

The research program of the Center for Economic Studies (CES) produces a wide range of theoretical and empirical economic analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of CES research papers. The papers are intended to make the results of CES research available to economists and other interested parties in order to encourage discussion and obtain suggestions for revision before publication. The papers are unofficial and have not undergone the review accorded official Census Bureau publications. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. Republication in whole or part must be cleared with the authors.

**ESTIMATING THE DISTRIBUTION OF PLANT LEVEL MANUFACTURING ENERGY
EFFICIENCY WITH STOCHASTIC FRONTIER REGRESSION**

by

Gale A. Boyd *
Duke University

CES 07-07 March, 2007

All papers are screened to ensure that they do not disclose confidential information. Persons who wish to obtain a copy of the paper, submit comments about the paper, or obtain general information about the series should contact Sang V. Nguyen, Editor, [Discussion Papers](#), Center for Economic Studies, Washington Plaza II, Room 206, Bureau of the Census, Washington, DC 20233-6300, (301-763-1882) or INTERNET address sang.v.nguyen@census.gov.

Abstract

A feature commonly used to distinguish between parametric/statistical models and engineering models is that engineering models explicitly represent best practice technologies while the parametric/statistical models are typically based on average practice. Measures of energy intensity based on average practice are less useful in the corporate management of energy or for public policy goal setting. In the context of company or plant level energy management, it is more useful to have a measure of energy intensity capable of representing where a company or plant lies within a distribution of performance. In other words, is the performance close (or far) from the industry best practice? This paper presents a parametric/statistical approach that can be used to measure best practice, thereby providing a measure of the difference, or “efficiency gap” at a plant, company or overall industry level. The approach requires plant level data and applies a stochastic frontier regression analysis to energy use. Stochastic frontier regression analysis separates the energy intensity into three components, systematic effects, inefficiency, and statistical (random) error. The stochastic frontier can be viewed as a sub-vector input distance function. One advantage of this approach is that physical product mix can be included in the distance function, avoiding the problem of aggregating output to define a single energy/output ratio to measure energy intensity. The paper outlines the methods and gives an example of the analysis conducted for a non-public micro-dataset of wet corn refining plants.

- * Work sponsored by the United States Environmental Protection Agency, Office of Atmospheric Programs under contract number W-31-109-Eng-38. Portions of this paper were presented at the *2004 Southern Economics Association Annual Meeting* in New Orleans, LA and the *2004 Illinois Economics Association Annual Meeting* in Chicago IL. The research in this paper has benefited from comments from participants at *Energy Star® Corn Refiners Industry Focus Meeting*, Washington, DC and the *2nd Annual ENERGY STAR corn refiners Focus Workshop*, Indianapolis, IN. The research in this paper was conducted while the authors were Census Bureau research associates at the Chicago Census Research Data Center. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Bureau of the Census or the sponsoring agency. This paper has been screened to insure that no confidential data are revealed. Support for this research at the Chicago RDC from NSF (awards no. SES-0004335 and ITR- 0427889) is also gratefully acknowledged.

Introduction

The concept of energy efficiency in manufacturing is complex. Energy efficiency is often analyzed in terms of *intensity*, or the ratio of energy to manufacturing output. This approach requires that a suitable aggregate measure of output be defined for the denominator of this ratio, as well as a measure of energy. Reduction over time in energy intensity is interpreted as improvement in energy efficiency. It has been demonstrated by a variety of researchers (Boyd, McDonald et al. 1987; Boyd, Hanson et al. 1988; Lui, Ang et al. 1992; Greening, Davis et al. 1997; Ang and Liu 2001; Boyd and Roop 2004) that the aggregation of manufacturing output and shifts in the composition of that output influences the measurement of interpretation of these trends in intensity. There are a number of different techniques that have been applied to the decomposition of energy intensity into trends of composition shift and “real” intensity change (for a survey of methods see Ang and Liu, 2001.)

This paper addresses a different perspective. Rather than simply describe the evolution of the energy output ratio over time, this research seeks to define energy efficiency relative to some benchmark, a notion of the “best practice” or lowest observable energy use. Just at the use of the energy output ratio is one simple approach applied to normalize energy use, the relevant benchmark must also normalize for factors that influence the amount of energy required to produce the particular goods or services. In a seminal paper on measuring production efficiency Farrell (1957) identifies two possible choices for information on this benchmark; *“Although there are many possibilities, two at once suggest themselves – a theoretical function specified by engineers and an empirical function based on the best results observed in practice.”* Engineering estimates of specific energy consumption have often been employed to estimate to compute the theoretical, practical, and economic level of energy use for a given process or production activity (for one example see Phylipsen, BloK et al. 1998). The concept used here for the benchmark is the second approach suggested by Farrell, an empirical estimate of the best observable performance, or “best practice”. This paper uses plant level data and stochastic frontier analysis (SFA) to estimate the best practice given those plant specific characteristics. SFA also provides a direct estimate of the distribution of energy efficiency measured relative to that estimated best practice.

SFA is one of many tools used to measure efficiency in general, not just energy efficiency². SFA uses a parametric statistical approach to estimating efficiency, but non-parametric approach using linear programming, commonly referred to as data envelopment analysis (DEA) are also popular Murillo-Zamorano (2004). Huntington (1995) discusses the potential application of the DEA approach to the question of energy efficiency. The DEA approach has been applied empirically to the study of energy production, for example Byrnes, Fare et al. (1987) in coal mining; (Coggins and Swinton 1996; Yaisawarng and Klein 1994) in electric utilities and associated pollution; (Pacudan and Guzmanb 2002; Hattori, Jamasb et al. 2005; Lowrey, Getachew et al. 2005) in electric distribution; and Kashani (2005) in petroleum production. DEA has also been applied to OECD level analysis of energy and productivity Murillo-Zamorano (2005).

² Studies applying DEA to general industries and productivity are too numerous to list.

The DEA approach has been applied to study energy use in buildings Ferrier and Hirschberg (1992) and the overall productivity of some energy intensive industries (Boyd 1997; Boyd and Pang 2000; Boyd, Tolley et al. 2002) but not focused on the energy efficiency, per se.

This approach has another benefit of eliminating the need to aggregate diverse industry outputs. For an industry that has a single, well defined product several researchers have argued that it is desirable to measure energy use relative to a physical measure of production (Freeman, Niefer et al. 1997; Worrell, Price et al. 1997). However, with the exception of industries like cement, few sectors satisfy this criterion. For example, the specific manufacturing sector which is the empirical focus of this paper, wet corn refining, is a capital intensive process/conversion industry that uses a common raw material, corn, to produce a mix of products ranging from animal feed, to glucose, to ethanol. The approach proposed here allows the energy use to be measured relative to diverse measures of physical production without any assumptions regarding aggregation across those products.

This research also supports the Environmental Protection Agency's (EPA) Energy Star program. Energy Star is an information and voluntary action program designed to give consumers and business more information about the energy efficiency of consumer products, commercial office space, and manufacturing. Energy Star considers a product or business eligible for Energy Star recognition if the energy efficiency is in the 75th percentile or above, for comparable products or businesses. In order to make this assessment, a method to assign a percentile ranking to any given plant is required. SFA is well suited for this task since it provides an estimate the distribution of the energy efficiency.

This research uses confidential data from the Census Bureau as the basis for the modeling. Census bureau rules prevent the release of individual plant results. Since Census permits the clearance of statistical parameters, SFA is again suited for this task. The end result is a model that can be used in a "out of sample" context, i.e. data other than that used in the estimation can be input into the model to compute an estimate for the "best practice" energy use and where the actual energy use lies within a statistical distribution of energy efficiency. This statistical distribution is the basis for ranking the plant on a percentile basis. We call the resulting model the Energy Performance Indicator (EPI) to avoid confusion with other types of engineering and managerial "benchmarks" of energy use.

Energy is a small but significant component of the cost of corn refining. However, the goal of the analysis is to identify the technical energy efficiency and not to address the issue of cost efficiency, i.e. we take a production rather than dual profit or cost perspective. It is presumed that if a level of energy efficiency is observed, that is because it was cost effective at some relevant prices, similar to the notion of revealed preferences. To the extent possible, the analysis focuses on physical, rather than dollar measure of inputs and outputs. This mirrors the focus on the production interpretation of technical efficiency.

This paper describes the methods underlying the EPI and gives an example of the specific form of the statistical model for the wet corn refining industry. The first section provides a motivation for the underlying analysis. The second section explains the details of the statistical frontier approach and its interpretation as an input distance function.

The third section gives estimates for wet corn refining. The last section discusses broader implications, caveats, etc. regarding applying these types of statistical models.

Motivation for the Stochastic Frontier Approach

The concept of the statistical frontier analysis that supports the EPI can be easily explained in terms of the standard linear regression model. This section provides a brief overview. A much more detailed discussion of the evolution of the statistical approaches for estimating efficiency can be found in Green (1993). Consider at first, the simple example of a production process that has a fixed energy component and a variable energy component. Given data on energy use, E , and production, y , for $i=1, N$ plants the parameters α and β can be fit via a simple linear regression model.

$$E_i = \alpha + \beta y_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad (1)$$

The linear regression gives the average relationship between production and energy use. If the departures from the average are due to energy inefficiency we would be interested in a version of (1) that gives the “best” or lowest observed energy use. One way to do this is to shift the line downward so that all the actual data points are on or above the line. This “corrected” ordinary least squares (COLS) regression is one way to represent the frontier.

While the COLS method has its appeal in terms of simplicity, a more realistic view is that not all the differences between the actual data and the frontier are due to efficiency. Since we recognize that there may still be random effects such as errors in data collection/reporting, we still wish to include the statistical noise, or “random error” term, in the analysis, v_i , but also add an second random component, u_i , which follows a one sided distribution, to reflect energy inefficiency. If we expand the simple example of energy use and production to include a range of potential effects we can write the more general version of the stochastic frontier model as

$$E_i = f(Y_i, X_i, Z_i; \beta) + \varepsilon_i \quad (2)$$

$$\varepsilon_i = u_i - v_i \quad v_i \sim N[0, \sigma_v^2],$$

Where

E is energy use, either electricity, non-electric energy, or total primary energy,

Y is production, measured by either physical production or total value

X includes systematic economic decision variables (i.e. non-energy production inputs),

Z includes systematic external factors, and

β includes all the parameters to be estimated.

We assume that energy (in)efficiency, u , is distributed according to some one-sided statistical distribution³, for example gamma, exponential, truncated normal, etc. It is then possible to estimate the parameters of equation (2), along with the distribution

³ We also assume that the two types of errors are uncorrelated, $\sigma_{u,v} = 0$.

parameters of u . The approach that is used to estimate these parameters depends on the type of distribution that is used to represent inefficiency. Gamma is a very flexible distribution, but also generates a model that is very difficult to estimate. Exponential and truncated normal frontier models can be estimated using relatively conventional maximum likelihood techniques available in many modern statistical packages. A wide range of additional distributional assumptions regarding the heteroscedasticity of either u or v are also possible. In addition, the treatment of panel data is a significant issue in the application of stochastic frontier, since the inefficiency term is likely to be correlated over time within a plant or firm. Greene (2002) presents an overview of panel treatments⁴.

Given data for any plant, we can compute the difference between the actual energy use and the predicted frontier energy use

$$E_i - f(Y_i, X_i, Z_i; \beta) + v_i = u_i \quad (3)$$

Since we have estimated the distribution of u , we can compute the probability that the plant inefficiency is greater than this computed difference

$$\Pr(\text{inefficiency} \geq E_i - f(Y_i, X_i, Z_i; \beta) + v_i) \quad (4)$$

We call this probability the EPI score and is equivalent to a *percentile ranking of the energy efficiency* of the plant. In practice we only can measure

$E_i - f(Y_i, X_i, Z_i; \beta) = u_i - v_i$, so this implies that the EPI score is effected by the random component of v_i , i.e. the score will reflect the random influences that are not accounted for by the function $f(*)$. This is different from the more common use of SFA to estimate u_i conditional on the observation of $E_i - f(Y_i, X_i, Z_i; \beta) = u_i - v_i$ within the estimation sample. Jondrow, Materov et al. (1982) have developed such a conditional estimator, but we do not have the ability to release results based directly on plant level observations used in the estimation process. The estimation of efficiency conditional on the random shocks is discussed more below.

The role of the function, $f(*)$ in the EPI is to normalize for exogenous effects, i.e. it controls for factors that influence energy use but are not decided on the basis of energy use alone. As was noted above, the types of production activities and structural factors that are included in the function $f(*)$, used to derive the EPI, are industry specific. However, there are a number of common factors that any industry analysis will likely consider. For simplicity we continue to assume that the function, $f(*)$, is a linear function of the parameters, β , but it may include non-linear forms for Y , X , and Z . There is no guidance on what mathematical form that $f(*)$ should take, so a substantial amount of judgment is exercised.

What variables to include (or exclude) for a given industry is driven by some knowledge and expectations about what factors will have significant influence on energy use in that sector. This choice for Y may be the value of total plant production, a physical production measure, or several physical production measures if an industry produces

⁴ We do not pursue these panel treatments here since we only have two years of data for any given plant.

different products. The ability to represent production as a vector instead of a scalar is very helpful, since the latter is the case for corn refining. X may include quantity and types of materials purchased, labor, or plant capacity. Z may include a variety of external factors like weather variables (e.g. heating and cooling degree days), capacity utilization, regulatory factors, etc. Since the statistical formulation allows us to conduct test of the estimated parameters, the decision to include any of these variables can be driven by the data and model estimates. Conventional statistical tests can determine which factors to include in the model.

The Energy Frontier as an Input Distance Function

The stochastic frontier can be viewed purely as a statistical tool, but also can be viewed from a production theoretic perspective. The stochastic frontier was first proposed for estimation of production functions (Aigner, Lovell et al. 1977). The stochastic frontier has also been applied as a tool to estimate distance functions when there are multiple inputs and outputs, (for example Morrison-Paul, Johnston et al. 2000). The application of the stochastic frontier and its interpretation as a distance function for multiple outputs is particularly useful in this context. In order to establish the connection between the frontier as a purely statistical tool and the interpretation of the distance function we need to connect the conceptual motivation provided above with some specific micro-economic concepts of the minimal energy factor requirements function, input distance function, and sub-vector input distance function. For more details on the underlying production theory of the distance function see Murillo-Zamorano (2004).

First define the minimal energy factor requirements function for the production vector, Y , and the non-energy inputs, X .

$$E^* = f_E(Y, X) = \inf\{E : \text{The input vector } (X; E) \text{ can produce the output vector } Y\} \quad (5)$$

This factor requirements function is analogous to equation (2) above.

The input distance function is defined as the largest scalar amount that one can proportionally reduce all inputs, energy and non-energy, and still be able to produce the output vector Y .

$$D_I(Y, X; E) = \sup\{\lambda : (X/\lambda; E/\lambda) \text{ can produce } Y\} \quad (6)$$

It is also possible to define the sub-vector input distance function by leaving some inputs fixed and only scaling a subset of inputs. In this application we scale a single input, energy.

$$D_{SI}(Y, X; E) = \sup\{\lambda : (X; E/\lambda) \text{ can produce } Y\} \quad (7)$$

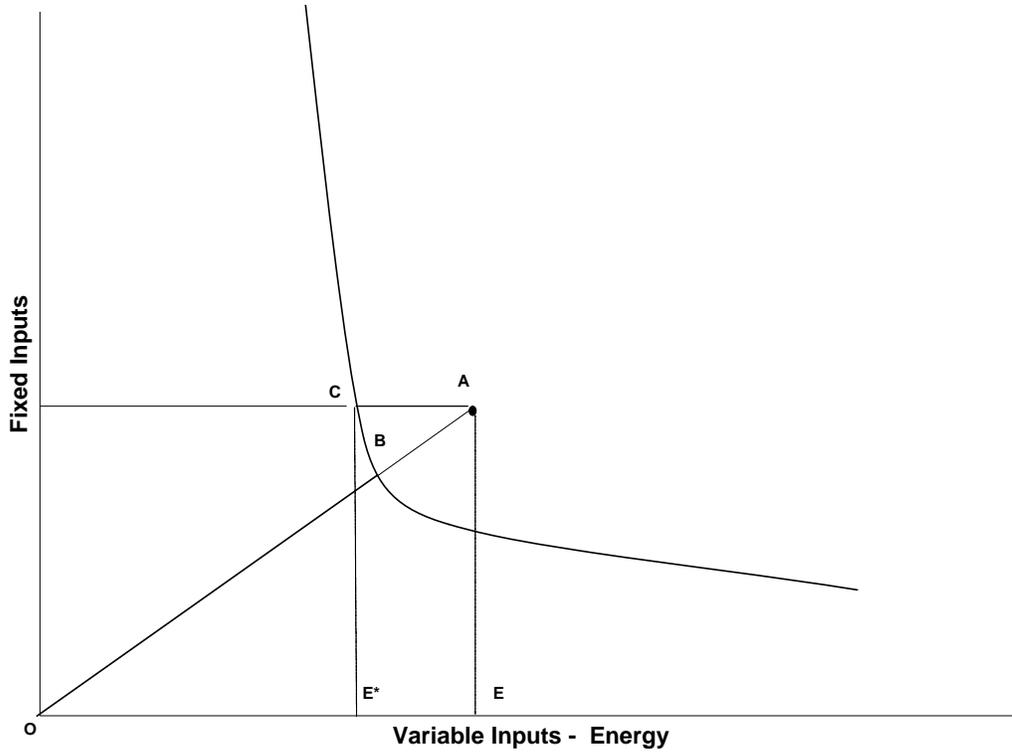


Figure 1 Graphical Example of the Input and Sub-vector Input Distance Functions

Each of these functions is illustrated in figure 1. A standard production isoquant for a fixed input vector Y is shown. In figure 1 the input distance function D_I is the ratio OA/OB .; the sub-vector energy input distance function is OA/OC . It is easy to see that the relationship between the factor requirements function and the sub-vector energy input distance function is,

$$D_{SI}(Y, X; E) = E^*/E = f_E(Y, X)/E \quad (8)$$

Since the distance function, including the energy sub-vector formulation, is homogenous of degree -1 in inputs we have,

$$\begin{aligned} D_{SI}(Y, X; E) &= 1/E D_{SI}(Y, X; E/E) \\ &= D_{SI}(Y, X; 1)/E \\ &= f_E(Y, X)/E \end{aligned} \quad (9)$$

The energy factor requirements function is equivalent to the sub-vector distance function evaluated at $E=1$.

The section above makes the factor requirements function an operational approach via (2) and adding the additional characteristics vector Z .

$$E_i = f_E(Y_i, X_i, Z_i; \beta) + u_i - v_i \quad (10)$$

Based on this empirically operational model we can show that,

$$\begin{aligned} 1 - \frac{u_i - v_i}{E_i} &= \frac{f_E(Y_i, X_i, Z_i; \beta)}{E_i} \\ &= D_{SI}(Y, X; E) \end{aligned} \quad (11).$$

Equation (11) provides a connection between the empirical approach using a factor requirements view and the sub-vector distance function.

It is also possible to use (9) to directly estimate the sub-vector distance function by taking the logs of both sides and suppressing subscripts.

$$\ln(D_{SI}(Y, X; E)) = -\ln(E) + D_{SI}(Y, X; 1) \quad (12).$$

Rearranging both sides yields and assuming D_{SI} can be estimated by a log linear (Cobb Douglas) model⁵ with a standard statistical error term, u , then we have,

$$\ln(E) = \left[\sum_{i=1}^m \beta_i^y \ln(Y_i) + \sum_{i=1}^{n-1} \beta_i^x \ln(X_i) + u \right] - \ln(D_{SI}(Y, X; E)) \quad (13).$$

In this case we interpret the efficiency error term, v , as a percentage difference, i.e. $\ln(D_{SI}(Y, X; E)) = v$. The major difference in interpretation between (13) and (11) is that the in (11) the subvector efficiency includes the random shocks and in (13) the sub-vector efficiency does not include random shocks. The latter is the more common interpretation in the literature. Jondrow, Materov et al. (1982) (JMLS) were the first to provide an estimate of the v_i conditional on the observed residual $u_i - v_i$ when the efficiency distribution is half-normal. Green (1993) summarizes this approach for the truncated normal, exponential, and gamma distributions. For our applications Census will not clear plant level estimates of the JMLS conditional estimator for efficiency, so we focus on the structural and distributional parameter estimates using the factor requirements interpretation.

Data Sources

This analysis uses confidential plant-level data from two sources: the Longitudinal Research Database (LRD) maintained by the Center for Economic Studies (CES), U.S. Bureau of the Census (Census), and Galitsky et al (2003). The LRD includes the non-public, plant-level data that is the basis of government-published statistics on manufacturing. CES has constructed a panel of plant-level data from the Annual Survey of Manufacturers (ASM) and the Census of Manufacturers (CM). The LRD includes economic activity — for example, labor, energy, plant and equipment, materials costs,

⁵ The same approach can be used if D_{SI} is approximated by a Translog.

and total shipment value of output — for a sample of plants during the survey years and complete coverage during the years of the economic census.

Under Title 13 of the U.S. Code, this data is confidential; however, CES allows academic and government researchers with Special Sworn Status to access these confidential micro-data, under its research associate program. The confidentiality restrictions prevent the disclosure of any information that would allow for the identification of a specific plant's or firm's activities. Aggregate figures or statistical coefficients that do not reveal the identity of individual establishments or firms can be released publicly. The ability to use plant-level data, rather than aggregate data, significantly enhances the information that can be obtained about economic performance, particularly when examining the issue of energy efficiency.

Industry Specific Results

This section gives background on the wet corn milling process, the industry specific estimates for the energy requirements function $f(*)$ and the distribution estimates for the error and efficiency terms. The model for corn refining uses a linear equation and a truncated normal distribution for efficiency. We take the approach here that the empirical energy factor requirements function can be estimated as a linear function. The model specification pays particular attention to the role of product mix. Capacity and/or capacity utilization also play important empirical roles in process industries.

Wet corn milling (SIC 2046 or NAICS 311221), which is also referred to as corn refining, is a relatively sophisticated process producing a variety of products for the paper, food, beverage and other industries. Wet corn milling plants require a large capital investment and are bound by large economies of scale. Typical plants in the US process at least 100,000 bushels per day (bu/day, or 2,500 tonne/day) and operate continuously for nearly 365 days per year.

The most important outputs of wet corn milling are starch, corn sweeteners and ethanol. Both corn sweeteners and ethanol are made from the starch in the corn. Sweeteners fall into three major categories: corn syrup, dextrose and fructose, often called glucose syrup. Ethanol is an increasingly important component of the U.S. fuel supply. About 60% of the ethanol produced in the U.S. currently comes from wet corn milling⁶, generally produced in the refining factories along with starches and syrups.⁷ Corn starch is another important corn refining product, with both food and industrial applications, such as the paper and corrugating industries. Corn oil, produced from the germ component, is the other main high value product. Corn refining also produces many byproducts that are used in animal feed. Table 1 gives an overview of the output from wet corn milling industries on a physical output basis and value basis for the last year information is available. The production process is illustrated below.

⁶ This percentage is based on value of output. The remaining amount is made mostly through 'dry corn milling', a similar process, which produces ethanol and animal feed byproducts, but not the other high-value products that wet corn milling produces.

⁷ The production of ethanol falls under a different industrial classification within the Chemicals industry. Wet corn milling falls into SIC 2046 and NAICS 311221. Ethanol production in the SIC system fall into the broad category 2869 Industry Organic Chemicals, Not Elsewhere Classified, but is separately classified in NAICS 325193, as ethyl alcohol.

Table 1: Wet Corn Milling Product Output

	Million tons, 2001 ^a	\$Billion, 1997 ^b
Corn Sweeteners	16.4	\$3.1
Starch Products	2.9	\$1.5
Corn Oil	0.6	\$1.0
Byproducts	7.2	\$1.6
Ethanol	N/A	\$1.4

^a These values are from the Corn Refining Association, reporting on the output from its member companies.

^b These values are from the Census and are reported based on *product* output, not industry output.

Source: as reported in Galitsky, et al (2003)

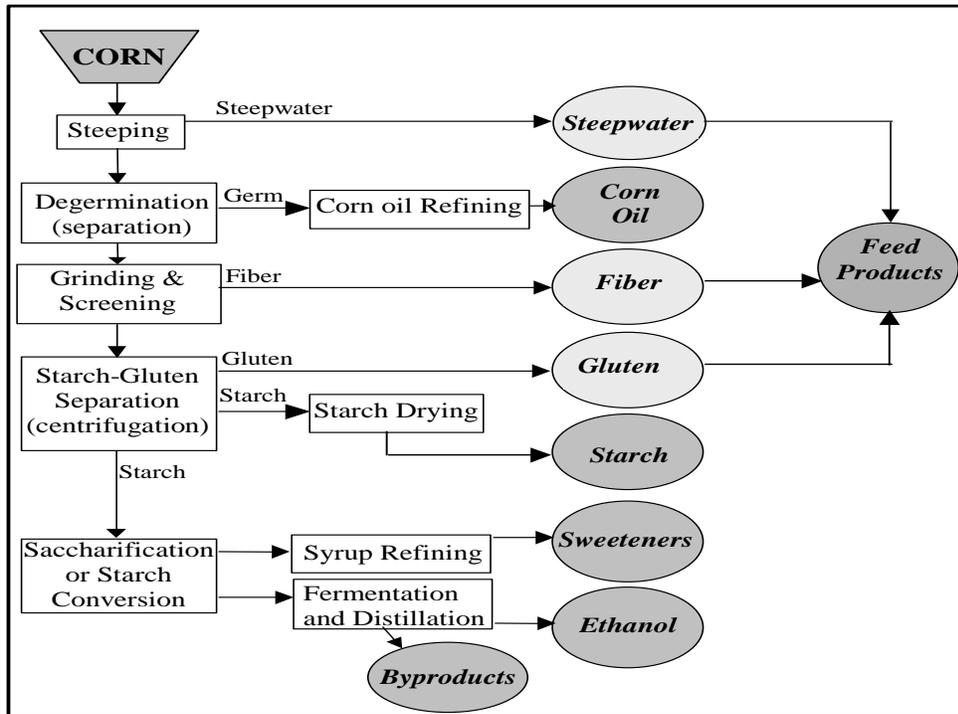


Figure 3 Corn refining process flow diagram

The corn refining model is based on total primary energy, defined as the total Btus of purchased/transferred fuels, steam, and hot water plus the total amount of purchased/transferred electricity converted from kWh to Btu at roughly the average rate of conversion efficiency for the entire U.S. electric grid, 10,236 Btu/kWh. Primary energy is used to more closely align our energy measure with the underlying goals of the EPA Energy Star program, pollution reduction at the source. For this reason a kWh of electricity is treated as the equivalent energy at the production source. The data are taken from the Census of Manufacturing for 1992 and 1997 from NAICS code 311221, this means that plants that only produce ethanol were excluded. Only those plants with capacity estimates either identified by Galitsky et al (2003) or though private

communication, which could be matched to the CM data, were included in the analysis. Plants that produced products in this NAICS category, but did not purchase corn as the primary input were assumed to be germ and corn oil processors and were not included in the analysis. Other plants may have been dropped from the analysis due to irreconcilable discrepancies in the data. The result is that 37 observations (plant years) were included in the analysis. Since there are 29 plants identified in Galitsky et al (2003), one of which is now closed, these number of plants-years in two years of data this sample seems fairly representative of this industry. Product mix is expected to be quite important in this analysis. The detailed product types collected by the Census for the Wet Corn Milling Industry are shown in table 2⁸. Sample means for selected variables are shown in table 3. Confidentiality restrictions prevent disclosure of means for production mix variables, but the total value of shipments is shown.

Table 2: List of Corn Refining Products

GLUCOSE SYRUP (Corn syrup)	
DEXTROSE MONOHYDRATE & ANHYDROUS HIGH FRUCTOSE CORN SYRUP (HFCS)	
CORN STARCH AND DEXTRIN	
	Modified
	Not modified
CORN OILS	
ETHYL ALCOHOL	
GLUTEN FEED & MEAL	

A variety of specifications, including various aggregations and disaggregation of the product categories were initially tried. The initially preferred model was cleared by Census for release to the industry review process. At first the comments focused on product specification. At the request of industry participants we disaggregated the monohydrate and anhydrous dextrose from the other categories of sweeteners. This product is a very high grade sugar that requires substantial processing. We also examined data on the sugar content of high fructose corn syrup (HFCS). Census data has three broad categories of HFCS sugar content. The disaggregation of the monohydrate and anhydrous dextrose made a substantial improvement to the model, while the sugar content of HFCS did not. We suspect that data quality of the HFCS categories to be responsible. We also tested whether location driven differences in climate, in the form of Heating Degree days (HDD) would have an impact on energy use. It did not.

In subsequent comments the issue of gluten feed was raised. One product produced by corn refining is gluten feed. This by-product of the initial milling and separation is sold at animal feed. The industry typically dries the gluten feed for shipment, but plants that are close to agricultural markets may ship some of the product in “wet” form. Companies voluntarily provided data on the moisture content of their shipments of gluten feed. Specifically they provided data on the dry solids content. By

⁸ This represents the entire product category list included by the U.S. Census Bureau on the Economic Census forms that companies are required by law to prepare every five years. Other data (not shown) from the Census is used to develop the EPI. Most plants will only produce a sub-set of these products. The product types shown here would be the basis for the production mix adjustments to the EPI.

taking the ratio of dry solids to the “commercial basis” as reported to Census we constructed a moisture variable. This variable was significant and including it improved the standard errors of the other variables in the model.

Table 3: Sample Means for Selected Variables

Variable name	Sample Mean
X	2.644
E	7.506
E (from fuels)	5.327
E (from electricity)	2.014
CU	0.757
Total Value of Shipments (million dollars)	289.9

For simplicity, we assume that the function form is linear in the parameters, but allow for non-linear transformations of the variables. A log transformation was not feasible since all plants do not produce every product. We found that non-linear (quadratic) terms for capacity utilization was appropriate. Several alternatives for the distribution of the inefficiency term, u , were tried. The distribution that performed best was half normal. The issue of panel data in the estimation was felt to be less important since we only have (at most) two observations per plant which are five years apart. The estimated coefficient of f^* are given in table 4. The last two variables in table 4 are the standard deviation of random error term, v , and the inefficiency term, u , which is assumed to be normally distributed and truncated at zero.

$$E_i = \beta_0 + \beta_1 \text{Corn} + \beta_2 Y_{\text{Modified Starch}} + \beta_3 Y_{\text{M A Glucose}} + \beta_4 Y_{\text{Glucose}} + \beta_5 Y_{\text{Alcohol}} + \beta_6 \frac{\text{Corn}}{\text{Capacity}} + \beta_7 \left(\frac{\text{Corn}}{\text{Capacity}} \right)^2 + \beta_8 (\text{Moisture}) + u_i - v_i \quad (6)$$

The variables included in the final model are:

Corn	Total amount of corn processed in a year, 10^9 lbs.
Y_{Modified}	Total amount of modified starch produced in a year, 10^9 lbs.
$Y_{\text{MAGlucose}}$	Total amount of monohydrate and anhydrous dextrose, 10^9 lbs.
Y_{Glucose}	Total amount of Glucose syrup sweeteners and solids 10^9 lbs.
Y_{Alcohol}	Total amount of alcohol, 10^9 gal,
Moisture	1 - %Moisture content of Gluten feed
CU	Plant capacity utilization (total corn processed / annual capacity)
E	Total Primary Energy (defined above), in trillion Btus.

The major effects in the corn refining model are total corn processed, the mix of products, and capacity utilization. Non-modified starch and high fructose corn syrup are not included in the model, since there is a mass balance relationship between the amount of corn process and the final products we can interpret all of the production variables

relative to a plant producing either of these two products⁹. For example, modified starch, alcohol, and anhydrous dextrose all have higher energy requirements, for a given level of corn processed into either starch or high fructose corn syrup, while Glucose sweeteners would have relatively lower energy requirements. Gluten products were excluded from the model since all plants produce them in quantities that are nearly proportional to the amount of corn processed, i.e. they are direct by products.

Table 4: Parameter Estimates - Total Primary Energy Frontier in Corn Refining

	Coefficient	Standard Error	z
Constant	2.78	0.0008	3275.1
Corn	2.90	0.0001	24568.1
Modified starch	2.12	0.0003	6551.4
Monohydrate and anhydrous	4.48	0.0006	7943.8
Glucose	-0.51	0.0001	-4622.9
Alcohol	16.12	0.0005	32559.1
Utilization	-11.50	0.0022	-5203.5
Utilization ²	5.37	0.0014	3899.0
Moisture content (%)	0.03	0.0005	66.0
σ_v	-30.77	178.67	-0.17
σ_u (Truncated Normal)	1.50	0.23	6.44

To illustrate how the product mix affects the frontier and a plants relative efficiency a hypothetical plant is constructed. The product mix shown in table 5 is not a specific plant, nor based on the sample means, but is consistent on a mass balance basis with the amount of corn processed. In this comparison case the non-modified starch production is shifted to modified starch production. This shifts the frontier. A plant with the same level of energy use would have lower levels of inefficiency and a higher percentile ranking based on the estimated variance of the truncated normal efficiency term.

⁹ Including either one of these two variables did not improve the model. The coefficients were both statistically insignificant from zero and each other.

Table 5: Example product mix and comparison case inputs

	Baseline	Comparison Case 1
Total Grind (10 ⁶ lbs)	2,644.50	2,644.50
Average Grind Rate (Bushels/Day):	131,176.00	131,176.00
Maximum Grind Rate (Bushels/Day):	131,176.00	131,176.00
Capacity Utilization:	100%	100%
HFCS Sweeteners (10 ⁶ lbs)	300	300
Crystalline & anhydrous glucose (10 ⁶ lbs)	0	0
Other non-HFCS Sweeteners (10 ⁶ lbs)	500	500
Modified Starch (10 ⁶ lbs)	743.8	943.8
Non-Modified Starch (10 ⁶ lbs)	200	0
Total Alcohol (10 ⁶ gals)	0	0

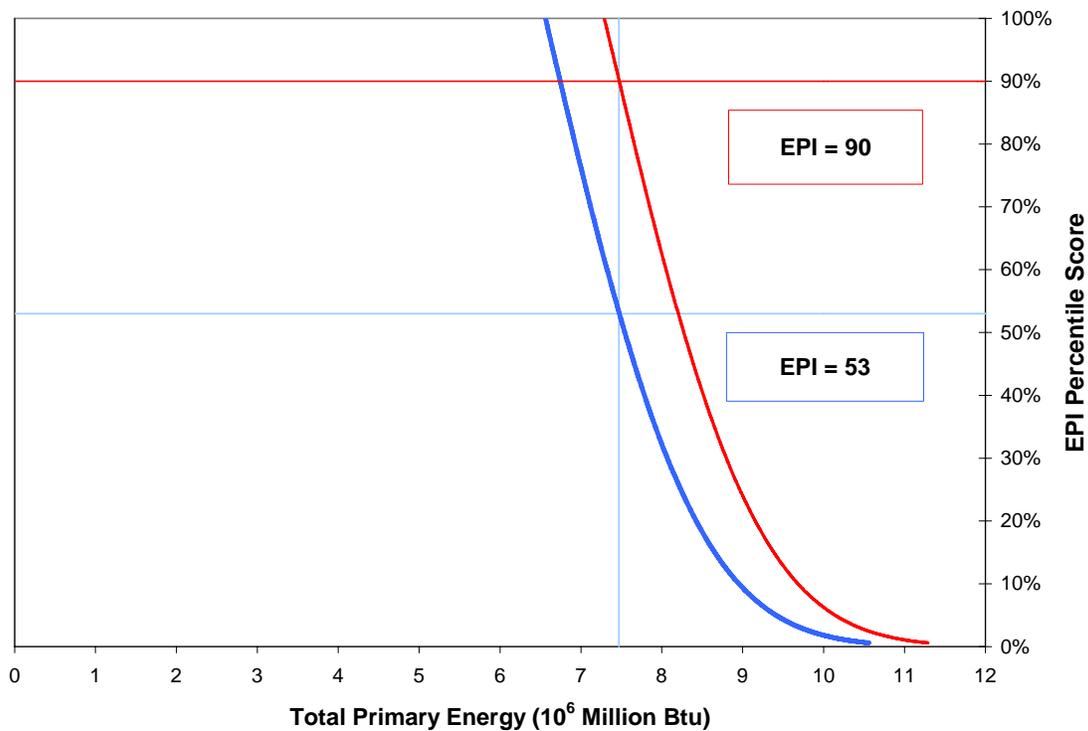


Figure 1: Comparison of Hypothetical Corn Refineries with the Same Energy Consumption.

Summary and Directions for Future Research

This paper presents the motivation for the application of stochastic frontier regression analysis to measure the distribution of efficiency in manufacturing sectors. The estimated models are intended for use by plant and corporate energy managers to provide information on their energy performance relative to their peers. In addition, the EPA ENERGY STAR program provides public recognition for companies with plants that perform in the 75th percentile.

The model that is developed can be viewed as a heuristic application or as a sub-vector distance function. The paper takes the first approach to the corn refining sector and the second approach to the cement sector. A number of natural extensions of this research present themselves, particularly regarding the latter interpretation of the analysis as a distance function. The first extension is to ask whether the distribution of energy (in)efficiency is similar to the distribution for total production efficiency. By estimating an input distance this comparison could be made to address the question “Are plants that are relatively more inefficient in the total factor sense more (or less) inefficient in the single factor of energy?” Boyd and Pang (2000) use a DEA application for the glass industry and found that plants with low total factor efficiency have relatively lower energy efficiency. Understanding the relationship between single and multi-factor efficiency in a stochastic frontier context would be important.

This paper also treats energy as an aggregate of electricity and fuels. The sub-vector interpretation would allow us to treat these energy forms without aggregation and could shed light on the possible differences between efficiency in the use of these two types of energy. The question of aggregation also comes up when the model is used to assess efficiency of multi-plant firms. The model can provide plant specific efficiencies, but does not address how these may be aggregated to a firm level efficiency measure.

Treatment of the panel data issues is another important extension, particularly if the time series component of the panel were expanded. Data from the 2002 Census of Manufacturing has just become available within the RDC network and extending the time series back to the eighties is feasible¹⁰. Battese and Coelli (1995) offer one approach to the panel data by treating the truncation of the normal inefficiency term as a function of firm level variables. The issue here is how to specify those effects. Greene (2002) considers fixed and random effects in the frontier estimation, but recognizes the difficulty in interpreting the results.

Extending the functional form beyond the simple linear to include flexible forms is desirable. This is not feasible for corn milling plants due to the small sample and wide range of product mix. Dashti (2003) considers a wide range of flexible forms using the Bayesian approach to the frontier estimation, although he focuses on comparing firm efficiency measures across the various functional forms. Census confidentiality would not permit the disclosure of those results, but the approach could be used internally at Census to evaluate and choose a functional form.

¹⁰ Major changes in energy prices occurred in the early eighties, so applying a production approach might not be as appropriate to a period with dramatically different energy prices.

References

- Aigner, D., C. A. K. Lovell, et al. (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models." Journal of Econometrics **6**(1): 53-66.
- Ang, B. and F. Liu (2001). "A new energy decomposition method: perfect in decomposition and consistent in aggregation." Energy **26**(6): 537-548.
- Battese, G. E. and T. J. Coelli (1995). "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data." Empirical Econometrics **20**: 325-332.
- Boyd, G. (1997). The Interrelationship between Environmental Goals, Productivity Improvement, and Increased Energy Efficiency in Integrated Paper and Steel Plants. Washington DC, U.S. Department of Energy.
- Boyd, G. and J. Pang (2000). "Estimating the Linkage between Energy Efficiency and Productivity." Energy Policy(28): 289-96.
- Boyd, G., G. Tolley, et al. (2002). "Productivity and Efficiency of Glass Industry: Analysis of Plant Level Environmental Performance." Environmental And Resource Economics **23**(1): 29-43.
- Boyd, G. A., D. A. Hanson, et al. (1988). "Decomposition of Changes in Energy Intensity: A Comparison of Divisia Index and Other Methods." Energy: 309-12.
- Boyd, G. A., J. F. McDonald, et al. (1987). "Separating the Changing Composition of U.S. Manufacturing Production from Energy Efficiency Improvements: A Divisia Index Approach." Energy Journal **8**(2): 77-96.
- Boyd, G. A. and J. M. Roop (2004). "A Note on the Fisher Ideal Index Decomposition for Structural Change in Energy Intensity." The Energy Journal **25**(1): 87-101.
- Byrnes, P., R. Fare, et al. (1987). "The Effect of Unions on Productivity: U.S. Surface Mining of Coal." Management Science **34**(9): 1037-1053.
- Coggins, J. and J. Swinton (1996). "The Price of Pollution: A Dual Approach to Valuing SO₂ Allowances." Journal of Environmental Economics and Management **30**: 58-72.
- Dashti, I. (2003). "Inference from concave stochastic frontiers and the covariance of firm efficiency measures across firms." Energy Economics **25**: 585-601.
- Farrell, M. J. (1957). "The Measurement of Productive Efficiency." Journal of the Royal Statistical Society Series A, General(120, part 3): 253-281.
- Ferrier, G. D. and J. G. Hirschberg (1992). "Climate Control Efficiency." Energy Journal **13**(1): 37-54.
- Freeman, S. L., M. J. Niefer, et al. (1997). "Measuring industrial energy intensity: practical issues and problems." Energy Policy **25**(7-9): 703-714.
- Green, W. (1993). The Econometric Approach to Efficiency Analysis. The Measurement of Productive Efficiency: Techniques and Applications. H. Fried, C. A. K. Lovell and S. Schmidt. New York, Oxford University Press: 68-119.
- Greene, W. H. (2002). Fixed and Random Effects in Stochastic Frontier Models. New York University Economics department working paper.
- Greening, L., W. Davis, et al. (1997). "Comparison of six decomposition methods: Application to aggregate energy intensity for manufacturing in 10 OECD countries." Energy Economics **19**(3): 375-390.

- Hattori, T., T. Jamasb, et al. (2005). "Electricity distribution in the UK and Japan: A comparative efficiency analysis 1985-1998." The Energy Journal **26**(2): 23-47.
- Huntington, H. (1995). "Been Top Down So Long It Looks Like Bottom Up." Energy.
- Jondrow, J., I. Materov, et al. (1982). "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." Journal of Econometrics **19**(2/3): 233-238.
- Kashani, H. A. (2005). "Regulation and efficiency: an empirical analysis of the United Kingdom continental shelf petroleum industry." Energy Policy **33**(7): 829-963.
- Lowrey, M., L. Getachew, et al. (2005). "Econometric Benchmarking of Cost Performance: The Case of US Power Distributors." The Energy Journal **26**(3): 75-92.
- Lui, X. Q., B. W. Ang, et al. (1992). "The application of the Divisia index to the decomposition of changes in industrial energy consumption." The Energy Journal **13**(4): 161-177.
- Morrison-Paul, C., W. Johnston, et al. (2000). "Efficiency in New Zealand Sheep and Beef Farming: The Impacts of Regulatory Reform." Review of Economics and Statistics **82**(2): 325-37.
- Murillo-Zamorano, L. (2004). "Economic efficiency and frontier techniques." Journal of Economic Surveys **18**(1): 33-77.
- Murillo-Zamorano, L. (2005). "The role of energy in productivity growth: A controversial issue?" The Energy Journal **26**(2): 69-88.
- Pacudan, R. and E. d. Guzmanb (2002). "Impact of energy efficiency policy to productive efficiency of electricity distribution industry in the Philippines." Energy Economics **24**(1): 41-54.
- Phylipsen, G. J. M., K. BloK, et al. (1998). Handbook on International Comparisons of Energy Efficiency in the Manufacturing Industry. Ministry of Housing, Physical Planning and the Environment. Utrecht, Dept. of Science, Technology and Society, Utrecht University.
- Worrell, E., L. Price, et al. (1997). "Energy Intensity in the Iron and Steel Industry: A Comparison of Physical and Economic Indicators." Energy Policy **25**(June/July): 7-9.
- Yaisawarng, S. and J. D. Klein (1994). "The Effects of Sulfur Dioxide Controls on Productivity Change in the U.S. Electric Power Industry." The Review of Economics and Statistics: 447-60.