modeling as the basis for quantifying the energy efficiency gap.

This paper contributes to the methods for and empirical quantification of energy efficiency in both a static and dynamic sense. For example, the CSC typically describes energy efficiency at a point in time. This paper also examines how efficiency may change over time. This is important because a popular conceptual method for modeling energy is the stock adjustment approach. This approach is based on

[1] Source: U.S. Energy Information Administration, *Monthly Energy Review* – Table 1.4b

the notion that energy use is tied to capital stock which changes over time in response to replacement, due to depreciation, and new expansion, to account for growth. This basic framework is used widely in the demand modules for the National Energy Modeling System (NEMS) by the U.S. Energy Information Administration (EIA). The Industrial Demand Model (IDM) is one such model that employs this underlying concept (EIA 2014).

This stock adjustment approach considers that the unit energy consumption (UEC) can be represented as a weighted average of the UEC for existing and new applications (equipment or capital stock, etc.)

$$UEC = \lambda * (UEC_{exsiting}) + (1-\lambda) * UEC_{new}$$

This approach shares features with the partial adjustment model that is commonly used in econometric studies to distinguish between long run and short run price elasticities for a wide range of macro and microeconomic phenomenon, including energy. The implications of such a model, particularly when interpreted in the context of a putty-clay approach, is that once the relevant piece of capital is put into use the UEC is constant (or nearly so) over its lifetime. This implies that there would be a distribution of UEC over different pieces of equipment. Since the distribution must have a minimum, this distribution can be thought of as the distribution of energy efficiency within a sector. The difference between the average UEC and the lowest can be thought of as a measure of the "energy gap,"

The putty-clay notion is an empirically strong one, particularly if the unit of observation is not a piece of equipment but a manufacturing facility. A facility is made of many difference pieces of capital, with different lifetimes. In addition, the UEC may not be entirely embodied in the capital, but respond either positively or negatively to other inputs; maintenance, management, etc. This means that the UEC of a manufacturing facility or plant may change over time, but the UEC will still within an observable distribution. The shape and position of this distribution may also change over time, either in response to technical change or other forces. The direction and magnitude is an important empirical question.

This paper estimates the distribution of energy efficiency in metal based durables (MBD) and the evolution over time in order to better understand the nature of the energy gap and how it interacts with technical change to drive industry level performance. This research can inform the stock-adjustment based energy forecasting of MBD in the NEMS IDM and other similar models. While the efficiency distribution could be estimated using Data Envelopment Analysis (DEA) estimation approaches, this study employs the stochastic frontier analysis (SFA) econometric approach. For a review of both the DEA and SFA methods, see (Murillo-Zamorano 2004) To measure the evolution of efficiency and technical change over time, a Malmquist index decomposition is conducted (Färe and Grosskopf 1992, Färe, Grosskopf et al. 1994). We also examine the efficiency of new plants entering the industry, relative to continuing plants. This provides insight into another aspect of the dynamics of energy efficiency as it relates to plant vintage.

Non-public plant level micro-data from the U.S. Census Bureau is used for this analysis. It allows us to control for differences in local market conditions (e.g. state level policies or energy prices) and exogenous plant specific characteristics (e.g. production volume, detailed NAICS, climate, etc.). Unlike the series of reports by (Glatt, Naranjo et al. 2016, Glatt, Naranjo et al. 2016, Glatt, Naranjo et al. 2016, Glatt, Naranjo et al. 2016) who use Census region aggregate data, these potential confounders that can lead to differences in plant energy use are precisely measured in our micro-data, and are not treated as differences in energy efficiency. By conducting the analysis at the plant level the model is capable of representing the efficiency distribution at various levels of aggregation, specifically in terms of geography and industry, without losing corresponding detail. In the context of NEMS, many of these plant/industry characteristics (e.g. size, 6-digit NAICS) are not amenable to forecasting. As such, they are likely to be assumed constant in the future. What is important is that differences in energy use that arise from these other characteristics, such as the distribution in plant size or detailed NAICS, is not confused with energy efficiency.

We conduct a plant level analysis of each of the five 3-digit NAICS that comprise MBD (see table 1).

TABLE 1 INDUSTRIAL DEMAND MODULE MBD INDUSTRY AGGREGATIONS

IDM Industry Code	Industry Description	NAICS Code
14	Fabricated Metal Products	332
15	Machinery	333
16	Computer and Electronic Products	334
17	Transportation equipment	336
18	Electrical Equip., Appliances, and Component	335

Electricity and total fuel use[2] are analyzed separately for each 3-digit sector, for a total of 10 sector-fuel combinations. The analysis time period includes the years 1987, 1992, 1997, 2002, 2007, and 2012; 6 time steps from the quinquennial Census of Manufacturing (CM).

The focus is to obtain measures of the frontier (or "best practice") energy use, conditional on production activity, prices, etc., to obtain the underlying distribution in plant level energy efficiency relative to the estimated frontier, over time. The SFA has the advantage that the estimated model parameters can be used by EIA for forecasting, while the DEA approach would only provide the plant level efficiency estimates. SFA allows the analysis to directly control for a variety of effects, such as plant size, labor intensity, detailed NAICS, etc. as described below. The SFA approach is also better suited to Census disclosure requirements.

This paper is organized as follows. A brief literature review of frontier methods as applied to energy efficiency, in particular manufacturing, is presented. Next we describe the sources for the data. A modeling section describes the general approach, a brief overview of the Malmquist decomposition, and the final version of the models we estimate. The results section includes parameter estimates for the year 2012, the Malmquist estimates over the 25 year time period, and some panel models with year, region and state fixed effects.

2. Review of Production Efficiency Measurement of Energy

This paper draws on the theoretical and applied literature on measuring efficiency. For a broad review see (Murillo-Zamorano 2004). The relationship between the "traditional" energy efficiency literature

[2] For simplicity fuels will be treated as all Natural Gas, with other fuel use treated as de-minimis. The detailed implementation is described below.

and the literature on productivity measurement was first discussed by (Huntington 1995). While we focus on the SFA approach applied to energy efficiency measurement, the DEA approach has also been applied to manufacturing energy efficiency. One of the earliest papers applied DEA to energy efficiency in buildings (Ferrier and Hirschberg 1992). (Boyd, Karlson et al. 1993) and (Boyd and Pang 2000) use DEA to compute a production efficiency and relate it to energy efficiency in steel and paper mills, respectively. (Mukherjee 2008) apply DEA to aggregate time series data. (Zhang, Lundgren et al. 2016) apply DEA to firm level manufacturing data in Sweden to examine the impact of the carbon emissions trading scheme in the European Union. (Bostian, Färe et al. 2016) use a network DEA model on similar Swedish data and also conduct a Malmquist decomposition. For a general review of energy and environmental application of DEA see (Zhou, Ang et al. 2008, Sueyoshi, Yuan et al. 2017).

The SFA approach used here falls within the general class of problems of non-radial, input specific efficiency measures; (Filippini and Hunt 2015) describes three similar approaches. The first two measures of technical efficiency via the factor requirements function (Boyd 2005, Boyd, Dutrow et al. 2008, Boyd 2008, Boyd 2014) which is shown by (Boyd 2008) to be equivalent to a directional distance function (Färe and Grosskopf 2000, Zhou, Ang et al. 2012). (Stern 2012) applies a similar concept as (Boyd 2008), in as much as he introduces the notion of the "energy only" direction in the context of a production frontier, but does not explicitly connect it to the directional distance function. The third approach expands the notion of efficiency to include both allocative and technical efficiency (Filippini and Hunt 2011). This paper also expands the concept of efficiency to include allocative efficiency via introducing energy prices in an ad-hoc energy demand equation as in (Filippini and Hunt 2011, Filippini and Hunt 2012) (Lundgren, Marklund et al. 2016). Other research on energy efficiency using SFA include (Adetutu, Glass et al. 2016) who apply SFA to OECD level data to estimate rebound effects. (Aranda-Usón, Ferreira et al. 2012) and (Feijoó, Franco et al. 2002) apply SFA to energy use in the Spanish manufacturing sector. (Lin and Long 2015) analyze the chemical sector in China using provincial level data.

2.1 SFA Representation of Efficiency

The SFA approach to energy may be broadly represented as

$$lnE_{j,i,t} = f(\,) + v_{j,i,t} + u_{j,i,t} \qquad (1)$$

Where

j = energy type (electricity and fuel)

i = individual observation (i.e. manufacturing plant)

t = time unit of the observations

The last two terms represent statistical noise, $v_{j,i,t}$, and inefficiency, $u_{j,i,t}$, respectively. There are wide number of choices to represent the inefficiency term $u_{j,i,t}$. Recent methods include Green's "true" fixed and random effects at the individual (plant) level (Greene 2002) which would separate plant specific heterogeneity (also called persistent efficiency) with some time varying component. However, Green observes that it isn't clear if one should treat heterogeneity as separate from efficiency.

The approaches available for exogenous inefficiency effects (including time, industry, plant size, etc.) are numerous. One approach is implemented in the Normal-Truncated Normal model, where it is possible to compute non-monotonic effects of the exogenous factors on both the variance and location (truncation point) of the efficiency distribution, for example (Filippini and Hunt 2011) (Filippini and Hunt 2012).   This form for representing the efficiency distribution is of particular interest in the proposed modeling of higher energy prices and other influences as putting "pressure" on plants to operate closer to the frontier. State dummy variables, representing possible region specific policy effects or climate, could also be included in this type of Normal-Truncated Normal. This approach would make energy use (efficiency) sensitive to prices, but making no attempt to differentiate between allocative and technical inefficiency.

The SFA parameters for the function *f(.)* could allow for scenario analysis and possible forecasting of exogenous effects on both the frontier and the underlying efficiency. In addition, the SFA would allow the generation of individual plant level estimates of efficiency. These plant level efficiency estimates cannot be cleared by Census[3], but a non-parametric kernel density of the plant level efficiency distribution(s), using a random number support points as an approximation to the estimated kernel, can be released. This alternative has the advantage as a simple way to compute the levels of inefficiency at the quartile or decile level without violating disclosure rules.

2.2 Price Effects

The applied and theoretical literature on efficiency measurement includes both production and cost efficiency approaches. To apply a cost efficiency measurement approach to energy, either a fully specified cost function, or a separable energy demand approach is needed. This paper is motivated by the latter.   Energy prices provide the basis for the frontier cost (minimizing) level of energy demand, given output and quasi-fixed inputs. The approach used by (Lundgren, Marklund et al. 2016) is the most similar to our analysis.

When one considers how energy prices effect energy use over time, the notion of induced technical change is important. In the case of induced technical change, it is possible that the frontier shifts in response to energy price shocks by the introduction of new technologies, i.e. high prices motivate the invention or commercialization of new energy efficient technologies, e.g. LED lighting or home air conditioning (Newell, Jaffe et al. 1999).   In this case, the asymmetric approach proposed by (Gately and Huntington 2002) might be appropriate, i.e. using the highest past price as an indicator of induced technical change. Other approaches to energy price induced technical change include (Lansink, Silva et al. 2000). We would expect induced change to be a global, not local effect, so national average prices might be appropriate. On the other hand, past prices at the plant level could reflect a putty-clay form of hysteresis, where high plant level prices in the past resulted in irreversible investments. In this case, Gately and Huntington used an asymmetric price variable that distinguished between rising and falling prices, to assess if there is hysteresis, i.e. rising prices may have different impact on demand than falling prices, particularly in the long run. (Steinbuks and Neuhoff 2014) estimate a dynamic stock adjustment model over OECD-level aggregate industry data to obtain price effects on the efficiency of the vintage of

[3] Census clearances rules regarding confidentiality prohibit release of individual plant level estimates, including the parameters of the kernel which are based on supports in the neighborhood of each observation.

the capital as well as operational energy price effects. Our approach is similar in spirit to (Steinbuks and Neuhoff 2014) but differ in terms of the level of data and our reduced form approach.

3. Data

Data for the study are non-public plant-level Census Bureau data available in the Federal Statistical Research Data Center. These data are protected under Title 13 and 26 of the US Code and used with permission from the Bureau. The analysis time period includes the years 1987, 1992, 1997, 2002, 2007, and 2012; 6 time steps from the quinquennial Census of Manufacturing (CM).

Data needed for the analysis include energy use and prices along with production activities and other location specific variables. While the Manufacturing Energy Consumption Survey (MECS) provides the most detailed data on energy use, particularly cost and quantity of fuels by type, the MECS is a stratified sample and not a balanced panel so the presence (absence) of an observation is not an indicator of entry (exit) in the industry. We need this information on entry/exit/continuing status for the Malmquist decomposition and comparison of the relative efficiency of entering versus continuing plants. Using the Census of Manufacturing (CM), part of the quinquennial Economic Census (EC) solves this problem.

The availability of plant level electricity use and prices in the EC is also a major advantage of this data set. (Davis, Grim et al. 2012) analyze the dispersion of those prices in detail. However, the EC only reports cost of fuels, not quantities, so Btu fuel consumption is imputed from fuel costs in the CM using the assumption that most fuel use in this sector is natural gas. This is a reasonable assumption for the metal based durables industries that we focus on, because publicly available MECS data from 2010 for these 5 sectors shows that 88% to 98% the purchased fuel in this sector is natural gas. We impute Btu consumption by taking the cost of fuels and dividing by the state level natural gas prices as published by the EIA's State Energy Data System (SEDS).[4] The CM provides plant level electricity consumption and costs, from which a plant level price can be computed. Plant level shipment values, adjusted for inventory changes are used to measure production. Labor is measured in production worker hours. Capital stock is based on a perpetual inventory method. All data in $ values are deflated using the (Bartelsman and Gray 1996) NBER 6-digit NAICS price deflators. The ZIP code location of the plant is merged with NOAA weather station data to get a plant specific heating and cooling degree day (HDD and CDD) measure as a control for the energy impact of location and time specific climate conditions. The section that follows provides an overview of the empirical models used to estimate plant-level energy efficiency using the EC data.

4. General Modeling Approach

The general approach is to estimate a stochastic frontier (SF) for energy and use the resulting plant level efficiency measures as an estimate of a distance function. The frontier we estimate has features in common with an energy factor requirements equation, but the presence of prices also reflect an ad-hoc

[4] SEDS data is available online at the following: http://www.eia.gov/state/seds/ (last accessed November, 2016).

energy demand approach. We interpret the efficiency estimates as a (energy sub-vector) directional distance function, with prices providing additional information regarding the optimal direction. These distance function estimates can then be used in the Malmquist decomposition. The following sections lay out the econometric and Malmquist approach in more detail.

4.1 Stochastic Frontier Applied to Energy Efficiency

We estimate a SF ad hoc energy demand equation for the two primary energy types in each of the five 3-digit NAICS that comprise this sector. The use of a log formulation will facilitate the NEMS stock adjustment modeling approach to measure relative energy intensity while controlling for a range of other effects, since the plant estimates of inefficiency, in log form, can be interpreted as percent differences from the frontier and can easily be transformed into percent differences from the mean. We consider models of the general form,

$$lnE_{j,i,t} = f\left(lnY_{i,t}, lnEmp_{i,t}, lnK, P_{j,t,s}, DYear_t, DNAICS_k, DState_s, CDD_{i,t}, HDD_{i,t}\right) + v_{j,i,t} + u_{j,i,t} \tag{2}$$

Where

j = energy type (electricity and fuel)

i = individual establishment (i.e. manufacturing plant)

t = year of the observations i.e. 1987, 1992, 1997, 2002, 2007, and 2012

k = 6-digit NAICS

s = state

The last two terms represent statistical noise, $v_{j,i,t}$, and inefficiency, $u_{j,i,t}$, respectively. We will return to specific approaches to the distributional assumptions of $u_{j,i,t}$ below.

Both total value of shipments (TVS) and value added (VA) are considered in the model as measures of activity, lnY. VA might be preferred since it partially controls for the "make/buy" element of the materials/energy substitution at the plant level, however, our results do not indicate any statistically significant differences among models dependent upon choice of production activity. As a result, our preferred models use TVS as the dependent variable, because this is the aggregate output variable employed in NEMS for forecasting purposes. lnK controls for productive capacity/capability. Labor, measured by $lnEmp$, controls for plant level utilization effects,[5] since labor may be sticky in the short run for The long run relationship between energy and plant scale is captured by the combined coefficient on production, capital and labor, lnK lnY and $lnEmp$. In a simple Cobb-Douglas specification the sum of the coefficients on lnY and $lnEmp$ reflect the economies of scale with respect to energy. In other words, we are interested in knowing if a plant producing twice the output and using twice the capital and labor will use twice the energy. If f(.) is log linear in parameters, and the sum of the associated coefficients is less than one then we can infer that larger plants will have proportionally

[5] Using the 5-year Economic Census also conveniently avoids the years of the Great Recession by including 2007 and 2012, but not the intervening years.

lower energy use than proportionally smaller plants. This means that the model will control for advantage due to economies of scale with respect to energy use directly in the plant specific frontier so that differences in plant scale (size) will not be subsumed into the energy efficiency measure.

Even within a 3-digit NAICS there can be a lot of heterogeneity with respect to energy services, so 6 digit NAICS industry controls are used for the primary analysis. One could consider 10 digit product level dummies as well, since the CM has such detail. (Boyd 2016) reviews industry specific case studies of energy use that employ some of this finer product detail. However, doing so would require very specific prior information about which product level NAICS are more/less intensive, since there are a very large number of 10-digit product NAICS. We believe that the 6 digit controls are quite sufficient and provide more detail than most industrial energy studies have employed before with the exception of (Boyd and Curtis 2014) who also use plant level Census micro-data.

The price variables, $P_{j,t,s}$, reflects the impact of the prices of both electricity and natural gas on the frontier level of energy use. Incorporating prices into the factor requirement function allows us to measure price responsiveness of the sectors. If we view the model in a production function context then higher energy prices could act as an exogenous shifter of the frontier, i.e. induced technical change. The prices of both types of energy (j = electricity and fuel) may impact either energy type. Variation in energy prices can be used to capture price incentives and allocative efficiency. Electricity and fuel have different data issues, so the treatment of prices will also have to be different. Census data collects plant level cost and quantity for electricity but only costs for fossil fuels. The problem with using plant level electric prices[6] directly in the model is that the plant may have some bargaining power or simply more choice over rate plans, with larger electricity users realizing lower average prices, resulting in an endogenous variable. Although not reported, we considered instruments for this using either state level prices from SEDS or county level average prices constructed directly from the Census data. Our IV estimates are generally similar to the SFA results presented herein, and Hausman tests for endogeneity of prices reject the null hypothesis that plant-level prices are endogenous in the MBD sector.

The "D" prefix in $DYear_t, DNAICS_k, \text{and } DState_s$above indicates dummy variables to control for various plant characteristics. State (or region) level dummies are used to capture a range of unobserved state characteristics. In particular, location dummies could capture differences in the regulatory / business environment, including state specific policies regarding energy efficiency. NAICS refers to the 6-digit industry code of the North American Industry Classification System within the corresponding 3-digit codes in table 3. This controls for industry level heterogeneity.

Heating Degree Days (HDD) and Cooling Degree Days (CDD) are used to control for ambient weather conditions on an annual basis using the zip-code location of the plant. "Weather" can impact building HVAC energy use, but also impact process energy via outside air to ovens and furnaces or chiller efficiencies, to the extent that the production requires these process. While one may expect that HDD would impact fuel use and CDD affects electricity use, both variables are included in each one of the energy regression equations.

Modeling electricity and fuel separately has advantages, since sector specific process needs will differ in terms of energy type. However, there may be opportunities to substitute electricity for fuel, combined

[6] These prices are not true marginal prices, but include demand charges, etc. They are total expenditures divided by total consumption.

heat and power being the most obvious. However, CHP in MBD is present in a very small percent of the plants. This variable was excluded in the results presented below since the number of non-zero observations were so small they would have presented confidentiality disclosure issues.

The initial focus was on several different versions of the SFA estimated in the cross-section; plant level random effects and models that combine fixed/random effects with SFA (i.e. the “true” fixed and random effects as presented by Green 2002). Despite the attractive nature of a panel data analysis with time varying efficiency and plant level heterogeneity, all the panel model specifications attempting to control for unobserved time invariant plant heterogeneity failed to converge. The problem of the fixed and random effect estimation approach when the number of establishment is very large relative to the number of time periods is well established. The panels suffered from the problem of large numbers of establishments (ranging from a few thousand to over 20 thousand depending on the NAICS) and a small number of time periods (6 five year increments). This is complicated by the large amount of entry and exit in these industries, so that in practice almost no plants are in the data for all six of the years included in the analysis.

After the initial analyses, the focus was shifted to obtaining parameter estimates and efficiency distributions from some of the simpler formulations for the SFA. We have focused on estimates from a sub-set of the available methods using year specific repeated cross sections.

The basic specification using repeated cross sections remains as above, but without the state dummies. It is not possible to use state dummies because the natural gas price is state level.

$$lnE_{j,i,t} = f\left(lnY_{i,t}, lnEmp_{i,t}\ , DNAICS_k, P_{Elec,t,s}, P_{Fuel,t,s}, CDD_{i,t}, HDD_{i,t},\right) + v_{j,i,t} + u_{j,i,t} \quad (3)$$

The SFA approach assumes that $u_{j,i,t}$ is follows a one-sided error distribution. Exponential and half-normal distributions, with variance σ_{u_j}. These are two of the distributions that have closed form likelihood functions and are commonly used in the SFA literature. Truncated normal and Gamma distributions require the estimation of two distributional parameters and were too computationally challenging for our data sets, so we focus on exponential and half-normal distributions. We choose to take advantage of the large number of plants in each year of the data and use a cross sectional approach for each 5 year period. We use the repeated cross sections and also examine these model for stability of the elasticities and efficiency distributions over time.[7]

To obtain an estimate of inefficiency we first take the estimate of the residual from the model

$$\widehat{\varepsilon_{j,i,t}} = v_{j,i,t} + u_{j,i,t} = lnE_{j,i,t} - f\left(lnY_{i,t}, lnEmp_{i,t}\ , DNAICS_k, P_{Elec,t,s}, P_{Fuel,t,s}, CDD_{i,t}, HDD_{i,t},\right) \quad (4)$$

Using standard methods developed in the SFA literature we compute the parameters for the function f() and the expectation of the efficiency component conditional on the residuals, using the mean or mode, and convert to a measure of inefficiency see (Murillo-Zamorano 2004).

$$\widehat{u_{j,i,t}} = \boldsymbol{E}\left(u_{j,i,t} \middle| \widehat{\varepsilon_{j,i,t}}\right) \ or\ \boldsymbol{M}\left(u_{j,i,t} \middle| \widehat{\varepsilon_{j,i,t}}\right) \quad (6)$$

$$Eff_{j,i,t}^{SF,m} = e^{-\widehat{u_{j,i,t}}} \quad (7)$$

[7] We retain the time subscript in the notation since we will get estimates for each time period, but not from a panel approach.

This estimate of efficiency is related to the distance function used as the basis for the Malmquist decomposition, as described in the next section.

4.2 Decomposing technical change with the Malmquist index

The general approach for measuring technical change is based on the Malmquist index. Following (Färe and Grosskopf 1992), to define the Malmquist index we must first define the input distance function.

$$D_i^t(y^t, x^t) = sup\left\{\lambda > 0: \left(\frac{x^t}{\lambda}\right) \epsilon\, L^t(y^t)\right\} \qquad (8)$$

Where $L^t(y^t)$ is the input correspondence of the production technology set

$$S^t = \{(y^t, x^t):\ x^t\ can\ produce\ y^t\}. \qquad (9)$$

The input correspondence follows the basic production axioms outlined in (Färe and Grosskopf 1992). The distance function is the largest scalar value, λ, by which we can reduce the input vector, x^t, and still produce the output vector, y^t.

Following (Balk 2001) the input oriented, Malmquist index of aggregate change between t=1 and t=0 is defined as the geometric mean of the ratios of the distance function in each time period, evaluated at the observed input-output combination in each time period.

$$M_i^1(y^1, x^1, y^0, x^0) = \left[\frac{D_i^0(y^0,x^0)}{D_i^0(y^1,x^1)}\frac{D_i^1(y^0,x^0)}{D_i^1(y^1,x^1)}\right]^{\frac{1}{2}} \qquad (10)$$

Not that the Malmquist index requires two distance functions defined over each time period. We can decompose this index of aggregate change into two components, efficiency change and frontier technical change. First, we define the Malmquist index of efficiency change as

$$ME_i^1(y^1, x^1, y^0, x^0) = \left[\frac{D_i^0(y^0,x^0)}{D_i^1(y^1,x^1)}\right] \qquad (11)$$

One can think of efficiency change as “observable” progress –the change in technical efficiency represents the change (ratio) between period 0 and period 1 in the distance between an observation and the frontier. Given this notion of efficiency change the measure of frontier technical change (here after just technical change) can be derived algebraically as

$$MT_i^1(y^1, x^1, y^0, x^0) = \left[\frac{D_i^1(y^1,x^1)}{D_i^0(y^1,x^1)}\frac{D_i^1(y^0,x^0)}{D_i^0(y^0,x^0)}\right]^{\frac{1}{2}} \qquad (12)$$

Such that

$$M_i^1 = ME_i^1 \cdot MT_i^1 \qquad (13)$$

Empirically the distance functions are often computed using DEA, but can also be computed using estimates from the stochastic frontier (Coelli, Rao et al. 2005). Specifically, $D_i^t(y^t, x^t) \geq 1$ and the efficiency score $Eff_{j,i,t}^{SF,m}$ranges between 0 and 1 so.

$$D_i^t(y^t, x^t) = 1/Eff_{j,i,t}^{SF,m} = 1/e^{-\widehat{u_{j,i,t}}} \qquad (14)$$

However, we are interested, not in the distance function defined over the entire vector of inputs, but on the energy sub-vector (directional) distance function as defined by (Boyd 2008). In addition, the interpretation of the directional distance function used by (Boyd 2008) needs to be modified in this context, due to the inclusion of prices in the efficiency measure and resulting distance. Consider the difference between the standard Shepard input distance function as shown in figure 1 and defined by the line AB. The sub-vector (energy directional) distance function used by (Boyd 2008) is defined by the line AC. Now consider the case shown in figure 2 where the isoquant and the cost minimizing input combinations are shown for two different sets of energy prices. The relevant direction for the distance function is either AD' or AD'', depending on the energy prices. At higher energy prices, the relative efficiency of point A would be $E\text{-}E_2^* < E\text{-}E_1^*$. In the DEA formulation of the distance function, one would need to specify the direction vector, but in the case of the SFA approach the inclusion of energy prices make the optimal direction implicit in the computed measure of efficiency.

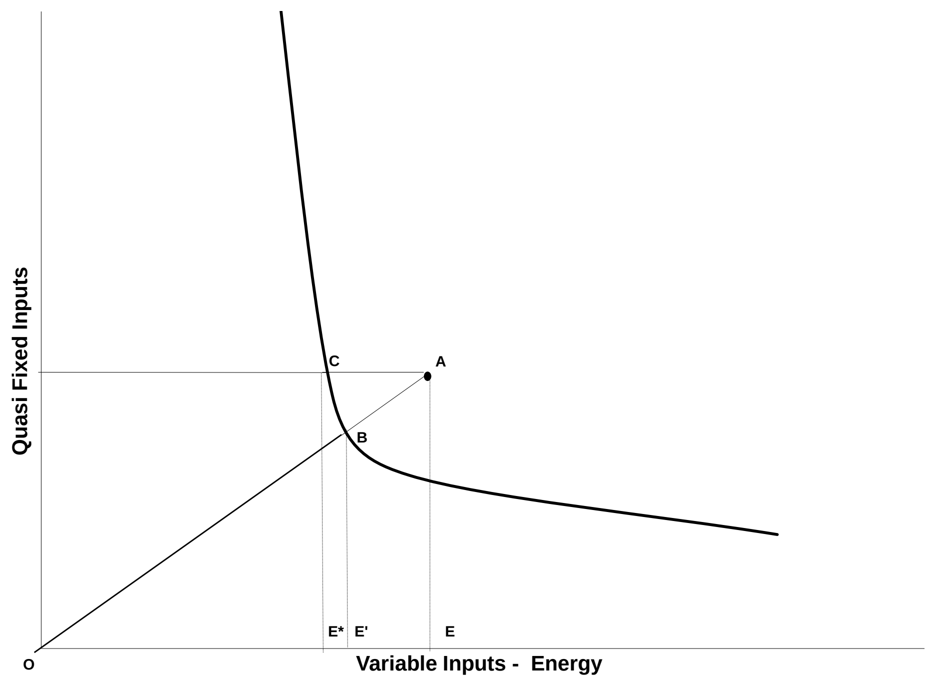

FIGURE 1 COMPARISON OF THE STANDARD INPUT DISTANCE FUNCTION AND THE SUB-VECTOR DISTANCE FUNCTION

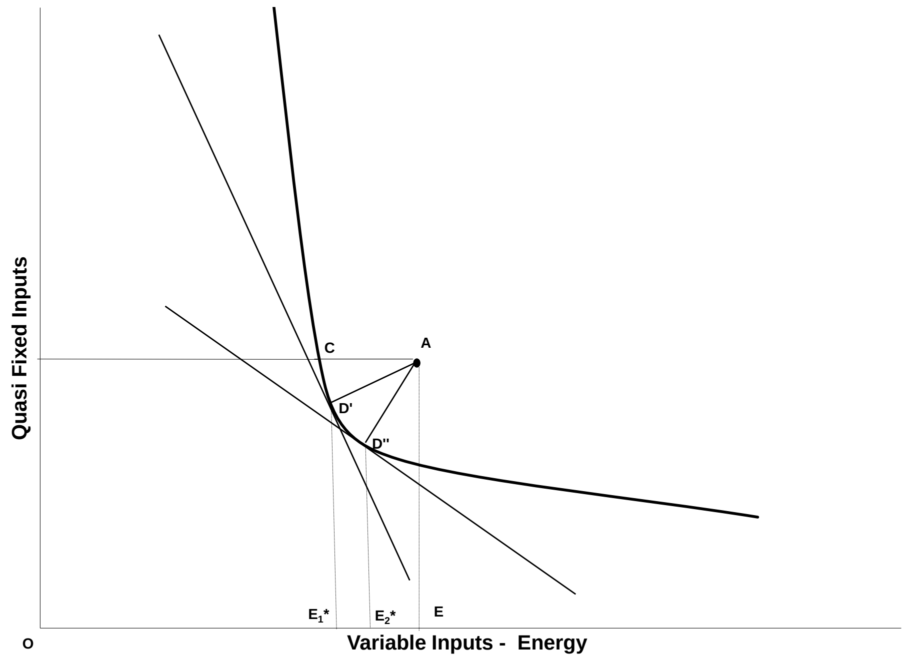

FIGURE 2 COMPARISON OF OPTIMAL PRICE DIRECTION DISTANCE FUNCTIONS

When using a non-parametric data envelopment analysis (DEA) model to calculate a Malmquist productivity measure and decomposing it into its technical change and change in technical efficiency components, mixed-period linear programming problems must be specified and solved. Our approach uses parametric SFA, similar to (Pantzios, Karagiannis et al. 2011) but with one important difference. They parameterize the time component in a pooled translog, model while we take advantage of the large sample size of the micro data and use repeated cross sections to estimate a year specific distance function. The mixed period results needed for the decomposition, e.g. $D_i^1(y^0, x^0)$, are generated by the out of sample prediction using the observed valued for (y^0, x^0) in the distance function estimated for year=1. This approach also allows us to avoid the econometric problems and additional assumptions needed for estimation of panel data frontier.

It is also important to recognize that the Malmquist index is generally use for output oriented distance functions; whether they are DEA or SFA based. (Kumbhakar and Tsionas 2006) show that SFA estimates using input and output orientation can differ. We are input oriented in a single factor direction so we are confident the Malmquist approach is still valid in our case.

5. Results

The results presented here are based on a repeated cross section estimation for each of the quinquennial EC years as described above. In addition, a pooled analysis was performed that assumes the concerns about time series correlation at the plant level is minimal, given the 5-year time steps. These panel results are not our preferred models, but are presented to provide an alternative view of technical change vis-à-vis the Malmquist analysis on the repeated cross sections.

One might argue that the five year time step is sufficiently long to assume that the error terms are uncorrelated and simply pool the entire sample and ignore the panel aspect of the data. We present those results based on that simplifying assumption in order to examine time fixed effects. Since the number of establishments is large we exploit the cross-sectional variation in each year to estimate a year specific model as the basis for the estimates of energy efficiency.

Our version of the more general model is estimated for each energy type, 3-digit NAICS and year.

$$lnE_{j,l,t,i} = \alpha + \beta^{y}_{j,l,t} lnY_{i,t} + \beta^{Emp}_{j,l,t} lnEmp_{i,t} + \beta^{K}_{j,l,t} lnK_{i,t} + \gamma^{E}_{j,l,t} lnP_{Elec,t,i} + \gamma^{F}_{j,l,t} lnP_{Fuel,t,s} + \delta^{CDD}_{j,l,t} lnCDD_{i,t} + \delta^{HDD}_{j,l,t} lnHDD_{i,t} + \sum_{k \in l} \alpha_k DNAICS_k + v_{j,K,i,t} + u_{j,K,i,t}$$

Where

j = the type energy (electricity and fuel respectively)

t = year of the observations i.e. 1987, 1992, 1997, 2002, 2007, and 2012

i = individual establishment (i.e. manufacturing plant)

s = state location for the individual establishment (i.e. manufacturing plant)

l = one of the 5, 3 digit NAICS comprising MBD

k = a 6 digit NAICS associated with the Kth 3-digit NAICS[8]

The last two terms represent statistical noise, $v_{j,i,t}$, and inefficiency, $u_{j,i,t}$, respectively. Half normal and exponential distributional assumptions for $u_{j,i,t}$ were used and are also compared to OLS parameter estimates.

5.1 SFA Parameter estimates

Parameter estimates for 2012 are presented in tables 2-6 for each of the 3-digit MBD NAICS. The parameter estimates are fairly stable over time; 2012 results are felt to be generally representative and most relevant to forecasting. SFA results under the two commonly used distributional assumption for the one-sided efficiency terms, half-normal and exponential distribution, are presented along with OLS for comparison purposes. All p-values are based on robust standard errors. Iteration counts above 20 generally indicate models that fail to converge resulting in estimates that are almost identical to OLS and with no estimates of efficiency.

[88] 6-digit NAICS parameter estimates were not requested to be cleared by Census; many of these industry fixed effects are significant based on simple t-tests.

Table 2 NAICS 332 Estimates for 2012

VARIABLES	OLS Electricity	SFA (Half Normal) Electricity	SFA (Exponential) Electricity	OLS Fuel	SFA (Half Normal) Fuel	SFA (Exponential) Fuel
lnY	0.381***	0.386***	0.391***	0.225***	0.234***	0.221***
$lnEmp$	0.551***	0.551***	0.553***	0.487***	0.464***	0.490***
lnK	0.145***	0.141***	0.135***	0.116***	0.0919***	0.0890***
lnP_{Elec}	-0.933***	-0.927***	-0.919***	-0.340***	-0.290***	-0.285***
lnP_{Fuel}	0.109**	0.105**	0.0972**	-0.719***	-0.719***	-0.723***
$lnHDD$	0.0318	0.0307	0.0289	0.165***	0.183***	0.190***
$lnCDD$	0.0510**	0.0499**	0.0478**	0.0576*	0.0488	0.0467
Constant	-0.202	-1.350***	-1.256***	2.329***	0.321	0.678
$ln\sigma_u^2$		-1.376***	-2.218***		0.683***	-0.492***
$ln\sigma_v^2$		-0.846***	-0.888***		-0.757***	-0.501***
Returns to scale	1.077	1.078	1.079	0.828	0.789	0.8
Observations	23063	23063	23063	23063	23063	23063
R-squared	0.823			0.625		
Iteration Count		47	7		8	6
*** p<0.01, ** p<0.05, * p<0.1						
mean efficiency		NA	81%		35%	48%
median efficiency		NA	86%		41%	60%
σ_u		0.0012	0.2138		1.3245	0.7382
σ_v		0.6980	0.6620		0.6873	0.7734
σ_u / σ_v		0.0017	0.3230		1.9271	0.9546

Table 3 NAICS 333 Estimates for 2012

VARIABLES	OLS Electricity	SFA (Half Normal) Electricity	SFA (Exponential) Electricity	OLS Fuel	SFA (Half Normal) Fuel	SFA (Exponential) Fuel
lnY	0.437***	0.450***	0.461***	0.366***	0.386***	0.386***
$lnEmp$	0.507***	0.499***	0.494***	0.475***	0.434***	0.449***
lnK	0.136***	0.133***	0.129***	0.0657***	0.0502***	0.0487***
lnP_{Elec}	-0.825***	-0.812***	-0.797***	-0.0802*	-0.0586	-0.0549
lnP_{Fuel}	-0.0704**	-0.0719***	-0.0767***	-0.907***	-0.914***	-0.916***
$lnHDD$	0.0470***	0.0507***	0.0574***	0.125***	0.146***	0.157***
$lnCDD$	0.0363**	0.0384**	0.0420***	-0.0128	-0.00966	-0.00675
Constant	-0.767***	-1.170***	-1.153***	2.851***	1.483***	1.673***
$ln\sigma_u^2$		-1.041***	-2.071***		0.626***	-0.547***
$ln\sigma_v^2$		-1.065***	-1.071***		-0.869***	-0.589***
Returns to scale	1.08	1.082	1.084	0.906	0.87	0.883
Observations	9987	9987	9987	9987	9987	9987
R-squared	0.815			0.619		
Iteration Count		10	7		7	5
*** p<0.01, ** p<0.05, * p<0.1						
mean efficiency		62%	70%		34%	47%
median efficiency		67%	78%		40%	59%
σ_u		0.5945	0.3550		1.3675	0.7607
σ_v		0.5871	0.5854		0.6476	0.7449
σ_u / σ_v		1.0126	0.6065		2.1117	1.0212

Table 4 NAICS 334 Estimates for 2012

VARIABLES	OLS Electricity	SFA (Half Normal) Electricity	SFA (Exponential) Electricity	OLS Fuel	SFA (Half Normal) Fuel	SFA (Exponential) Fuel
lnY	0.381***	0.386***	0.391***	0.225***	0.234***	0.221***
$lnEmp$	0.551***	0.551***	0.553***	0.487***	0.464***	0.490***
lnK	0.145***	0.141***	0.135***	0.116***	0.0919***	0.0890***
lnP_{Elec}	-0.933***	-0.927***	-0.919***	-0.340***	-0.290***	-0.285***
lnP_{Fuel}	0.109**	0.105**	0.0972**	-0.719***	-0.719***	-0.723***
$lnHDD$	0.0318	0.0307	0.0289	0.165***	0.183***	0.190***
$lnCDD$	0.0510**	0.0499**	0.0478**	0.0576*	0.0488	0.0467
Constant	-0.202	-1.350***	-1.256***	2.329***	0.321	0.678
$ln\sigma_u^2$		-1.376***	-2.218***		0.683***	-0.492***
$ln\sigma_v^2$		-0.846***	-0.888***		-0.757***	-0.501***
Returns to scale	1.077	1.078	1.079	0.828	0.789	0.8
Observations	3409	3409	3409	3409	3409	3409
R-squared	0.846			0.588		
Iteration Count		12	7		7	6
*** p<0.01, ** p<0.05, * p<0.1						
mean efficiency		67%	72%		33%	46%
median efficiency		71%	80%		39%	58%
σ_u		0.5041	0.3299		1.4071	0.7819
σ_v		0.6551	0.6415		0.6849	0.7784
σ_u / σ_v		0.7695	0.5143		2.0544	1.0045

Table 5 NAICS 335 Estimates for 2012

VARIABLES	OLS Electricity	SFA (Half Normal) Electricity	SFA (Exponential) Electricity	OLS Fuel	SFA (Half Normal) Fuel	SFA (Exponential) Fuel
lnY	0.470***	0.475***	0.488***	0.351***	0.355***	0.356***
lnK	0.445***	0.445***	0.445***	0.419***	0.413***	0.426***
$lnEmp$	0.167***	0.160***	0.147***	0.164***	0.147***	0.134***
lnP_{Elec}	-0.881***	-0.893***	-0.951***	0.0336	-0.0423	-0.148
lnP_{Fuel}	0.0248	0.0302	0.0402	-0.763***	-0.751***	-0.734***
$lnHDD$	0.0973**	0.0961**	0.0858**	0.205***	0.196***	0.184***
$lnCDD$	0.146***	0.150***	0.144***	0.0457	0.0157	0.00223
Constant	-2.983***	-3.051***	-3.023***	1.671*	0.198	0.5
$ln\sigma_u^2$		-1.319**	-2.200***		0.413***	-0.890***
$ln\sigma_v^2$		-0.979***	-1.028***		-0.397***	-0.211**
Returns to scale	1.082	1.08	1.08	0.934	0.915	0.916
Observations	2067	2067	2067	2067	2067	2067
R-squared	0.875			0.648		
Iteration Count		9	6		7	6
*** p<0.01, ** p<0.05, * p<0.1						
mean efficiency		69%	71%		43%	57%
median efficiency		73%	79%		49%	68%
σ_u		0.4677	0.3364		1.0544	0.5641
σ_v		0.6825	0.6486		0.8963	0.9422
σ_u / σ_v		0.6852	0.5187		1.1764	0.5987

TABLE 6 NAICS 336 ESTIMATES FOR 2012

VARIABLES	OLS Electricity	SFA (Half Normal) Electricity	SFA (Exponential) Electricity	OLS Fuel	SFA (Half Normal) Fuel	SFA (Exponential) Fuel
lnY	0.344***	0.340***	0.322***	0.191***	0.180***	0.161***
$lnEmp$	0.539***	0.549***	0.585***	0.596***	0.612***	0.640***
lnK	0.157***	0.153***	0.145***	0.154***	0.148***	0.145***
lnP_{Elec}	-1.010***	-1.004***	-0.979***	-0.220***	-0.211***	-0.205***
lnP_{Fuel}	-0.0231	-0.0264	-0.0364	-0.839***	-0.821***	-0.821***
$lnHDD$	0.0666***	0.0637***	0.0573***	0.171***	0.177***	0.179***
$lnCDD$	0.0571***	0.0561**	0.0543**	-0.0267	-0.0191	-0.0212
Constant	-0.799	-0.991**	-0.737**	2.660***	1.987***	2.354***
$ln\sigma_u^2$		-1.525**	-2.179***		0.106	-1.145***
$ln\sigma_v^2$		-0.764***	-0.866***		-0.219***	-0.119**
Returns to scale	1.04	1.042	1.052	0.941	0.94	0.946
Observations	4383	4383	4383	4383	4383	4383
R-squared	0.873			0.679		
Iteration Count		12	6		7	6
*** p<0.01, ** p<0.05, * p<0.1						
mean efficiency		69%	71%		43%	57%
median efficiency		73%	79%		49%	68%
σ_u		0.4677	0.3364		1.0544	0.5641
σ_v		0.6825	0.6486		0.8963	0.9422
σ_u / σ_v		0.6852	0.5187		1.1764	0.5987

In these results, total value of shipments is used in the model as measures of activity, lnY. Since labor is assumed to be sticky in the short run relative to energy use, $lnEmp$ is treated as a quasi-fixed input relative to energy. It is unlikely that employment would adjust to differences in energy, since energy is a very small cost element in this sector. Labor can also control for some plant-level utilization effects. The long run relationship between energy and plant scale is captured by the combined coefficient on production, capital, and labor, lnY lnK, and $lnEmp$. As described above, we are interested in knowing if a plant producing twice the output and using twice the capital and labor will use twice the energy. Generally, this energy returns to scale for electricity is close to unity or slightly larger, while for fuel it tends to be less than unity.

Own energy price elasticities range from -0.7 to -1.0. With electricity tending to have slightly higher elasticity than fuel. This is higher than the energy price elasticity of -.47 estimated by (Bentzen 2004) using aggregate data for the U.S and higher than the aggregate energy price elasticities reported by (Parker and Liddle). However, the estimates estimated here are more in line with (Steinbuks and Neuhoff 2014) who use a KLEM cost function approach for OECD countries and estimate energy price elasticities for Electrical and optical equipment between -0.38 and -0.66 and for Basic metals and fabricated metal products between -0.86 and -1.06.[9] These two industry categories are also closer to the MBD industries that are the focus of this study. As a cross section estimate, we would tend to interpret these as "long run" price elasticities, and would therefore expect them to be larger than short run elasticity estimates from time series estimates.

One concern regarding the electricity results is possible endogeneity of prices, since plant level electric prices are used. This is not the case for fuel elasticities, because state level prices are used in the estimation.[10] If large electricity users can negotiate better prices then this would likely bias our elasticity estimates upwards. Using both region and state level prices as an instrument for plant electric prices found no evidence of this type of endogeneity.[11] This may be due to the relatively low share of energy cost to value added for these sectors. We also note that while cross price elasticities were not constrained to be equal, but are not significantly different from zero for NAICS 332 (Fabricated Metal Products) and 335 (Electrical Equip., Appliances, and Components). We have no a-priori notions of electricity – fuel substitution in these sectors. On the other hand, the empirical pattern is for some statistically significant *complementarity* in 333, 334, and 336, mostly in the fuel equation. (Steinbuks 2012) points out that there are only a few categories of energy processes, i.e. production activates hat generate a "service demands" for energy, where there appear to be technical possibility for substitution. He finds that electricity is "poor substitute" for fuel in both the aggregate and for the "heating process" where substitution should be more feasible. His results are based on a cross section of all manufacturing. Ours is for assembly and fabrication and does not include industries that process primary raw materials like steel, glass, etc.

The weather effects are mostly as expected. HDD is large and significant in fuel use; CDD is not significant except in 332 (Fabricated Metal Products). Both CDD and HDD have significant impacts on electricity use in every NAICS but 334 (Computer and Electronic Products). Since electricity is used in

[9] Range reflect different modeling assumptions.
[10] Plant level fuel prices are only in the MECS and we are currently using CM data. CM data does allow us to compute plant prices.
[11] The instrumental variables analysis was done using 2-stage least squares, not SFA.

both heating and cooling, this result is consistent with our apriori expectations of the impact of weather on energy use.

The R-squared from the OLS models suggest a better fit for electricity than fuel. This may be due to the reliance on imputed fuel use (i.e. cost of fuels divided by average natural gas price). By the same token the SFA analysis shows much higher levels of inefficiency in the fuel model, so it may simply be that there is more variation in fuel use relative to the production and climate variables in the estimated frontier. The Half normal specification of the frontier model attributes more variance to (in)efficiency than noise, when compared to the exponential specification in 9 out of 10 fuel-NAICS equations;[12] in one case the half normal model did not converge and the estimates are basically equivalent to OLS. This finding may simply be a statistical artifact of the alternative distributional assumptions for the inefficiency component. It is worth noting that the average efficiency estimates are generally similar across the half normal and exponential estimates in the electricity models. These average efficiency estimates differ by as little as 6 percentage points in the transportation equipment sector (NAICS 336), and are within 9 percentage points of one another across the remaining sectors.

If we average across all 5 NAICS and model specifications the efficiency is shown below.

TABLE 7 AVERAGE MEAN AND MEDIAN ENERGY EFFICIENCY ACROSS NAICS AND SFA MODELS

	Electric Efficiency	Fuel Efficiency
Mean	67%	42%
Median	73%	51%

The mean and median efficiency estimates (based on the estimated error variances) range from a low of 32% (fuel use 334 - Computer and Electronic Products) to 86% (electricity use in 332 - Fabricated Metal Products). The exponential model has higher efficiencies than half normal. Fuel efficiency is consistently worse than electric efficiency for all NAICS. The estimates across industries in electric versus fuel efficiency is remarkable similar, with the exception of electricity use in NAICS 332. Plotting a kernel density for the plant level estimates of each type of energy efficiency, using the exponential model, are shown in Figures 3 and 4.

[12] Based on computing the ratio of σ_u to σ_v.

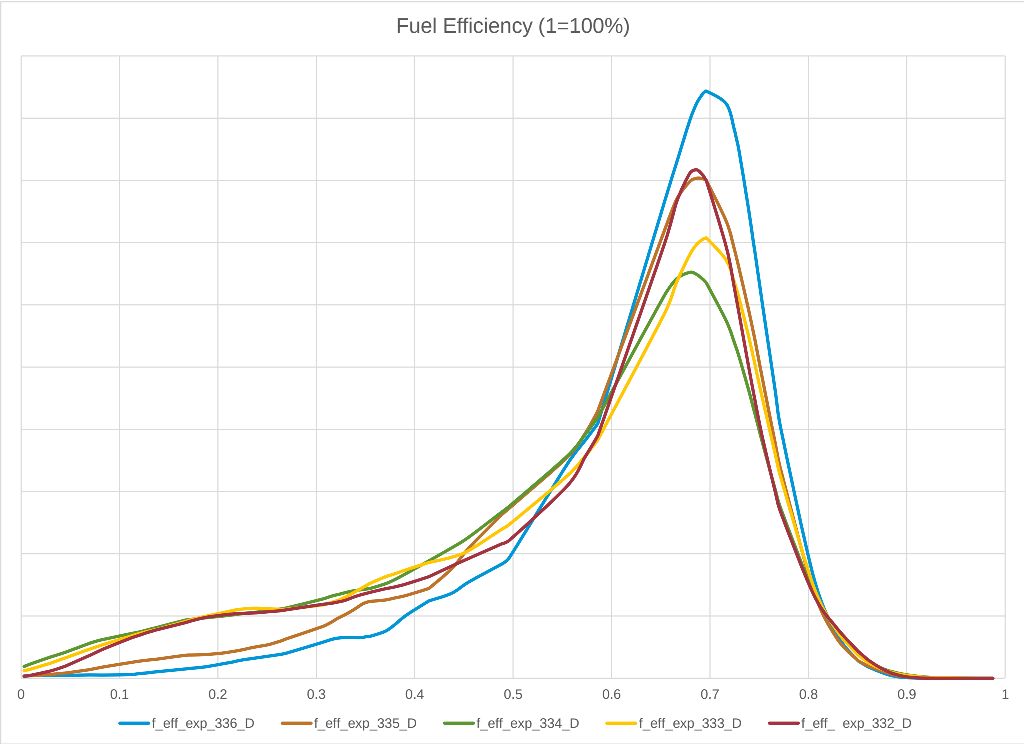

FIGURE 3 KERNEL DENSITY OF PLANT LEVEL FUEL EFFICIENCY ESTIMATES BY INDUSTRY (YEAR=2012, EXPONENTIAL DISTRIBUTION)

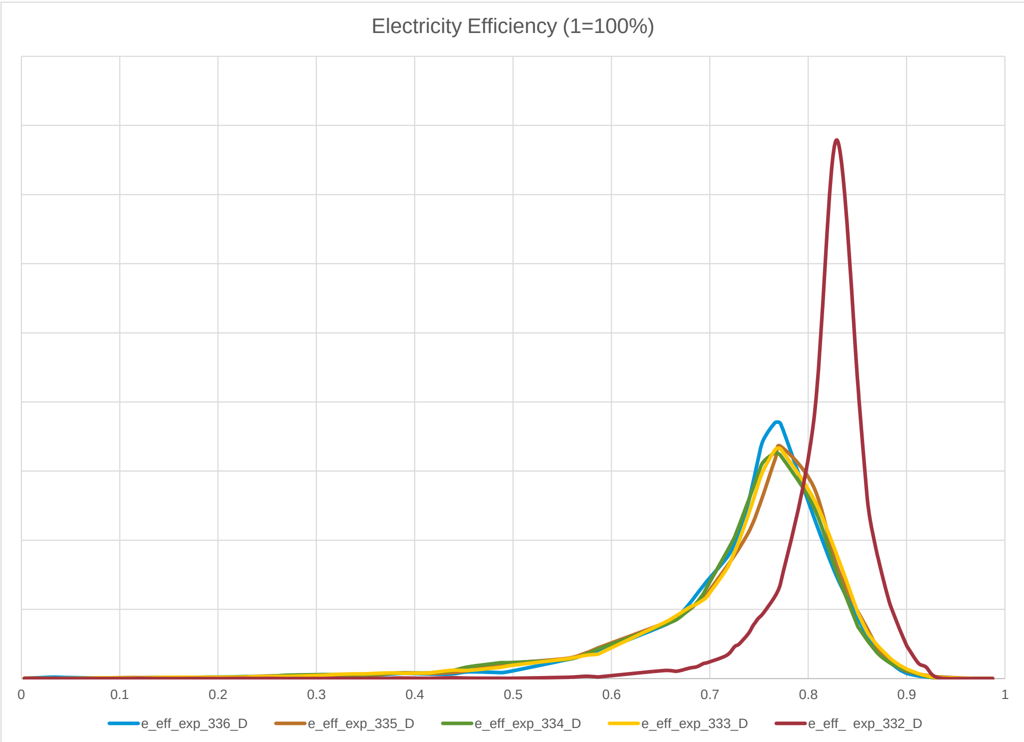

FIGURE 4 KERNEL DENSITY OF PLANT LEVEL ELECTRICITY EFFICIENCY ESTIMATES BY INDUSTRY (YEAR=2012, EXPONENTIAL DISTRIBUTION)

The kernels provide a non-parametric representation of the plant level efficiency distribution. They give more insight into the efficiency distribution than what is obtained by the half-normal and exponential parameter estimates presented above. If we compute the median efficiency from the non-parametric kernel density and compare it to the estimate of median efficiency directly from the variance parameters of the SFA distributions in tables 1-5 there is fairly close correspondence, but not perfect. Non-parametric median estimates of efficiency are between 1.5% and 3% points *higher* for electricity. For Fuels it is 0.5% to 3% points *lower* for all but NAICS 336. Another thing that is striking about the kernel density is that an extremely small number of plants have efficiency that exceed 90%. In fact, for fuel use only 2% of plants exceed 80% efficiency. This may be an artifact of the SFA analysis combined with the relatively high amount of noise versus efficiency in these model estimates.

Another way to look at these kernel estimate of efficiency is from a policy / energy management perspective. One can easily convert these distributions into cumulative densities and consider the quartile containing the least efficient plants; those plants consume a disproportionate share of the sector's total energy use. One might views these plants as the "low hanging fruit" in terms of potential energy savings. While this paper does not *identify* why those plants are the most inefficient, we can use the estimated distributions to *quantify* how much energy use would be involved if they could improve. We compute the change in energy use under the assumption that all plants in the most inefficient quartile improve to the median level of performance, for each industry and energy type, under the assumption that the variables on the right hand side of our estimating equation are randomly distributed relative to the estimated efficiency levels. We compute electricity reductions ranging from a low of 2.6% (NAICS 336) to around 8.5%-10% for the other industries. For fuel use, even though we set the target level of performance as the median, which is lower for fuel than electricity, the reductions are much higher due to the shape of the fuel efficiency distributions. Fuel efficiency has a much fatter tail. Reductions range from 20% to 65% for fuel use. The total across all industries and energy types is 21.4%

5.1 Malmquist Decomposition

The Malmquist index and decomposition results presented below are based on the repeated cross sectional estimates, without regional or state effects, for each NAICS and year from 1992 through 2012. Many of the frontier models did not converge for 1997 so we begin the Malmquist analysis in 1992 relative to the prior year. Models for NIACS 334, 335, and 336 do not converge for the year 2007, so the 2012 index for those NAICS are computed relative to 2002, i.e., a ten year time step.

The formula for the Malmquist index requires the distance function to be able to be computed for each time period and evaluated at the plant data point for the current and prior time period. Since we estimate a separate frontier for each year, computing $D_i^0(y^0, x^0)$ and $D_i^1(y^1, x^1)$ is straightforward using the within year sample residuals, $\widehat{\varepsilon_{J,\iota,t}}$, from the estimated year specific frontier (equation 6). And using the formula from equation (14) above,

$$\widehat{u_{J,\iota,t}} = \boldsymbol{E}\left(u_{j,i,t} \middle| \widehat{\varepsilon_{J,\iota,t}}\right) \; or \; \boldsymbol{M}\left(u_{j,i,t} \middle| \widehat{\varepsilon_{J,\iota,t}}\right) \qquad (6)$$

$$D_i^t(y^t, x^t) = 1/Eff_{j,i,t}^{SF,m} = 1/e^{-\widehat{u_{J,\iota,t}}} \qquad (14)$$

Computing $D_i^1(y^0, x^0)$ and $D_i^0(y^1, x^1)$ is computed for each continuing plant by calculating the out of sample residuals, i.e. plugging in data from the current year (y^1, x^1) into the frontier estimated for the prior year and taking the residuals to get the distance function. This implies the relevant data are from a balanced panel, which our data are not. (Kerstens and Van de Woestyne 2014) discuss the impact of unbalanced vs. balanced panel data. They point out that the approach of "balance on a 2-years by 2-years basis," which is the same as only computing the index for "continuing plants" (in our case over a five year time step) has some shortcomings. In particular it does not account for the role entry and exit plays, so the index is only a measure of industry change for surviving plants. We do examine empirically examine new (entering) plants, but do not attempt to integrate entry (or exit) into an encompassing index.

This is done for each NAICS and fuel types for every plant, using the repeated cross section exponential frontier estimates. We report a simple average of the three plant level indices, chained to 1992=1.0. The average rates of growth over the entire 20 year time period is reported in table 8 below. MI is the overall Malmquist index; ME is the Malmquist efficiency component; MT is the Malmquist technical change component. The overall rates of change for electricity are small with the exception of 335. All of the improvement is due to technical change, but in 336 the declines in efficiency completely offsets the higher rate of technical change. In other words, the frontier is shifting but on average the industry is not keeping up with the frontier. For fuels, the average level of efficiency is lower and the rates of change are higher. The failure to keep up with the frontier technical change, which ranges from less than 1% to over 4%, results in a decline in energy performance for sectors 334, 335, and 336, while 332 and 333 show a net improvement.

TABLE 8 AVERAGE ANNUAL GROWTH RATES FROM 1992-2012 FOR THE PLANT LEVEL GEOMETRIC AVERAGE MALMQUIST INDICIES

	Electric growth rate			Fuel growth rate		
	M	ME	MT	M	ME	MT
332 Fabricated Metal Products	0.32%	-0.06%	0.39%	0.64%	-2.73%	3.37%
333 Machinery	0.09%	-0.86%	0.94%	0.95%	-3.40%	4.34%
334 Computer and Electronic Products	0.18%	-0.27%	0.45%	-0.95%	-1.33%	0.38%
335 Electrical Equip., Appliances, & Component	0.92%	-0.18%	1.13%	-0.07%	-1.32%	1.25%
336 Transportation equipment	-0.04%	-0.72%	0.67%	-0.31%	-1.07%	0.75%

5.2 New entrants

The Malmquist index is only computed for continuing plants. In addition, there is a lot of turnover in these industries. Few plants exist for the entire 25 year time period and the great recession results in significant net exit between 2007 and 2012. One might expect that entrants would have newer technology and possibly be more efficient. To examine this we take each year for which we have a year specific frontier and compute the efficiency for entrants versus continuing plants. We test to see if the mean efficiency is significantly different between these groups. The results are shown in Table 9. While some differences are as small as 0.1%, most are statistically significant and the average entrant is more efficient than the average existing plant. The average difference over all the NAICS-year estimates is about 0.6% for electricity and 1.7% for fuels.

TABLE 9 EFFICIENCY OF NEW VERSUS EXISTING PLANTS (* INDICATES THE NEW PLANTS ARE HAVE SIGNIFICANTLY HIGHER EFFICIENCIES AT A 90% CONFIDENCE LEVEL OR HIGHER)

		Electricity Efficiency (percent)				Fuel Efficiency (percent)			
		New	Existing	Difference	t-test	New	Existing	Difference	t-test
332	1997	82.0%	81.9%	0.1%	1.20	74.8%*	74.2%	0.6%	6.38
	2002	75.1%	75.0%	0.1%	0.49	61.6%*	60.7%	0.9%	3.97
	2007	70.1%*	69.5%	0.5%	3.50	69.3%*	68.2%	1.1%	8.43
	2012	82.7%*	82.3%	0.5%	6.76	58.2%*	57.0%	1.1%	4.90
333	1997	79.4%	79.3%	0.0%	0.18	68.3%*	66.9%	1.4%	6.36
	2002	73.2%*	72.7%	0.6%	2.25	61.0%*	58.0%	2.9%	8.45
	2007	73.8%*	73.4%	0.3%	1.78	77.9%	78.0%	0.0%	-0.08
	2012	73.8%	74.0%	-0.2%	-0.76	56.6%	56.8%	-0.2%	-0.57
334	1997	78.5%*	77.9%	0.6%	2.81	62.8%*	60.8%	2.1%	5.06
	2002	73.3%*	70.8%	2.5%	6.96	58.8%*	54.6%	4.2%	7.68
	2007								
	2012	75.6%*	74.4%	1.3%	3.64	56.7%*	54.7%	2.0%	3.03
335	1997	75.1%	75.0%	0.1%	0.31	71.7%*	70.1%	1.6%	4.67
	2002	79.6%*	78.9%	0.7%	2.42	66.5%*	63.5%	3.0%	5.96
	2007								
	2012	75.6%*	74.3%	1.3%	2.97	61.8%*	59.7%	2.1%	3.25
336	1997	78.3%*	77.9%	0.4%	1.71	69.5%*	67.7%	1.8%	6.17
	2002	73.9%*	72.5%	1.4%	4.48	67.0%*	64.6%	2.4%	6.86
	2007								
	2012	75.3%*	74.4%	0.8%	2.97	64.5%*	63.1%	1.4%	3.66

5.3 Pooled models

Pooled models are used to explore year specific dummy variables as an alternative measure technical change. We set aside statistical concerns regarding the serial correlation of the plant level observations. Since the time steps are 5-year intervals, it is possible that the correlation concerns are lessened. Since the cross sectional models that accommodate time series correlations fail to converge, this is the only way to econometrically explore the time series dimension. We also incorporate state and regional fixed effects in these pooled models. Interestingly, the pooled electric model for 332 and 333 do not converge without the region and state fixed effect, but converge easily with them. We also examine the impact of state and region level fixed effects on the parameter estimates.

The pooled models from 1987-2012 in five year time steps are shown in tables 10-14. The first two columns are for electricity and fuels with year dummies and no region effects; the second pair of estimates include 4 Census region level dummies relative; the third pair of estimates include state fixed effects (regional coefficients are suppressed for Census disclosure purposes).

The models with and without region and state fixed effects have similar output, labor, and price elasticities. There are a few changes in significance of the gas price in the electric equation when using state fixed effects, although sometimes the coefficient is larger and significant than the model without state effects and other times the reverse. The main coefficients that change are for climate, i.e., HDD and CDD, which are clearly location related characteristics. The HDD and CDD coefficients typically are smaller with region and state fixed effects; in a few cases the smaller coefficients for HDD in the electricity equations become insignificant, similarly for CDD and fuel use. In two cases (333 and 334) the CDD coefficient changes sign in the gas equation, and is statistically significant. A possible explanation is that there is heterogeneity in plant fuel use across different quantiles of CDDs and the state demeaned estimates are estimating the effects of CDD for plants with relatively similar climates.

Figures 5 and 6 show that the pattern of the year effects is generally downward (declining energy intensity) in the first 3 time steps until 2002. In all cases there is a lot of volatility in 2007 and 2012, particularly for fuels. In 2006 natural gas prices peaked in the US and between 2007 and 2012 includes the great recession. More research on the dynamics of this period would be useful.

FIGURE 5 TREND IN 5-YEAR FIXED EFFECTS GROWTH RATES BY SECTOR (ELECTRICITY)

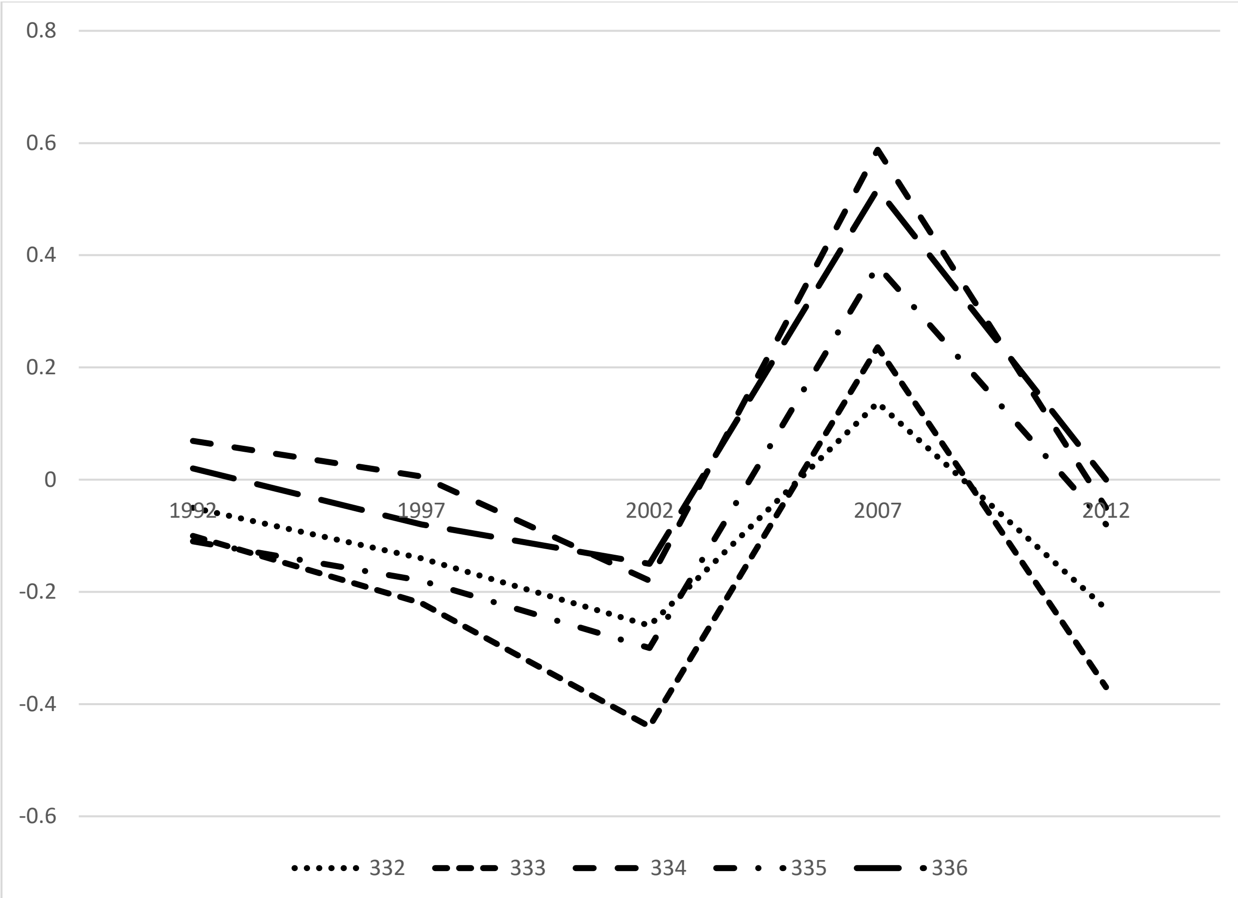

FIGURE 6 TREND IN 5-YEAR FIXED EFFECTS GROWTH RATES BY SECTOR (FUELS)

TABLE 10 NAICS 332 POOLED MODEL ESTIMATES

	No region effects		Census Region Fixed Effects		State fixed effects	
Variable	Electric	Fuel	Electric	Fuel	Electric	Fuel
lnY	0.702***	0.701***	0.701***	0.621***	0.621***	0.621***
$lnEmp$	0.272***	0.271***	0.268***	0.236***	0.236***	0.236***
lnK	0.118***	0.117***	0.118***	0.0819***	0.0810***	0.0815***
lnP_{Elec}	-0.721***	-0.684***	-0.628***	-0.0132*	-0.0217**	-0.0109
lnP_{Fuel}	-0.0413***	-0.0129	-0.0860***	-0.893***	-0.953***	-0.991***
$lnHDD$	0.0741***	0.0450***	0.0237***	0.120***	0.0343***	0.0196*
$lnCDD$	0.0537***	0.00760*	0.0180***	0.0398***	-0.00399	-0.00222
Year=1992	-0.0280***	-0.0620***	-0.0720***	-0.0521***	-0.0594***	-0.0584***
Year=1997	-0.103***	-0.137***	-0.136***	-0.176***	-0.156***	-0.146***
Year=2002	-0.147***	-0.179***	-0.163***	-0.330***	-0.279***	-0.260***
Year=2007	-0.503***	-0.564***	-0.498***	-0.00738	0.0893***	0.137***
Year=2012	-0.221***	-0.293***	-0.287***	-0.333***	-0.262***	-0.235***
constant	0.118***	0.117***	0.118***	0.0819***	0.0810***	0.0815***
$ln\sigma_u^2$	-0.721***	-0.684***	-0.628***	-0.0132*	-0.0217**	-0.0109
$ln\sigma_v^2$	-0.0413***	-0.0129	-0.0860***	-0.893***	-0.953***	-0.991***

legend: * p<0.05; ** p<0.01; *** p<0.001

TABLE 11 NAICS 333 POOLED MODEL ESTIMATES

	No region effects		Census Region Fixed Effects		State fixed effects	
Variable	Electric	Fuel	Electric	Fuel	Electric	Fuel
lnY	0.482***	0.482***	0.483***	0.450***	0.451***	0.450***
$lnEmp$	0.412***	0.411***	0.409***	0.327***	0.326***	0.327***
lnK	0.158***	0.158***	0.158***	0.0918***	0.0906***	0.0915***
lnP_{Elec}	-0.719***	-0.703***	-0.684***	0.00452	-0.00604	-0.000601
lnP_{Fuel}	-0.00857	0.00324	-0.0793***	-0.768***	-0.833***	-0.774***
$lnHDD$	0.0495***	0.0350***	0.0114	0.166***	0.0603***	0.0147
$lnCDD$	0.0754***	0.0428***	0.0341***	0.0368***	-0.0121	-0.0349***
Year=1992	0.0610***	0.0387***	0.0345***	-0.0765***	-0.0881***	-0.104***
Year=1997	-0.0711***	-0.0897***	-0.0734***	-0.223***	-0.203***	-0.226***
Year=2002	-0.109***	-0.121***	-0.0811***	-0.465***	-0.413***	-0.446***
Year=2007	-0.0471**	-0.0716***	0.0277	0.216***	0.317***	0.236***
Year=2012	-0.0508***	-0.0778***	-0.02	-0.383***	-0.311***	-0.377***
constant	0.158***	0.158***	0.158***	0.0918***	0.0906***	0.0915***
$ln\sigma_u^2$	-0.719***	-0.703***	-0.684***	0.00452	-0.00604	-0.000601
$ln\sigma_v^2$	-0.00857	0.00324	-0.0793***	-0.768***	-0.833***	-0.774***

legend: * p<0.05; ** p<0.01; *** p<0.001

TABLE 12 NAICS 334 POOLED MODEL ESTIMATES

	No region effects		Census Region Fixed Effects		State fixed effects	
Variable	Electric	Fuel	Electric	Fuel	Electric	Fuel
lnY	0.300***	0.298***	0.294***	0.208***	0.213***	0.205***
$lnEmp$	0.505***	0.506***	0.510***	0.445***	0.441***	0.448***
lnK	0.224***	0.225***	0.225***	0.159***	0.158***	0.159***
lnP_{Elec}	-0.756***	-0.744***	-0.772***	-0.0939***	-0.0915***	-0.148***
lnP_{Fuel}	0.257***	0.263***	0.332***	-0.455***	-0.497***	-0.363***
$lnHDD$	0.00322	0.0159*	0.000266	0.156***	0.0775***	0.0676**
$lnCDD$	0.0724***	0.0531***	0.0321***	0.0572***	0.00769	-0.0519***
Year=1992	0.127***	0.117***	0.114***	0.0868***	0.0776***	0.0691***
Year=1997	0.123***	0.114***	0.102***	0.0195	0.0287	0.0063
Year=2002	0.0194	0.0115	-0.014	-0.172***	-0.140***	-0.184***
Year=2007	0.00502	-0.00944	-0.0802**	0.663***	0.721***	0.588***
Year=2012	-0.0504**	-0.0570***	-0.0748***	-0.0448*	-0.0265	-0.0512*
constant	0.224***	0.225***	0.225***	0.159***	0.158***	0.159***
$ln\sigma_u^2$	-0.756***	-0.744***	-0.772***	-0.0939***	-0.0915***	-0.148***
$ln\sigma_v^2$	0.257***	0.263***	0.332***	-0.455***	-0.497***	-0.363***

legend: * p<0.05; ** p<0.01; *** p<0.001

TABLE 13 NAICS 335 POOLED MODEL ESTIMATES

	No region effects		Census Region Fixed Effects		State fixed effects	
Variable	Electric	Fuel	Electric	Fuel	Electric	Fuel
lnY	0.446***	0.448***	0.451***	0.391***	0.398***	0.400***
$lnEmp$	0.379***	0.379***	0.375***	0.320***	0.316***	0.314***
lnK	0.223***	0.221***	0.221***	0.170***	0.167***	0.167***
lnP_{Elec}	-0.821***	-0.829***	-0.845***	-0.128***	-0.175***	-0.179***
lnP_{Fuel}	0.105***	0.0896***	0.0256	-0.668***	-0.851***	-0.766***
$lnHDD$	0.0224**	0.0359***	0.0596**	0.193***	0.0592***	0.0635
$lnCDD$	0.0972***	0.0752***	0.0570***	0.0671***	0.00664	-0.00256
Year=1992	0.0581***	0.0538***	0.0563***	-0.119***	-0.108***	-0.116***
Year=1997	-0.0834***	-0.0814***	-0.0590**	-0.226***	-0.154***	-0.181***
Year=2002	-0.185***	-0.171***	-0.125***	-0.401***	-0.261***	-0.307***
Year=2007	-0.104**	-0.0792*	0.0184	0.214***	0.492***	0.381***
Year=2012	-0.121***	-0.0938**	-0.0062	-0.232***	-0.00764	-0.0897
constant	0.223***	0.221***	0.221***	0.170***	0.167***	0.167***
$ln\sigma_u^2$	-0.821***	-0.829***	-0.845***	-0.128***	-0.175***	-0.179***
$ln\sigma_v^2$	0.105***	0.0896***	0.0256	-0.668***	-0.851***	-0.766***

legend: * p<0.05; ** p<0.01; *** p<0.001

TABLE 14 NAICS 336 POOLED MODEL ESTIMATES

	No region effects		Census Region Fixed Effects		State fixed effects	
Variable	Electric	Fuel	Electric	Fuel	Electric	Fuel
lnY	0.401***	0.400***	0.396***	0.336***	0.338***	0.335***
$lnEmp$	0.447***	0.447***	0.452***	0.433***	0.433***	0.435***
lnK	0.196***	0.196***	0.193***	0.161***	0.158***	0.159***
lnP_{Elec}	-0.796***	-0.772***	-0.791***	-0.139***	-0.157***	-0.168***
lnP_{Fuel}	0.0521***	0.0716***	0.0284	-0.644***	-0.758***	-0.604***
$lnHDD$	0.0454***	0.0305***	0.019	0.136***	0.0554***	0.0233
$lnCDD$	0.0679***	0.0290***	0.0329***	0.0152	-0.0119	-0.0339*
Year=1992	0.136***	0.110***	0.124***	0.0309	0.0394*	0.0201
Year=1997	0.0773***	0.0537***	0.0815***	-0.0762***	-0.0318	-0.0819***
Year=2002	-0.0417**	-0.0614***	-0.0198	-0.150***	-0.0657**	-0.152***
Year=2007	0.322***	0.280***	0.360***	0.549***	0.723***	0.518***
Year=2012	0.225***	0.183***	0.251***	0.0116	0.132***	-0.000733
constant	0.196***	0.196***	0.193***	0.161***	0.158***	0.159***
$ln\sigma_u^2$	-0.796***	-0.772***	-0.791***	-0.139***	-0.157***	-0.168***
$ln\sigma_v^2$	0.0521***	0.0716***	0.0284	-0.644***	-0.758***	-0.604***

legend: * p<0.05; ** p<0.01; *** p<0.001

6. Conclusions

6.1 Summary

This paper utilizes a rich plant level data source to estimate the evolution of energy efficiency in metal based durables over a 25 year period using a stochastic frontier regression analysis approach applied to an ad-hoc energy demand framework. We apply the analysis to 5 3-digit NAICS in the MBD manufacturing industries. We control for 6-digit NAICS, location based climate, and energy prices to estimate frontier demand and the distribution of efficiency relative to the frontier in 5-year time steps from the quinquennial EC.

We find that median efficiency for electricity is higher (73%) than for fuel use (51%) averaged across all five sectors. The Malmquist decomposition implies generally small overall improvements in electricity efficiency over time, ranging from 0.0 to 0.9 percent annually; overall fuel performance was more variable, including declines and improvements ranging from -1.0 to 1.0 percent. In both cases, improvement is from technical change, i.e. shifts in the frontier. Moreover, failure of plants "keeping up," as measured by the Malmquist efficiency index, retarded the overall improvement and in some sectors making it negative. The failure to "keep up" eroded gains in both electricity and fuel use. While the difference between plants entering the industry and those that are continuing tends to be small, we find it does represent a statistically significant improvement in energy efficiency.

6.2 Using the SFA results in NEMS

This research is not intended to be a wholesale replacement of the NEMS structure. However, the choice of 3-digit pooled NAICS with detailed 6-digit industry controls is a decision based on the NEMS forecasting environment. By the same token, 3-digit NAICS (2-digit SIC) is a common choice of aggregation in the energy literature so it is fairly natural one to make. We have not fully examined how best to make use of the proposed SFA approach within the NEMS forecasting environment, but rather focus the analysis to develop a set of estimates of energy efficiency in MBD that could be used by EIA to inform changes in the NEMS parameters. The underlying efficiency distributions should be more than adequate to develop "simple" NEMS efficiency scenarios, e.g. assuming the lower quartile moves to the median, etc.

The separation of process and building energy in NEMS is not something that this approach readily supports. The plant level data are for the entire establishment and therefore the model incorporates building and process energy. This is one motivation to attempt to control for weather/climate in the form of plant specific HDD and CDD. We presume that EIA would either have to assume that the efficiency distributions that are estimated apply equally to process and building energy or develop some ex-post method to reconcile the total energy efficiency distribution with a separate distribution of building energy efficiency.

The kernel density estimated of the efficiency distributions provide a basis for a variety of efficiency related scenario analysis. For example, one could envision a scenario where the lowest performing plans (e.g. lowest quartile) improve to some target level (e.g. 2nd or 3rd quartile) as we have computed above. The fact that few plants have efficiency exceeding 90% suggests that scenarios of this type might be more "realistic" compared to ones that assume 100%, or even 90%, efficiency targets. In essence, 100% efficiency may be achievable in theory, but rarely observed empirically on this data.

7. REFERENCES

Adetutu, M. O., A. J. Glass and T. G. Weyman-Jones (2016). "Economy-wide Estimates of Rebound Effects: Evidence from Panel Data." The Energy Journal 37(3): 251-269.

Allcott, H. and M. Greenstone (2012). "Is There an Energy Efficiency Gap?" The Journal of Economic Perspectives 26(1): 3-28.

Aranda-Usón, A., G. Ferreira, M. D. Mainar-Toledo, S. Scarpellini and E. Llera Sastresa (2012). "Energy consumption analysis of Spanish food and drink, textile, chemical and non-metallic mineral products sectors." Energy 42(1): 477-485.

Balk, B. M. (2001). "Scale Efficiency and Productivity Change." Journal of Productivity Analysis 15(3): 159-183.

Bartelsman, E. and W. Gray (1996). The NBER Manufacturing Productivity Data Base. Cambridge MA, National Bureau of Economic Research.

Bentzen, J. (2004). "Estimating the rebound effect in US manufacturing energy consumption." Energy Economics 26(1): 123-134.

Blumstein, C. and S. Stoft (1995). "Technical Efficiency, production functions, and conservation supply curves." Energy Policy 23(9): 765-768.

Bostian, M., R. Färe, S. Grosskopf and T. Lundgren (2016). "Environmental investment and firm performance: A network approach." Energy Economics 57: 243-255.

Boyd, G., E. Dutrow and W. Tunnesen (2008). "The Evolution of the Energy Star Industrial Energy Performance Indicator for Benchmarking Plant Level Manufacturing Energy Use." Journal of Cleaner Production Volume 16(Issue 6): 709-715.

Boyd, G., S. Karlson, M. Neifer and M. Ross (1993). "Energy Intensity Improvements in Steel Minimills." Contemporary Policy Issues XI(3): 88-99.

Boyd, G. and J. Pang (2000). "Estimating the Linkage between Energy Efficiency and Productivity." Energy Policy(28,): 289-296.

Boyd, G. and G. Zhang (2013). "Measuring improvement in energy efficiency of the US cement industry with the ENERGY STAR Energy Performance Indicator." Energy Efficiency 6(1): 105-116.

Boyd, G. A. (2005). "A Method for Measuring the Efficiency Gap between Average and Best Practice Energy Use: The ENERGY STAR Industrial Energy Performance Indicator." Journal of Industrial Ecology 9(3): 51-65.

Boyd, G. A. (2008). "Estimating Plant Level Manufacturing Energy Efficiency with Stochastic Frontier Regression." The Energy Journal 29(2): 23-44.

Boyd, G. A. (2014). "Estimating the changes in the distribution of energy efficiency in the U.S. automobile assembly industry." Energy Economics 42(0): 81-87.

Boyd, G. A. (2016). "Comparing the statistical distributions of energy efficiency in manufacturing: meta-analysis of 24 Case studies to develop industry-specific energy performance indicators (EPI)." Energy Efficiency: 1-22.

Boyd, G. A. and E. M. Curtis (2014). "Evidence of an “Energy-Management Gap” in U.S. manufacturing: Spillovers from firm management practices to energy efficiency." Journal of Environmental Economics and Management 68(3): 463-479.

Coelli, T. J., D. S. P. Rao, C. J. O'Donnell and G. E. Battese (2005). An Introduction to Efficiency and Productivity Analysis, Springer US.

Davis, S. J., C. Grim, J. Haltiwanger and M. Streitwieser (2012). "Electricity Unit Value Prices and Purchase Quantities: U.S. Manufacturing Plants, 1963–2000." Review of Economics and Statistics 95(4): 1150-1165.

Färe, R. and S. Grosskopf (1992). "Malmquist Productivity Indexes and Fisher Ideal Indexes." The Economic Journal 102(410): 158-160.

Färe, R. and S. Grosskopf (2000). "Theory and application of directional distance functions." Journal of Productivity Analysis 13(2): 93-103.

Färe, R., S. Grosskopf, M. Norris and Z. Zhang (1994). "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries." American Economic Review 84(1): 66-83.

Feijoó, M. a. L., J. F. Franco and J. M. Hernández (2002). "Global warming and the energy efficiency of Spanish industry." Energy Economics 24(4): 405-423.

Ferrier, G. D. and J. G. Hirschberg (1992). "Climate Control Efficiency." The Energy Journal 13(1): 37-54.

Filippini, M. and L. C. Hunt (2011). "Energy Demand and Energy Efficiency in the OECD Countries: A Stochastic Demand Frontier Approach." The Energy Journal 32(2): 59-80.

Filippini, M. and L. C. Hunt (2012). "US residential energy demand and energy efficiency: A stochastic demand frontier approach." Energy Economics 34(5): 1484-1491.

Filippini, M. and L. C. Hunt (2015). "Measurement of energy efficiency based on economic foundations." Energy Economics 52, Supplement 1: S5-S16.

Gately, D. and H. G. Huntington (2002). "The asymmetric effects of changes in price and income on energy and oil demand." The Energy Journal 23(1): 19-56.

Gerarden, T. D., R. G. Newell and R. N. Stavins (2015). "Assessing the Energy-Efficiency Gap." National Bureau of Economic Research Working Paper Series No. 20904.

Glatt, S., R. Naranjo, N. Swalnick, G. Shields, J. McVeigh, L. Wright and V. Kraus (2016). Midwest Region Energy Efficiency: Potential Manufacturing Energy And Cost Savings Report. Washington DC, U.S. Department of Energy.

Glatt, S., R. Naranjo, N. Swalnick, G. Shields, J. McVeigh, L. Wright and V. Kraus (2016). Northeast Region Energy Efficiency: Potential Manufacturing Energy And Cost Savings Report. Washington DC, U.S. Department of Energy.

Glatt, S., R. Naranjo, N. Swalnick, G. Shields, J. McVeigh, L. Wright and V. Kraus (2016). South Region Energy Efficiency: Potential Manufacturing Energy And Cost Savings Report Washington DC, U.S. Department of Energy.

Glatt, S., R. Naranjo, N. Swalnick, G. Shields, J. McVeigh, L. Wright and V. Kraus (2016). West Region Energy Efficiency: Potential Manufacturing Energy And Cost Savings Report. Washington DC, U.S. Department of Energy.

Greene, W. H. (2002). Fixed and Random Effects in Stochastic Frontier Models. New York University Economics department working paper.

Howarth, R. B. and B. Andersson (1993). "Market Barriers to Energy Efficiency." Energy Economics (October): 262-272.

Huntington, H. (1995). "Been Top Down So Long It Looks Like Bottom Up." Energy.

Jaffe, A. B. and R. N. Stavins (1994). "The energy-efficiency gap: What does it mean?" Energy Policy 22(10): 804-810.

Kerstens, K. and I. Van de Woestyne (2014). "Comparing Malmquist and Hicks–Moorsteen productivity indices: Exploring the impact of unbalanced vs. balanced panel data." European Journal of Operational Research 233(3): 749-758.

Kumbhakar, S. C. and E. G. Tsionas (2006). "Estimation of stochastic frontier production functions with input-oriented technical efficiency." Journal of Econometrics 133(1): 71-96.

Lansink, A., E. Silva and S. Stefanou (2000). "Decomposing productivity growth allowing efficiency gains and price-induced technical progress." European Review of Agricultural Economics, 27(4): 497-518.

Lin, B. and H. Long (2015). "A stochastic frontier analysis of energy efficiency of China's chemical industry." Journal of Cleaner Production 87: 235-244.

Lundgren, T., P.-O. Marklund and S. Zhang (2016). "Industrial energy demand and energy efficiency – Evidence from Sweden." Resource and Energy Economics 43: 130-152.

Mukherjee, K. (2008). "Energy use efficiency in U.S. manufacturing: A nonparametric analysis." Energy Economics 30(1): 76-96.

Murillo-Zamorano, L. (2004). "Economic efficiency and frontier techniques." Journal of Economic Surveys 18(1): 33-77.

Murillo-Zamorano, L. R. (2004). "Economic Efficiency and Frontier Techniques." Journal of Economic Surveys 18(1): 33-77.

Newell, R. G., A. B. Jaffe and R. N. Stavins (1999). "The Induced Innovation Hypothesis and Energy-Saving Technological Change." The Quarterly Journal of Economics 114(3): 941-975.

Pantzios, C. J., G. Karagiannis and V. Tzouvelekas (2011). "Parametric decomposition of the input-oriented Malmquist productivity index: with an application to Greek aquaculture." Journal of Productivity Analysis 36(1): 21-31.

Parker, S. and B. Liddle "Energy Efficiency in the Manufacturing Sector of the OECD: Analysis of price elasticities." Energy Economics.

Rosenfeld, A., C. Atkinson, J. Koomey, A. Meier, R. J. Mowris and L. Price (1993). "Conserved Energy Supply Curves For U.S. Buildings." Contemporary Economic Policy 11(1): 45-68.

Steinbuks, J. (2012). "Interfuel Substitution and Energy Use in the U.K. Manufacturing Sector." The Energy Journal Volume 33(1).

Steinbuks, J. and K. Neuhoff (2014). "Assessing energy price induced improvements in efficiency of capital in OECD manufacturing industries." Journal of Environmental Economics and Management 68(2): 340-356.

Stern, D. I. (2012). "Modeling international trends in energy efficiency." Energy Economics 34(6): 2200-2208.

Stoft, S. E. (1995). "The Economics of Conserved-Energy “Supply” Curves." The Energy Journal 16(4): 109-137.

Sueyoshi, T., Y. Yuan and M. Goto (2017). "A literature study for DEA applied to energy and environment." Energy Economics 62: 104-124.

Worrell, E., S. Ramesohl and G. Boyd (2004). "Advances In Energy Forecasting Models Based On Engineering Economics." Annual Review of Environment and Resources 29(1): 345-381.

Zhang, S., T. Lundgren and W. Zhou (2016). "Energy efficiency in Swedish industry: A firm-level data envelopment analysis." Energy Economics 55: 42-51.

Zhou, P., B. W. Ang and K. L. Poh (2008). "A survey of data envelopment analysis in energy and environmental studies." European Journal of Operational Research 189(1): 1-18.

Zhou, P., B. W. Ang and D. Q. Zhou (2012). "Measuring economy-wide energy efficiency performance: A parametric frontier approach." Applied Energy 90(1): 196-200.