

Energy Efficiency and Price Responsiveness in Energy Intensive Chemicals Manufacturing¹

Gale Boyd¹ and Jonathan M. Lee²

¹*Social Science Research Institute, Duke University, Box 90989, Durham NC 27708*

²*Department of Economics, East Carolina University, Greenville, NC 27858*

April 28th, 2017

Abstract

This paper presents estimates of the distribution of energy efficiency and price elasticities in the four major energy-using sectors of the upstream, energy-intensive portions of the Chemical industry (inorganic, organic, resins & plastics, and fertilizers). To obtain the estimates, we analyze non-public plant-level data from the Census of Manufacturing (CM) and the Manufacturing Energy Consumption Survey (MECS) separately, since these two data sources have their own strengths and weaknesses. The basic approach is to use an ad-hoc (reduced form) stochastic frontier energy demand function for electricity and fuel use separately. A two-stage approach is used to control for plant-level energy price endogeneity and plant-level heterogeneity. This approach provides a decomposition of the total energy efficiency into a persistent (plant specific) and time-varying component. We find that the dispersion of efficiency is relatively small, consistent with other studies of energy intensive sectors. The CM analysis implies, that if all plants were to perform at the 90 percentile of their corresponding efficiency distribution, the reduction in energy use would range between 4% and 13%. Persistent efficiency is smaller than time-varying efficiency, and new plants tend to have higher persistent efficiency but enter the industry with lower time-varying efficiency that subsequently improves. The CM analysis finds higher energy price elasticities than the MECS analysis; many are near or above unity. The MECS analysis finds elasticities in the range of -0.2 to -1.0. A logit analysis of energy price elasticities in the MECS data using a logit energy share framework a logit analysis finds lower own price elasticities and modest evidence of fuel-electricity substitution. This lower own price effect is expected because it does not account for the overall impact of aggregate energy prices on total energy demand.

¹ This paper was funded by the U.S. Energy Information Administration (EIA), Office of Energy Consumption and Efficiency Analysis while working with Leidos Inc. under contract with EIA (DE-EI0000564) and was prepared while the authors were Special Sworn Status researchers at the Triangle Research Data Center, a member of the Federal Statistical Research Data Center Network. The paper benefited from comments by attendees at the *2017 FSRDC Annual Conference in UCLA* and from ongoing discussions and comments from Kelly Perl and Paul Otis on an earlier draft. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau or EIA. All results have been reviewed to ensure that no confidential information is disclosed.

Introduction

The Chemical Industry, as defined by the North American Industry Classification System (NAICS) code 325, is a diverse collection of sectors ranging from commodity chemicals (e.g., ammonia, chlor-alkalies, ethylene) to consumer products (e.g., paint, pharmaceuticals, cosmetics, etc.) The former are the upstream process industries that encompass some of the most energy-intensive chemical conversions of feedstock into intermediate chemicals, which are used primarily by other industries. The latter uses and produces a wide range of downstream chemicals to make, package, and distribute final consumer goods. Of the more than 5 quadrillion Btu (quads)² of energy reported by the 2010 Manufacturing Energy Consumption Survey (MECS) that is used in NAICS 325, about 4.2 quads are used in the 13 energy intensive six-digit NAICS listed below.

These energy-intensive chemical sectors can be grouped into four chemical industry classifications that mimic the four-digit NAICS hierarchical groups with some minor exceptions; Inorganic Chemicals, Organic Chemicals, Plastics and Resins, and Fertilizers. These are the same industry sector groupings used by EIA's National Energy Modeling System (NEMS) Industrial Demand Module (IDM) (Energy Information Administration 1994).

Inorganic Chemicals

- 325120 Industrial Gases
- 325181 Alkalies and Chlorine
- 325182 Carbon Black
- 325188 Other Basic Inorganic Chemicals

Organic Chemicals

- 325110 Petrochemicals
- 325192 Cyclic Crudes and Intermediates
- 325193 Ethyl Alcohol
- 325199 Other Basic Organic Chemicals

Plastics and Resins

- 325211 Plastics Materials and Resins
- 325212 Synthetic Rubber
- 325222 Noncellulosic Organic Fibers

Fertilizers

- 325311 Nitrogenous Fertilizers
- 325312 Phosphatic Fertilizers

A popular conceptual modeling method for energy forecasting is the stock adjustment approach. This approach is based on the notion that energy use is tied to the capital stock that changes over time in response to replacement, as a result of depreciation and new expansion, to account for growth. This

² This 5 quads includes feedstocks as well as energy for heat and power. All energy data are from the 2010 MECS *Table 1.2 First Use of Energy for All Purposes (Fuel and Nonfuel)*. These data are measured at end-use; i.e., electric generation losses are not included.

basic framework is used widely in the demand modules for the NEMS. The IDM is one such model that employs this underlying concept (Energy Information Administration 2014).

This modeling approach considers that the unit energy intensity (UEI), measured relative to physical production units, dollar shipments, value added, etc., can be represented as a weighted average of the UEI for existing and new applications (plants or individual capital stock, etc.)

$$UEI = \lambda * (UEI_{existing}) + (1 - \lambda) * UEI_{new} \quad (1)$$

This approach shares common 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 UEI is constant (or nearly so) over its lifetime. This implies that there would be a distribution of UEI 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 UEI and the lowest UEI can be thought of as a measure of the “energy gap,” which has been the subject of numerous studies (Jaffe and Stavins 1994, Huntington 1995, Allcott and Greenstone 2012, Boyd and Zhang 2013, Boyd and Curtis 2014, Boyd 2016). Equation 1 should be viewed as a general representation to illustrate that UEI is a mix of existing and new equipment. This is not exactly how the IDM implements the concept. For end use models, the IDM look at UEI at a facility level. UEI for existing facilities can also improve over time because of capital replacement or adoption of energy management procedures ([Assumptions 2016, p. 57](#)).

This paper provides estimates of energy efficiency and energy price response in the energy-intensive chemical manufacturing sector. This report shares important features with the methodology presented in and applied to analyze metal-based durables (Boyd and Lee 2016) in that it measures the distribution of energy efficiency of demand relative to local (plant-level) energy prices. Both studies use a stochastic frontier analysis (SFA), but this paper uses a two-stage variant of SFA to account for both plant-level energy price endogeneity and plant-specific heterogeneity in energy use. See (Amsler, Prokhorov et al. 2016) a review of endogeneity in SFA. This paper takes a different approach. In particular, we are concerned with *both price endogeneity and plant-level heterogeneity* in energy use. The two-stage method developed by (Kumbhakar, Lien et al. 2014) and modified here to account with the plant-level price endogeneity controls for both. The two-stage approach also allows the decomposition of efficiency into a plant-specific (persistent) and time-varying component. We conduct a parallel analysis of two sources of plant-level data, as detailed below. We also contrast the results obtained by the SFA with a logit analysis of fuel shares.

This report is organized as follows. The first sections describe the two plant-level data sources at the core of the analysis. The next sections describe concerns over plant-level energy price endogeneity and plant-specific heterogeneity. A two-stage approach is presented as a solution. This section also introduces the notion of time-varying and persistent inefficiency. Finally, the parameter estimates for the elasticities and the distribution(s) for efficiency are discussed for both data sources and for the logit fuel share analysis that does not account for efficiency.

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. Since these sectors are energy intensive, a parallel approach regarding the

data sources is used. These data sources are the Manufacturing Energy Consumption Survey (MECS) and the quinquennial Census of Manufacturing (CM). MECS is a sample-based survey conducted by EIA in 1985, 1988, 1991, 1994, 1998, 2002, 2006, and 2010³. The CM is part of the quinquennial Economic Census (EC); in principle it includes all establishments operating during the analysis time period of 5 five-year time steps, 1992, 1997, 2002, 2007, and 2012. Both data span similar time periods but, for the most part, different years. The MECS and CM each have advantages and disadvantages, which is why a parallel analysis approach was used.

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 relative efficiency of entering vs 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 CM is one advantage of this data set. (Davis, Grim et al. 2012) analyze the dispersion of those prices in detail. However, the CM only reports cost of fuels, not quantities, so Btu fuel consumption is imputed from fuel costs in the CM using the assumptions-regarding the state-level average price of fossil fuels. In analysis by (Boyd and Lee 2017), fossil fuel use is imputed from the price of natural gas. This was seen as a reasonable assumption for the metal-based durables industries, because publicly available MECS data from 2010 for these five sectors suggests that 88% to 98% the purchased fuel in this sector is natural gas. This is less true for energy-intensive chemicals. MECS reports that in 2010 natural gas was only 77% of fossil fuels used for heat and power in these energy-intensive sectors. We impute Btu consumption by taking the cost of fuels and dividing by a weighted average of the state-level fossil fuel prices as published by the EIA's State Energy Data System (SEDS)⁴, where the weights are computed from the published MECS data for each six-digit NAICS above and applied to the closest year between the MECS and CM. The CM provides plant-level electricity consumption and costs, from which a plant-level price can be computed. Plants that generate part of their own electricity, not uncommon in this industry, will likely purchase more fossil fuel and less electricity. To account for this, the ratio of generated power to the total net consumption is computed.

Plant-level shipment values, adjusted for inventory changes, are used to measure production. Labor is measured in production worker hours. Capital stock is the total of plant and equipment. Non-energy material costs are computed by subtracting total material expenditures from the cost for electricity and fuels. 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 National Oceanic and Atmospheric Administration (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 MECS provides the most detailed data on energy use, particularly cost and quantity of fuels by type. The MECS is a sub-sample of the Census Bureau's Annual Survey of Manufactures (ASM) that targets mainly large plants⁵. MECS provides detail on a wide range of fossil fuel types, including the quantity of fossil fuels used as feedstocks. The plant-level cost and quantity can be used to compute plant-level

³ 2014 was the most recent year, but not yet available to external researchers.

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

⁵ In later years of the MECS, the sample design is not strictly a sub-sample of the ASM, but we need data from the ASM on production and employment, so in those years we use the overlap between MECS and ASM.

average fossil fuel prices, as well as plant-level electricity prices. While the MECS is a sub-sample, its primary advantage is in the fossil fuel detail, a major component of energy use in this industry. MECS data on fossil fuel consumption is obtained by directly aggregating over all fossil energy types, excluding those used as feedstocks⁶. Costs are similarly aggregated, and plant-level average fuel prices are constructed and deflated to constant dollars using a GDP price deflator. MECS also indicates the amount of fossil fuels, mostly natural gas and gas liquids, as chemical feedstocks. Since plants using these feedstocks are likely to be more energy intensive, an indicator variable is created to reflect a plant is a feedstock-using plant. All other economic variables in the MECS sample analysis are the same as those constructed for the CM.

Methodology

This section briefly presents the ad-hoc demand model specification. This is done by adding energy prices to the energy factor requirement function described by (Boyd and Delgado 2012), which is equivalent to a directional input distance function. (Boyd and Lee 2016) motivate this by considering the energy prices as a modification of the direction of the distance function, but do not make that connection explicit. A review of stochastic frontier applications for energy use can be found in (Filippini and Hunt 2015). The report then discusses concerns regarding price endogeneity and plant-level heterogeneity. A two-stage estimation approach is presented as a solution that addresses both of these concerns in the first stage. This approach also allows for the decomposition of efficiency into two components; one is plant specific and constant over time (persistent efficiency), and one that is time varying. The price elasticity results are compared to a logit fuel share analysis that does not account for aggregate demand or efficiency.

Stochastic Frontier approach to Energy demand

Following (Boyd and Lee 2016), we consider an SF ad hoc energy demand equation for the two primary energy types in each of the four sectors, with a few modifications, which are discussed below. We consider log linear models (KLEM Cobb-Douglas) of the general form,

$$\ln E_{j,i,t} = f(\ln Y_{i,t}, \ln K_{i,t}, \ln Emp_{i,t}, \ln NEM_{i,t}, \ln P_{j,t,s}, DYear_t, DNAICS_k, GERATIO_{i,t}) + \varepsilon_{j,i,t} \quad (2)$$

Where

$\ln E_{j,i,t}$	=	log of energy use
$\ln Y_{i,t}$	=	log of production or output
$\ln Emp_{i,t}$	=	log employment or other measure of labor
$\ln K_{i,t}$	=	log capital stock
$\ln NEM_{i,t}$	=	log of non-energy material use
$\ln P_{j,t,s}$	=	ln price of energy ⁷
$DYear$	=	dummy for the year

⁶ Data from the CM on fuel use states that this is for heat and power and should not include feedstocks, making these definitions comparable.

⁷ The subscript 's' refers to state level, but we use both state, and plant-level prices as detailed below.

$DNAICS_k$ = dummy for the 6-digit NAICS code

$GERATIO_{i,t}$ = ratio of self generated electricity to total of purchased plus generated less sold

j = energy type (electricity and fuel)

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

s = state

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

k = six-digit NAICS

The standard SF approach is to treat $\varepsilon_{j,i,t}$ as the sum two terms representing statistical noise, $v_{j,i,t}$, and inefficiency, $u_{j,i,t}$, respectively. We will return to specific approaches to the distributional assumptions of $\varepsilon_{j,i,t}$ below.

Total value of shipments (TVS), deflated and adjusted for inventory changes, is used as the measure of productive output. Labor, measured by number of employees, controls for plant-level utilization effects⁸, since labor may be sticky in the short run. To better control for upstream and downstream plants within the sectors, we include capital stock and non-energy materials. The most energy-intensive chemical processes tend to be very capital intensive and have very simple material feedstocks. Downstream plants may purchase chemicals produced by upstream plants and may involve simpler, less energy-intensive production processes. To account for this, we consider non-energy material use. We estimate models with and without capital stock, because the capital stock variable is not available in our final CM year, 2012. Non-energy material use is the deflated costs of material purchases, less energy costs, which are included in the Census (material costs variables). The long-run relationship between energy and plant scale is captured by the combined coefficient on production, capital, non-energy materials, and labor. In a simple Cobb-Douglas specification, the sum of the coefficients reflects the economies of scale with respect to energy. If the sum of the coefficients is less than one, then we can infer that larger plants will have lower frontier *energy intensity* than smaller plants. This means that the model will control for scale differences with respect to the energy efficiency measure.

Even within our four chemical sectors there can be a lot of heterogeneity of products and corresponding energy services, so six-digit NAICS industry controls (fixed effects) are used. 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 sufficient and are more detailed than other industrial energy studies have employed before. One exception is (Boyd and Curtis 2014) which also use plant-level Census micro-data at the six-digit level.

The price variables, $\ln P_{j,t,s}$, 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

⁸ Using the five-year Economic Census also conveniently avoids the years of the Great Recession by including 2007 and 2012, but not the intervening years.

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 prices⁹ 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.

We considered the possibility that heating degree days (HDD) and cooling degree days (CDD) could be used to control for ambient weather conditions on an annual basis using the zip-code location of the plant. Weather can impact building heating, ventilation, and air-conditioning (HVAC) energy use, but also it also can affect process energy via outside air to ovens and furnaces or chiller efficiencies, to the extent that the production requires these process. Preliminary analysis found these to be insignificant. This is not surprising given the small role for HVAC in this sector. These variable were not used in the final model results.

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 heat and power (CHP) being the most obvious. Since Census data does include on-site generation we include a variable to control for this. We compute the ratio of self-generated power to the sum of self-generated power and purchased power minus sales to the grid and include it as a control variable. In the electric equation we would expect the coefficient to be negative (i.e., less purchased electricity), but in the fuel equation the coefficient would be positive to account for the amount of extra fuel consumed in the CHP.

Directly estimating the model above faces some issues due to particular concerns in these sectors. These concerns are the endogeneity of energy prices and plant-level heterogeneity that should be separated from efficiency. The next two sections describe these concerns, followed by our approach to account for them.

Price endogeneity

Large energy users, either by virtue of sheer size or by virtue of having energy-intensive production processes, have good reasons to get the lowest possible energy prices. This means that lower plant-level prices would be correlated with higher energy demand for reasons other than pure price responsiveness, i.e., estimated price elasticities would be biased upwards in absolute terms. Preliminary analysis using plant-level electricity prices from both the CM and MECS found extremely high own-price elasticities of demand and complementarity between fuel and electricity, i.e., negative cross-price elasticities in fossil fuel demand. This result mirrored what was found in non-energy intensive chemical manufacturing, so the analysis focused on methods to control for price endogeneity, using state-level prices. Endogeneity concerns were limited to electricity in the CM analysis, which includes plant-level electric but not fuel prices. The MECS data analysis allow for plant-level fuel price endogeneity to be considered as well.

Concerns regarding and methods to control for endogeneity in the SF context is reviewed by (Amsler, Prokhorov et al. 2016). One approach is the control function. In this method, plant-level energy prices are regressed against the instrument, in this case state level price, and all the independent variables of the SF model. The residuals of this first stage regression are included in the SF estimation. A significant

⁹ These prices are not true marginal prices, but include demand charges, etc. They are total expenditures divided by total consumption.

coefficient on the residuals indicates the prices are endogenous. We take a slightly different approach, one that can also account for plant-level heterogeneity.

Plant Heterogeneity

Even within these four sectors of the chemical industry, we anticipate plant-level differences in processes and products that can require very different levels of energy. In organic chemicals, the production of ethylene is much more energy intensive than subsequent downstream product. Ethylene is a component of many plastics, so if a plastics plant is fully integrated and produces its own ethylene, then that plant would be much more energy intensive. Another example is ammonia production for fertilizers. This is a primary chemical input to other fertilizer chemical and is also produced as a final product. Ammonia production is a very energy-intensive chemical to produce, but fertilizer plants may buy it instead of making it on site. There are other example of producing sulphur-related chemicals where the process is exothermic, i.e., since the reaction generates useable energy rather than requiring energy to sustain it.

One approach to account for plant heterogeneity would be use detailed material and product codes. This has been done by (Boyd and Delgado 2012, Boyd and Guo 2014, Boyd 2016) for some selected industries, but requires a large amount of knowledge regarding which specific material and product types are most relevant. Use of capital stock and material purchases might partially account for these plant-level differences, since energy-intensive plants are likely to have less expensive feedstocks since they may make, rather than buy, some intermediate product. Making an intermediate product is more likely to be more energy intensive and more capital intensive both. Even though we include capital stock and material purchases in the specification, additional methods to account for plant-level heterogeneity are desirable.

The desire to distinguish between efficiency and heterogeneity requires an extension of the SFA framework. The standard treatment for plant-level heterogeneity in panel data is to include either a plant-specific fixed or a random effect. Equation (3) represents the non-stochastic frontier implementation of plant-level heterogeneity by the inclusion of ω_i , for the i^{th} plant. ω_i may be estimated by either a fixed or random effects estimator. In our application below we focus on results generated from a random effects estimator.

$$E_{i,t} = f(X_{i,t}; \theta) + \omega_i + \varepsilon_{i,t} \quad (3)$$

In the SF approach the typical error term is hypothesized to be made up of two parts,

$$\varepsilon_{i,t} = u_{i,t} + v_{i,t} \quad (4)$$

Where $u_{i,t}$ is a one-sided efficiency error term, and $v_{i,t}$ is noise. (Greene 2002) shows that this extension of the SF framework is econometrically tractable via maximum likelihood estimation (MLE). This approach has been labeled Greene's true fixed effect (TFE) and true random effect (TRE) estimators. In the TRE model, the estimates of ω_i are the basis for an estimate of persistent efficiency, and $u_{i,t}$ is time varying efficiency. (Filippini and Hunt 2011, Filippini and Hunt 2012) employ this approach on panels of US states and OECD countries, respectively. However, these models can be difficult to obtain convergence in the MLE when the number of time periods is relatively small and the number of plants is relatively large. This was the same problem encountered by (Boyd and Lee 2016).

An alternative approach is to estimate these error components in a two-stage process (Kumbhakar, Lien et al. 2014). The next section describes the two stage process. The advantages are that the convergence problems are ameliorated and both are treated heterogeneity and the price endogeneity is treated in the first stage using a random effects, instrumental variable approach.

Two stage model for persistent and time varying efficiency

The plant-level efficiency estimates are obtained by a two-stage approach. The first stage uses a plant-level random effects estimator with state-level electricity prices as an instrument for plant-level electricity prices. The general form for the random effects estimate is

$$E_{i,t} = f(X_{i,t}; \theta) + \omega_i + \varepsilon_{i,t} \quad (5)$$

where ω_i is the plant-level random effect for the i^{th} plant and $\varepsilon_{i,t}$ is Gaussian error. These two error components are not directly observable, but the residual of the regression, $E_{i,t} - f(X_{i,t}; \hat{\theta})$, can be decomposed into an estimate of the plant specific effect, $\widehat{\omega}_i$ that is constant over time for each plant and the time varying noise component, $\widehat{\varepsilon}_{i,t}$, based on the estimated parameters, $\hat{\theta}$.

$$\widehat{\omega}_i = E[\omega_i; E_{i,t} - f(X_{i,t}; \hat{\theta}), \hat{\theta}] \quad (6a)$$

$$\widehat{\varepsilon}_{i,t} = E[\varepsilon_{i,t}; E_{i,t} - f(X_{i,t}; \hat{\theta}), \hat{\theta}] \quad (6b)$$

The second stage is used to further extract efficiency estimates from the decomposed error terms using the stochastic frontier. Using the two plant-level estimates from the first stage, a frontier analysis is conducted on each estimated error component

$$\widehat{\omega}_i = \alpha + u_i^{per} + v_i \quad (7a)$$

$$\widehat{\varepsilon}_{i,t} = \alpha + u_{i,t}^{tv} + v_{i,t} \quad (7b)$$

Where the “usual” stochastic frontier model assumptions apply; u_i^{per} and $u_{i,t}^{tv}$ follow a one-sided exponential distribution and v_i and $v_{i,t}$ are noise. We are not interested in the estimate, $\hat{\alpha}$, per se, but in the estimates of $\widehat{u}_{i,t}^{tv}$ and \widehat{u}_i^{per} based on the residuals, $\widehat{\omega}_i - \hat{\alpha}$ and $\widehat{\varepsilon}_{i,t} - \hat{\alpha}$, from each regression. The standard JMLS (Jondrow, Materov et al. 1982) frontier estimates from STATA of $\widehat{u}_{i,t}^{tv}$ and \widehat{u}_i^{per} are obtained from these two 2nd stage regressions. The exponent of these JMLS estimates represent time-varying (*tv*) and persistent (*per*) efficiency.

$$tv_{i,t} = \exp(\widehat{u}_{i,t}^{tv}), \quad (8a)$$

$$per_i = \exp(\widehat{u}_i^{per}), \quad (8b) \text{ and}$$

$$\widehat{tot}_{i,t} = \exp(\widehat{u}_{i,t}^{tv} + \widehat{u}_i^{per}) \quad (8c).$$

Where $tot_{i,t}$ is the combined total efficiency estimate.

Empirical Results

The model and estimation approach described above is applied to a panel dataset for the CM and MECS in separate analysis. In principle, these data sets could be pooled, but we do separate analyses for two reasons. The first is that the MECS is a stratified sample and the CM is a Census, i.e., includes all plants. We wish to explore how these two data collections might impact the results. The second is that MECS has much more detail on energy, including a physical measure of fossil fuels and the corresponding detail needed to compute fuel specific, plant-level prices. The detailed nature of the MECS also might result in different persons within a firm/plant to being tasked with filling it out, compared to the CM. In some sense the MECS might include better or more accurate data on energy use, and on fuels in particular. While there are differences in the parameter estimates, results, particularly for the efficiency measures, the broad pattern is similar. The summary section compares the stylized results from the two data sets.

The next section presents the results from the CM analysis for each of the four sectors and two energy types. The impact of the instrumental variables on the price elasticities is discussed. The efficiency estimates are discussed in some detail, including the decomposition of efficiency into persistent and time varying and the comparison of efficiency for existing and new plants, i.e., whether new plants that enter the industry are more efficient than their counterparts. Finally we explore the aggregate implications for the estimated distribution of total efficiency in these sectors.

The subsequent sections highlight some differences that arise from using the MECS sample. These include the ability to instrument for fossil fuel prices using the same approach as employed in the CM dataset and a companion analysis that focused on fuel-electricity substitution based on a logit analysis of energy shares. Logit statistical models account for the fact that fuel share as independent variables will sum to unity, requiring a different statistical modeling approach.

CM Two Stage Parameter Estimates

In the first stage of the two-stage process the ad hoc energy demand model described in equations (2) and (3) using a random effects estimator for ω_i , the plant level random effect for the i^{th} plant¹⁰. In addition we also instrument for endogeneity of plant level electricity prices using state level prices are reported in the EIA SEDS data. Two sets of analysis were done; one includes capital stock and the other does not. Capital stock data is not available in the CM for 2012. Results for the models without capital stock, but including 2012, are included in the appendix. Estimates of the price elasticities and efficiency measure are very similar.

Tables 1-4 show the estimates for inorganic, organic, resins and plastics, and fertilizers, respectively, for each energy type. Results are for the random effects estimator (RE) and with and without the instrumental variables for Price (IV-RE). The use of instrumental variables for electricity price results in lower electric price elasticities in all four sectors, but still exhibits relatively high price elasticities, ranging from -0.75 to -1.3. In three of the four sectors the use of instruments also eliminates the significant estimates of complementarity for electricity price in the fossil fuel equation. In two cases the coefficient of concern changes sign, and in all cases the coefficient is no longer significant. Fossil fuel price elasticities are all greater than unity, ranging from -1.2 to -1.3. For the most part there isn't significant evidence of substitution (significant cross price coefficients) between electricity and fuels. However, in Resins and Plastics there is significant complementarity of fossil fuels in the electric equation, but not the reverse. While we do not model the dynamics, we interpret the high elasticities as reflecting a long-run phenomenon due to cross sectional variation in our data.

Non-energy materials is significant in only one sector and for electricity use. Inorganic chemicals is a very diverse collection of products and processes, some of which are quite electric intensive. One is industrial gases. It may be that some plants in this sector primarily mix or bottle gases made elsewhere for delivery. If that is the case, then those plant might have high non-energy material shares. Examining the CM micro data found that this was often the case; this sector had the highest level and variation in non-energy material shares. Even with the six-digit NAICS, control we believe that the negative and significant coefficient reflects this underlying phenomenon.

The self-generation ratio is always significant and has the expected sign, positive for fuel and negative for electricity, with exception of fuel use in fertilizers. The mean time varying and persistent efficiencies are almost all above 0.8 and have very small standard deviations. Since total efficiency is the product of

¹⁰ A fixed effect first stage was also estimated, but those results are not presented here. The random effects stage one resulted in better convergence of the SF second stage, to the random effects estimator is our preferred approach.

the two components the mean for total efficiency is smaller. We will take a closer look at the efficiency distribution, focusing on results from the IV model, the preferred specification.

Table 1 Stage one: Random Effects Instrumental Variable Estimates for Inorganic Chemicals, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
<i>lnEmp</i> Log Employment	0.0741*	0.0711*	0.314***	0.308***
<i>lnK</i> Log Capital	0.339***	0.347***	0.265***	0.275***
<i>lnNEM</i> Log non-energy Materials	-0.318***	-0.323***	0.0442	0.0397
<i>lnY</i> Log Total Value of Shipments	0.899***	0.910***	0.417***	0.431***
<i>GERATIO</i> Self Generation Ratio	-0.494***	-0.524***	2.086***	2.058***
<i>lnP_{NG}</i> Log Natural Gas Price	-0.0399	-0.0690	-1.181***	-1.242***
<i>lnP_{Elec}</i> Log Electricity Price	-1.271***	-1.098***	-0.131	0.176
Constant	-2.711***	-2.484***	3.649***	4.692***
Observations	2400	2400	2400	2400
Number of Firms	1400	1400	1400	1400
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.983	0.835	0.972	0.789
Std Dev	0.0203	0.0448	0.0698	0.119
Persistent Efficiency	0.980	0.971	0.975	0.951
Std Dev	0.0405	0.0855	0.0707	0.0773
Overall Efficiency	0.964	0.811	0.948	0.751
Std Dev	0.0445	0.0790	0.0201	0.119
*** p<0.01, ** p<0.05, * p<0.1				

Table 2 Stage one: Random Effects Instrumental Variable Estimates for Organic Chemicals, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
<i>lnEmp</i> Log Employment	0.306***	0.309***	0.153***	0.158***
<i>lnK</i> Log Capital	0.196***	0.200***	0.368***	0.377***
<i>lnNEM</i> Log non-energy Materials	-0.0207	-0.0209	0.0298	0.0313
<i>lnY</i> Log Total Value of Shipments	0.652***	0.650***	0.483***	0.478***
<i>GERATIO</i> Self Generation Ratio	-1.839***	-1.850***	1.006***	0.965***
<i>lnP_{NG}</i> Log Natural Gas Price	-0.0646	-0.0841	-1.262***	-1.318***
<i>lnP_{Elec}</i> Log Electricity Price	-0.973***	-0.917***	-0.405***	-0.244
Constant	-1.973***	-2.169***	3.503***	2.867***
Observations	1900	1900	1900	1900
Number of Firms	1100	1100	1100	1100
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.865	0.859	0.980	0.807
Std Dev	0.0710	0.0417	0.0990	0.0864
Persistent Efficiency	0.843	0.833	0.969	0.965
Std Dev	0.0514	0.0721	0.0971	0.0990
Overall Efficiency	0.730	0.715	0.950	0.779
Std Dev	0.0524	0.0695	0.000610	0.0660
*** p<0.01, ** p<0.05, * p<0.1				

Table 3 Stage one: Random Effects Instrumental Variable Estimates for Resins and Plastics, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
<i>lnEmp</i> Log Employment	0.295***	0.296***	-0.0428	-0.0435
<i>lnK</i> Log Capital	0.254***	0.260***	0.356***	0.375***
<i>lnNEM</i> Log non-energy Materials	0.0112	0.0250	0.0888*	0.132***
<i>lnY</i> Log Total Value of Shipments	0.534***	0.524***	0.559***	0.528***
<i>GERATIO</i> Self Generation Ratio	-1.399***	-1.411***	2.177***	2.147***
<i>lnP_{NG}</i> Log Natural Gas Price	-0.209***	-0.258***	-1.058***	-1.204***
<i>lnP_{Elec}</i> Log Electricity Price	-0.922***	-0.754***	-0.326***	0.179
Constant	-1.247***	-0.801**	1.228***	2.558***
Observations	2300	2300	2300	2300
Number of Firms	1300	1300	1300	1300
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.889	0.890	0.981	0.816
Std Dev	0.0459	0.0621	0.000608	0.0353
Persistent Efficiency	0.985	0.847	0.974	0.965
Std Dev	0.0551	0.0395	0.0451	0.0864
Overall Efficiency	0.876	0.754	0.956	0.788
Std Dev	0.0360	0.0608	0.0460	0.0843
*** p<0.01, ** p<0.05, * p<0.1				

Table 4 Stage one: Random Effects Instrumental Variable Estimates for Fertilizers, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
<i>lnEmp</i> Log Employment	0.341***	0.333***	0.195	0.183
<i>lnK</i> Log Capital	0.244***	0.259***	0.0728	0.107
<i>lnNEM</i> Log non-energy Materials	0.0372	0.0373	0.208	0.186
<i>lnY</i> Log Total Value of Shipments	0.603***	0.607***	0.565***	0.600***
<i>GERATIO</i> Self Generation Ratio	-1.515***	-1.543***	-0.00885	-0.0526
<i>lnP_{NG}</i> Log Natural Gas Price	-0.0605	-0.0696	-1.173***	-1.302***
<i>lnP_{Elec}</i> Log Electricity Price	-1.411***	-1.291***	-0.583**	-0.0333
Constant	-4.177***	-3.972***	1.045	2.373
Observations	300	300	300	300
Number of Firms	200	200	200	200
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.987	0.885	0.968	0.783
Std Dev	0.000363	0.0651	0.101	0.212
Persistent Efficiency	0.979	0.972	0.967	0.886
Std Dev	0.0522	0.0524	0.0739	0.254
Overall Efficiency	0.966	0.860	0.936	0.693
Std Dev	0.0528	0.0469	0.0739	0.0791
*** p<0.01, ** p<0.05, * p<0.1				

Including capital stock allows us to look at the short-run and long-run effect of scale on energy use. We define the long-run elasticity of energy with respect to scale as the sum of the coefficients for labor, non-energy materials, total value of shipments, and capital stock. This would measure the percentage impact on energy use of a larger plant for a percent change in both variable and fixed inputs and outputs. In the short run, capital is fixed. The short-run elasticity is the sum of the coefficients for the variable inputs and output. Table 5 shows that, in the long run, the elasticity of scale is close to or slightly greater than unity. The values greater than unity may reflect the tendency for energy-intensive activities to be located in larger plants. The smaller short-run elasticities reflect that, as variable inputs and production fall relative to the fixed capital stock, energy use falls less than proportionally. This is consistent with observations that the energy output ratio tends to rise as plants produce at less than full capacity over the business cycle.

Table 5 Elasticities of Scale with respect to Energy Use

	Inorganic	Organic	Resins and Plastics	Fertilizer
	Electric			
Long Run	1.0051	1.1381	1.105	1.2363
Short Run	0.6581	0.9381	0.845	0.9773
	Fuel			
Long Run	1.0537	1.0443	0.9915	1.076
Short Run	0.7787	0.6673	0.6165	0.969

Efficiency Results - Census of Manufacturing

This section explores the efficiency estimates from the CM analysis in more detail, focusing on the estimates from the IV-RE model. While the mean and standard deviations for the plant-level efficiencies show that the level of efficiency is fairly high and tightly distributed. For electricity, fertilizers have the highest efficiencies, followed by inorganics, resins, plastics, and organic chemicals. For fossil fuels resins, plastics are the most fuel efficient, followed by organics, inorganics and fertilizers.

The kernel densities for overall efficiency, shown in Figures 1 and 2, reveal even more. The distributions for fossil efficiency are more tightly clustered than for electricity. More importantly, there are virtually no plants that might be called “highly inefficient”; the left tail is very thin. In addition, there are few plants that are considered 100% efficient; this puts the mean efficiency estimates into a different light. One way to view the overall level of efficiency in each industry would be to compute the reduction in aggregate energy use if all plants were “efficient”, i.e., achieve some empirically relevant level of performance other than 100%. Since there are empirically few plants that are 100% efficient, we define an “efficient plant” as one that performs at the 90th percentile of the corresponding efficiency distribution. We take the plant-level energy use and reduce it by the amount needed to put it at the 90th percentile. If the plant is already at the 90th or greater, then the plant is already efficient. The ratio of the sum of the “efficient” energy consumption to the sum of the actual energy use reflects the potential level of energy use if all plants were efficient. Table 6 shows the potential percent reduction in energy from eliminating inefficiency as measured by one minus the above ratio. The average is about -9%. This may seem like a small percentage, but since the base level of energy use in this sector is large, this could be to a substantial amount of energy reduction. This result is also consistent with what (Boyd 2016) reports in a meta-analysis of 2 dozen industry case studies; energy intensive sectors have much tighter distributions of estimated efficiency than non-energy intensive sector. It is possible that competitive industries that produce commodities and have a high proportion of energy in their costs will not tolerate as much energy inefficiency.

Table 6 Potential Reduction in Energy Use if All Plants Were Efficient (90th percentile) by sector and energy type

	Inorganic	Organic	Resins & Plastic	Fertilizers
Electricity	-7%	-14%	-13%	-4%
Fuels	-9%	-7%	-8%	-7%

Another consideration is whether new plants that enter the industry might be more efficient than their existing counterparts. On the one hand, a new plant can have more advanced technology, but it may initially exhibit operations management that poor. Over time, with learning, a new plant may become more efficient. We compare the mean time-varying (TV) and persistent (PER) efficiency of new vs. existing plants in table 7. A pattern along the lines described above emerges. In almost every sector-energy combination, the time varying efficiency of new plants is worse (lower) and statistically significant via a t-test for difference in group means. The exception is fuel use in fertilizer. The pattern is opposite for persistent efficiency, except for fuel use in inorganics. Fewer of these differences are statistically significant. We interpret this to mean that, while new plants may have slight advantages in technology, when they first enter the industry, those advantages are not fully realized, i.e., start out with lower time-varying efficiency. Over time this difference in time-varying efficiency goes away, i.e., these plants learn by doing as they become existing plants five years later (the next CM year of the data). This analysis does not explore these dynamics in detail, so this is a hypothesis to examine in future research.

Table 7 Comparison of Efficiency of New vs Existing Plants by sector and energy type

		Inorganic		Organic		Resins & Plastic		Fertilizers	
		New	Existing	New	Existing	New	Existing	New	Existing
Electric	TV	0.812***	0.823	0.850***	0.860	0.867***	0.880	0.865*	0.873
	PER	0.971*	0.970	0.977***	0.965	0.979	0.978	0.974	0.971
Fuel	TV	0.782***	0.792	0.779***	0.800	0.796***	0.813	0.767	0.762
	PER	0.948	0.949	0.966*	0.962	0.965*	0.964	0.912	0.884

*** p<0.01, ** p<0.05, * p<0.1

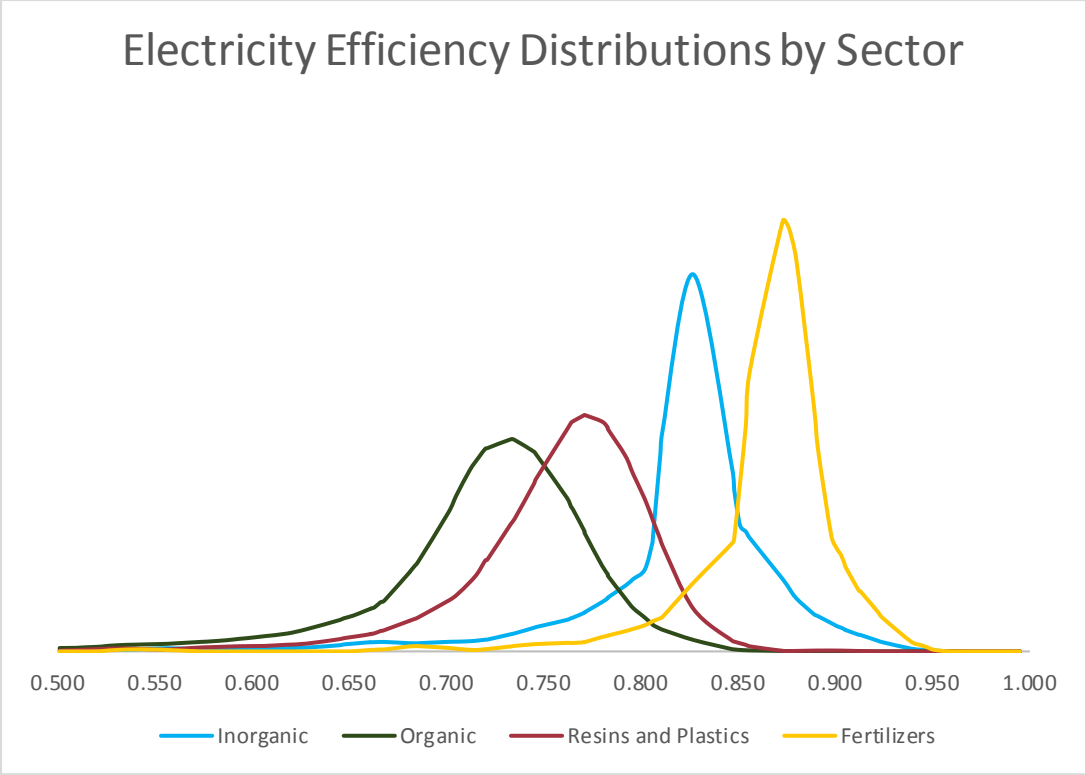


Figure 1 Kernel Density for Plant level Electricity Efficiency by sector

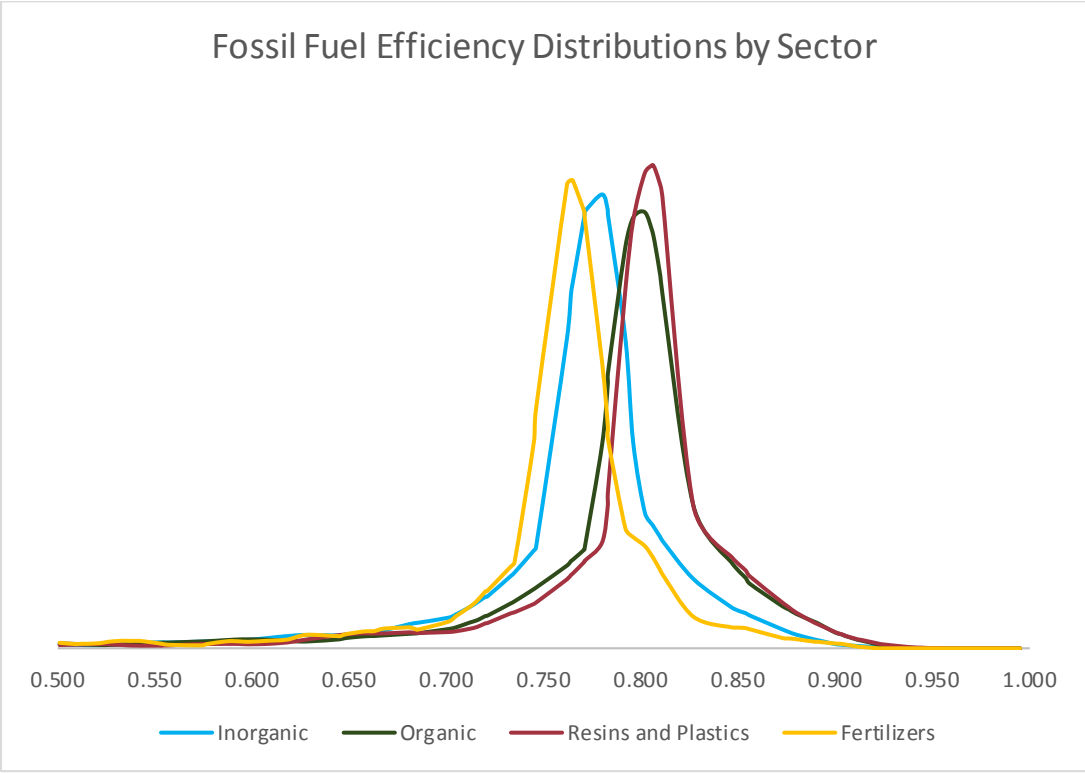


Figure 2 Kernel Density for Plant level Fossil Fuel Efficiency by sector

MECS Two Stage Parameter Estimates

We conduct a parallel analysis to the analysis using the CM data, but using the MECS sample. The major difference is that the MECS data allows us to use plant-level fuel and electric prices, with fuel prices begin further broken down into natural gas and all other. The use of plant-level fuel prices raises the same price endogeneity issues as the plant-level electric prices does in the CM data. The same two-stage estimation strategy is used, but we instrument the plant-level energy prices with the corresponding state-level prices for electric, natural gas, and other fuels.

Results are shown in table 8-12. The IV models generally result in smaller but still significant electricity price elasticities; the exception is organics. For natural gas prices, the results of the IV is that natural gas elasticities are similar in magnitude but no longer significant. This suggests that endogeneity is less of an issue for natural gas. Natural gas is a much more national market, with access to pipeline gas markets for large users. The electric market is fragmented, and local plant-level prices possibly are subject to more heterogeneity and local influence. Complementarity, i.e., negative cross price elasticities, are prevalent in the MECS analysis, but are usually not significant in the IV models with the exception of organic chemicals.

Table 8 Stage one - MECS: Random Effects Instrumental Variable Estimates for Inorganics, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
Log total value of shipments	0.562***	0.593***	0.416***	0.434***
Log employment	0.268***	0.278***	0.503***	0.511***
Log non-energy material cost	-0.162***	-0.180***	0.0935	0.0845
Log capital	0.264***	0.309***	0.260***	0.255***
Self-Generation Ratio				
Log Price of "other" fuels	0.0235	0.0412	-0.149***	-0.279*
Log Price of Natural Gas	-0.0726	0.270	-0.583***	-0.740
Log Price of Electricity	-1.321***	-1.055***	-0.000948	0.0464
Observations	1300	1300	1300	1300
Number of firm	700	700	700	700
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.876	0.881	0.767	0.757
Persistent Efficiency	0.974	0.796	0.963	0.945
Overall Efficiency	0.853	0.701	0.739	0.716
*** p<0.01, ** p<0.05, * p<0.1				

Table 9 Stage one - MECS: Random Effects Instrumental Variable Estimates for organics, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
Log total value of shipments	0.317***	0.322***	0.329***	0.322***
Log employment	0.429***	0.436***	0.384***	0.377***
Log non-energy material cost	0.0306	0.0284	0.0588	0.0584
Log capital	0.309***	0.316***	0.357***	0.347***
Self-Generation Ratio				
Log Price of "other" fuels	0.0110	-0.0736	0.0927	-0.0688
Log Price of Natural Gas	-0.105**	-0.0143	-0.480***	-0.329
Log Price of Electricity	-0.658***	-0.758***	-0.630***	-0.967***
Observations	1100	1100	1100	1100
Number of firm	600	600	600	600
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.991	0.920	0.978	0.798
Persistent Efficiency	0.974	0.973	0.962	0.960
Overall Efficiency	0.966	0.895	0.941	0.767
*** p<0.01, ** p<0.05, * p<0.1				

Table 10 Stage one - MECS: Random Effects Instrumental Variable Estimates for Resins & Plastics, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
Log total value of shipments	0.389***	0.408***	0.418***	0.427***
Log employment	0.330***	0.333***	0.525***	0.521***
Log non-energy material cost	-0.00264	0.00276	0.106	0.112*
Log capital	0.358***	0.377***	0.284**	0.298**
Self-Generation Ratio				
Log Price of "other" fuels	-0.0151	0.0318	0.0117	0.0135
Log Price of Natural Gas	-0.0236	-0.0591	-0.202**	-0.231
Log Price of Electricity	-0.913***	-0.642***	-0.529***	-0.333
Observations	1300	1300	1300	1300
Number of firm	500	500	500	500
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.913	0.913	0.788	0.763
Persistent Efficiency	0.807	0.785	0.956	0.952
Overall Efficiency	0.737	0.717	0.753	0.727
*** p<0.01, ** p<0.05, * p<0.1				

Table 11 Stage one - MECS: Random Effects Instrumental Variable Estimates for Fertilizers, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
Log total value of shipments	0.250**	0.289***	0.294*	0.375**
Log employment	0.678***	0.714***	0.586**	0.736***
Log non-energy material cost	0.0723	0.0542	0.167	0.135
Log capital	0.328***	0.358***	0.310**	0.275*
Self-Generation Ratio				
Log Price of "other" fuels	-0.0320	-0.177	-0.0448	0.146
Log Price of Natural Gas	0.166	0.239	-0.0499	0.463
Log Price of Electricity	-0.906***	-0.562**	-1.010***	-0.482
Observations	300	300	300	300
Number of firms	100	100	100	100
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.900	0.896	0.977	0.981
Persistent Efficiency	0.971	0.968	0.951	0.933
Overall Efficiency	0.874	0.866	0.929	0.914
*** p<0.01, ** p<0.05, * p<0.1				

Fuel switching¹¹

A natural extension of the Cobb-Douglas stochastic frontier fuel demand models specified in equation (2) is to consider the pure substitution effects of fuel price changes that can be attributed to intra-plant fuel switching. Note that if higher energy prices are encouraging plants to invest in unobserved energy efficiency, then the fuel price coefficients in equation (2) will capture both the fuel switching effects of price changes as well as the additive inverse of the impact of fuel prices on energy efficiency. Indeed, the latter of these two effects may explain why the stochastic frontier results presented in Tables 1 – 4 generally do not find statistically significant evidence of fuel substitutability, and in several sectors actually find evidence of a complementary relationship among fuel use, which is statistically significant in Resins & Plastics (CM analysis) and in Organics (MECS analysis).

One possible estimation strategy to address this issue is to use instrumental variables (IV) estimators, but the IV approach relies on instruments for fuel prices that are uncorrelated with unobserved investments in energy efficiency. This approach may be suspect because any exogenous shocks to fuel prices are also likely to affect energy efficiency investments unless we make additional restrictive assumptions such as permanent efficiency investments that only respond to upward fuel price shocks (Gately and Huntington 2002).

Alternatively, a structural fuel share model can be estimated that relies on the same basic Cobb-Douglas specification given in equation (2) in order to isolate the pure fuel price substitution effects. Fractional response logit models are a natural choice for estimating fuel shares because the shares are bounded between zero and one, and the sum of all fuel shares in a plant must sum to one (see Papke and Wooldridge, 1996 for an overview of the fractional response logit model and the extension to a multinomial setting provided by Buis, 2012). From a theoretical econometric perspective, the fractional response multinomial logit model adopted herein is equivalent to the multi-step logit demand share estimators employed by several previous authors with the added benefit that our approach is capable of handling zero shares for plants that only use one fuel type (see, for example, Pindyk 1977, Considine and Mount, 1984, and Urga and Walters, 2003 for an overview of the multi-step logit demand share models used in industry-level aggregate fuel share estimation).

The following fractional response multinomial logit model is estimated in order to isolate the fuel switching behavior among chemical plants in response to fuel price changes:

$$s_{j,K,t,i} = \frac{E_{j,K,t,i}}{\sum_j E_{j,K,t,i}} = g(X_{j,K,t,i} \beta_{j,K}) = \frac{e^{X_{j,K,t,i} \beta_{j,K}}}{\sum_j e^{X_{j,K,t,i} \beta_{j,K}}}, \quad j \in \{E, N, O\}, \quad (9)$$

where the share of fuel j of plant i in sector K at time t , $s_{j,K,t,i}$, is simply defined as the ratio of energy demand for fuel j measured in millions of British thermal units (MMBtu), $E_{j,K,t,i}$, divided by the total energy demand for the plant, $\sum_j E_{j,K,t,i}$. Because the fractional response multinomial logit model is

estimated over the MECS subsample that includes detailed fuel input at each plant we consider fuel demand for three different fuel types: electricity (E), natural gas (N), and a composite “other” fuel (O).

¹¹ In this section we use the term “fuel switching” as it has been commonly been used in the economic literature to describe the switching between different types of energy, including electricity and not just among fossil fuels.

$X_{j,K,t,i}$ is a vector plant characteristics, and the fuel-specific demand coefficients, β_j , are estimated assuming a logit distribution such that they maximize the log-likelihood function given by the following:

$$LL(\beta) = \sum_i \sum_t \sum_j s_{j,K,t,i} * g(X_{j,K,t,i} \beta_{j,K}). \quad (10)$$

The specification for $X_{j,K,t,i} \beta_j$ is assumed to follow the same specification used in the stochastic demand models in equation (2) as follows:

$$X_{j,K,t,i} \beta_{j,K} = \alpha + \beta_{j,K,t}^y \ln Y_{i,t} + \beta_{j,K,t}^{Emp} \ln Emp_{i,t} + \gamma_{j,K,t}^E \ln P_{E,t,i} + \gamma_{j,K,t}^F \ln P_{F,t,s} + \sum_{k \in K} \alpha_k DNAICS_k + v_{j,K,i,t} + u_{j,K,i,t} \quad (11)$$

Note that if we drop the j subscript on the inefficiency component, $u_{j,K,i,t}$, then the inefficiency term drops out of both the numerator and denominator in equation (9). This simplifying assumption therefore allows us to interpret the price coefficients, γ , in equation (11) as measuring the pure fuel-switching price elasticities net of any energy efficiency effects of energy prices. The assumption that energy efficiency effects are netted out in the logit fuel share models is tantamount to assuming that investments in electricity efficiency have 100% spillover for fuel efficiency and vice-versa. This may be a reasonable assumption unless there are significant production processes within a plant where fuels are not substitutable and considering that we focus on fuel used for heat and power, not feedstocks. Indeed, if we compare the estimated fuel and electricity inefficiency terms in Table 1-4 or 8-12 the results suggest that average intra-sector fuel and electricity efficiencies are generally comparable in magnitude.

Results from the unrestricted fractional response multinomial logit model given in equations (9) and (10) are presented in Table 12. Electricity shares are the omitted fuel category in Table 12, and all coefficient estimates can be interpreted as measuring the impact of a change in $X_{j,K,t,i}$ on the log odds ratio relative to the omitted electricity category. The sign of the estimated coefficients in Table 12 provides some intuition regarding the impact of covariate X on the fuel share relative to electricity. As expected, increased electricity prices generally increase natural gas and other fuel shares relative to electricity shares, and increases in the own fuel prices for natural gas and other fuels typically reduce their respective shares relative to electricity. These patterns generally hold across sectors, and suggest the presence of fuel switching between electricity, natural gas, and other fuels.

Table 12 Unrestricted Estimates of Logit Energy Share Equations, by sector and energy type

	Inorganic		Organic		Fertilizers		Resins and Plastics	
VARIABLES	N. gas	Other	N. gas	Other	N. gas	Other	N. gas	Other
ElectricPr.	1.378***	1.849***	-0.0803	0.0982	-0.462	0.465	0.379**	0.143
N.G. Price	-0.523***	-0.254	-0.456***	0.117	-0.358*	0.382	-0.481***	0.700***
Other Price	-0.190***	-0.471***	0.238***	-0.419***	0.0478	-0.384*	0.104**	-0.380***
Log TVS	-0.460***	-0.261*	-0.0594	-0.0974	0.521**	-0.587	-0.0250	0.130
Log Employ	0.273***	0.619***	-0.0870	0.396***	-0.397**	1.100***	-0.137*	0.268*
Log Material	0.536***	0.353***	0.0675	0.0347	-0.129	0.216	0.238***	-0.105
Log Capital	-0.0425	-0.156	0.0302	-0.222**	-0.0695	-0.278	-0.0659	0.178
geratio	W	W	W	W	W	W	W	W
Constant	-3.414***	-7.011***	1.365*	0.878	0.376	0.381	-1.320	-5.732***
Observations	1300	1300	1100	1100	300	300	1300	1300

*** p<0.01, ** p<0.05, * p<0.1, W=withheld for disclosure purposes

However, the coefficient estimates on fuel prices do not directly reveal the price elasticities of interest, γ , in equation (11) because all three fuel share coefficients are not simultaneously identified. For example, the estimated coefficient on electricity prices in the natural gas share equation simply provides an estimate of the difference of the cross-price elasticity of natural gas for electricity and the own-price elasticity of electricity, $\gamma_{N,K,t}^E - \gamma_{E,K,t}^E$. In order to recover the price elasticities it is necessary to impose traditional neoclassical constraints of homogeneity ($\gamma_{j,K,t}^j = \sum_z \gamma_{j,K,t}^z, \forall z \neq j$) and symmetry ($\gamma_{z,K,t}^j = \gamma_{j,K,t}^z, \forall z \neq j$). Letting γ^* denote the differenced elasticities estimated in the multinomial logit model implies that homogeneity is formally imposed in the logit model by constraining $\gamma_{j,K,t}^{j*} = \sum_z \gamma_{j,K,t}^{z*}, \forall z \neq j, j \in (N, O)$. Symmetry requires one additional constraint that $\gamma_O^{E*} + \gamma_N^{O*} = \gamma_O^{N*} + \gamma_N^{E*}$.

Results from the constrained fractional response multinomial logit model are presented in Table 13. It is useful to note that the fuel share coefficient estimates from the constrained model are generally comparable to the results from the unconstrained models in Table 13 for most sectors in both sign and magnitude. This feature of the results suggests that the assumptions of symmetry and homogeneity are reasonable within the heavy chemicals sector. As before, the coefficient estimates simply provide some intuition on the sign of the impact of any covariate X on the share of a fuel relative to electricity shares. In the restricted models, however, it is possible to recover the elasticity estimates due to the imposition of symmetry and homogeneity.¹² Price elasticity estimates are provided in Table 14 and suggest that electricity, natural gas and other fuels are generally net substitutes in production (i.e. positive cross price elasticities) once the impact of investments in energy efficiency are netted out of the price elasticities. As expected, the own price elasticities (diagonal elements) are all negative and much smaller (-0.07 to -.65) than the counterparts from the MECS data analysis that incorporate general demand reduction (-.2 to -1.1). Comparing sectors, some patterns emerge with respect to sector specific fuel-switching capabilities with respect to electricity. The largest elasticity estimates in absolute terms are in the Inorganic sector, suggesting that this sector may be more responsive to changes in electricity use for fuel prices due to switching capabilities. This estimate of the substitution is still only about 0.3 for natural gas and other fuels with respect to electricity use. All other substitution elasticities between fuel(s) and electricity are 0.1 or less. We also note that non-energy materials is significant in this sector and nowhere else. This was true for both the CM and MECS frontier models. This might bear further examination.

The fertilizer sector is the least responsive to changes in own-natural gas prices but similar for other fuels. Natural gas is used as a feedstock. While the MECS (and CM) forms are very explicit that natural gas used for feedstock is not reported for heat and power, there may be issues with this reporting in fertilizers. However, the same can be said for some organic chemical sub-sectors. Furthermore, if we consider the difference in elasticity estimates from the stochastic frontier models in Table 1-4 and 8-12 and the fractional response multinomial logit models in Table 14 as a crude approximation to the impact of fuel price changes on energy efficiency, we see that fuel price increases have the largest impact on efficiency improvements in the Inorganic sector (as estimated by the MECS analysis) where fuel switching capabilities are the highest and plants are more likely to experience larger spillover effects due to the substitutability of fuel in related production processes.

¹² For example adding respective coefficients across fuel types for natural gas and other fuels yields the following: $\gamma_N^{N*} + \gamma_O^{N*} = -3 * \gamma_E^N$. The result suggests that the cross price elasticity of electricity for natural gas is simply the sum of the coefficient estimates on natural gas prices in the natural gas and other fuel share equations divided by negative 3. Remaining cross-price elasticities can be derived in a similar fashion, and the own-price elasticity of electricity is recovered from the homogeneity assumption.

Table 13 Restricted Estimates of Logit Energy Share Equations, by sector and energy type

VARIABLES	Inorganic		Organic		Fertilizers		Resins and Plastics	
	N. gas	Other	N. gas	Other	N. gas	Other	N. gas	Other
Electric Pr.	1.024***	0.929***	0.0888	0.130	-0.0335	0.125	0.364***	0.317***
N.G. Price	-0.752***	-0.367***	-0.333***	0.286***	-0.0835	0.275**	-0.493***	0.0810
Other Price	-0.272***	-0.562***	0.245***	-0.416***	0.117	-0.400**	0.128***	-0.398***
Log TVS	-0.522***	-0.384***	-0.0544	-0.0967	0.570***	-0.627	-0.0290	0.0977
Log Employ	0.273***	0.599***	-0.0794	0.402***	-0.353*	1.106***	-0.131*	0.305**
Log Material	0.572***	0.433***	0.0665	0.0353	-0.185	0.247	0.236***	-0.124
Log Capital	-0.0722	-0.224**	0.0524	-0.209**	-0.0221	-0.315	-0.0675	0.165
geratio	W	W	W	W	W	W	W	W
Constant	-1.454***	-2.866***	0.404	0.400	-1.834**	1.941	-1.251**	-4.935***
Observations	1300	1300	1100	1100	300	300	1300	1300

*** p<0.01, ** p<0.05, * p<0.1, W=withheld for disclosure purposes

Table 14 Own and cross price elasticities - Restricted Logit model by energy type and sector

Inorganics			
	Electricity	Naturalgas	Other
Electricity	-0.651		
Naturalgas	0.373	-0.379	
Other	0.278	0.00595	-0.284
Organics			
	Electricity	Naturalgas	Other
Electricity	-0.0729		
Naturalgas	0.0159	-0.317	
Other	0.0570	0.302	-0.359
Fertilizers			
	Electricity	Naturalgas	Other
Electricity	-0.0304		
Naturalgas	-0.0638	-0.147	
Other	0.0942	0.211	-0.305
Resins and Plastics			
	Electricity	Naturalgas	Other
Electricity	-0.227		
Naturalgas	0.137	-0.355	
Other	0.0898	0.218	-0.308

Summary

This paper presents estimates of the distribution of energy efficiency and price elasticities in the four major energy using sectors of the upstream, energy-intensive portions of the Chemical Industry. We analyze data from the CM and MECS separately, since these data sources have their own strengths and weaknesses. If we compare the mean efficiency estimates between the two data sets (table 15) the mean fuel efficiency are fairly similar. Electricity efficiency in inorganics and organics and fuel use in

fertilizer differ the most. There is no evidence of bias, in the sense that one data source uniformly has higher or lower mean efficiency. None of the mean efficiencies are particularly large. When comparing price elasticities, the fuel price elasticities show the least similarity across the data sets. It should be noted that the MECS fertilizer elasticity estimate for fuel use is positive and not significant in the IV model, but negative and significant for the non-IV version. Non-IV estimates for the MECS fuel elasticities might be the preferred estimates for fuels, but not necessarily for electricity due to the aforementioned ability of large energy users to contract lower electricity prices and relative inability to contract lower natural gas prices (where national spot markets determine prices). Recall that the CM data require the fuel use to be imputed from fuel expenditure and state average natural gas prices and MECS has more detailed data on physical consumption, but only for a sample of plants. We would expect the two data sources to be the most different for fuel (natural gas) use. Electricity elasticities are more similar for all but fertilizers. All the MECS elasticities are smaller in absolute magnitude. The logit share analysis has much lower own-price elasticities, but these are net of the “efficiency effect” of higher prices generally so are expected to be lower.

Table 15 Comparison of Mean Efficiency Estimates

	Inorganic		Organic		Resin & Plastics		Fertilizer	
	Electric	Fuel	Electric	Fuel	Electric	Fuel	Electric	Fuel
MECS	0.701	0.716	0.895	0.767	0.717	0.727	0.866	0.914
CM	0.811	0.751	0.715	0.779	0.754	0.788	0.860	0.693

Table 16 Comparison of Own Price Elasticity Estimates

	Inorganic		Organic		Resin & Plastics		Fertilizer	
	Electric	Fuel	Electric	Fuel	Electric	Fuel	Electric	Fuel
MECS	-1.055	-0.740	-0.758	-0.329	-0.642	-0.231	-0.562	0.463
CM	-1.098	-1.242	-0.917	-1.318	-0.801	-1.204	-1.291	-1.302

The CM analysis, since it is not a sample, is the preferred source for the aggregate analysis of the potential savings from efficiency, since all plants are included in the data. That analysis shows that the range of efficiency difference is quite narrow, and the total savings is small in percentage terms, ranging from a low of 4% to a high of 14%, depending on the sector and energy type. The average is about 9%. The relatively small percentage difference in efficiency is consistent with other studies that find energy-intensive sectors e.g., steel, cement, paper, etc. (Boyd and Zhang 2013, Boyd and Guo 2014, Boyd, Doolin et al. 2017) have a much narrower range of efficiency than less energy-intensive ones, e.g., metal based durables, auto assembly, etc. (Boyd 2014, Boyd and Lee 2016). We find that new plants have slightly higher persistent efficiency than existing plants, but enter the industry with lower time-varying efficiency. We interpret this as a new plant learning phenomenon, but this analysis doesn’t model this explicitly. The results for fertilizers might bear further examination since this sector was the most sensitive to the data source (CM vs MECS) and model specification.

Appendix A: Stage-one estimates without Capital Stock

Table 17 Stage one: Random Effects Instrumental Variable Estimates for Inorganic Chemicals, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
Log Employment	0.109***	0.107**	0.345***	0.338***
Log Capital				
Log non-energy Materials	-0.332***	-0.338***	0.0359	0.0285
Log Total Value of Shipments	1.152***	1.177***	0.620***	0.650***
Self-Generation Ratio	W	W	W	W
Log Natural Gas Price	-0.0796	-0.131	-1.201***	-1.270***
Log Electricity Price	-1.402***	-1.106***	-0.228**	0.154
Constant	-2.233***	-1.730***	3.924***	5.131***
Observations	2500	2500	2500	2500
Number of Firms	1400	1400	1400	1400
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.981	0.818	0.972	0.788
Std Dev	0.0197	0.0275	0.0890	0.127
Persistent Efficiency	0.981	0.971	0.969	0.949
Std Dev	0.0197	0.0828	0.000914	0.123
Overall Efficiency	0.963	0.794	0.942	0.748
Std Dev	0.0273	0.0821	0.0916	0.0760
*** p<0.01, ** p<0.05, * p<0.1 W= withheld for disclosure purposes				

Table 18 Stage one: Random Effects Instrumental Variable Estimates for Organic Chemicals, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
Log Employment	0.409***	0.418***	0.304***	0.318***
Log Capital				
Log non-energy Materials	0.0597*	0.0612*	0.169***	0.174***
Log Total Value of Shipments	0.655***	0.657***	0.552***	0.554***
Self Generation Ratio	W	W	W	W
Log Natural Gas Price	-0.0952	-0.142*	-1.332***	-1.421***
Log Electricity Price	-1.091***	-0.944***	-0.656***	-0.370**
Constant	-1.826***	-1.550***	3.231***	3.485***
Observations	2100	2100	2100	2100
Number of Firms	1200	1200	1200	1200
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.866	0.856	0.979	0.792
Std Dev	0.0432	0.101	0.000643	0.0558
Persistent Efficiency	0.848	0.970	0.970	0.964
Std Dev	0.0497	0.0702	0.0721	0.0976
Overall Efficiency	0.735	0.830	0.950	0.763
Std Dev	0.0634	0.0877	0.0736	0.0902
*** p<0.01, ** p<0.05, * p<0.1 W= withheld for disclosure purposes				

Table 19 Stage one: Random Effects Instrumental Variable Estimates for Resins and Plastics, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
Log Employment	0.341***	0.343***	0.0366	0.0356
Log Capital				
Log non-energy Materials	0.00669	0.0177	0.0114	0.0362
Log Total Value of Shipments	0.744***	0.746***	0.921***	0.928***
Self Generation Ratio	W	W	W	W
Log Natural Gas Price	-0.178***	-0.233***	-0.892***	-1.019***
Log Electricity Price	-1.006***	-0.803***	-0.506***	-0.0197
Constant	-1.385***	-0.872***	0.709**	1.951***
Observations	2800	2800	2800	2800
Number of Firms	1600	1600	1600	1600
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.881	0.875	0.980	0.806
Std Dev	0.0469	0.0667	0.000650	0.0849
Persistent Efficiency	0.985	0.979	0.973	0.964
Std Dev	0.000457	0.0186	0.0413	0.0838
Overall Efficiency	0.868	0.856	0.953	0.777
Std Dev	0.0475	0.0673	0.0405	0.0317
*** p<0.01, ** p<0.05, * p<0.1 W= withheld for disclosure purposes				

Table 20 Stage one: Random Effects Instrumental Variable Estimates for Fertilizers, by type of Energy

VARIABLES	Electricity	Electricity	Fuels	Fuels
Log Employment	0.341***	0.343***	0.0366	0.0356
Log Capital				
Log non-energy Materials	0.00669	0.0177	0.0114	0.0362
Log Total Value of Shipments	0.744***	0.746***	0.921***	0.928***
Self Generation Ratio	W	W	W	W
Log Natural Gas Price	-0.178***	-0.233***	-0.892***	-1.019***
Log Electricity Price	-1.006***	-0.803***	-0.506***	-0.0197
Constant	-1.385***	-0.872***	0.709**	1.951***
Observations	2800	2800	2800	2800
Number of Firms	1600	1600	1600	1600
Model	RE	IV-RE	RE	IV-RE
Time-varying Efficiency	0.881	0.875	0.980	0.806
Std Dev	0.0469	0.0667	0.000650	0.0849
Persistent Efficiency	0.985	0.979	0.973	0.964
Std Dev	0.000457	0.0186	0.0413	0.0838
Overall Efficiency	0.868	0.856	0.953	0.777
Std Dev	0.0475	0.0673	0.0405	0.0317
*** p<0.01, ** p<0.05, * p<0.1 W= withheld for disclosure purposes				

REFERENCES

- Allcott, H. and M. Greenstone (2012). "Is There an Energy Efficiency Gap?" The Journal of Economic Perspectives **26**(1): 3-28.
- Amsler, C., A. Prokhorov and P. Schmidt (2016). "Endogeneity in stochastic frontier models." Journal of Econometrics **190**(2): 280-288.
- Bartelsman, E. and W. Gray (1996). The NBER Manufacturing Productivity Data Base. Cambridge MA, National Bureau of Economic Research.
- Boyd, G., M. Doolin, E. Dutrow and S. Zhang (2017). "A New Benchmark of Energy Performance for Energy Management in U.S. and Canadian Integrated Steel Plants." Iron and Steel Technology(May): 53-61.
- Boyd, G. and J. M. Lee (2016). Measuring Plant Level Energy Efficiency and Technical Change in the U.S. Metal-Based Durable Manufacturing Sector Using Stochastic Frontier Analysis, Center for Economic Studies, U.S. Census Bureau.
- Boyd, G. and J. M. Lee (2017). Energy Efficiency, Technical Change and Price Responsiveness in Non-Energy Intensive Chemicals Manufacturing.
- 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. (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.
- Boyd, G. A. and C. Delgado (2012). Measuring Improvement in the Energy Performance of the U.S. Corn Refining Industry. DUKE ENVIRONMENTAL ECONOMICS WORKING PAPER SERIES. Durham NC, NICHOLAS INSTITUTE FOR ENVIRONMENTAL POLICY SOLUTIONS: 1-18.
- Boyd, G. A. and Y. Guo (2014). "An Energy Performance Indicator for integrated paper and paperboard mills: A new statistical model helps mills set energy efficiency targets." Paper360 **March/April**: 26-28.
- Buis, Maarten L. 2012. "Fmlogit: Stata Module Fitting a Fractional Multinomial Logit Model by Quasi Maximum Likelihood." Statistical Software Components.
- Considine, Timothy J and Timothy D Mount. 1984. "The Use of Linear Logit Models for Dynamic Input Demand Systems." The Review of Economics and Statistics, 434-43.
- 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.

Energy Information Administration (1994). NEMS Industrial Module Documentation Report, U.S. Department of Energy.

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.

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

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.

Jondrow, J., I. Materov, K. Lovell and P. Schmidt (1982). "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." Journal of Econometrics **19**(2/3): 233-238.

Kumbhakar, S. C., G. Lien and J. B. Hardaker (2014). "Technical efficiency in competing panel data models: a study of Norwegian grain farming." Journal of Productivity Analysis **41**(2): 321-337.

Papke, Leslie E and Jeffrey M Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates." Journal of Applied Econometrics, **11**(6), 619-32.

Pindyck, Robert S. 1979. "Interfuel Substitution and the Industrial Demand for Energy: An International Comparison." The Review of Economics and Statistics, 169-79.

Urga, Giovanni and Chris Walters. 2003. "Dynamic Translog and Linear Logit Models: A Factor Demand Analysis of Interfuel Substitution in Us Industrial Energy Demand." Energy Economics, **25**(1), 1-21.