Measuring Plant Level Energy Efficiency When Production Activities Are Not Homogeneous: The ENERGY STAR Energy Performance Indicator

Gale A. Boyd,[1] *Argonne National Laboratory*

ABSTRACT

Energy efficiency is the comparison of actual energy used to the lowest amount of energy required to perform some level of service or produce some amount of product. Energy intensity, also called specific energy consumption (SEC), is the simple normalization of energy use to a measure of service or production. This energy intensity can then be compared to the "best practice." This simple description belies the complexities that lie in industrial activity. When there is no single measure of activity for the denominator, the energy intensity may not be well defined. There is a substantial literature that examines what the "best measure" of production to use for various sectors; physical volumes, market value, value-added, etc. Creating aggregate measures of activity, e.g. total dollars, is often an inadequate approach. When sectors are aggregated, the changing mix of production in various sectors influences the aggregate energy intensity index. The difference between the observed and "best practice" is the level of (in)efficiency. The estimate of "best practice" can be an optimum configuration derived from engineering, possibly a notion of the theoretical minimum, or comparison to a real world application with the lowest observed energy intensity, i.e. "best *observed* practice." Since production/activity mix can be quite varied, it may be difficult to find a real world application that is sufficiently similar to the observed plant to conduct the needed comparison. This reflects an oft-expressed view that "every plant is unique."

A statistical model which provides a functional relationship between the level of energy use and the level and type of various production activities, material inputs quality, and external factors, e.g. climate, is one solution to these problems. The stochastic frontier regression estimates the lowest observed plant energy use, given these factors. This statistical model also provides a distribution of (in)efficiency across the industry, which allows the user to answer the hypothetical but very practical question, "How would my plant compare to everyone else in my industry, *if all other plants* were similar to mine?" The result is a tool that can be used by corporate and plant energy managers to estimate the energy efficiency of their portfolio of plants.

---

[1] Work sponsored by the U.S. 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 the *Energy Star® Corn Refiners Industry Focus Meeting*, Washington, DC (December 11, 2003), the *2nd Annual ENERGY STAR corn refiners Focus Workshop*, Indianapolis, IN (September 9th, 2004), and an anonymous referee. The research in this paper was conducted while the author was a Census Bureau research associate at the Chicago Census Research Data Center. Research results and conclusions expressed are those of the author and do not necessarily indicate concurrence by the Bureau of the Census, Argonne National Laboratory, 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.

This paper presents recent empirical results prepared for ENERGY STAR for corn refining. This sector typifies the issue of product mix impact on plant energy use. Accounting for these factors allows construction of a normalized distribution of energy efficiency for the entire sector and a percentile performance score, for any individual plant. The results of the analysis can be used by energy managers in these various sectors to assess plant energy performance, set targets, and track progress.

Introduction

Energy efficiency is the comparison of actual energy used to the lowest amount of energy required to perform some level of service or produce some amount of product. Energy intensity, also called specific energy consumption (SEC), is the simple normalization of energy use to a measure of service or production in the form of a ratio of energy use to a measurable service or product output. This energy/output measure of energy intensity can then be compared to the "best practice" measure of energy intensity. The estimate of "best practice" can be an optimum configuration derived from engineering, possibly a notion of the theoretical minimum, or comparison to a real world application with the lowest observed energy intensity, i.e. "best *observed* practice." The difference between the observed and "best practice" is the level of (in)efficiency.

This simple description belies the complexities that lie in industrial activity. While at the process level it may be feasible to define the specific energy service or production activity, modern manufacturing plants encompass many type of activities and possibly different products. If there is no single measure of activity for the denominator, the energy intensity may not be well defined. In addition production/activity mix can be quite varied, so it may be difficult to find a real world application that is sufficiently similar to the observed plant to conduct the needed comparison. This reflects an oft-expressed view that "every plant is unique."

Battles (1996) provides an overview of the literature that examines what the "best measure" of production to use for various sectors; physical volumes, market value, value-added, etc. Freeman, Niefer et al. (1997) argue that creating aggregate measures of activity, e.g. total dollars, is often an inadequate approach and that physical units are often preferred. Differences between plants that produce a variety of products make the goal of a single meaningful measure of physical production difficult to achieve. A statistical model which provides a functional relationship between the level of energy use and the level and type of various production activities is one solution to these problems. The statistical model can be conceptualized as a method of estimating weights that are appropriate for aggregating various production activities into an overall production index appropriate for energy analysis. While most statistical model are based on estimating averages, stochastic frontier regression is statistical tool that estimates the lowest observed plant energy use, given these factors.

This paper explains this proposed statistical approach and presents recent empirical results, prepared for ENERGY STAR, using stochastic frontier analysis on plant level data for corn refining. This sector typifies the issue of product mix impact on plant energy use. Accounting for these factors allows construction of a normalized distribution of energy efficiency for the entire sector and a percentile performance score, for any individual plant. The results of the analysis can be used by energy managers in these various sectors to assess plant energy performance, set targets, and track progress.

Stochastic Frontier Approach

The concept of the stochastic frontier analysis can be easily described in terms of the standard linear regression model. A more detailed and complete discussion of stochastic frontier analysis may 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. A simple linear equation for this can be written as,

$$E_{t,i} = \alpha + \beta_y y_{t,i} \qquad (1)$$

where

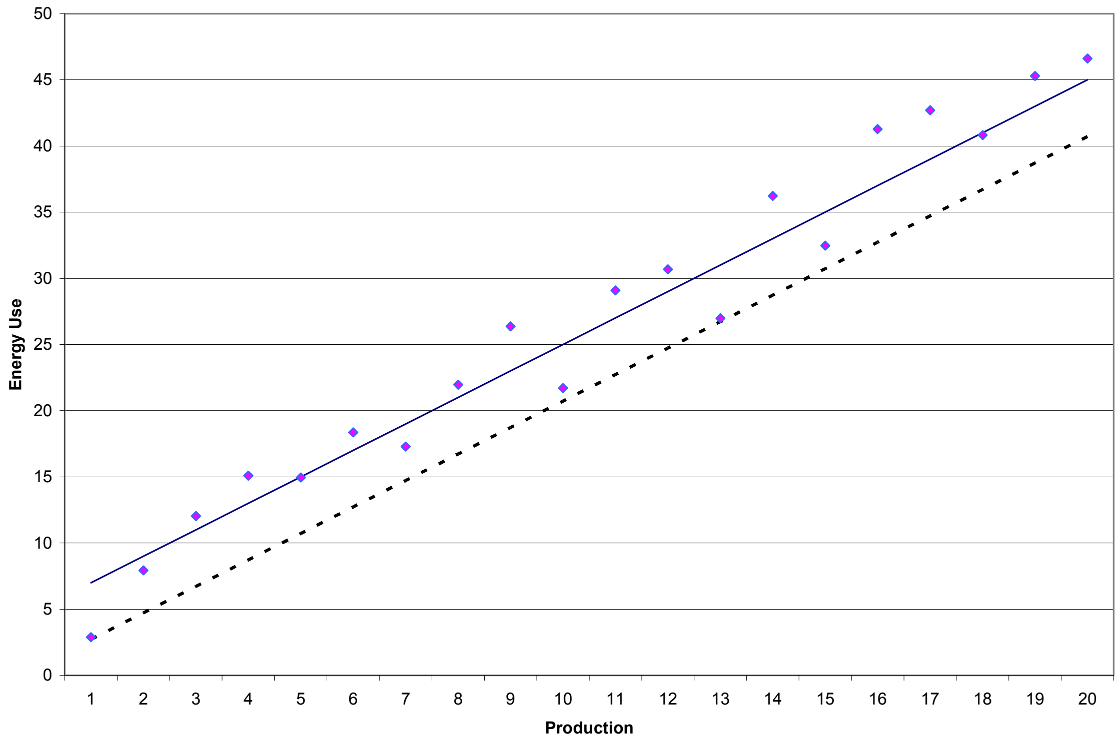
t is the smallest time period of interest,
i is the i^{th} plant, and
y is production.

Given data on energy use and production the parameters α and β_y can be fit via a linear regression model. Since the actual data may not be perfectly measured and this simple relationship between energy and production may only be an approximation of the "true" relationship, linear regression estimates of the parameters relies on the proposition that any departures in the plant data from equation (1) are "random." This implies that the actual relationship includes a random error term that is follows a normal (bell shaped) distribution with a mean of zero and variance of σ^2, i.e. that about half of the actual values of energy use are less than what (1) would predict and half are greater.

$$E_{t,i} = \alpha + \beta_y y_{t,i} + \varepsilon_{t,i},\ \varepsilon \sim N(0,\sigma^2) \qquad (2)$$

The linear regression gives the average relationship between production and energy use. If the departures from the average, particularly those that are above 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 (see Figure 1). This "corrected" regression is one way to represent the frontier.

Figure 1. Average and Corrected Linear Regression of Production and Energy



While the corrected regression 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 errors in data collection/reporting, effects that are unaccounted for in the analysis, and that a linear equation is an approximation of the complex factors that determine manufacturing energy use we still wish to include the statistical noise. or "random error." term in the analysis, $v_{t,i}$, but also add an second random component, $u_{t,i}$, 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

$$E_i = f(Y_i, X_i, Z_i; \beta) + \varepsilon_i \qquad (3)$$
$$\varepsilon_i = u_i - v_i \qquad v \sim N\left[0, \sigma_v^2\right],$$

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,
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 distribution[2], for example gamma, exponential, truncated normal, etc. It is then possible to estimate the parameters of equation (3), along with the distribution 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 techniques available in most modern statistical packages.

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(*),* 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 the *Y*, *X*, and *Z*. This means that *f(*)* may be log linear or include second order (quadratic) terms. There is no guidance on what mathematical form that *f(*)* should take so an substantial amount of judgment, as well as trial and error is exercised.

What variables to include (or exclude) for a given industry is driven by some prior 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 different products. The flexibility to choose the latter is important for many applications, in particular when there are a wide range of products produced at any given plant. In the case of corn refining there are a range of products that may be produced, some of which are byproducts that are always produced. Some products may or may not be produced in a given plant, resulting in a very non-homogeneous mix of plants to compare.

X may include quantity and types of materials purchased, labor, or plant capacity. In the case of corn refining, total corn processed and total plant capacity are obvious choices. *Z* may include a variety of external factors like energy prices, weather variables, e.g. heating degree days (HDD) and cooling degree days (CDD), capacity utilization, regulatory factors, etc. Since the statistical formulation allows us to estimate the standard error of the estimated parameters, the decision to include any of these variables can be driven by the data and model estimates. For example, one can test whether weather variables have a statistically significant impact on energy use. This is likely to be the case for fabrication and assembly industries, like automobile manufacturing where the building HVAC is a large percentage of the total energy consumed, but is less likely to be the case for process industries, like steel or cement. Conventional statistical tests can determine which factors to include the model, so if HDD and CDD do not have a measurable influence in corn refining then these variables need not be included in the final version. Conventional statistical tests provide confidence intervals for any effect in the model. The next section presents an application of this approach to corn refining and discuses the choice of variables in more detail.

---

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

Corn Refining[3]

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 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 milling[4], generally produced in the refining factories along with starches and syrups.[5] 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.

Table 1. Wet Corn Milling Product Output

	Million tons, 2001[6]	$Billion, 1997 [7]
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

Source: Galitsky 2003

Version 09.09.04 of the corn refining EPI 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.

The data are taken from the Census of Manufacturing for 1992 and 1997 from NAICS code 311221, this means that plant that only produce ethanol were excluded. Only those plants

[3] This background discussion of the corn refining industry is taken from Galitsky (2003).
[4] 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.
[5] 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.
[6] These values are from the Corn Refining Association, reporting on the output from its member companies.
[7] These values are from the Census and are reported based on *product* output, not industry output.

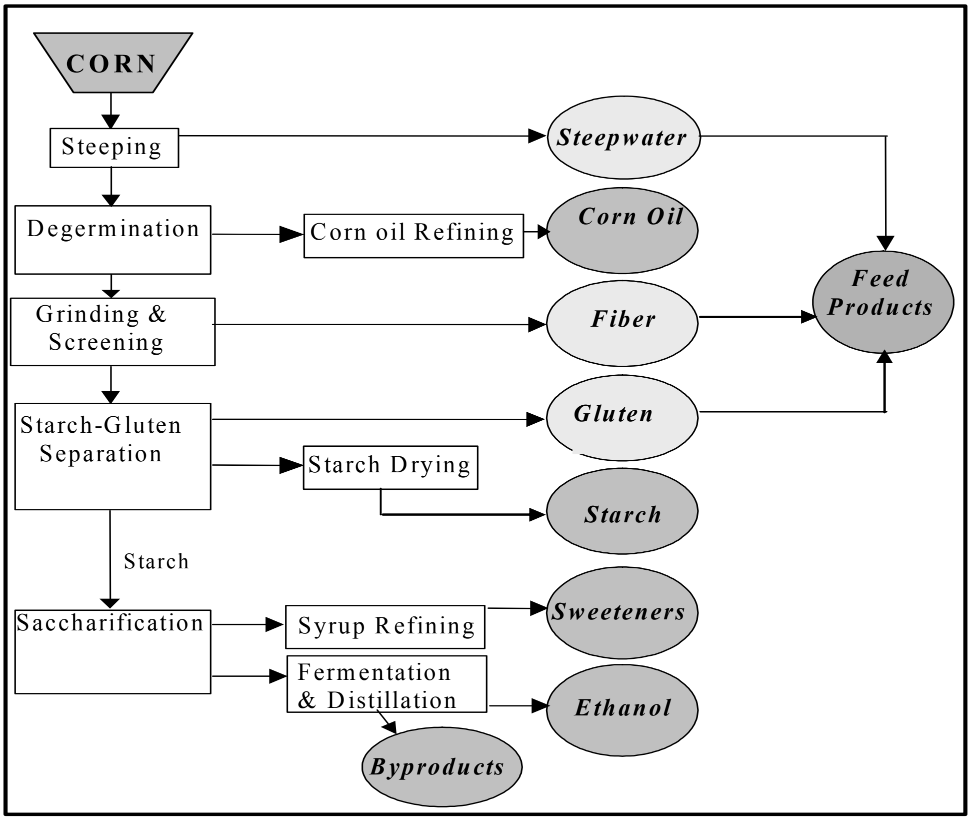
with capacity estimates either identified by Galitsky et al (2003) or though private communication, which could be matched to the CM data, were included in the analysis. Plants that produced products in this NAICS category, but did not purchase corn as the primary input were assumed to be germ and corn 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.

The variables included in *f(*)* are:

Corn	Total corn processed in a year, in billions of pounds,
$Y_{Mod\ Starch}$	Total modified starch produced in a year, in billions of pounds,
$Y_{MADextros}$	Total monohydrate and anhydrous dextrose, in billions of pounds,
$Y_{Glucose}$	Total Glucose syrup sweeteners and solids, in billions of pounds,
$Y_{Alcohol}$	Total alcohol, in billions of gallons,
CU	Capacity utilization (total corn processed divided by annual capacity), and
E	Total Primary Energy (defined above), in trillion Btus.

The choice of variables to include in the model are based on a combination of understanding of the production process (see Figure 2) and statistical test conducted at earlier stages of the model development. All plants produce feed products as a result of the basic separation stage with little variation in yields. This means that these products need not be include in the output mix variable, so long as the amount of corn processed is included in the model. Starch is produced either as a final product or as a feedstock to the saccharification. Starch may be unmodified or have additional processing for specialty products. Our model implicitly assumes that some combination of unmodified starch or HFCS is produced at the plant. The impact on energy use resulting from a departure from this “default” product mix is represented by the other four product variables.

Figure 2. Corn Refining Process Flow Diagram



Source: Galitsky 2003

Sample means for selected variables are shown in Table 2. Confidentiality restrictions prevent disclosure of means for production mix variables, but the total value of shipments is shown.

Equation (4) gives the linear equation model for corn refining. A truncated normal distribution is used to represent inefficiency.

$$E_i = \beta_0 + \beta_1 Corn + \beta_2 Y_{Modified\,Starch} + \beta_3 Y_{MA\,Glucose} + \beta_4 Y_{Glucose} + \beta_5 Y_{Alcohol} + \beta_6 \frac{Corn}{Capacity} + \beta_7 \left(\frac{Corn}{Capacity}\right)^2 + u_i - v_i \quad (4)$$

The estimated coefficients are given in Table 3.

Table 2. Sample Means for Selected Variables

Variable name	Sample Mean
Corn	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

Table 3. Parameter Estimates for the Total Primary Energy Frontier in the Wet Corn Refining Industry

	Coefficient	Standard Error	z-test
Constant	2.70	0.667	4.1
Corn	2.89	0.230	12.5
$Y_{\text{Modified starch}}$	3.31	0.935	3.5
$Y_{\text{Monohydrate and anhydrous}}$	3.64	0.521	7.0
Y_{Glucose}	-0.66	0.019	-34.1
Y_{Alcohol}	14.49	0.357	40.6
Capacity Utilization	-11.15	3.389	-3.3
Capacity Utilization ^{2}	5.13	2.006	2.6
σ_v	7.51E-07	6.74E-05	0.01
σ_u (truncated normal)	2.17	0.252	8.6

The last two variables in Table 1 are the standard deviation of random error term, *v*, and the inefficiency term, *u*, the latter of which is assumed to be normally distributed and truncated at zero. The z test indicates the confidence with which the variable is estimated. A z test of greater than 2.0 indicates confidence of 99% or higher. A value of z less than 1.2 indicate confidence of 90% or lower.[8] Our estimates have extremely low error variance (0.000000075), essentially attributing all of the variation from the fitted regression model to differences in efficiency.

The major effects in the corn refining model are total corn processed, the mix of products, and capacity utilization. Animal feed products, e.g. gluten, fiber, etc., are byproducts of the initial stage of processing and do not vary significantly across plants, so these outputs can be omitted from the analysis. Non-modified starch and high fructose corn syrup are also not included in the equation so we interpret all of the other product variables relative to a plant producing either of these two products. For example, modified starch, alcohol, and anhydrous dextrose all have higher energy requirements relative to a given level of corn processed into either starch or high fructose corn syrup. Glucose sweeteners would have relatively lower energy requirements.

Translating Energy Efficiency into a Percentile Score

In addition to accommodating a range of products, the frontier approach allows us to translate the absolute energy efficiency, i.e. the difference between actual observed energy use and the predicted best practice energy use, into a percentile score. This is possible since we estimate the underlying statistical distribution of energy efficiency. Absolute inefficiency of energy use is estimated to follow a truncated normal distribution, so we can also translate the absolute level of inefficiency into a percentile score, or ranking, by using equation (5).

$$EPI = \Pr(inefficiency \le E_i - f(Y_i, X_i, Z_i; \beta)) \qquad (5)$$

This probability can be computed directly by equation (6), where $\boldsymbol{F}$ is the cumulative distribution function for a normal distribution with mean zero and variance σ_u^2. This measure is a percentile ranking of the energy efficiency of the plant.

$$EPI = 1 - 2 * \left[F(E_i - f(Y_i, X_i, Z_i; \beta), 0, \sigma_u^2) - 0.5 \right] \qquad (6)$$

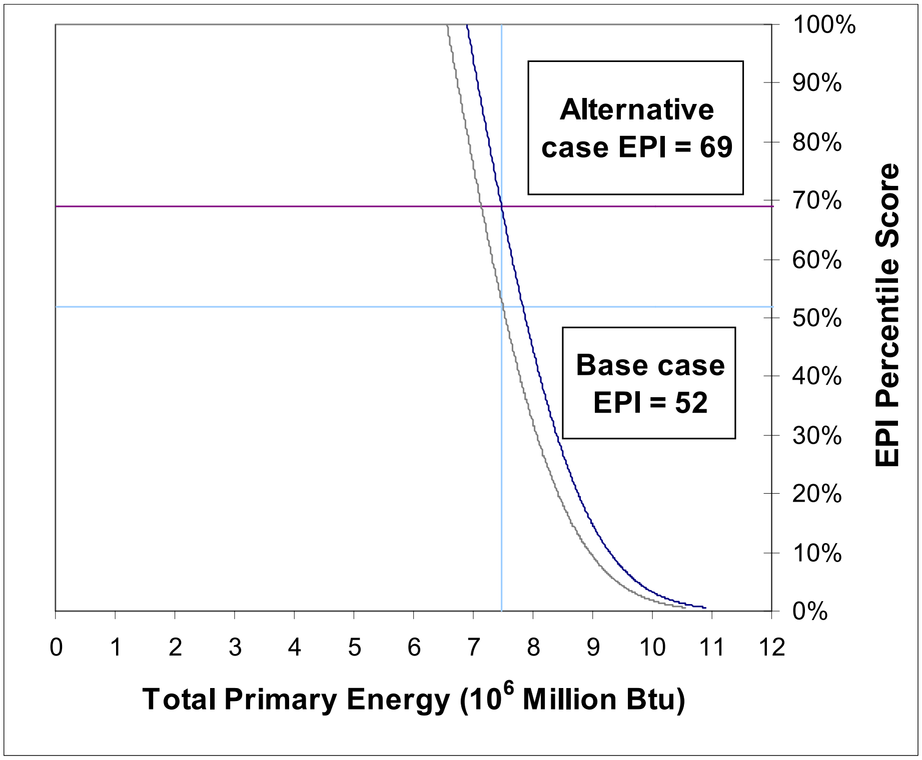
To illustrate how this may be applied we show a hypothetical plant with the following production values.

Total Grind:	2,644.5
Maximum Grind Rate (Bushels/Day):	131,176
Capacity Utilization:	100%
HFCS Sweeteners:	300 million lbs
Crystaline and anhydrous glucose:	none
Other non-HFCS Sweeteners:	500 million lbs
Modified Starch:	743 million lbs
Non-Modified Starch:	200 million lbs
Total Primary Energy	7,469,527 mMBtu

[8] More precisely, the z test provides the probability that the variable is significantly different from zero.

Figure 3 shows that this plant would have a percentile score of 52, which is slightly above the median performance. However, if that plant produced all HFCS and no glucose, i.e. 800 million lbs HFCS, and used the same level of energy its percentile energy efficiency would be 17 points higher, since HFCS has higher energy requirements.

Figure 3. Comparison of Percentile (EPI) for Plants Producing a Mix of Products



Summary

This paper presents an approach to statistically measure plant specific energy efficiency when there are multiple outputs, making the choice of a single normalizing factor to measure energy efficiency difficult. The stochastic frontier predicts the lowest energy use for any plant, given information on the capacity and amount of total corn processed with the mix of final products produced. The difference between the actual energy use and the predicted "best observed practice" is the level on inefficiency.

The model presented extends the notion of energy efficiency, commonly measured by the specific energy consumption or ratio of energy to unit product, to the case where there are multiple products that are produced in a variety of quantities and may have different energy requirements. Economic models may aggregate products based on their market value, but the approach here is more flexible and specific to the question at hand, since the weights are based on best practice energy. Market values are the result of a variety of economic drivers. The approach presented here also translates the estimated inefficiency into a percentile score which can be interpreted as a plant level ranking.

Since this ranking is based on distribution of inefficiency for the entire industry, but normalized to the specific systematic factors of the given plant, this statistical model allows the user to answer the hypothetical but very practical question, "How does my plant compare to everyone else's in my industry, *if all other plants* were similar to mine?" The results of this analysis have been provided by ENERGY STAR to representatives of the corn refining industry in a series of workshops to test and provide comments. The final version of this model can be used by the industry to benchmark industrial energy performance within there industry and to set goals for improvement.

References

Battles, S. (1996). Measuring Energy Efficiency in the United States Industiral Sector. Workshop on Methodologies for International Comparisons of Industrial Energy Efficiency, Vancouver, BC, CIEEDAX/Simon Fraser University.

Freeman, S. L., M. J. Niefer, et al. (1997). "Measuring industrial energy intensity: practical issues and problems." Energy Policy **25**(7-9): 703-714.

Galitsky, C., E. Worrell, et al. (2003). Energy efficiency improvement and cost saving opportunities for the Corn Wet Milling Industry: An ENERGY STAR Guide for Energy and Plant Managers, Lawrence Berkeley National Laboratory

Green, W. (1993). The Econometric Approach to Efficiency Analysis. The Measurement of Productive Efficiency: Techniques and Applications. S. Schmidt. New York, Oxford University Press**:** 68-119.