

# Comparing the statistical distributions of energy efficiency in manufacturing: meta-analysis of 24 Case studies to develop industry-specific energy performance indicators (EPI)

Gale A. Boyd

Received: 30 June 2015 / Accepted: 18 April 2016 / Published online: 13 May 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** There is growing interest among policy makers and others regarding the role that industrial energy efficiency can play in mitigation of climate change. For over 10 years, the US Environmental Protection Agency (EPA) has supported the development of sector-specific industrial energy efficiency case studies using statistical analysis of plant level on energy use, controlling for a variety of plant production activities and characteristics. These case studies are the basis for the ENERGY STAR<sup>®</sup> Energy Performance Indicators (EPIs). These case studies fill an important gap by estimating the distribution of efficiency using detailed, sector-specific, plant-level data. These estimated distributions allow Energy Star to create “energy-efficient plant benchmarks” in a variety of industries using the upper quartile of the estimated efficiency distribution. Case studies have been conducted for 14 broad industries, 2 dozen sectors, and many more detailed product types. This paper is a meta-analysis of the approach that has been used in this research and the general findings regarding the estimated distribution of performance within and across industries. We find that there are few sectors that are well represented by a simple “energy per

widget” benchmark that less energy-intensive sectors tend to exhibit a wider range of within-industry efficiency than energy-intensive sectors, but changes over time in the level and range of energy efficiency do not reveal any single pattern.

**Keywords** Efficiency · Energy performance indicator · Manufacturing · Normalized statistical distributions

## Introduction

ENERGY STAR is a voluntary program launched by the Environmental Protection Agency (EPA) in 1992 to identify and promote energy-efficient products, buildings, homes, and manufacturing facilities. The program was established to find cost-effective ways to reduce greenhouse gas emissions associated with energy use. Initially focused on consumer products, the program expanded into the commercial building market in 1995 and released its first energy-efficiency benchmark for office buildings in 1998. In 2000, the EPA expanded the program to include manufacturing plants (US Environmental Protection Agency 2015). ENERGY STAR focuses on filling critical knowledge gaps regarding industrial energy efficiency by providing tools, in the form of the 24 industry-specific case studies that are described in this paper, which encourage better corporate energy management through the development of sector-specific energy performance benchmarks. These 24 case studies, which ENERGY STAR calls energy performance indicator (EPI), are a statistical analysis of

---

G. A. Boyd  
Social Science Research Institute and Department of Economics,  
Duke University, Box 90989, Durham, NC 27708, USA

G. A. Boyd (✉)  
Social Science Research Institute, Duke University, Box 90989,  
Durham, NC 27708, USA  
e-mail: gale.boyd@duke.edu

energy plant level energy use, conducted with input from energy managers in the respective industries selected by EPA, which result in a tool that energy managers can use to assess energy performance.

One knowledge gap that these case studies fill is to provide companies within a manufacturing sector an objective measure of how a plant compares to the rest of the industry. Most companies lack sufficient information on the relative efficiency of their plants within the broader industry because that information is confidential. Consequently, many companies did not know if their plants are operating efficiently or how much they might be able to improve relative to existing best practices. These case studies address this by conducting statistical analysis of confidential data from a variety of sources and then making it available in an accessible manner that answers these questions but does not reveal the underlying confidential information. The second knowledge gap that these case studies fill of these studies is to establish criteria for determining which plants would be considered “efficient” and qualify for recognition by the Energy Star program. In order to provide recognition for energy-efficient manufacturing plants, EPA needed an objective and transparent means to determine which plants are “efficient,” which is defined by ENERGY STAR as “*the upper quartile of performance for similar production facilities.*” The use of “*upper quartile*” as the definition of efficient means that a distribution of energy performance is required. The qualifier “*similar production facilities*” also requires some way of identifying and controlling for differences in plant characteristics so that the resulting estimate of the distribution of efficiency, used to measure the quartiles, is an apples-to-apples comparison.

The statistical approach used in the case studies meet these requirements by providing (1) industry access to estimates of efficiency with (2) a high level of confidentiality protection where no individual plant (company) performance is revealed or can be reidentified while (3) estimating the quartiles of the distribution of performance and (4) controlling for statistically testable, observable differences between plants’ energy use in the available data. There are many sources of estimates of the—average or best practice—energy for a manufacturing plant, a specific process, or an industry. However, it is not enough to estimate that a plant is, for example, 2 % better than average, since that statistic does not reveal how many plants in the industry satisfies that criteria; that level of performance may place the plant

in the top decile or only slightly more than the 50th percentile. Similarly, an engineering estimate of best practice may reveal what represented the 99th percentile of performance, but not the extent to which those technologies have been adopted.

To fill these knowledge gaps, these case studies address a variety of issues:

- How to define energy efficiency?
- How to control for differences between plants and which ones?
- What data are available?
- What statistical approaches can be used to measure the quartiles?

This paper is a meta-analysis, or overview, of how these case studies of efficiency in a range of industries have addressed these questions and knowledge gaps. This meta-analysis should not be confused with studies that employ meta-frontier approaches (e.g. Battese and Rao 2002). The paper first defines the notion of efficiency used in the case studies, followed by a discussion of what differences should be (ideally) controlled for in the analysis and how this relates to other energy benchmarking approaches in the literature. The next section lays out the basic statistical methodology and its advantages and disadvantages. Then, data and the associated confidentiality issues are presented. The last two sections are an overview of the results from these case studies.

## Defining Energy Efficiency<sup>1</sup>

Efficiency is a measure of relative performance, but relative to what? Defining energy efficiency requires a choice of a reference point against which to compare energy use. Energy efficiency measures can be developed through a variety of means, such as engineering and theoretical estimates of performance or through observing the range of actual levels of performance. The choice of method used to define efficiency depends on the need to define a reference point for energy efficiency. One of the challenges with using energy efficiency measures based on engineering or theoretical estimates is that they may be dismissed by the industry as being economically infeasible. Consequently, these

<sup>1</sup> This section draws from Boyd (2012)

case studies have focused on developing energy efficiency measures based on actual or observed operational performance rather than theoretical estimates of potential efficiency levels. Additionally, the EPA needed to identify a method that would be perceived by users as providing economically feasible performance targets.

The reference point for economic potential (observed practice) depends, in part, on the reason for measuring efficiency as well as the available information to create a reference. Generally, the *ceteris paribus* principle (“all other things being equal or held constant”) is usually desired in creating the reference point. From a practical perspective, there is a hierarchy of measures and methods by which one can “hold constant” things that influence *energy use* that are not part of *energy efficiency*. The first is some measure of production activity, either production of the final product or, alternatively, a ubiquitous input into the production process. This is most commonly done by computing the ratio of energy use to activity, a measure of energy intensity. Energy intensity is a common metric that controls for changes in production and is commonly confused with energy efficiency, as in the statement “*the industry or plant’s energy efficiency has improved based on the fact that the corresponding energy intensity has declined over time.*” This type of statement brings us to the second way that one may approach the *ceteris paribus* principle for measuring efficiency, comparing energy intensity of a particular plant, firm, or industry to itself over time. This approach is a plant<sup>2</sup>-specific *baseline* comparison or *intraplant* efficiency measure. The baseline approach has the advantage of controlling for some plant-specific conditions that do not change during the comparison period.

The next level of this *ceteris paribus* principle is an *interplant* comparison that may include a variety of factors that influence energy use, but may not be viewed as efficiency. Factors may include difference in the types of product and materials used, as well as location-specific conditions. Intraplant comparisons within an industry also get us closer to the notion of an observed best-practice measure of economic energy efficiency, since by definition, there is some group of plants that are the best performers. This was the notion introduced

by Farrell (1957) and has been the basis for measuring production efficiency in economics (Huntington 1995). A modified approach has been adopted (Boyd 2005a, b) and its evolution is discussed by Boyd et al. (2008).

#### Intensity metric selection

Murray (1996) discusses a variety of intensity metrics for efficiency measurement, including thermodynamic approaches. Intensity ratios provide a basic metric for measuring energy efficiency and performance compared to a baseline. To measure intensity, you need a measure of energy and something for the denominator. For the numerator, these case studies use total source energy, defined as the net Btu total of the fuels (Btu) and electricity (Kwh) with electricity converted to Btu based on the level of efficiency of the US grid for delivered energy, i.e., including generation and transmission losses. A net measure is needed for when energy is transferred offsite, most commonly in the form of steam or electricity.

The choice of the denominator is a major issue for measuring intensity. Ideally, the denominator should capture some measure of production. Freeman et al. (1997) show that industry level trends in energy intensity based on value, both total and value added, can differ dramatically from those based on physical quantities. As Freeman et al. have observed, there are many challenges with creating efficiency measures based on price indexes, cost, and other value measurements.<sup>3</sup>

Given issues with linking energy use with price indexes, these studies have focused on using metrics based on physical quantities. For physical production to be meaningful, it needs to be at a high level of industry specificity. For example, the “Dairy” industry produces many products that cannot be aggregated, but “Fluid Milk” can. Therefore, within industries, it is necessary to differentiate between specific types for plants and manufacturing operations.

<sup>3</sup> As Freeman et al. (1997) note, “For an industry producing a single, well-defined, homogeneous good, it is relatively easy to construct an accurate price index. Most industries, however, produce many poorly defined, heterogeneous goods. For a variety of reasons, the more diverse the slate of products produced by an industry, the more difficult it becomes to construct an accurate price index. ...the accuracy of industrial price indexes is of extreme importance to industrial energy analysts and policy makers who use value-based indicators of energy intensity.”

<sup>2</sup> Throughout the paper, we will refer to the plant level as the unit of observation, but the concept may also apply to more aggregate levels such as firms and industries, and sometimes, to less aggregate levels such as process units.

Similarly, measures of building energy use commonly use physical size (ft<sup>2</sup>) as the main denominator for energy intensity, but for most industrial facilities, this is not appropriate.<sup>4</sup> While commonly used for commercial buildings where energy use is primarily tied to plug loads, lighting and heating, ventilation and air-conditioning (HVAC) systems, energy intensity based on size (sq. ft) does not correspond well with manufacturing process energy uses.

### Multifactor efficiency measures

While energy intensity ratios are commonly used for intraplant level baseline comparisons in an industrial energy management setting, their value for developing interplant comparisons may be limited. For interplant comparison, there are multiple factors that must be considered. To make an interplant comparison for energy management, one wants to know which plant is performing better, so the company can learn from the top performers and focus efforts on the bottom performers. The comparison needs to be “fair”; i.e., an “apples-to-apples” comparison. To do so, we must “normalize” or “remove” the differences between the plants that influence energy use, but we do not consider being part of the “energy efficiency” of the plant. This raises the question of what should be included, or “normalized for,” in the efficiency measure and what should not. Other research has tried to address this question. May et al. (2015) lays out a proposal to develop multiple energy key performance indicators. Tanaka (2008) develops a framework to address measures of energy efficiency performance, ranging from simple intensity to more inclusive approaches as are proposed here.

The first basic principle proposed is to normalize for things “beyond the control of the energy manager.” There are obvious things, like weather, that meet this criterion. However, should one also consider long-lived capital investment in a particular energy using process? A business is unlikely to relocate to a favorable climate for energy reasons, but they may also be unlikely to replace a large piece of capital *solely for reasons of energy efficiency*. This first principle is viewed as

<sup>4</sup> The one exception is pharmaceutical manufacturing where energy intensity is expressed as MMBTU/SQ FT. This metric was chosen largely because of the huge impact of HVAC systems in pharmaceutical manufacturing.

necessary, but not sufficient condition to be part of the normalization.

Consider the example of location and the influence of weather (climate) on energy. The location of the plant is an example of the second principle, a “*market-driven business decision*.” The location decision may be to best serve a local market, have access to raw materials at reasonable prices, etc. It is unlikely to consider the energy use related to the climate location, unless those cost are very large. Like the location of the plant, there are other basic categories of *market-driven business decision* that should be included in the normalization if they have significant impact on energy use. These should include<sup>5</sup> the following:

- Materials selection
- Plant capacity and utilization
- Product mix

Materials selection can influence energy use most commonly in the “make vs buy decision.” If a plant manufactures an intermediate input then it is using energy that would have been used elsewhere. To compare two plants that make juice, one from concentrate and one directly from fruit, it is necessary to account for the material choice. The final product may be virtually indistinguishable, but the plant-level energy use will differ. This “make or buy” business decision to be vertically integrated up the supply chain is market driven, has impacts on energy, and should be in the normalization. This issue is a common problem of defining the “boundary” for a benchmark, i.e., what plants to include only fully integrated plants or only one with identical inputs. The approach in these case studies is to include input choice in the analysis and allow the statistical models to normalize for those decisions. There are also issues of material quality that may imply more or less energy. Some materials have lower energy intensity but may impact product quality or be in short supply. In this case, the normalization should be to the average industry practices regarding input quality. Sand or cullet in glass making is an example. In this case, the appropriate comparison for normalization is to what the industry average rate of cullet use is.

<sup>5</sup> All considerations regarding normalization variables require adequate available data, since we are considering an empirical application.

Plant size and utilization rates can have impacts on energy use. Large plants may have advantages in terms of energy use due to economies of scale. The size of the plant will also be a market-driven decision regarding investment, market size/share, etc. A “small” plant cannot become “big” to reap the energy advantage if the company does not have the market share to sell higher product volumes. It is important to conduct the energy benchmark based on size, to the extent that size is a large impact on energy use, and there are different-sized plants in the industry that reflect differing market conditions. During economic downturn, energy intensity may rise for many reasons, one of which is the inherent levels of “fixed” vs “variable” energy use. While it is important for energy management to minimize “fixed” energy, the efficiency measure needs to account for this. Not all plants will experience the same market conditions at the same time and the normalization of utilization on energy due to fixed energy use should depend on industry practice. If industry practice on fixed energy management improves over time, then the efficiency measure should reflect this and normalize “less” for utilization.

The most ubiquitous component of normalization is product mix. Even within an industry like cement, the product is not entirely homogeneous. Some products that are in demand in the market may be more energy intensive. Masonry cement is more energy intensive than ASTM 1; milk in small cardboard cartons are more intensive (per gallon) than milk in gallon plastic jugs; orange juice is more intensive than apple juice; complex cast steel components are more intensive than simple cast steel item. In some cases, there is no way to substitute one energy-intensive product for another less-intensive product. In other cases, market preferences may change toward “green” or “low carbon” products over time, but companies have limited influence over these consumer preferences but react to those market demands. Milk in smaller containers (typically cardboard) is a different product than milk in gallon jugs from a consumers’ perspective. A company will not try to convince the consumer to drink the less energy-intensive apple juice instead of orange juice, but the efficiency measure for juice manufacturing can allow an “apples-to-apples” comparison for juice plants. The efficiency measure can then be used to inform the energy efficiency of the product that consumers demand.

Some aspects of manufacturing may appear to satisfy the two principles above, particularly when one

considers differences in underlying processes. For example, a small-scale, wet-process (energy intensive) cement kiln installed 30 years ago may be (1) “*beyond the control of the energy manager*” now and (2) “*market driven*,” at least at the time the decision was made. Similar examples exist for the choice automotive painting processes and paint booth design. In these cases of process choice, the 3rd, “*same product*,” principle needs to be applied.

Making products to serve the market is the *raison d’être* for manufacturing companies. If the product is, for all practical purposes, identical to another, then the plant using the more efficient process should be the reference point. Some examples may appear on the surface to satisfy this principle. Steel “mini-mill” plants are touted as being more energy efficient, using scrap and electricity to make steel. The products they produce, re-bar, etc. are not the same as sheet steel from more energy-intensive, integrated steel mills. In fact, the emergence of the mini-mill drove traditional steel out of those product lines to specialize in higher quality steel products (Boyd and Karlson 1993). The product demanded in the market and the process to make it became linked in the competitive markets; the efficiency measure should treat these two types of mills separately, not because of process, but product.

When making intraplant comparisons, it necessary to consider a variety of factors that do not neatly fit under the denominator of an energy intensity ratio. All plants may make a common product and other differences can significantly affect energy intensity. The difficulty with estimating an industry-level interplant efficiency measure is controlling for interplant difference other than production volume. While the things that differ between plants are numerous, there is a common thread across the case studies that the primary difference that have the most impact on energy fall into the following categories.

- Product mix
- Process input choices (i.e., “make or buy” upstream integration)
- Size—physical or productive capacity and utilization rates
- Climate (and other location-specific factors)

The choice of factors to include in the analysis depends upon the nature of the production process, the configuration of the industry (e.g., is upstream integration common or rare?), the availability of data to

represent these factors, and the outcome of the statistical testing. In order to address these types of factors, these studies use a multivariate approach to normalization where multiple effects are simultaneously considered (Boyd and Tunnessen 2007). The next sections discussed the four basic categories of effects that are commonly considered. There is further elaboration on the way this is implemented in the section on the specific case studies.

### *Product mix*

Not all plants produce exactly the same product. In fact, many plants produce multiple products. The diversity between plants gives rise to a mix of derived demands for specific processes and energy services. To the extent that the final product is the result of a series of energy using steps, the energy use of the plant will depend on the level and mix of products produced. Rather than specifying each process step individually, the approach used here is to identify those products that use significantly more (or less) energy and then use statistical analysis to estimate the different energy requirements of these products.

One approach to controlling for product mix is to segment the industry into cohorts based on product categories. This works best when there is no overlap between plants that produce the various basic products and there are sufficient numbers of plants to conduct the statistical comparison between those resulting groups. This means each subgroup is effectively treated as a separate industry for evaluation purposes. A good example is the glass industry where containers and flat glass are distinct industry segments.

When such natural subsectors do not exist and multiple products are produced within a plant, additional approaches are needed. The statistical approach is well suited to testing if a particular grouping of products is appropriate for benchmarking differences in energy. When industries produce a mix of products that differs across plants, then the product mix (share of activity) of distinct products is needed. This approach was first used in wet corn mills (Boyd 2008) and was later applied to other sectors.

In the absence of meaningful data on discrete product classes, an alternative is a continuous measure of product differentiation. Price is often taken as a measure of quality difference. To the extent that such quality differences arise for additional energy using processes, then

value of shipments may be an appropriate proxy for product mix. Differences in value may not involve higher energy use, as in luxury cars or specialty beers, but may be the case in creating different types of glass bottles or more complex cast metal products. Given available data, the link between energy and value (price) can be treated as largely an empirical issue, but preferably with some underlying hypothesis about the industry in the case study. Value of shipments might be used instead of a physical production variable or in conjunction with physical outputs. In the latter case, the ratio of value to physical product is price and becomes an implicit variable in the analysis. Other measures of energy-related product differences are industry specific; as in the case of vehicle size in automobile assembly.

### *Size*

Size and associated capacity utilization rates may directly impact energy use. Size may impact specific engineering and managerial advantages to energy use. If there is a substantial “fixed” level of energy use in the short run, the utilization rates may have a non-linear impact on energy intensity. In order to include size (and utilization) as a normalizing factor a meaningful measure of size or capacity is needed. It may be measured on an input basis, output basis, or physical size. In some cases, there may be advantages to larger scale of production, i.e., economies of scale. If it is the case that a larger production capacity or larger physical plant size has less than proportionate requirements for energy consumption, then there are economies of scale with respect to energy use. For example, in the cement industry, the scale is quite important. The larger size of the kiln (rather than several smaller kilns) has advantages in terms of energy use.

### *Process inputs*

There are three ways that process inputs are important for benchmarking. The first is that inputs such as materials, labor, or production hours may be good proxy measures of overall production activity when measures of production output are not available or have specific shortcomings.<sup>6</sup> The second is in the identification for upstream (vertical) integration, i.e., whether a plant

<sup>6</sup> As discussed in Freeman et al. (1997)

makes an intermediate product or purchases some preprocessed input. This is an important “boundary” issue for the energy footprint of a plant, even when two plants produce identical outputs. The third way is a variation of the second, relating to material “quality.” If there are alternative input choices that differ qualitatively and also with respect to energy use, then input quality measures can be introduced into the benchmark.

The first way process inputs can be helpful in developing a statistical benchmark of energy use is that inputs such as materials, labor, or production hours may be good proxy measures of overall production activity when measures of production output are not available or have specific shortcomings. If a physical measure of output is not readily available and pricing makes the value of shipments a questionable measure of production, then physical inputs can be a useful proxy. For some industries, the basic material input is so ubiquitous that it makes sense to view energy use per unit of basic input rather than (diverse) outputs. Process inputs may also be useful in measuring utilization, either directly or indirectly. Corn refining is a good example of this approach. The industry uses a ubiquitous input, corn. In some industries, physical production data may not be available but material flows are and can be used instead. For example, sand, lime, soda ash, and cullet (scrap glass) are the primary inputs to glass manufacturing.

The second way that process inputs are important for interplant benchmarking is when vertical integration is common in a sector but varies in degree from plant to plant. Industries are categorized by the products they produce, but some sectors may face a “make-or-buy” decision in the way they organize production. A plant may purchase an intermediate product or produce it at the plant as part of a vertically integrated plant. For example, an auto assembly plant may stamp body panels or ship them in from a separate facility. The energy use of these two facilities is not directly comparable. The interplant benchmarking approach must account for those “make-or-buy” decisions in the specific plant configurations. Examples range from food processing, where plant may make juice from concentrate or fresh fruit or paper mills which may purchase market pulp or recycled fibers.

The third way that process inputs are important for interplant benchmarking is when differences in material quality may also be related to energy use. For example, if the materials mix to produce a product directly impacts energy uses, then the

statistical model can apply different weights to the material mix in the same manner that it does with product mix. In other words, product/process level differences in energy use can be inferred from the volume and types of materials used in production. To be considered in the statistical normalization, they must be measured on a consistent plant-level basis across the industry. For cement plants, the hardness and moisture content of the limestone is hypothesized to influence energy use, but no consistent data is available for this, leaving it the subject of future analysis if data can be collected.

One ubiquitous input is labor. Labor may be helpful in capturing the quasi-fixed nature of energy if there are production slow-downs or non-production periods of operation, but when both labor and energy are being used. In this way, labor captures a plant activity level that is related to energy use, even when product output is not being generated. As an empirical issue, the statistical significance, or lack thereof, of labor in the analysis can capture this potentially industry-specific phenomenon.

### *Climate*

There are many things under the control of a plant or energy manager, but one they cannot control is “the weather.” In most manufacturing plants, HVAC contributes to energy demand and weather determines how much is required to maintain comfort. Since the approach used here is annual, seasonal variation does not enter into the analysis, but differences due to the location of a plant and annual variation from the location norm will play a role. The approach that has been taken for all sectors under study is to include heating and cooling degree days (HDD and CDD) into the analysis to determine how much of these location-driven differences in “weather” impact energy use.

In principle, all plants have some part of energy use that is HVAC related, but when the HVAC component of energy use is small relative to total plant consumption, the statistical approach may not be able to measure the effect accurately enough to meet tests for reliability. For some sectors, weather is a factor in the process, like auto assembly. It is a factor because of paint booths and the climate control technology needed for those systems. Pharmaceutical manufacturing, where “clean room” production environment is common, is another

good example. The climate impact in this sector is only applicable to the “finish and fill” production stage. The more energy-intensive chemical preparation stage is not sensitive to climate. Even in industries where the HVAC component is not an obvious or large part of energy use, there may be production process-related effects that analysis needs to test for. For example, processes that use chillers may be sensitive to CDD (summer) loading. Process heat furnaces may be sensitive to cold outside air so HDD (winter) effects might be included in the model.

### Relationship to other benchmarking approaches

Every industry is unique and these guiding principles have to be carefully weighed *against the specific industry conditions and available data*. These case studies endeavor to do so. The above discussion specifically avoids the term “benchmark,” although these case studies fall within the broad literature on benchmarking. The problem with the term benchmark is that it is very context specific, although research on energy (or emissions) benchmarks often identifies some of the same issues discussed above.

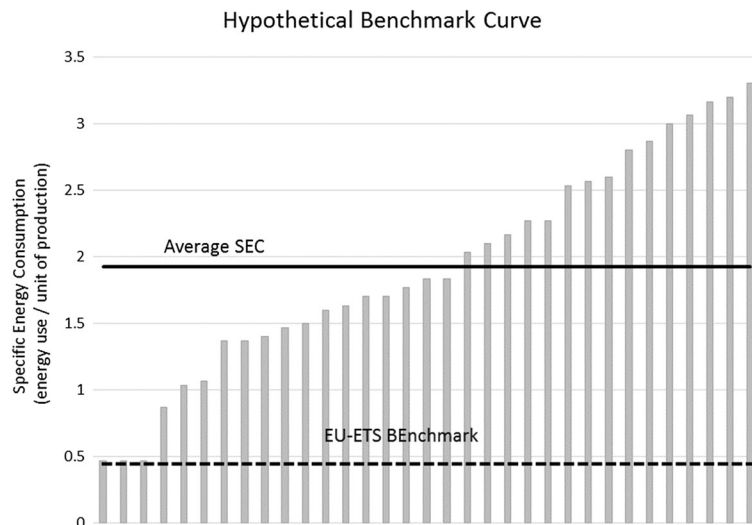
Aarons et al. (2013) introduced the concept of benchmarking as follows: “*The term “benchmarking” can refer to a wide variety of concepts* (emphasis added). *For the purposes of this paper, “benchmarking” refers to developing and using metrics to compare the energy or emissions intensity of industrial facilities. Benchmarks are primarily used to compare facilities within the same sector, but can also be used to identify best practices across sectors where common process units, such as boilers, are used.*” They consider using benchmarks as the basis for GHG allowance allocations and prefer a “one product, one benchmark” approach to promote systematic efficiency. They suggest, for this application of benchmarking, variables such as production technology, fuel choice, feedstock, climate, resource availability, or capacity should not be corrected for when a benchmark is set. The case studies in this paper control for feedstock, climate, and capacity when statistically appropriate, but not these others. The difficulty with a one product, one benchmark approach is that it requires facilities to

have single common physical product. This is rarely the case and is an example of how the approach is used in the case studies can address multiproduct manufacturing.

Erickson et al. (2010) consider a range of issues for benchmarking, also in the context of GHG emissions and regulations, considering both data availability and the context of related air regulatory processes in the USA. Egenhofer (2007) describe the evolution of the European Union Emissions Trading System (EU-ETS) and the role that benchmarking has played. The development of the EU-ETS benchmarks is described in more detail by Neelis et al. (2009). In particular, they develop 11 principles to guide the development. The one product, one benchmark method was developed to address the allocation of emission permits under the EU-ETS. The average of the 10 % most greenhouse gas efficient installations, in terms of metric tons of CO<sub>2</sub> emitted per ton of product produced at European level in the years 2007–2008, was selected as the guiding concept for emission allocations to individual companies and facilities. Fifty-two benchmarked sectors cover 75 % of industrial emissions in EU ETS. As is the case recommended by Aarons, Golding et al. and Neelis, Worrell et al., these EU-ETS product benchmarks are not differentiated by technology, fuel mix, size, age, climatic circumstances, or raw material quality of the installations producing the product. The choice of what products would receive a separate benchmark is given particular attention. The purpose of the EU-ETS benchmarks is to set the level of the allocation of free allowances which take into account the emissions performance of the most efficient installations. The allocation then incentivizes other installations to reduce emissions.

Similar to the applications of benchmarks to GHG regulations, most studies of energy benchmarks use specific energy consumption (SEC) which is an energy intensity usually defined over physical production units (Saygin et al. 2010) defines a *benchmark curve* as on that “*plots the efficiency of plants as a function of the total production volume from all similar plants or as a function of the total number of plants that operate at that level of efficiency or worse,*” so are concerned with the cumulative distribution of performance defined by SEC using data from a variety of sources, including plant, company, or region-level aggregates to compute averages. A hypothetical benchmark curve is shown in Fig. 1. To construct such a “curve,” one needs plant (company or region)-level data on energy use and





**Fig. 1** Hypothetical benchmark curve

production. The y-axis could be plants or represent cumulative production. A single “benchmark” for an industry might be set at the mean (as shown), median, or some percentile of the distribution (as was done for the ETS, but using GHG in the numerator).

These approaches to energy (or GHG) benchmarking requires plant (company or region) data on energy (GHG) and a denominator in terms of physical production. These approaches also have all the attendant problems identified by Freeman et al. regarding the choice of denominator, as discussed above. These examples above also do not distinguish between feedstock and product difference that might influence energy use. The question about what one *does or does not* wish to normalize for depends on the intended applications for the benchmark. Depending on the goals of the benchmark, what is viewed as “efficiency” may have multiple interpretations. However, empirically identifying the drivers for energy use, the underlying distribution (quartiles, etc.) is the first step.

An alternative but more data-intensive approach of benchmarking is exemplified by Phylipsen et al. (2002) and Worrell and Price (2006). This approach calculates an Energy Efficiency Index (EEI) for each plant by determining the difference between actual energy intensity of *each process unit* and *that of a hypothetical state-of-the-art unit* and then aggregating units within an entire plant. Their approach is comparable to the Solomon’s Energy Intensity Index<sup>®</sup> (as discussed in Phylipsen, Blok et al.)

Another energy benchmarking approach which draws from concepts in the process stage methods but employs data envelopment analysis (DEA) is Azadeh et al. (2007). They use DEA to account for the “structural factors” of energy use, such as input and product mix to create energy efficiency rankings but apply it to country-level OECD data rather than plant-level data. Oh and Hildreth (2014) apply both statistical and DEA methods to auto industry data.

Data is a driving concern in benchmarking. The one industry, one benchmark approach for regulatory purposes may have advantages in terms of transparency and when data provision is compulsory. The one industry, one benchmark approach also must have very specific product boundaries to avoid an “apples-to-oranges” comparison. This specificity may result in very small numbers of comparable production units, i.e., manufacturing establishments or process units. Process stage models have a strong appeal but are data intensive, since sub-establishment level information is required. The approach used in these case studies can be seen as a compromise between the needs in the one industry, one benchmark to very narrowly define industry sub-sector boundaries and the detailed data requirements of the process stage approach.

These case studies deal with these *scope* or *boundary* issues by selecting an industry segment (group of plants) with manageable ranges of comparable products and levels of upstream integration. This is typically based on a single 6-digit code within the North American

Industry Classification System (NAICS). However, industries within even a single 6-digit NAICS may differ in terms of upstream integration and downstream product mix. The approach used here means that the plants within an industry-level NAICS do not need to be further segmented by product mix or upstream integration, which is often impossible given the small number of facility data available. The next section discusses the data used in the case studies and also how difficult issues of confidentiality are managed.

## Data

Data is key to the development of these case studies. The case studies are based on non-public, plant-level data from the US Census Bureau (hereafter the Bureau) and other public and non-public sources. Data from the Manufacturing Energy Consumption Survey (MECS) and the Census of Manufacturing (CM) form is the starting point of all the case studies and the core data for many. The plant-level versions of the data are confidential and protected by Title 13 and 26 of the US Code. For these case studies, the data are used via permission from the Center for Economic Studies at the Bureau and are accessed at the Triangle Research Data Center, a secure computer lab operated by Duke University in cooperation with the Bureau as part of the Federal Statistical Research Data Center program.<sup>7</sup>

The CM provides a detailed plant-level snapshot of energy use vis-à-vis other material and labor inputs as well as the level and mix of a variety of products for *every plant in the USA* in the respective CM quinquennial years.<sup>8</sup> MECS provides even more detail on energy use (types of fuels etc.) for a subsample of plants in the MECS sample years. The Bureau collects these data to publish aggregate economic statistics but also provides access to researchers on a project-by-project basis. These data from the Bureau form the basis for the underlying statistical analysis in the case studies. The Bureau provides strict oversight of confidentiality of the data. No results from the cases studies can be used to reidentify any plant or company. This is a requirement for permission to use the data and an advantage of these case studies, ameliorating industry concerns regarding potentially sensitive information. Similar establishment

level energy and production data in the UK were used by Bloom et al. (2010) and in Germany by Falck (2007), but the latter study is not focused on energy per se.<sup>9</sup>

In some case studies, in consultation with energy managers in companies with operations in these case study industries, data is *voluntarily provided* to either supplement or supplant the data from the Bureau. In those case studies, Duke University provides each company legal non-disclosure protection for the information used in the analysis. Statistical protection methods, similar to those required by the Bureau, are also used to prevent reidentification of any plant's or company's data. The fact that these case studies use statistical methodology is what provides this confidentiality protection. The next section describes the general statistical methods that are used.

## Methodology

The methodology underlying these case studies can be motivated by the concept that there is some reduced form relationship between plant-level energy use and the various plant input and output characteristics described above. We assume that this relationship can be approximated by a functional form that is amenable to statistical estimation using data from a cross-section or panel of plants within some “reasonably defined” industry group. Depending on the form of the statistical model, discussed in more detail below, the actual plant energy use can then be compared to the predicted average or best practice energy use, given the plant's characteristics. How far the actual energy use is above or below the predicted average, or how far the actual energy use is from predicted best practice, is the plant's measure of efficiency. In statistical terms, the difference between actual and predicted energy use is equal to the residual of the statistical model for plants that are in the sample; alternatively, this difference is an out-of-sample prediction when the statistical model is applied to other data. It is in this out-of-sample context that we expect the model to be most often used, i.e., to compute energy efficiency using data for plant-level operations that were not in the statistical analysis, possibly from a different year. If that is the case, then the model is measuring current performance to a prior “benchmark year.” If we

<sup>7</sup> For more details see <http://www.census.gov/fsrdc>

<sup>8</sup> Conducted in years ending in “2” and “7”

<sup>9</sup> Other countries may have programs similar to the FSRDC program in the USA, but the author is not aware of them.

further assume that the estimated distribution of efficiency from the statistical model, i.e., the error variance, is static, then the out-of-sample prediction of efficiency can be converted to a percentile (ranking) of efficiency based on the estimated distribution. This is similar to guidance from ISO 50001 regarding the creation of EnPI.<sup>10</sup> Other application of this approach includes estimation of energy savings postretrofit (Kelly Kissock and Eger 2008) but using time series data for a single plant.

The interpretation of the comparison of actual and predicted values as a measure of efficiency is conditional on the assumption that the function used in the analysis is a reasonable approximation of the plant-level determinants of energy use and that the characteristics included in the analysis are reasonably appropriate and complete. These are admittedly strong assumptions. An alternative interpretation is that the difference between the actual energy use and the predictions of the statistical model are an *indicator of energy performance* that *normalizes*, i.e., *controls for*, multiple plant characteristics in a more sophisticated manner than a simple SEC, i.e., energy per unit of product.

The underlying approach of these case studies is a regression model of the general form

$$E = f(Y, X; \beta) + \epsilon \quad (1)$$

where  $E$  is the measure of *total source energy* (total Btu of fuels and electricity use converted to Btu based on average US thermal plant efficiency including line losses),  $Y$  is either production or a vector of production-related activities,  $X$  is a vector of plant characteristics,  $\beta$  is a parameter vector (the normalization factors), and  $\epsilon$  is the measure of relative plant efficiency. This is similar to studies of total factor productivity where the left hand side of Eq. (1) is production rather than energy and the right hand side of Eq. (1) is a production function, then  $\epsilon$  is the Solow residual of total factor productivity. Syverson (2011) provides a review of this productivity dispersion literature. Boyd (2008) provides some of the theoretical connections between the function  $f(\cdot)$ , the *energy factor requirements function*, and the subvector distance function (see Murillo-Zamorano (2004) for a review of the distance function approaches). The error term is interpreted as a measure

of *relative energy efficiency*, since it is relative to the data used to estimate the function in (1). The estimates of the distribution of  $\epsilon$  are used to summarize the range of energy performance within a sector.

The case studies can be characterized by whether (i) the functional form of the analysis is linear (in either energy use or energy intensity) or log linear and (ii) the assumption about the distribution of the error term,  $\epsilon$ . One variation in the case studies is one imposing the relationship be homogeneous of degree one in a single production variable,  $y$ .

$$E = y(f(X; \beta) + \epsilon), \quad (2)$$

or rearranging to an intensity form

$$E/y = f(X; \beta) + \epsilon \quad (3)$$

This approach ignores many issues about the assumption of the distribution of the error term but simply posits the regression model in the intensity form. When the function  $f(\cdot)$  is linear in  $X$ , then this represents the *linear functional form*. This approach may be appropriate if the activities represented by  $X$  are additively independent, e.g., in corn refining the moisture content of one of the primary byproducts influences the energy use but has no impact on the energy use of other products.

An alternative approach is the *log-log functional form*, where  $E$  and  $Y$  are in natural logarithms.

$$\ln(E) = (f(\ln(Y), X; \beta) + \epsilon) \quad (4)$$

$$\ln(E) = a + \sum_{i=1}^n b_i \ln(y_i) + \sum_{i=1}^m c_i X_i + \epsilon \quad (5)$$

In either case of a linear or log-linear functional form, standard measure of statistical significance provide test for whether or not to include a particular characteristic. In other words, one can test if two products have similar energy intensity or if weather influences energy use and a statistically identifiable way. In the log-linear form, the error term is interpreted as a percent efficiency, rather than absolute levels or intensities as in the linear approach. Whether this assumption of linearity is one of convenience or not, there are only two case studies that use the linear approach; all others use log linear. One advantage of the log linear approach is that it provides an easy estimate of energy returns to scale. In other words, the sum of the coefficients,  $b$ , of the activity vector,  $Y$ , is a measure of whether energy use scales

<sup>10</sup> Both Energy Star and ISO 50001 use the term Energy Performance Indicator. Since Energy Star began publically using the term first, ISO adopted the acronym “EnPI” to limit confusion.

proportionally to total activity. If the sum of the coefficient is close to unity, then larger plants do not have an “advantage” in terms of lower energy use. If the sum is less than unity, then larger plants use less than a proportional amount of energy. If capacity is included in the  $Y$  vector, then the case study provides both a short and long run estimate of returns to scale.

The distributional assumptions for the error term can reflect whether the efficiency in the industry is approximately (log) normal or follows a skewed behavior. All the case studies consider the possibility that efficiency is skewed and test whether a stochastic frontier (SF) energy model is appropriate (Boyd 2008). The SF approach assumes that the error term is composed of two parts

$$\varepsilon = u - v \tag{6}$$

where

$$v \sim N(0, \sigma_v^2) \tag{7}$$

and energy (in)efficiency,  $u$ , is distributed according to some one-sided statistical distribution,<sup>11</sup> for example, gamma, exponential, half normal, and truncated normal. It is then possible to estimate the parameters of Eq. (2), along with the distribution parameters of  $u$  and  $v$  using maximum likelihood methods. The approach that is used to estimate these parameters depends on the type of distribution that is used to represent inefficiency. Exponential and truncated normal frontier models can be estimated using relatively conventional maximum likelihood (ML) techniques available in many modern statistical packages. Gamma is a very flexible distribution, but also generates a model that is very difficult to estimate since there is no closed form for the likelihood function, and a simulated maximum likelihood has been used. 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 treatments.<sup>12</sup> If there is no empirical evidence of skewness of the error term, which can be tested via a likelihood ratio test, then

<sup>11</sup> We also assume that the two types of errors are uncorrelated, i.e.,  $\sigma_{u,v} = 0$ .

<sup>12</sup> No case studies of these panel treatments here since they only have, at most, a few years of data for any given plant in an unbalanced panel.

the ML estimate of  $\sigma_v^2$  will be close to zero and the estimates are equivalent to ordinary least squares (OLS) regression. The SF approach is also used by Lundgren et al. (2016) to measure energy efficiency in the Swedish pulp and paper industry, by Filippini and Hunt (2012) on US residential energy data and by Filippini and Hunt (2011) on aggregate OECD country energy data.

Since the distribution of the error term, composed or normal, is taken as a measure of efficiency dispersion, then heteroscedasticity takes on a specific interpretation beyond the usual concerns for parameter standard errors. If the heteroscedasticity is related linearly to some variable in the  $X$  vector of the form

$$\sigma_{i,v} = \gamma X_{k,i} \tag{8}$$

or to production,  $Y$

$$\sigma_{i,v} = \gamma Y_i \tag{9}$$

or to the inverse of production

$$\sigma_{i,v} = \gamma \frac{1}{Y_i} \tag{10}$$

then the efficiency distribution and the associated quartiles depends on the size of  $X$  or  $Y$ . In the last example, the distribution of efficiency is wider for low production plants than for high production plants.

Assuming we are using a model that has been estimated in one of the case studies in the out-of-sample context described above and we have data for a plant  $i$  in year  $t$  different from the study data year, we can estimate the difference between the actual energy use and the predicted average energy use from Eq. (1).

$$\ln(E_{t,i}) - f(\cdot; \beta) = \hat{\varepsilon}_{t,i} \tag{11}$$

For the models using OLS,<sup>13</sup> we have also estimated the variance of the error term of Eq. (1) and we can compute the probability that the difference between actual energy use and predicted average energy use is no greater than this computed difference under the assumption that the error,  $\varepsilon$ , is normally distributed with zero mean and variance  $\sigma^2$ .

<sup>13</sup> Similar formulas are used when the model is a frontier, see Boyd and Delgado (2012)

$$\Pr(\varepsilon \leq \hat{\varepsilon}_{t,i}) \quad (12)$$

This probability is the *energy performance score* (EPS) and is the same as a *percentile ranking of the energy efficiency of the plant*.

## Overview of case studies

Drawing from the general approach above, Table 1 summarizes the factors that have been included in each of the industry case studies to explain difference in interplant energy use. In all case studies, the potential for differences in product mix, materials use, and weather on energy use was examined. Specific issues that were raised by industry reviewers (energy managers in various companies) were also investigated given available data. Statistical test for significance were used to include or exclude the specific variables in the case study models. The types of characteristics that were included in the analysis are grouped according to product mix, production unit measurement, inputs, size or capacity, external factors, and others.

It is clear from the diversity in Table 1 that each industry is unique in the characteristics that “matter” for energy benchmarking. Twenty of the studies use physical units for activity; of those, 15 have between 2 and 8 different product types that influence energy, while 3 industries are characterized by mix of production activities that is unique to that sector. Four studies use the product price to characterize product differences, i.e., higher value products require more energy to produce. One, motor vehicles, uses product size (wheelbase) to capture differences in energy use. Five studies include measure of plant size (and indirectly utilization), but the small number is due primarily to data limitations, i.e., available plant level capacity information. About half of the sectors include process inputs, either as a ubiquitous measure of input, e.g., corn in corn refining or scrap in mini mills, or in the form of raw vs preprocessed inputs, e.g., fresh fruit vs concentrate in juice production or virgin vs recycled fiber in paper production. The selection of inputs is based in part on data availability, but then only included when the estimated effect is of reasonable size and statistically significant. Person or operation hours are included in eight industries. In some cases, the labor hours *may* be playing a similar role to utilization, i.e., capturing non-

production activity that uses energy. Climate, in the form of heating degree days (HDDs) or cooling degree days (CDDs), were found to influence energy use in eight of the studies. Finally, six of the case studies include industry-specific issues that impact energy use, such as the use of onsite freezers for storage or production; if there is an onsite water treatment plant; the moisture content of one of the products; and the use of bleaching chemicals to account for the “whiteness” of the paper that is produced.

Table 2 further describes the statistical and functional form of the models; the years, number of observations (plants), and data sources; and results of economies of scale and the range of performance in the industry. Seven sectors exhibit a skewed distribution of energy intensity and are modeled as stochastic frontiers; the rest are estimated as log normal OLS, i.e., the percentage difference from average performance are “bell shaped.” Standard statistical tests (likelihood ratio) are used to decide if the frontier or ordinary least squares are appropriate. Two of the log normal models exhibit heteroscedastic distributions with variance declining with respect to production volumes.

The earliest year of data for a study is 2002. This is largely driven by the data available when the analysis was conducted. That of the year 2007 is the most recently available data from Economic Census.<sup>14</sup> Sectors that use industry or trade association-provided data tend to have more recent benchmark years. For the less energy-intensive industries using CM data, the energy content of the fuels is imputed using cost data and state-level energy prices. This is done since the sample sizes in the MECS are too small to meet disclosure requirements to protect confidentiality. For industries with sufficiently large MECS samples, the more detailed energy information is used directly. Sample sizes vary depending on the industry; some data is dropped due to missing variables from incomplete reporting or other data quality screens such as for extreme outliers. However, the sample sizes shown should be viewed as a fairly complete count of all the plants in that subsector. Even though the CM data is a *census*, i.e., *includes all the plant in an industry*, depending on the industry, the sample size can be quite small. The adequacy of the sample size needs to be considered against the complexity of the model, i.e., how many variables are included in

<sup>14</sup> As of this draft, the 2012 EC was not yet available in the RDC network

**Table 1** ES-EPI benchmarks inputs, by industry and subsector

Focus industries	Product mix	Units	Inputs	Size or capacity	Climate	Other
Cement	3 product types	Tons	–	Capacity and no. of Kilns	–	Person hours
Corn refining	5 product types	Bushels	Corn	Capacity	–	Feed moisture
Dairy—fluid milk <sup>a</sup>	6 product types	Gallons	Whole milk	–	CDD	Person hours
Dairy—ice cream <sup>a</sup>	4 product types	Gallons	2 types	–	CDD	Person hours
Ethyl alcohol <sup>b</sup>	Single	Gallons	–	–	–	–
Food—juice (canned)	8 product types	Gallons	2 types	–	–	–
Food—frozen fried potatoes	Single	Pounds	–	–	–	Warehouse (frozen)
Food—tomato products <sup>b</sup>	2 product types	Pounds	2 types	–	–	Person hours
Baking—cookies and crackers	3 product types	Pounds	–	–	–	–
Baking—bread and rolls	5 product types	lbs	Raw dough	–	HDD, CDD	Freezers
Glass—flat	Single	Pounds	Sand	–	–	–
Glass—container	Single—price	Pounds	Sand, Cullet	–	–	–
Iron and steel—integrated <sup>a</sup>	4 stages <sup>c</sup>	Tons	–	Furnace capacity	–	–
Iron and Steel—mini mills <sup>a</sup>	Single—price	Tons	Scrap	Furnace capacity	–	–
Metal casting—iron <sup>a</sup>	4 product types—price	Tons	–	–	HDD	Person hours
Metal casting—investment steel <sup>a</sup>	Single	\$ Value	–	–	–	Person hours
Metal casting—“other” steel <sup>a</sup>	3 product types	\$ Value	–	–	–	–
Metal casting—aluminum <sup>a</sup>	3 product types—price	Tons	–	–	HDD	–
Motor vehicle—assembly	Vehicle size	Number of Vehicles	–	Production capacity	HDD, CDD	Air tempering
Pharmaceuticals	3 activity types <sup>d</sup>	Share of Floor Space	–	Facility size (ft2)	HDD, CDD	Operation hours
Printing—lithograph <sup>a</sup>	6 product types	\$ Value	6 types	–	HDD, CDD	–
Pulp mills	3 product types	Tons	2 types	–	–	Water treatment
Paper and board—integrated mills	3 product types	Tons	3 types	–	–	Water treatment, bleaching chemicals
Ready mix concrete <sup>a</sup>	2 activities	Tons, Miles	–	–	–	–

HDD heating degree days, CDD cooling degree days

<sup>a</sup> Under industry review

<sup>b</sup> Preliminary

<sup>c</sup> Liquid iron, casting, hot mill, and finishing

<sup>d</sup> R&D, chemical synthesis, Fill and finish

the analysis. The degrees of freedom, difference between the number of observation, and the number of coefficients in the model impact the statistical tests used to include or exclude variables in the model. The smallest number of plants is for integrated steel, but we have multiple years of data and this increases degrees of freedom. The next smallest is Frozen Fried Potatoes (27); that model only includes two simple variables, both of which meet the statistical significance criteria for the available degrees of freedom. We feel the data are adequate of these purposes and typically are for all plants in the USA that completely report data to the Bureau.

Estimates of elasticity of returns to scale (RTS) give insight into whether larger plants have an advantage over small ones in terms of energy. The elasticity estimate is how much energy would increase given a 1 % increase in scale of operations. The elasticity is computed as the sum of the coefficient estimates of the log activity variables, RTS estimates can be further broken down into short run (SR), where only production activities are considered, and long run (LR) when increases in production and plant capacity are included. In some cases, no single RTS measure can be computed from the functional form and is labeled *variable*. In other cases, the RTS is not estimate in the analysis but assumed to be *constant*. The studies that use linear forms impose constant returns in the long run and allow for variable returns in the short run. Some studies use a variable returns specification by including second-order terms in the activity variables or a measure of capacity utilization, i.e., production/capacity, in the analysis. Values close to one indicate that no size advantage is found. This is the case for eight studies. For ten studies, size is found to be important. For example, the RTS estimate for Baking—Cookies and Crackers is 0.71 which implies that a 10 % increase in production would result, on average, in only a 7.1 % increase in energy use.

The last column labeled 75 to 50th represents the third quartile range, i.e., percent difference of the 75th percentile, i.e., the ENERGY STAR certified plant level, and the average or median performance, the 50th percentile. This ranges from as low as 4.5 % to nearly 44 %. Figure 2 compares this third quartile range to the industry average share of energy cost to value added. This cost share reflects how “important” energy is in the sector. We see a clear correlation between high cost shares and the range of performance. This makes sense since industries with higher relative energy costs would

put more effort into management of those costs. There are outliers in this relationship, however. They include pulp mills and ethanol (dry mill) plants. The result for these may suggest the need for additional scrutiny. However, the study for pulp mills measures the efficiency of net purchased energy and pulp mills provide a large amount of internally generated power from black liquor and CHP. There may actually be a wide range of practices in terms of net purchased energy in this sector than for other energy-intensive ones.

### Updates for benchmark year for three ES-EPI

As is seen above, the analysis is typically done for a particular year. When a group of years is used, the analysis measures the energy efficiency change for the final year relative to the earlier year. Since the case studies are typically used in an out-of-sample context, we call this the benchmark or reference year. In the 2010, three previously completed case studies, auto assembly, cement, and corn refining (see (Boyd 2005a, b, Boyd 2006, Boyd 2008) for detailed descriptions of the earlier models), began to be updated. Comparing the old benchmark with the new benchmark reveals information about how these three, very different industries have changed over time. Since the analysis reveals both the general level and range of energy performance, the comparison focuses on how much the change in the “best practice” and the change in the range of performance contribute to the overall reduction in energy use in the sector (see Boyd and Zhang (2012), Boyd and Delgado (2012), and Boyd (2010) for the details of the updates).

For the cement industry, if one computes the ratio of total energy costs to total value of shipments (adjusted for inflation) in 1997 and 2007 from data collected in the Economic Census, one would conclude that this measure of energy intensity has fallen ~16 %, from 0.184 to 0.158. Aggregate data may also give the impression that all plants have made the same steady improvements. Figure 3 shows the energy use per short ton of clinker against the energy performance score (EPS) defined above in Eq. (12). The picture that emerges from our plant-level statistical analysis is somewhat different and more subtle; poorer-performing plants from the late 1990s have made efficiency gains, reducing the gap between themselves and the top performers, whom have changed only slightly. The results from this study focus

**Table 2** ES-EPI benchmarks model details, by industry and subsector

Focus industries	Model	Year	No. of plants	Data source	Returns to Scale	75th to 50th
Cement	Log normal (heteroscedastic OLS)	2000–2008	96	Industry	Variable SR 0.92 LR	-6.1 %
Corn refining	half normal frontier	2004–2009	37	Industry	Variable SR constant LR	-14.5 %
Dairy—fluid milk <sup>a</sup>	Log normal OLS	2002	258	CM	0.85	-29.0 %
Dairy—ice cream <sup>a</sup>	Log normal OLS	2002	89	CM	1.05	-23.9 %
Ethyl alcohol <sup>b</sup>	Log normal OLS	2007	111	CM	0.70	-35.4
Food—juice (canned)	Log normal OLS	2002	44	CM	0.84	-41.8 %
Food—frozen fried potatoes	Log normal OLS	2002	27	CM	0.91	-16.0 %
Food—tomato products <sup>b</sup>	Log normal OLS	2002	40	CM	1.11	-43.7 %
Baking - cookies and crackers	Log normal OLS	2002	64	CM	0.71	-30.9 %
Baking—bread and rolls	Log normal OLS	2010–2012	135	Industry	0.78	-15.8 %
Glass—flat	Log half normal frontier	2002	38	CM, MECS	Variable	-16.3 %
Glass—container	Log normal OLS	2002	62	CM, MECS	1.03	-11.6 %
Iron and steel—integrated <sup>a</sup>	Log exponential frontier	2005–2012	14	Industry	0.72 SR 0.99 LR	-4.5 %
Iron and steel—mini mills <sup>a</sup>	Log normal OLS	2002	39	CM, MECS		-12.1 %
Metal casting—iron <sup>a</sup>	Log normal OLS	2006	83	CM, MECS	1.06	-23.2 %
Metal casting—investment steel <sup>a</sup>	Log half normal frontier	2007	51	CM	1.03	-32.8 %
Metal casting—“other” steel <sup>a</sup>	Log normal OLS	2007	59	CM	Variable	-25.8 %
Metal casting—aluminum <sup>a</sup>	Log normal (heteroscedastic OLS)	2007	290	CM, MECS	1.05	-28.3 %
Motor vehicle—assembly	Gamma frontier	2003–2005	33	Industry	Variable SR Constant LR	-21.4 %
Pharmaceuticals	Log half normal frontier	2004–2006	61	Industry	Variable SR 0.98 LR	-30.1 %
Printing—lithograph <sup>a</sup>	Log half normal frontier	2007	775	CM	1.0	-35.0 %
Pulp mills	Log normal OLS	2002	28	CM, MECS	1.05	-36.1 %
Paper and board integrated mills	Log normal OLS	2002	99	CM, MECS	0.71	-19.5 %
Ready mix concrete <sup>a</sup>	Log normal	2008–2009	62	Industry	0.83 SR 0.89 LR	-35.5 %

CM Census of Manufacturing, MECS Manufacturing Energy Consumption Survey, SR short run, LR long run, OLS Ordinary Least Squares

<sup>a</sup> Under Industry Review

<sup>b</sup> Preliminary

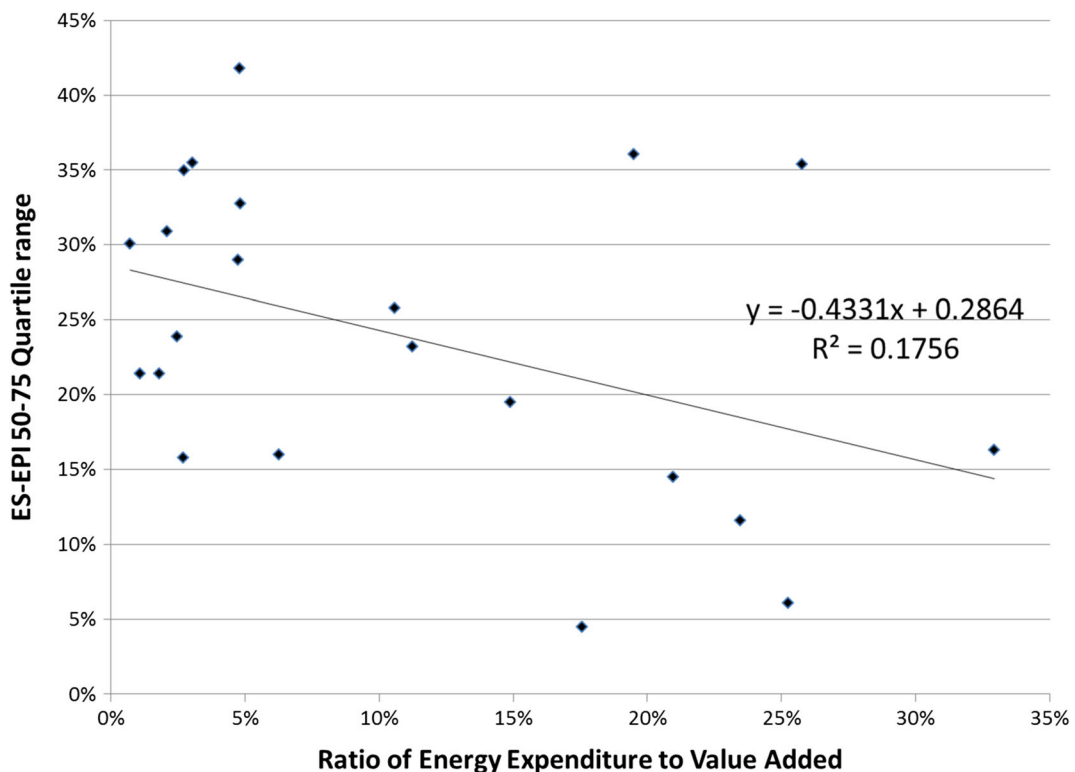


on energy efficiency and controls for other structural changes in the industry, e.g., increases in average plant size, which also tend to lower energy use. Our estimate of the overall energy efficiency improvement in the 96 plants in our database represents a 13 % change in total source energy, and the source of these changes is clearly not uniform. The case study does not have the details needed to identify the exact sources of this shift but is consistent with the shift away from wet kilns to dry, the addition of precalcine and preheater stages, and in some cases, complete facility retrofits. Dutrow et al. (2009) describe a retrofit by the Salt River Materials Group cement plant in Clarksdale, Arizona plant that reduced various process energy use by 37–67 %. There have not been any major cement technology “breakthroughs,” so the lack of shift at the highest levels of performance should not be surprising.

Results for the auto assembly industry are similar, but less dramatic (Fig. 4). There are two sources of improvement, the changes in the industry energy frontier, i.e., “Best Practices” and technology, and the changes in efficiency, i.e., whether plants are catching up or falling behind. The results suggest that slightly more than half

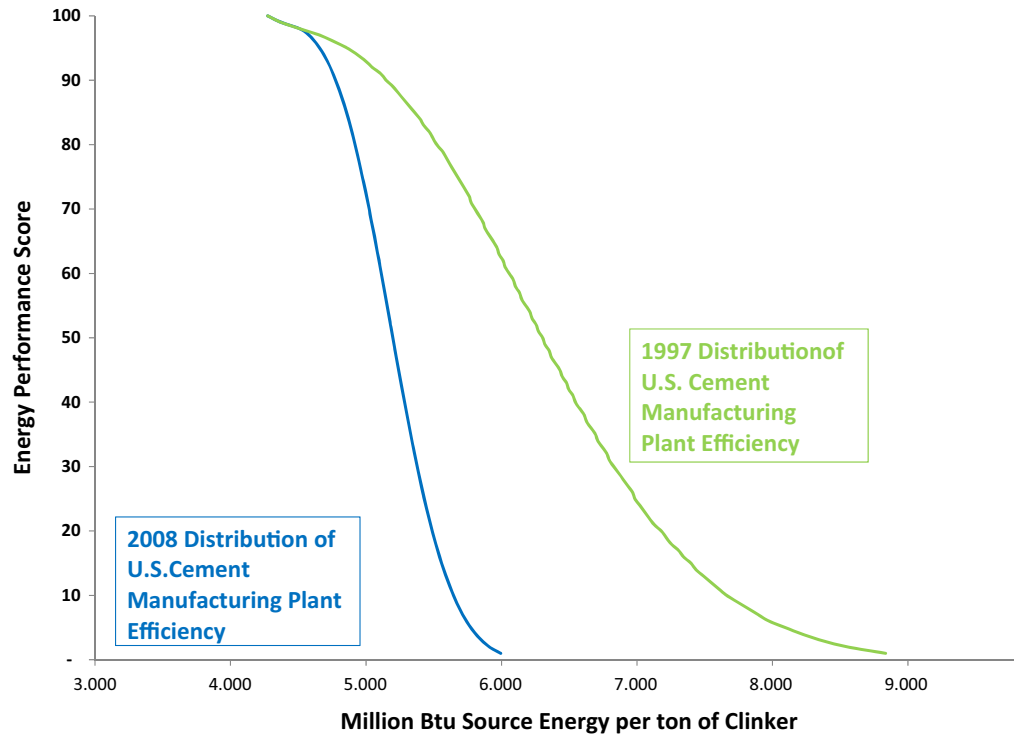
of the improvement is changes in efficiency, which have slightly outpaced changes in the frontier. The combined effect when evaluated against the over 7 million vehicles produced in 2005 by the plants in our study implies in a reduction of 11.6 % or 1462 million lbs of CO<sub>2</sub>, attributable to changes in observed industry energy efficiency practices. No single technological source of efficiency improvements can be pointed to in auto assembly, but the update analysis does imply that reductions in fuel use dominated the shift.

The change in the distribution of energy efficiency for a representative corn refining plant is shown in Fig. 5. If we multiply this plant-specific change in energy intensity by the level of corn input production for each plant operating in the industry in 2009, and total across all plants, we compute a reduction of 6.7 trillion Btu in annual energy use. Relative to an average annual total source energy consumption of 155 trillion Btu in 2009 for all the plants in our data set, this represents about a 4.3 % reduction in overall energy use by this industry. When energy-related greenhouse gas emissions are considered, this represents an annual reduction of 470 million kg of energy-related CO<sub>2</sub> equivalent



**Fig. 2** Relationship between industry average energy cost to value added and the estimated 50–75 quartile range

**Fig. 3** Comparison of two benchmark distributions of energy efficiency in cement manufacturing (source: Boyd and Zhang 2012)

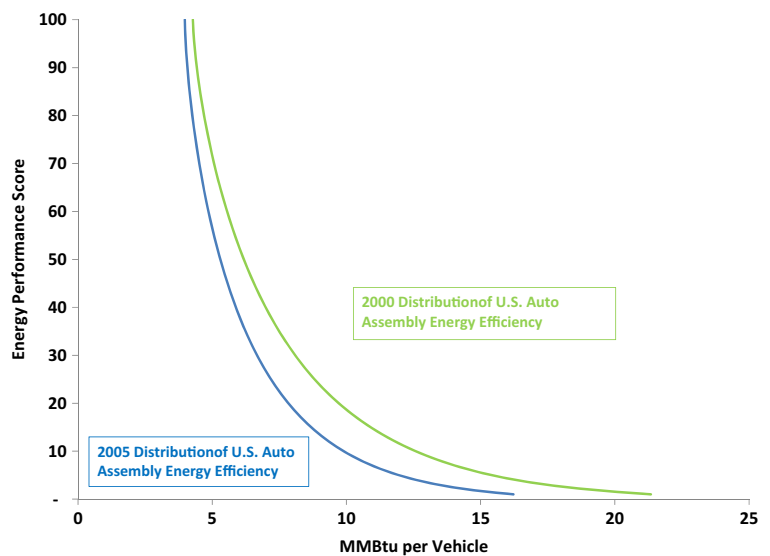


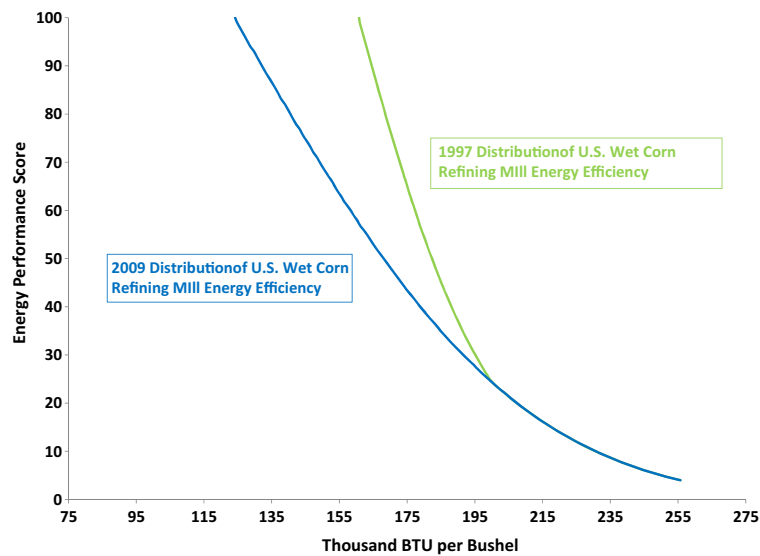
emissions from improved energy efficiency. The case study update was not able to point to specific technological or operational changes that resulted in this shift, so more work in this sector is needed.

The change in performance from these three industries is all quite different. Cement reflects the case where

best practice has changed very little, but “catching up” comprises the main source of improvements. Corn refining is at the opposite end of this spectrum, where there are substantial changes in the best plants, but laggards remain or in some sense are even falling behind by failing to keep pace. The auto assembly plants are a

**Fig. 4** Comparison of two benchmark distributions of energy efficiency in auto assembly (source: Boyd 2010)





**Fig. 5** Comparison of two benchmark distributions of energy efficiency in wet corn refining (source: Boyd and Delgado 2012)

mixture of changes in best practice and some modest “catching up.” These benchmark updates also reflect different time periods. When we compute the average annual change from the total reduction in energy use for each sector, we see that the auto industry has made the greatest strides (see Table 3).

## Conclusions

The objective of these case studies is for developing sector-specific energy performance benchmarks and to create a tool that would motivate companies to take actions to improve the energy efficiency of their plants and ultimately help reduce greenhouse gas emissions in the industrial sectors benchmarked. These case studies accomplish this goal in three major ways. This first is by developing a statistical approach that effectively accounts, or normalizes, for differences in product mix, upstream integration, and external factors such as

climate, effectively expanding the concept for efficiency beyond a single measure of intensity. The side effect of the statistical approach also that it is highly effective in handling complex boundary issues. The third is to provide this information to the public, including industry, in a form that utilizes the most detailed data possible, while maintaining strict confidentiality. These case studies address each industry-specific energy issues within a common framework, recognizing that the variance of simple energy intensity would overstate the range of efficiency. Meta-analysis of these case studies reveal some additional insights, in particular that all industries exhibit unique drivers to the energy underlying the range of energy use, but that the range of performance is inversely correlated to the industry level of energy use. In other words, industries that must use more energy to make their product appear to “pay more attention” to energy use; competitive pressures likely push them closer together in terms of energy use. Boyd and Curtis (2014) find similar results when examining how management practices influence energy use. In the future, these tools can be used to explore more deeply how efficiency changes over time.

The approach used in the studies is not the only method to measure energy efficiency; other uses have taken different approaches to meet different objectives. The approach used here focused on providing useful information to energy managers and policy makers about the potential for improvement and a tool to motivate that change. One way Energy Star uses to motivate

**Table 3** ES-EPI benchmarks updates: rate of change by industry

Sector	New benchmark Year	Old Benchmark Year	Time period	Total reduction	Average annual change
Auto	2005	2000	5	12.0 %	2.3 %
Cement	2008	1997	11	13.0 %	1.2 %
Com	2009	1997	12	4.3 %	0.4 %

change is to recognize top performance. As of December 2015, EPA had published 12 EPIs, awarded 141 ENERGY STAR plant certifications, and engaged an additional 12 industrial sectors and subsectors in the EPI development process (see Table 4). Compared to average plants (EPS score of 50) EPA estimated in 2015 that plants earning the ENERGY STAR have saved an estimated 544 trillion Btus (US Environmental Protection Agency 2015). Companies using these tools report that they find them valuable and beneficial for evaluating current performance and setting efficiency goals. Many companies report they have incorporated these tools into their energy management programs and have made achieving ENERGY STAR certification as an objective.

The case studies are not done in an academic vacuum; all are conducted with consultation from industry energy managers and sometimes industry-provided data. Initially, there was industry skepticism that a whole-

plant benchmark could be developed using statistical case studies. Skeptics largely believed that each plant is too “unique” for whole plant comparisons to be made. However, both the process and method used to develop the case studies has helped change skeptics participating in the industrial focus process into supporters. The *process* of engaging the industry in the development of the case studies has been critical. By developing the case studies in a transparent, objective, and collaborative process, industry participants were directly involved in the design and review from the beginning. This process enabled the identification of potential factors for inclusion in the regression analysis, receive timely feedback on draft results, quickly address concerns, and ultimately ensure a high degree of support and “buy-in” for the tool. By using a benchmarking method based on actual operational data and that allowed for controls to address industry-specific differences between plants, concerns were overcome that industrial plants are too heterogeneous, even within a specific subsector, to be able to benchmark.

The availability of sector-wide energy and production data through the US Census Bureau was critical for the analysis. One of the greatest barriers to any benchmarking exercise is inadequate or unrepresentative data. The case studies has benefited from the robust industrial energy and production data collected by the US Census through the Census of Manufacturing (CM) and the Manufacturing Energy Consumption Survey (MECS). The availability of this data for use in developing the statistical models has been critical to ensuring the early success of the ENERGY STAR industrial benchmarking program. First, it provided EPA with the ability to develop the benchmarks without having to undertake a data collection. Second, by working with Census data, which has strict confidentiality requirements, the ENERGY STAR team was able to build trust among industry participants that the company-specific data used for benchmarking would be kept confidential and would not be shared with either focus participants and the EPA. While some of the more recent case studies have drawn on data provided by the industry, the availability and quality of the CM and MECS data enabled ENERGY STAR to successfully develop the first case studies and demonstrate that whole-plant energy performance benchmarking is possible.

The process of developing EPIs has uncovered new insights into energy use and the drivers of efficiency within the sectors benchmarked. Additionally, the establishment of industry baselines has enabled EPA to visualize the range of performance within a sector.

**Table 4** Status of case studies (published or under review)

Focus industries	Status
Cement	Boyd and Zhang (2012)
Com refining	Boyd and Delgado (2012)
Dairy—fluid Milk	Under industry review
Dairy—ice cream	Under industry review
Ethyl alcohol	Under industry review
Food—juice (canned)	Boyd (2011)
Food—frozen fried potatoes	Boyd (2011)
Food—tomato products	Under industry review
Baking—cookies and crackers	Boyd (2011)
Baking—bread and rolls	Under EPA final review
Glass—flat	Boyd (2009a, b)
Glass—container	Boyd (2009a, b)
Iron and Steel—integrated	Under industry review
Iron and Steel—mini mills	Under industry review
Metal casting—iron	Under EPA final review
Metal casting—investment steel	Under industry review
Metal casting—“other” steel	Under industry review
Metal casting—aluminum	Under EPA final review
Motor Vehicle—assembly	Boyd (2014)
Pharmaceuticals	Boyd (2009a, b)
Printing—lithograph	Under industry review
Pulp mills	Boyd and Guo (2012)
Paper and board integrated mills	Boyd and Guo (2014)
Ready mix concrete	Under industry review

Visualizing the distribution of performance offers important information for policy makers and others interested in promoting efficiency or reducing GHG emissions from specific industrial sectors. The slope of the baseline curve generated by the EPI can help policy makers and others evaluate what action is needed to improve the performance of the industry. For example, sectors with steep baseline curves and distributions indicate that the opportunities for improving energy efficiency through existing measures may be limited. These sectors should be considered for R&D investments to develop new technology that can create a step change in the level of performance. Additionally, these sectors may face greater difficulties reducing their GHG emissions through existing energy management measures. On the other hand, sectors with flatter curves indicate that more opportunities are available through existing technologies and practices. In these sectors, there is a greater distribution of performance, which usually suggests that existing energy management measures and investments can improve performance.

The process of benchmarking and rebenchmarking a sector provides further insights into the improvement potential of the industry over time. Understanding how the distribution of energy performance in a sector is changing or not changing can provide valuable information for policy makers as well as business leaders in developing strategies to drive future performance gains.

The approach and method used by ENERGY STAR to benchmark whole-plant energy performance has potential applicability to other sustainability metrics, such as water and waste, as well as subsystems within plants. While developing such benchmarks is beyond the scope of the ENERGY STAR program, several companies participating in the Industrial Focus process have recently initiated an independent effort that applies the ENERGY STAR benchmarking approach to process lines within the plant and to non-energy measures such as water. If successful, the results of this effort will break new ground in advancing the field of energy performance and sustainability benchmarking.

**Acknowledgments** This paper was developed at Duke University with funding from the US Environmental Protection Agency's Office of Atmospheric Programs, ENERGY STAR for Industry. The paper and associated analysis would not have been possible without the input of all the industry participants in the ENERGY STAR Focuses on Energy Efficiency. Their willingness to provide data, guidance on important issues affecting manufacturing energy use, and time and energy in reviewing the case studies was

invaluable. I would also like to thank Walt Tunnessen, Betsy Dutrow, Josh Smith, and the reviewers and participants of the ACEEE *Summer Study on Energy Efficiency in Industry*, July 2013, Niagara Falls, NY for their comments on an earlier version of this paper "Plant Energy Benchmarking: A 10-Year Retrospective of the ENERGY STAR Energy Performance Indicators (ES-EPI)". Some results in this paper were prepared while the author was a special sworn status research associate at the Triangle Census Research Data Center. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the US Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Any errors are the sole responsibility of the author. Many research assistants have been involved in the preparation of these case studies. While some are cited as coauthors in the reference section, other contributors not so cited including Jeremy Chiu, Matt Doolin, Chris Geissler, Songman Kang, Tatyana Kuzmenko, Béla Személy, Shouyue Yu, and Su Zhang.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Aarons, K., Golding, C., & Clark S. (2013). Key considerations for industrial benchmarking in theory and practice
- Azadeh, A., Amalnick, M. S., Ghaderi, S. F., & Asadzadeh, S. M. (2007). An integrated DEA PCA numerical taxonomy approach for energy efficiency assessment and consumption optimization in energy intensive manufacturing sectors. *Energy Policy*, 35(7), 3792–3806.
- Battese, G. E., & Rao, D. S. P. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business and Economics*, 1(2), 87–93.
- Bloom, N., Genakos, C., Martin, R., & Sadun, R. (2010). Modern management: good for the environment or just hot air?\*. *The Economic Journal*, 120(544), 551–572.
- Boyd, G. A. (2005a). *Development of a performance-based industrial energy efficiency indicator for automobile assembly plants*. Argonne: Argonne National Laboratory.
- Boyd, G. A. (2005b). A method for measuring the efficiency gap between average and best practice energy use: the ENERGY STAR industrial energy performance indicator. *Journal of Industrial Ecology*, 9(3), 51–65.
- Boyd, G. A. (2006). *Development of a performance-based industrial energy efficiency indicator for cement manufacturing plants*. Argonne: Argonne National Laboratory.
- Boyd, G. A. (2008). Estimating plant level manufacturing energy efficiency with stochastic frontier regression. *The Energy Journal*, 29(2), 23–44.
- Boyd, G. (2009a). *Development of a performance-based industrial energy efficiency indicator for glass manufacturing plants* (p. 26). Washington: US EPA.

- Boyd, G. A. (2009b). *Development of a performance-based industrial energy efficiency indicator for pharmaceutical manufacturing plants*. Durham: Duke University. report to the Environmental Protection Agency.
- Boyd, G.A. (2010). *Assessing Improvement in the energy efficiency of U.S. auto assembly plants*. Duke Environmental Economics Working Paper Series, Nicholas Institute for Environmental Policy Solutions.
- Boyd, G. (2011). *Development of performance-based industrial energy efficiency indicators for food processing plants*. Washington: US EPA.
- Boyd, G. (2012). *A statistical approach to plant-level energy benchmarks and baselines: the energy star manufacturing-plant energy performance indicator*. Orlando: Carbon Management Technology Conference.
- Boyd, G. A. (2014). Estimating the changes in the distribution of energy efficiency in the U.S. automobile assembly industry. *Energy Economics*, 42, 81–87.
- Boyd, G. A., & Curtis, E. M. (2014). Evidence of an “Energy-Management Gap” in U.S. manufacturing: spillovers from firm management practices to energy efficiency. *Journal of Environmental Economics and Management*, 68(3), 463–479.
- Boyd, G. A., & Delgado, C. (2012). *Measuring improvement in the energy performance of the U.S. corn refining industry*. Duke Environmental Economics Working Paper Series (pp. 1–18). Durham: Nicholas Institute For Environmental Policy Solutions.
- Boyd, G., & Guo, Y. (2012). *Development of energy star® energy performance indicators for pulp, paper, and paperboard mills* (p. 30). Washington: US EPA.
- Boyd, G.A., & Guo, Y. (2014). *An energy performance Indicator for integrated paper and paperboard mills: a new statistical model helps mills set energy efficiency targets*. Paper360.
- Boyd, G. A., & Karlson, S. H. (1993). The impact of energy prices on technology choice in the United States steel industry. *Energy Journal*, 14(2), 47–56.
- Boyd, G. A., & Tunnessen, W. (2007). *Motivating industrial energy efficiency through performance-based indicators*. White Plains: ACEEE Summer Study on Energy Efficiency in Industry Improving Industrial Competitiveness: Adapting To Volatile Energy Markets, Globalization, And Environmental Constraints.
- Boyd, G., & Zhang, G. (2012). Measuring improvement in energy efficiency of the US cement industry with the ENERGY STAR Energy Performance Indicator. *Energy Efficiency*:1–12.
- Boyd, G., Dutrow, E., & Tunnessen, W. (2008). The evolution of the ENERGY STAR® energy performance indicator for benchmarking industrial plant manufacturing energy use. *Journal of Cleaner Production*, 16(6), 709–715.
- Dutrow, E., Boyd, G., Worrell, E., & Dodendorf, L. (2009). Engaging the U.S. cement industry to improve energy performance. *Cement World*.
- Egenhofer, C. (2007). The making of the EU emissions trading scheme: status, prospects and implications for business. *European Management Journal*, 25(6), 453–463.
- Erickson, P., Lazarus, M., Adair, J., Brant, J., Levitt, E., Hermann, H., Larson, T., Ross, B., & Wheelless, A. (2010). *Issues and options for benchmarking industrial GHG emissions*. Stockholm Environment Institute.
- Falck, O. (2007). *Emergence and survival of new businesses*. Berlin and Heidelberg: Springer.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), 253–290.
- Filippini, M., & Hunt, L. C. (2011). Energy demand and energy efficiency in the OECD countries: a stochastic demand frontier approach. *The Energy Journal*, 32(2), 59–80.
- Filippini, M., & Hunt, L. C. (2012). US residential energy demand and energy efficiency: a stochastic demand frontier approach. *Energy Economics*, 34(5), 1484–1491.
- Freeman, S. L., Niefer, M. J., & Roop, J. M. (1997). Measuring industrial energy intensity: practical issues and problems. *Energy Policy*, 25(7–9), 703–714.
- Greene, W.H. (2002). *Fixed and random effects in stochastic frontier models*. New York University Economics Department Working Paper.
- Huntington, H. (1995). *Been top down so long it looks like bottom up*. Energy.
- Kelly Kisssock, J., & Eger, C. (2008). Measuring industrial energy savings. *Applied Energy*, 85(5), 347–361.
- Lundgren, T., Marklund, P.-O., & Zhang, S. (2016). Industrial energy demand and energy efficiency—evidence from Sweden. *Resource and Energy Economics*, 43, 130–152.
- May, G., Barletta, I., Stahl, B., & Taisch, M. (2015). Energy management in production: a novel method to develop key performance indicators for improving energy efficiency. *Applied Energy*, 149, 46–61.
- Murillo-Zamorano, L. (2004). Economic efficiency and frontier techniques. *Journal of Economic Surveys*, 18(1), 33–77.
- Murray, G. P. (1996). What is energy efficiency? Concepts, indicators and methodological issues. *Energy Policy*, 24(5), 377–390.
- Neelis, M., Worrell, E., Mueller, N., Angelini, T., Cremer, C., Schleich, J., & Eichhammer, W. (2009). *Developing benchmark criteria for CO2 emissions*. Netherlands: Ecofys Netherlands and The Fraunhofer Institute for Systems and Innovation research.
- Oh, S.-C., & Hildreth, A. (2014). Estimating the technical improvement of energy efficiency in the automotive industry—stochastic and deterministic frontier benchmarking approaches. *Energies*, 7(9), 6196–6222.
- Phylipsen, D., Blok, K., Worrell, E., & Beer, J. d. (2002). Benchmarking the energy efficiency of Dutch industry: an assessment of the expected effect on energy consumption and CO2 emissions. *Energy Policy*, 30(8), 663–679.
- Saygin, D., Patel, M.K., & Gielen, D.J. (2010). Global industrial energy efficiency benchmarking: an energy policy tool. *E. a. C. C. a. Research and Branches*.
- Syverson, C. (2011). What determines productivity? *Journal of Economic Literature*, 49(2), 326–365.
- Tanaka, K. (2008). Assessment of energy efficiency performance measures in industry and their application for policy. *Energy Policy*, 36(8), 2887–2902.
- U.S. Environmental Protection Agency. (2015). *Reducing greenhouse gas emissions by advancing industrial energy efficiency 2000–2015*. Washington: O. A. P. C. P. Division.
- Worrell, E., & Price, L. (2006). An integrated benchmarking and energy savings tool for the iron and steel industry. *International Journal of Green Energy*, 3(2), 117–126.