A Method for Measuring the Efficiency Gap between Average and Best Practice Energy Use

The ENERGY STAR Industrial Energy Performance Indicator

Gale A. Boyd

Address correspondence to:
Dr. Gale A. Boyd
Center for Energy, Environmental, and
  Economic Systems Analysis
9700 South Cass Avenue
Argonne National Laboratory
Argonne, IL 60439 USA
<gboyd@anl.gov>
<www.dis.anl.gov>

Summary

A common feature distinguishing between parametric/statistical models and engineering economics models is that engineering models explicitly represent best practice technologies, whereas parametric/statistical models are typically based on average practice. Measures of energy intensity based on average practice are of little use in corporate management of energy use or for public policy goal setting. In the context of company- or plant-level indicators, it is more useful to have a measure of energy intensity that is capable of indicating where a company or plant lies within a distribution of performance. In other words, is the performance close to (or far from) the industry best practice? This article 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 used by the ENERGY STAR™ industrial energy performance indicator (EPI) to energy intensity. Stochastic frontier regression analysis separates energy intensity into three components: systematic effects, inefficiency, and statistical (random) error. The article outlines the method and gives examples of EPI analysis conducted for two industries, breweries and motor vehicle assembly. In the EPI developed with the stochastic frontier regression for the auto industry, the industry median "efficiency gap" was around 27%.

Introduction

The word TANSTAAFL, an acronym for the expression "There Ain't No Such Thing As A Free Lunch," was coined by the author Robert Heinlein in his book, *The Moon is a Harsh Mistress*. It was the way the inhabitants of the Moon expressed the notion that nothing was free, even the air they breathed. In the end, there is always a cost. In the context of industrial ecology, one could interpret this as the admonition that every action has a reaction, at least in the larger system. In particular, TANSTAAFL implies that waste has a cost and in the long run may not be sustainable. In the context of neoclassical economics, the term is commonly used to reflect the view that you can't reduce resource use or pollution without incurring an economic loss. This might be true if all economic decisions were fully optimal and rational. Panayotou and Zinnes (1994) discuss how economic decision processes can break down to produce decisions that are suboptimal from an industrial ecology perspective.

But operating a business is about discovering profitable opportunities. This discovery process may have its own cost, for example, the search cost. Two important dimensions of this search cost are knowing both whether to look and where. Porter and Linde (1995, 121) provide some perspective on where to look by equating pollution with inefficiency:

> Is it naive to expect that reducing pollution will often enhance competitiveness? We think not, and the reason is that pollution often is a form of economic waste . . . a sign that resources have been used incompletely, inefficiently, or ineffectively. (p. 121)

Such opportunities are described as "win-win" situations because business and social goals are both met. Many public policy efforts, such as ENERGY STAR, seek to identify and eliminate the production inefficiencies and regulatory disincentives that prevent simultaneous improvement in both productivity and environmental quality. Similarly, the private sector has engaged in its own search for win-win solutions. Nelson (1994) describes a classic case study of the success of Dow Chemical's energy conservation program. Whether these types of private-sector or public-policy initiatives are successful depends on the extent to which such inefficiencies are widespread, particularly in the energy/pollution-intensive industries. In other words, both policymakers and business owners would like to know whether to look for, and reasonably expect to find, opportunities for energy reductions. The answer requires information on how widespread the opportunities are in any given sector.

Industrial ecology provides a number of approaches to apply to these questions. Panayotou and Zinnes (1994) categorize these approaches as either descriptive or prescriptive. Descriptive approaches describe qualitative relationships and summarize them quantitatively. Prescriptive approaches seek to optimize a given system. This article focuses on a specific feature distinguishing between a common descriptive approach (parametric/statistical) and a common prescriptive approach (engineering/economic). The distinguishing feature is that engineering/economic models explicitly represent "best practice" technologies, whereas parametric/statistical models are typically based on "average practice." "Best practice" is defined here to refer to the existence and use of multiple technologies having different levels of performance for a specific application or circumstance. When this variety or variation in performance for a specific application or circumstance is not explicit, but represented by a single aggregate, that is called "average practice." Intuitively, the differences between average and best practice might be observed in the performance ranges of specific appliances in existing homes (for example, refrigerators) or plant-level performance differences between various facilities in an industry.

Variations between observed practices can exist for a number of reasons, including those based on economic decisions, for example, energy prices, and those that are structural, for example, different production processes or energy service requirements. Those economic or structural differences are not considered as differences between best and average practice. The differences between best and average practice are defined for a specific application or circumstance, so observable economic and structural differences must be accounted for in measuring differences between best and average practice. Statistical models are well suited to accounting for these

types of measurable differences, but the parametric/statistical approach is commonly based on aggregate data, so cross-sectional industry differences or time-series price differences are included in a model, but there is no explicit treatment of any remaining difference between best and average practice.

This distinction between best and average practice is important because the description of an average industrial plant may have little meaning. The average body temperature of a person with his or her head in the freezer and feet in the oven is not very revealing. Measures of energy intensity based on average practice are of little use in corporate management of energy use or for public policy goal setting. In the context of company- or plant-level indicators, it is more useful to have a measure of energy intensity capable of representing where a company or plant lies within the entire distribution of performance within an industry. In other words, is the performance of a particular company or plant close to (or far from) the industry best practice? The answer to this question tells the company or plant manager whether to look more closely for a specific opportunity to reduce energy use.

This article presents a parametric/statistical approach that can be used to measure best practice and average practice, thereby providing a measure of the difference, or "efficiency gap." This method is designed to measure the magnitude of win-win opportunities where energy can be reduced relative to a set of plants operating under similar conditions. It also provides the distribution of these inefficiencies. It is important to understand that the modeling does not evaluate win-win in the context of retrofit decisions, but identifies feasible performance for comparable activities. This information on feasible levels of performance may then help policymakers and private business owners to set up targets to reduce these inefficiencies. The approach presented here requires plant-level data and applies a method called stochastic frontier regression analysis to describe the energy intensity of a group of plants. This approach has been used to develop the ENERGY STAR industrial energy performance indicator (EPI) for several industries. Stochastic frontier regression analysis separates the energy intensity into three components: systematic effects, inefficiency, and statistical (random) error. The article outlines the approach and the data used and gives example results.

The Frontier Approach

The frontier approach is a descriptive analysis that quantifies the relationship between resource use and production. It is specifically focused on defining and measuring various types of inefficiency. The approach has been generalized to include both the desirable and undesirable outcomes of production, so it has seen growing application to questions of environmental efficiency that concern industrial ecology. Tyteca (1998) applies this approach as a necessary condition for sustainability. Previous empirical research on the linkage between environmental performance and productivity utilizing frontier approaches includes work by Boyd and McClelland (1999), Brannlund and colleagues (1995), and Chung and colleagues (1997) studying the paper industry; work by Yaisawarng and Klein (1994), Boyd and colleagues (1996), and Corio and Boyd (1996) examining the electric power industry; and work by Jeon and Sickles (2001), Zofio and Prieto (2001), and Boyd and colleagues (1998) examining global CO_2 emissions. These studies apply a nonparametric version of the frontier approach. The literature of efficiency measurement using frontier techniques is quite vast. One bibliography (Seiford 1999) lists over 1,500 empirical references, many of which are empirical applications. Although these applications do not all focus on energy or environmental applications, many do.

The concept of statistical frontier analysis that supports the EPI can be motivated easily in terms of the standard linear regression model. This section provides a brief overview of this motivation. A much more detailed discussion of the evolution of the statistical approaches for estimating efficiency can be found in a book chapter by Green (1993). Consider 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 situation can be written as

$$E_{t,i} = \alpha + \beta_y \, y_{t,i} \tag{1}$$

where t is the smallest time period of interest, i is the ith 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. Because the actual data may not be perfectly measured and this simple relationship between energy and production may only approximate the true relationship, linear regression estimates of the parameters rely 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 follows a normal (bell-shaped) distribution with a mean of zero and variance, σ^2; that is, that about half of the actual values of energy use are less than what equation (1) would predict and half are greater.

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

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 equation (1) that gave the best, or lowest, observed energy use. One way to do this is to shift the line downward so that all actual data points are on or above the line (see figure 1). This corrected ordinary least squares (COLS) regression is one way to represent the frontier.

Although the COLS method has its appeal in terms of simplicity, a more realistic view is that not all differences between the actual data and the frontier are due to efficiency. Because 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, but also add an second random component to reflect energy inefficiency.

The stochastic frontier regression approach is a parametric/statistical modeling approach. It requires plant-level or similar disaggregate data for the analysis of energy intensity. This statistical method was developed to identify overall production or cost efficiency, but is adapted here to identify energy efficiency. Stochastic frontier regression analysis separates the energy intensity
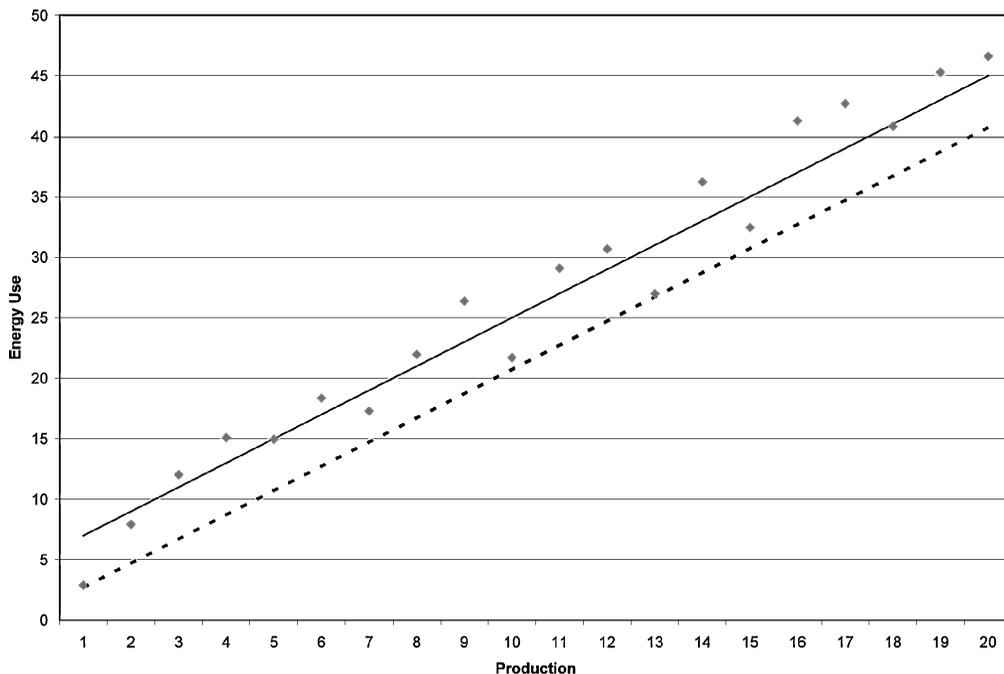

Figure 1 Average (solid line) and corrected (dotted line) linear regression of production and energy.

into three components: systematic effects, inefficiency, and statistical (random) error. Standard linear regression analysis includes only systematic effects and random error. As in standard regression, stochastic frontier regression requires that a linear model of systematic effects on energy intensity be specified. Variables in the model could include economic decision variables, such as energy prices and capital costs, or structural variables, such as plant size, location, utilization, and technology. The stochastic frontier regression analysis assumes there are two types of error terms. The first error term is the usual statistical, or random noise, error term. The second error term has a one-sided distribution and represents inefficiency, that is, the degree of departure from best practice energy intensity. Various assumptions can be made regarding the distribution of this inefficiency component. It can follow an exponential, truncated normal, or gamma distribution. Maximum likelihood techniques exist for when the efficiency distribution is exponential or truncated normal, and are available in many econometrics packages, for example, LIMDEP (Greene 1995). For an efficiency distribution that is assumed to follow the gamma distribution, a method of moments and a new method of simulated maximum likelihood (SML) make this approach computationally feasible (Greene 2000).[1]

If we modify equation (2) so that the right-hand side is energy intensity, that is, the ratio of energy use to output, the stochastic frontier regression approach can be represented by

$$
E/Y_i = f(X_i, Z_i; \beta) + \varepsilon_i \\
\varepsilon_i = u_i\text{-}v_i
\tag{3}
$$

where E is energy use, either electricity, nonelectric energy, or total primary energy; Y_i is output, measured either by physical production or as total value of output (that is, value of shipments corrected for inventory changes); X_i is the vector of systematic economic variables; Z_i is the vector of systematic external factors; β is the vector of parameters to be estimated; v_i is a typical random error term, $v \sim N\lfloor 0, \sigma_v^2 \rfloor$; u_i is distributed according to some one-sided error distribution, for example, gamma, exponential, half normal, or truncated normal; and $\sigma_{u,v} = 0$.

The flexibility of the gamma distribution makes this approach highly desirable, but difficult to estimate. The gamma distribution can generate the exponential distribution as a special case, as well as a more "general" one-sided distribution (see figure 2). No maximum likelihood techniques are available when u_i is distributed following a gamma distribution, but a consistent moments estimator exists. Conceptually, the regression intercept is shifted so that (nearly) all
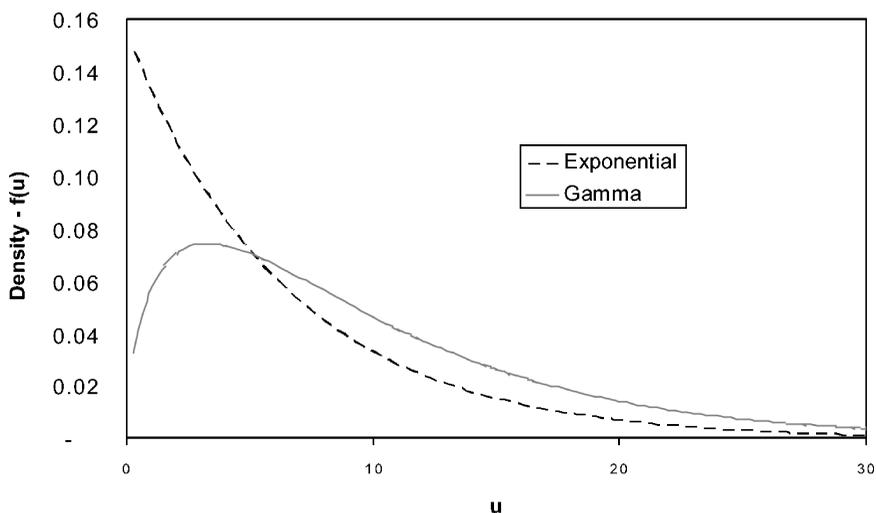

Figure 2 Example of exponential and gamma distributions.

residuals are positive, and the parameters of the normal error and gamma efficiency distributions are computed from those residuals. The method of moments computes the two parameters, θ and P, of the gamma distribution, $f(u)$, and provides a consistent adjustment to the linear regression intercept, a, based on the higher order moments of the residuals, e_i, from the ordinary least squares regression. The formulae for this estimator are

$$f(u) = [\theta^P/\Gamma(P)]e^{-\theta u}u^{P-1}, \ u, P, \theta > 0$$
$$\hat{\theta} = -3m_3/(m_4 - 3m_2^2)$$
$$\hat{P} = -\hat{\theta}^3 m_3/2$$
$$\hat{\sigma}_v^2 = m_2 - \hat{P}/\hat{\theta}^2 \qquad (4)$$
$$\hat{\alpha} = a - \hat{P}/\hat{\theta}$$
$$m_r = (1/N)\sum_i e_i^r$$

A recently developed method of SML for the gamma distribution is an improvement on the method of moments, because the parameters of the gamma distribution are estimated simultaneously with the systematic effects, β. We report examples of both the moment estimator and SML[2] approaches.

Data

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 data provided to the Argonne National Laboratory by automobile manufacturing companies.[3] The LRD includes the nonpublic, plant-level data that are 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, 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.[4] Under Title 13 of the U.S. Code, these data are confidential; however, CES allows academic and government researchers with Special

Table I Sample means for brewery EPI (monetary units are in thousands of 1997 $ and all other units are as shown)

Variable	Sample mean
Number of plants	42
Total value of shipments	$365,689
Production (thousand barrels)	4,408
Fuel costs	$2,374
Fuel (MMBtu)	618,394
Electricity costs	$3,353
kWh (thousands)	63,576

Note: 1 millon (MM) British thermal units (Btu) $\approx 1.055 \times 10^9$ joules (J, SI) $\approx 2.522 \times 10^7$ kilocalories (kcal). 1 barrel = 31.5 gallons (gal) ≈ 119.2 liters (L). 1 kilowatt-hour (kWh) $\approx 3.6 \times 10^6$ joules (J, SI) $\approx 3.412 \times 10^3$ British thermal units (Btu).

Sworn Status to access these confidential microdata under its research associate program. The confidentiality restrictions prevent the disclosure of any information that would allow 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 the issue of energy efficiency is examined.

In the examples presented below, the results for breweries, North American Industry Classification System (NAICS) code 312120, are based on the Census LRD data, and the results for motor vehicle assembly are from NAICS codes 33611, 336112, and 336120, based on the Census LRD data and also on data provided directly from five manufacturing companies with operations in the United States. This analysis of the Census data focused on detailed information available from the 1997 Census of Manufacturing. Sample statistics for the breweries analysis are shown in table 1, and sample statistics for the auto assembly analysis, based on the directly provided data, are shown in table 2.

For the analysis using the Census data, the choice of the variables in the X vector includes energy prices, plant scale (measured by output), and other input/output ratios, for example,

Table 2 Sample means for motor vehicle assembly EPI

Variable	Sample mean
kWh (thousands)	156,641
Fuel use (MMBtu)	1,273,120
Total site energy (MMBtu)	1,783,842
Production (vehicles)	227,615
Capacity (vehicles)	285,751
Heating degree days	4,970
Number of plant years	139

capital, materials, and labor. Because energy consumption may be influenced by climate, as well as production activity, state-level data on heating and cooling degree-days are included in the energy intensity regressions in the Z vector. This data are matched to plant location. Plants that manufacture certain energy-intensive inputs, rather than purchase them, will have different patterns of efficient energy use. This level of upstream integration in motor vehicle assembly was measured by the fraction of material costs for stampings and engines. For breweries, those that manufacture malt from grain were similarly identified by the share of costs.

The data from the five automotive manufacturing companies were provided after a review of the results based on the Census data (described below). The data provided by these companies were more limited in scope than that available at the Census, but included numbers of vehicles produced and plant capacity, which were not available at the Census, as well as electricity and nonelectric energy consumption. Plants with upstream parts operations at a plant, for example, stamping and engines, had energy use for those areas of the plant subtracted from the plant total. In principle, this allows a more direct comparison of energy use in an assembly plant.[5]

Results

Initial model results for the energy intensity equations for both sectors and two forms of energy, electricity and nonelectric energy, were cleared by the Census for public review. Nonelectric energy is principally purchased fossil fuels, for example, coal or natural gas. The functional form was a simple linear model. Cross

product and squared terms were evaluated, but were rarely significant, with the exception of the square of production. This term was left in the model.

To facilitate review by company energy managers, spreadsheets were constructed to display the gamma density and distribution functions for the models that were estimated, which were evaluated at user-selected values of the X and Z vectors. This aided in comparing the magnitude of the systematic effects (changes in X and Z) with the gamma efficiency distribution by graphically displaying the results. The spreadsheets were provided to energy managers from companies in the brewery and auto assembly sectors. Examples of the spreadsheets are shown in figures 3 and 4 for the brewery and auto assembly sectors, respectively, with the data based on sample means. The energy managers from the companies were encouraged to input data for their own plants and then provide comments.

During the review process, the point was made that product pricing in both of these industries reflects a number of market characteristics (oligopoly pricing, rebates, and so forth) such that shipment value of production might not be an appropriate measure of output, particularly for measuring energy efficiency. These observations led to an additional analysis that significantly influenced our estimates of the efficiency distribution.

For breweries, we constructed a physical measure of production, gallons[6] of malt beverages, using detailed Census product code data, which report shipments identified by packaging type (cans, bottles, and so forth) and size. We compared this estimate of 5.7 billion gallons to the total industry gallons of production published by the U.S. Bureau of Alcohol, Tobacco, and Firearms (BATF) of 6.2 billion gallons. Our data included only large breweries. Small breweries produced 9% of the total value of output in 1997, which is consistent with the difference between the physical volume that we estimated and the data published by BATF.

The brewery model was re-estimated using physical production. The results from the electricity equation illustrate the difference that this output product price variation makes in the estimates of efficiency distribution. For the electricity model illustrated in figure 3, the estimates of the
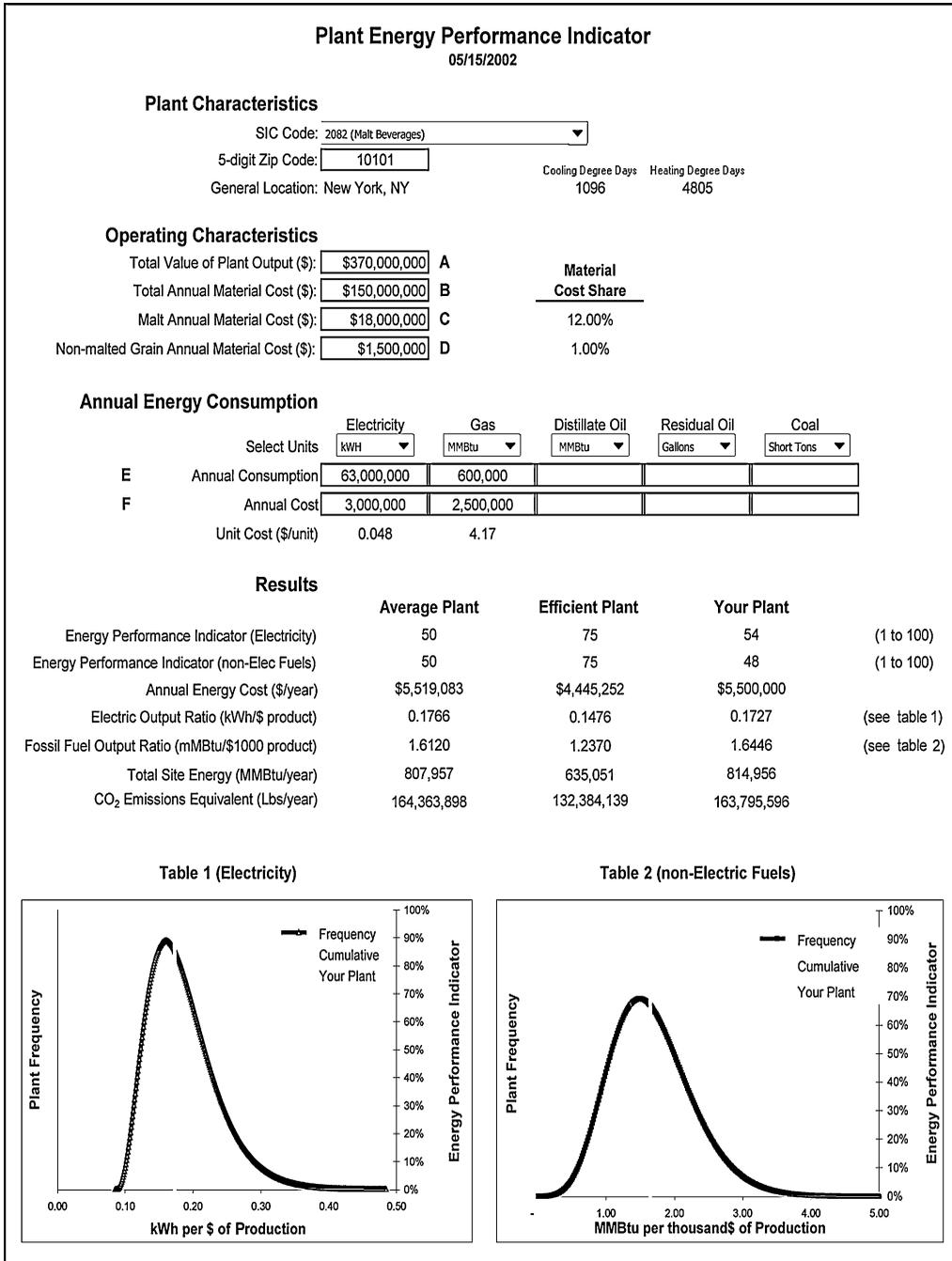
Plant Energy Performance Indicator
05/15/2002

Plant Characteristics

SIC Code: 2082 (Malt Beverages) ▼

5-digit Zip Code: 10101

General Location: New York, NY

	Cooling Degree Days	Heating Degree Days
	1096	4805

Operating Characteristics

			Material
Total Value of Plant Output ($):	$370,000,000	A	**Cost Share**
Total Annual Material Cost ($):	$150,000,000	B	
Malt Annual Material Cost ($):	$18,000,000	C	12.00%
Non-malted Grain Annual Material Cost ($):	$1,500,000	D	1.00%

Annual Energy Consumption

		Electricity	Gas	Distillate Oil	Residual Oil	Coal
	Select Units	kWH ▼	MMBtu ▼	MMBtu ▼	Gallons ▼	Short Tons ▼
E	Annual Consumption	63,000,000	600,000			
F	Annual Cost	3,000,000	2,500,000			
	Unit Cost ($/unit)	0.048	4.17			

Results

	Average Plant	Efficient Plant	Your Plant	
Energy Performance Indicator (Electricity)	50	75	54	(1 to 100)
Energy Performance Indicator (non-Elec Fuels)	50	75	48	(1 to 100)
Annual Energy Cost ($/year)	$5,519,083	$4,445,252	$5,500,000	
Electric Output Ratio (kWh/$ product)	0.1766	0.1476	0.1727	(see table 1)
Fossil Fuel Output Ratio (mMBtu/$1000 product)	1.6120	1.2370	1.6446	(see table 2)
Total Site Energy (MMBtu/year)	807,957	635,051	814,956	
CO_2 Emissions Equivalent (Lbs/year)	164,363,898	132,384,139	163,795,596	

Table 1 (Electricity) **Table 2 (non-Electric Fuels)**

(Table 1 chart: Plant Frequency vs kWh per $ of Production; Energy Performance Indicator 0%–100%; legend: Frequency, Cumulative, Your Plant; x-axis 0.00, 0.10, 0.20, 0.30, 0.40, 0.50)

(Table 2 chart: Plant Frequency vs MMBtu per thousand$ of Production; Energy Performance Indicator 0%–100%; legend: Frequency, Cumulative, Your Plant; x-axis 1.00, 2.00, 3.00, 4.00, 5.00)

Figure 3 Spreadsheet for breweries model.

parameters of the gamma distribution, θ and P, are 39.6 and 3.99, respectively. For the model with physical production as the denominator, estimates of θ and P are 7.27 and 0.441, respectively. When $P = 1$, the gamma model is equivalent to the exponential distribution. Figure 5 illustrates that when production is measured in barrels, the efficiency distribution is substantially

Plant Energy Performance Indicator
05/20/2002

Plant Characteristics

SIC Code: 3711 (MOTOR VEHICLE ASSEMBLY) ▼

5-digit Zip Code: 10101

General Location: New York, NY

Cooling Degree Days 1096   Heating Degree Days 4805

Operating Characteristics

			Material Cost Share
Total Value of Plant Output ($):	$3,000,000,000	A	
Total Annual Material Cost ($):	$2,000,000,000	B	
Automotive Stamping Annual Material Cost ($):	$180,000,000	C	9.00%
Gas Engines & Parts Annual Material Cost ($):	$300,000,000	D	15.00%
Painting Annual Material Cost ($):	$20,000,000	E	1.00%
Drive Train & Parts Annual Material Cost ($):	$250,000,000	F	12.50%

Annual Energy Consumption

		Electricity	Gas	Distillate Oil	Residual Oil	Coal
	Select Units	kWH ▼	MMBtu ▼	Gallons ▼	Gallons ▼	Short Tons ▼
G	Annual Consumption	162,715,141	1,028,690			
H	Annual Cost	8,034,483	4,047,090			
	Unit Cost	0.0494	3.93			

Results

	Average Plant	Efficient Plant	Your Plant	
Energy Performance Indicator (Electricity)	50	75	66	(1 to 100)
Energy Performance Indicator (non-Elec Fuels)	50	75	74	(1 to 100)
Annual Energy Cost ($/year)	$14,210,225	$11,087,757	$12,081,573	
Electric Output Ratio (kWh/$ product)	0.065	0.049	0.055	(see table 1)
Fossil Fuel Output Ratio (MMBtu/ $1,000 product)	0.41	0.34	0.35	(see table 2)
Total Site Energy, All Fuels (MMBtu/year)	1,406,638	1,159,645	1,028,690	
Estimated CO_2 Emissions (Tons/year)	213,558	166,632	181,568	

Table 1 (Electricity)

Plant Frequency vs. kWh per $ of Production; Energy Performance Indicator. Legend: Frequency, Cumulative, Your Plant.

Table 2 (Fossil Fuels)

Plant Frequency vs. MMBtu per $1,000 of Production; Energy Performance Indicator. Legend: Frequency, Cumulative, Your Plant.

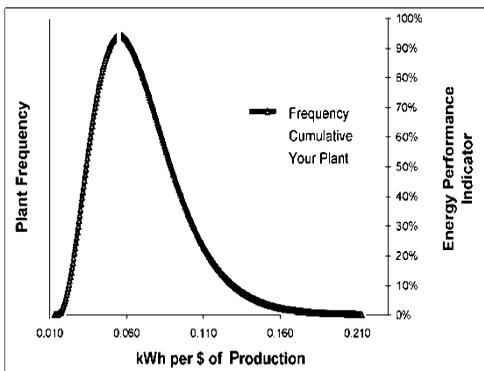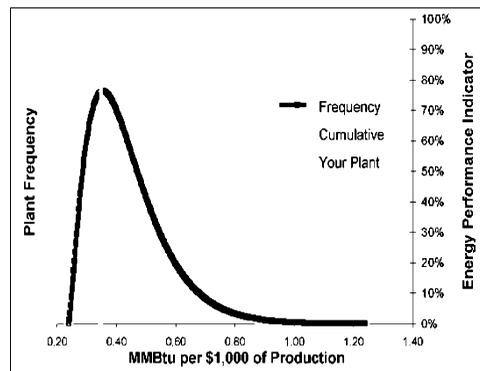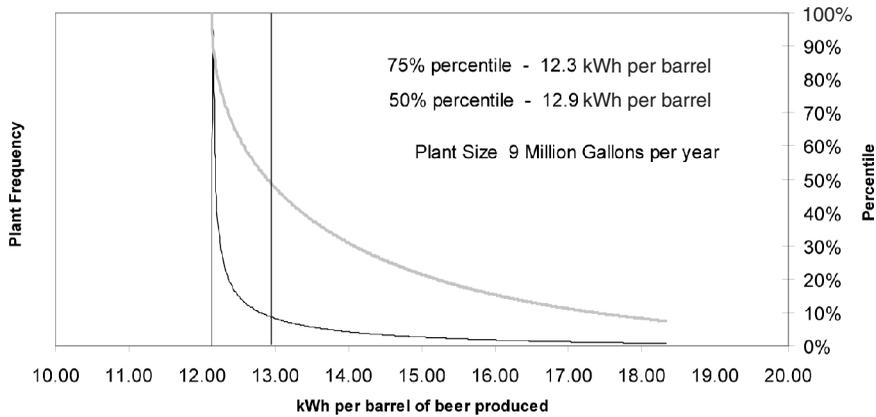**Figure 4** Spreadsheet motor vehicle assembly model.

Figure 5 Efficiency distributions for electricity use per barrel show that many brewery plants are near the best practice, but with a smaller number of extreme outliers.

different from that in the previous model, based on total value of output, and is much closer to an exponential distribution.

For automobile assembly, participating companies provided plant-level production and energy data, which were used to develop a model estimate. Total site energy is the variable used to aggregate energy; that is, kilowatt-hours are converted to British thermal units (Btus) using 3,412 Btu/kWh. Auto assembly was defined to include only the production activities of welding, painting, and final assembly. Either the data used were for an entire plant with only those activities present, or the energy data for those activities were isolated from the energy associated with other activities. Three years of data were used for plants from five motor vehicle manufacturing companies with assembly plants in the United States. The final EPI model for motor vehicle assembly was based on 101 plant-years of data.[7] Sample means of the data are shown in table 3.

The model estimated for total site energy is

$$E/Y = A + \beta_2 Util + \beta_3 Util^2 + \beta_4 HDD$$
$$+ \beta_5 CDD + \beta_6 CDD^* AT$$
$$+ \beta_7 WB + u\text{-}v \quad (5)$$

where E = total site energy use in MMBtu (1 kWh = 3,412 Btu); Y = number of vehicles produced; Util = plant utilization rate, defined as output/capacity; $Util^2$ = the square of utilization; HDD = heating degree days for the plant

location and year (thousands); CDD = cooling degree days for the plant location and year (thousands) regardless of whether or not a plant is air–tempered; CDD^*AT = cooling degree days for the plant location and year (thousands) if a plant is air–tempered; WB = wheelbase (in.) of the primary product; and β = the vector of parameters to be estimated, $v \sim N(0, \theta_v^2)$, and $u \sim \theta (\theta, P)$.

The method of simulated maximum likelihood is used to jointly estimate the structural and error distribution parameters. The final parameter estimates are presented in table 4. All variables are significant at the 1% level or less, except those noted with an asterisk. The signs of all variables are as expected.

Using the gamma distribution results in an almost exponential distribution for the efficiency distribution of total site energy per vehicle (see figure 6). The gamma distribution becomes

Table 3 Sample means for motor vehicle assembly EPI

Variable	Sample mean
kWh (thousands)	156,641
Fuel use (MMBtu)	1,273,120
Total site energy (MMBtu)	1,783,842
Production (thousands vehicles)	227,615
Capacity (thousands vehicles)	285,751
Cooling degree days	1,501
Heating degree days	4,283
Number of plant years	101

Table 4 Model results for motor vehicle assembly EPI

Variable	Coefficient	t-ratio	Sample mean
Constant	0.597*	0.532	
Util	−17.27	−16.546	1.08
Util^2	6.982	11.425	1.31
CDD	1.962	65.906	1.59
HDD	1.299	24.183	4.28
CDD*AT	0.723	6.731	–
WB	0.046	13.658	122
θ	0.154	2.966	
P	0.527	4.492	
θ_v	4.98E-06*	0.286	

Note: *Not significant.

exponential when $P = 1.0$, and our estimate is $P = 0.53$. Plant utilization rates were included to capture the quasi-fixed relationship between energy capital equipment and production throughput. We did not expect these effects to be linear, so we used a specification that included a linear second-order term. The impact of CDD and HDD indicates that a plant has increased heating and cooling loads according to climate. The impact of plants that air-temper (that is, cool the outside air) is based on the combined effect of both CDD and CDD*AT. In other words, the adjustment for cooling degrees for air-tempered plants is $(1.962 + 0.723)$ MMBtu/vehicle for every thousand cooling degree-days. WB measures the effects that producing vehicles of varying sizes have on energy requirements and best practice energy use. The mean of the WB variable indicates that the average vehicle is 122 in.

The impact of utilization rates is best illustrated graphically (see figure 7). Because capacity is defined as two shifts, greater than 100% utilization is possible. Best practice declines at a declining rate until the plant is at three shifts (150%), where the impact on best practice energy use is negligible. Because the capacity utilization adjustment is based on a quadratic curve, we can illustrate the impact of utilization from
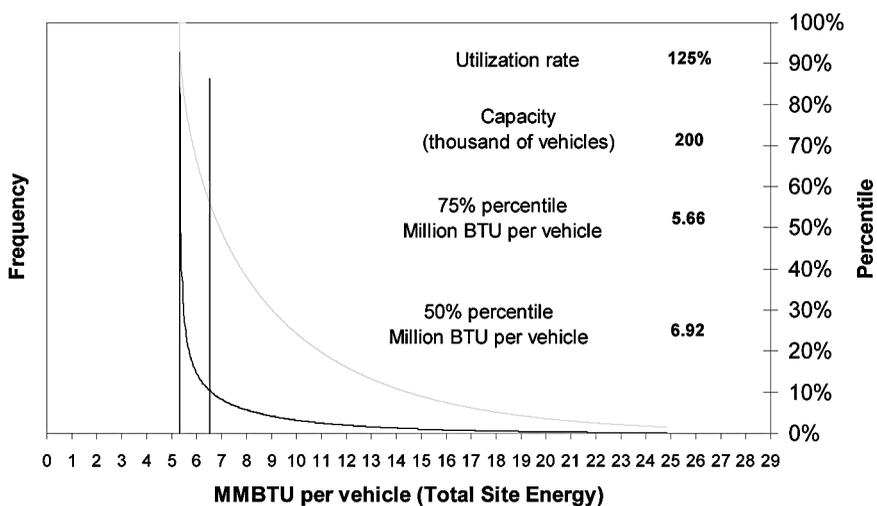

Figure 6 Efficiency distributions for total site energy use per vehicle follow a distribution similar to that of breweries.
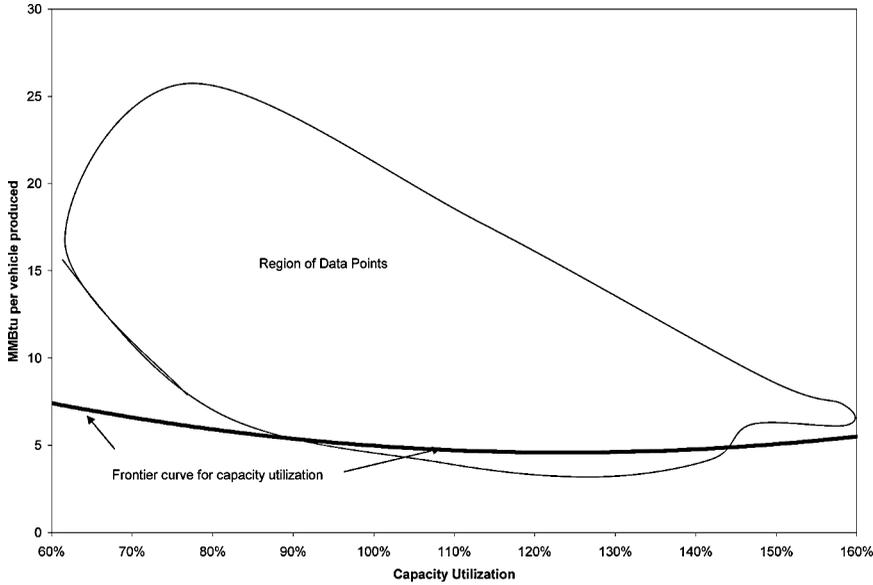
Figure 7 Impact of utilization on best practice total site energy per vehicle.

the change in best practice energy relative to a 100% utilized plant. The change in MMBtu per vehicle when utilization rises to 125% is −0.4, but is 1.3 when utilization falls.

A normalized distribution is constructed to illustrate how this model can be applied to estimating the overall distribution of efficiency performance relative to the frontier. To do this, the absolute level of efficiency—that is, the difference between actual plant performance and the plant-specific best practice—is computed. The plant-specific best practice is the level of energy use that could be obtained, given the plant's observed capacity, utilization, weather, and so forth.

We then compute the best practice level of energy use for a "typical plant" producing 200,000 automobiles per year at full utilization. Figure 8 shows the results. On this normalized basis, about half of the plants use less than 6.5 million Btu per vehicle, with a rather long tail on the distribution. The average MMBtu per vehicle is 8.1, and the median MMBtu per vehicle is 6.9. The estimated best practice, although influenced by external forces, varies far less; average best practice is 5.1 MMBtu per vehicle, and the median is 4.7. Another way to look at the model results is to compute the percentage energy inefficiency, (actual energy—best practice)/actual energy (see
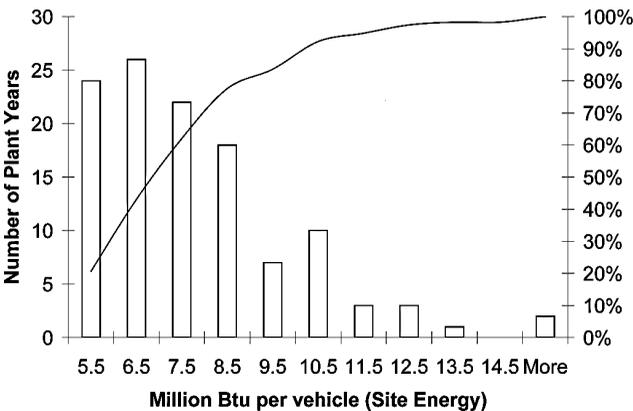

Figure 8 Normalized distribution of total site energy per vehicle. Bars indicate frequency in plant years. Solid line indicates cumulative distribution (percentage).
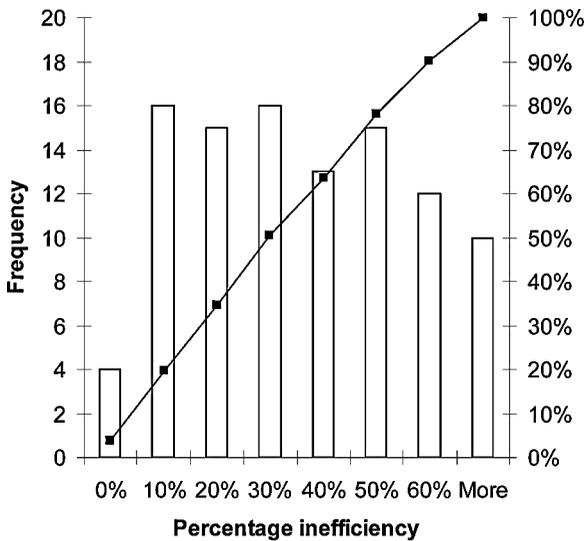
Figure 9 Distribution of percentage energy inefficiency. Bars indicate frequency in plant years. Solid line indicates cumulative distribution (percentage).

figure 9). The median of this percentage taken over all 101 plant years is an energy efficiency gap of 27%. On the other hand, the figure shows how widely this varies across plants.

Conclusion

This article presents a statistical method that can be used to compute best observed practice and a normalized distribution of energy efficiency. It is based on plant-level data and a stochastic frontier regression approach. The use of the gamma distribution is found to be very flexible in representing efficiency. During a review of the results with industry energy managers, it was found that using dollar measures of plant output might obscure some energy relationships, so physical production was used to compute the energy/output ratio. Detailed results from the auto sector show the importance of capacity utilization in best practice energy use. Large vehicles and more extreme heating and cooling loads all contribute to higher energy requirements. This approach accounts for these differences when measuring best practice; hence, the estimate of efficiency. The median efficiency is estimated to differ from the plant-specific best practice for auto assembly by 27%.

The primary use of this analysis, however, is not to identify "average" levels of inefficiency, but rather to assist each plant or corporate-energy manager in understanding how individual plants are performing relative to their peers. Without information on achievable levels of energy performance, the manager may have difficulty in setting goals or targeting resources. Although the claim that "every plant is different" is shown to be correct, the model can normalize for many of these differences and provide information on the potential to improve the environment and the business's bottom line by reducing energy consumption. The close involvement of this research with ENERGY STAR aims to develop a tool that is usable by these and other industries to improve their information regarding achievable levels of energy performance.

Acknowledgments

in Washington, DC in June 2002, participants in the 2nd Annual Energy Star Motor Vehicle Industry Focus Meeting, held in Detroit, MI, in August 2003, and participants in the 3rd Annual Energy Star Auto Industry Focus Meeting, held in Georgetown, KY, in August 2004. Portions of the research presented in this article represent analysis undertaken while the author was a research affiliate at the Center for Economic Studies at the U.S. Census Bureau. The research has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. Research results and conclusions expressed are those of the author and do not necessarily indicate concurrence by the Census Bureau, the Argonne National Laboratory, or the sponsoring agency. Results have been screened to ensure that no confidential information is revealed.

Notes

1. SML replaces a complex term in the likelihood function with a consistent estimate based on random draws from the appropriate distribution. SML techniques have been applied to other likelihood functions, for example, multinomial probit and other discrete choice models.
2. The simulated maximum likelihood has only recently been made operational in the LIMDEP statistical package, so this article only presents the SML for the auto industry data.
3. The auto manufacturing industry data are proprietary business information and were provided to ANL under a nondisclosure agreement with the respective companies.
4. They are the years ending in "2" or "7," for example, 1982 or 1997.
5. For the purposes of this article, "assembly" encompasses painting, body welding, and the final assembly process.
6. One gallon ≈ 3.79 L.
7. Thirty-four plants were included in the analysis. For one plant, missing data resulted in one of the three years being dropped, so the number of plant years was 101.

References

Boyd, G. A. and J. D. McClelland. 1999. The impact of environmental constraints on productivity im-

provement in integrated paper plants. *Journal of Environmental Economics and Management* 38(2): 121–142.

Boyd, G., J. Molburg, and R. Prince. 1996. Alternative methods of marginal abatement cost estimation: Nonparametric distance functions. Paper presented at the 17th Annual North American Conference of the United States Association for Energy Economics and the International Association for Energy Economics, October 27–30, Boston.

Boyd, G. A., R. Färe, and S. Grosskopf. 1998. *International productivity growth with CO_2 as an undesirable output*. Boston: Eastern Economic Association.

Brannlund, R., R. Färe, and S. Grosskopf. 1995. Environmental regulation and profitability: An application to Swedish pulp and paper mills. *Environmental and Resource Economics* 6(1): 23–36.

Chung, Y. H., R. Färe, and S. Grosskopf. 1997. Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management* 51(3): 229–240.

Corio, M. and G. Boyd. 1996. Applying the competitive business equation to power generation operations and economics. Paper presented at the EPRI Fossil Plant Heat Rate Improvement Conference, May, Dallas, TX.

Green, W. 1993. The econometric approach to efficiency analysis. In *The measurement of productive efficiency: Techniques and applications*, edited by H. Fried, C. A. K. Lovell, and S. Schmidt. New York: Oxford University Press.

Greene, W. 1995. *LIMDEP version 7.0 user manual*. Plainview, NY: Econometric Software.

Greene, W. 2000. *Simulated likelihood estimation of the normal-gamma stochastic frontier function*. Working paper, New York University Economics Department, New York.

Jeon, B. M. and R. C. Sickles. 2001. The role of environmental factors in growth accounting: A nonparametric analysis. Paper presented at the 2001 North American Winter Meetings of the Econometric Society, 5–7 January, New Orleans, LA.

Nelson, K. 1994. Finding and implementing projects that reduce waste. In *Industrial ecology and global change*, edited by R. Socolow, C. Andrews, F. Berkhout, and V. Thomas. New York: Cambridge University Press.

Panayotou, T. and C. Zinnes. 1994. Free lunch economics for industrial ecologists. In *Industrial ecology and global change*, edited by R. Socolow, C. Andrews, F. Berkhout, and V. Thomas. New York: Cambridge University Press.

Porter, M. E. and C. v. d. Linde. 1995. Green and competitive. *Harvard Business Review* 73(5): 120–134.

Seiford, L. M. 1999. A cyber-bibliography for data envelopment analysis (1978–1999). In *Data envelopment analysis: A comprehensive text with models, applications, references, and DEA-solver software,* edited by W. W. Cooper, L. M. Seiford, and K. Tone. Boston: Kluwer Academic Publishers.

Tyteca, D. 1998. On sustainability indicators at the firm-level: Pollution and resource efficiency as a necessary condition toward sustainability. *Journal of Industrial Ecology* 2(4): 61–77.

Yaisawarng, S. and J. D. Klein. 1994. The effects of sulfur dioxide controls on productivity change in the U.S. electric power industry. *Review of Economics and Statistics* 76(3): 447–460.

Zofio, J. L. and A. M. Prieto. 2001. Environmental efficiency and regulatory standards: The case of CO_2 emissions from OECD industries. *Resource and Energy Economics* 23(1): 63–83.

About the Author

Gale A. Boyd is a staff economist at the Argonne National Laboratory in Argonne, Illinois, USA and the Chairman of the Chicago Census Research Data Center.