Article

The Assessment of Car Making Plants with an Integrated Stochastic Frontier Analysis Model

Seog-Chan Oh [1] and Jaemin Shin [2,*]

[1]   Manufacturing Systems Research Laboratory, General Motors R&D, Warren, MI 48090, USA;
      seog-chan.oh@gm.com
[2]   Department of Mathematical Sciences, Hanbat National University, Daejeon 34158, Korea
[*] Correspondence: jaemin.shin@hanbat.ac.kr

Abstract: As global competition has intensified in the automotive industry, there is a strong need for management teams to develop methods that allow accurate and objective assessments of plant productivity and to identify productivity improvement opportunities for the best manufacturing practices. Stochastic frontier analysis (SFA) models have been used as a statistical benchmarking tool to provide a bird's-eye view of an industrial sector. SFA models can also be adapted for plant productivity assessment. However, owing to the problem of multicollinearity, the general form of SFA is difficult to apply to the assessment of complex manufacturing systems in the automotive industry, which is characterized by many control and external factors that are intercorrelated to each other. This study proposes a method for applying SFA to vehicle manufacturing plants with a focus on gaining high accuracy in model parameter estimation, by decomposing a plant into components (i.e., shops), building an SFA model for each shop, and reintegrating the general plant system through the appropriate combination of shop-level inefficiency distributions. In particular, this study focuses on documenting the derivation of a new probability density function that integrates three different inefficiency distributions. For illustration of the proposed approach, hypothetical vehicle assembly plants are assessed as examples, where the total labor hours are split into Bodyshop, Paintshop, and General Assembly, exclusively and collectively. Finally, this study offers a solution process to clarify the reasons for underperforming plants in terms of labor productivity and identify the course of actions to cure the issues with some managerial insights emphasizing the balanced approach, incorporating people, process and technology.

Keywords: productivity assessment; benchmarking; stochastic frontier analysis

1. Introduction

In the automotive industry, productivity is a primary indicator for measuring vehicle assembly plant performance (along with quality and work-in-process inventory). Therefore, one of the overriding concerns for management teams in the industry is how to accurately assess productivity and identify productivity improvement opportunities to close the competitive gap between their actual practice and competitive best practices.

Regarding productivity assessment in the automotive industry, the idea of using hour per unit (HPU) as a performance metric is relatively well-accepted (HPU-based productivity may not be an ideal performance measure if the financial performance of manufacturing systems is available because financial performance is more directly aligned with the underlying goal of the company, that is, profit maximization. However, as financial performance is generally reported to the public at the company level, it is difficult to estimate the profitability of an individual manufacturing shop or line. Given the lack of publicly-available financial performance data, HPU as a metric of productivity is a reasonable index for assessing manufacturing performance.) and has been used in similar studies [1,2]. HPU is the total working hours spent at a plant divided by the total produced vehicle volume. The total working hours includes paid lunches, breaks, leaves and overtime.

Note that HPU is an inverse measure of productivity, that is, a greater value for HPU indicates lower productivity, and vice-versa. One distinct advantage of using HPU is that it can account for production volume increases or decreases (denominator part) and labor efficiency changes (numerator part) simultaneously.

The current practice for assessing productivity is a simple HPU-based comparison between competitive plants without consideration of different manufacturing operation conditions among them, which can cause misleading assessments in identifying the competitive productivity gap. As an alternative to potentially inaccurate comparisons, a statistical regression model can be created for manufacturing plants within a particular industrial sector, which is used to benchmark a particular plant in relation to competing plants that perform the best practice in regard to HPU. This study uses stochastic frontier analysis (SFA) as a base model to account for different operation conditions exposed by influencing factors of HPU.

Since Aigner et al. (1977) [3] and Meeusen and Broeck (1977) [4] proposed SFA models independently, SFA has been used as a statistical benchmarking tool to provide a bird's-eye view of industry or plant-level efficiency analysis. Through SFA-enabled benchmarking, inefficient distributions of productivity can be identified, from which the best theoretical level of productivity can be estimated for a given plant, thereby providing an answer to the practical question: "how competitive would my plant be compared with everyone else's, if all other plants wear the same shoes as mine?" Although the original specification of the SFA model has been maintained by preserving its error term divided into two components–one to account for random effects (or measurement errors) and the other to account for technical inefficiency–the inefficiency component has been altered or extended in several ways. The survey on alternation and extension will be discussed in Section 2. It is worth noting that Data Envelopment Analysis (DEA) [5] is also proven an extremely useful tool based on non-parametric and deterministic modeling concept in the benchmarking study. However, this study suggests to use SFA because it provides parametric properties required to integrate shop-level stochastic models. Further, SFA provides useful modeling settings to address measurement errors that are caused by missing variables like cultural difference by company, varying weather condition by region and diverse local union contract which are likely to play a significant role in labor productivity study.

A general form of SFA for HPU-based productivity benchmarking is as follows:

$$HPU_i = f(X_i, Z_i; \beta) + \epsilon_i, \quad \epsilon_i = u_i + v_i, \quad v_i \sim N(0, \sigma_v^2), \tag{1}$$

where HPU_i is hour per unit of plant i, which is the fraction of total labor hours over total vehicle production; X_i and Z_i are control and external variables (Control variables typically consist of the variables that a plant can manipulate by their tactical strategies. External variables are factors lying outside the decision scope of a plant, such as a new vehicle model allocated by headquarters as per the company's long-term manufacturing capacity planning. Examples of control variables include automation levels in the Bodyshop and Paintshop, in/outsourcing levels in the Bodyshop and General Assembly (GA). Examples of external variables are the number of models (or market segments) that the plant produces, frame styles (unibody vs body-on-frame), welding contents in the Bodyshop, or total parts attached in GA. The choice of variables to include (or exclude) for the model is determined by subject matter experts' priory knowledge and expectations about what factors will have significant influence on productivity and statistical causality analysis results.), respectively; β includes the parameters to be estimated; u_i is the inefficiency or competitiveness gap in HPU use, which is distributed according to one-sided distribution; and v_i denotes the measurement error. Note that the sign of u_i is plus because the actual HPU_i will be higher than the "best practice" plant. The role of the function $f(X_i, Z_i; \beta)$ is to normalize for different operation conditions by the plant, converting an "apples-to-oranges" comparison

into an "apple-to-apple" comparison. Given a set of plant data, the distribution of the competitiveness gap (u_i) can be estimated as follows:

$$u_i = HPU_i - f(X_i, Z_i; \beta) + v_i. \tag{2}$$

From the distribution of u_i, the percentile ranking of i-plant's productivity can be calculated as

$$Pr(u \geq u_i) = \int_{u_i}^{\infty} \phi(u)du$$
$$= 1 - \Phi(u_i) \tag{3}$$

where $\phi(u)$ is the probability density function (pdf) for the inefficiency, which is assumed to be supported in $[0, \infty)$, and $\Phi(u) = \int_0^u \phi(s)ds$ is the cumulative distribution function (cdf).

Note that the percentile ranking depicted by (3) becomes 100% if $u_i = 0$, meaning that plant i is ranked as the best in the automotive industry for a given operation condition. If the productivity of plant i is poor, then the value of (3) is close to 0%. This study uses the percentile ranking as a percentage performance metric for measuring plant productivity. For clarification between productivity and efficiency, what is usually analyzed in SFA is the efficiency score (as being $e^{-inefficiency}$), where efficiency is defined relative to some benchmark of performance as known as "best practice". However, in the labor study context, this concept must also consider the minimum labor resource required to produce the particular goods or services. From the managerial viewpoint, it must also help set a realistic goal based on best observable performance. Therefore, this paper defines the percentile ranking of i-plant's productivity from the inefficiency distribution. Using the defined productivity, it is straightforward for a plant to set a percentile target from an industrial performance distribution that has a specific curve shape representing what the industrial competition for operational efficiency looks like and then find a HPU target according to the percentile target. To highlight the difference between HPU competitiveness gap and the labor producibility concept, in fact, the interpretation of HPU competitiveness gap depends on how well a plant performs. The lower ranked plants (i.e., lower percentage performance) have higher potentials to further improvement so that the reducing one unit of HPU competitiveness gap is relatively easy for the lower ranked plants. Meanwhile, the plants that are already closer to their best practice level (i.e., higher percentage performance) tend to reach a point of diminishing returns and so the higher ranked plants should spend more efforts to lessen one unit of HPU competitiveness gap.

However, there is a modeling challenge in applying a multivariate SFA model to assess complex manufacturing systems, such as the vehicle assembly process. A typical automobile manufacturing process consists of many processes that can be grouped into the three main processes of Bodyshop, Paintshop, and GA, as depicted in Figure 1. The Bodyshop transforms raw materials into the structure of a vehicle. Next, the Paintshop applies a protective and visual coating to the product. Finally, GA assembles all subcomponents, such as the engine, tires, and seats, into the vehicle. Although these three processes can be executed sequentially and simultaneously in the same plant, they are practically independent of each other in terms of technology, operation and staffing. This is the reason many car companies are able to leverage their inner network of manufacturing facilities that are located remotely, to produce (or paint) parts and pieces alternatively and then, ship them to another plants for other processes like final assembly.
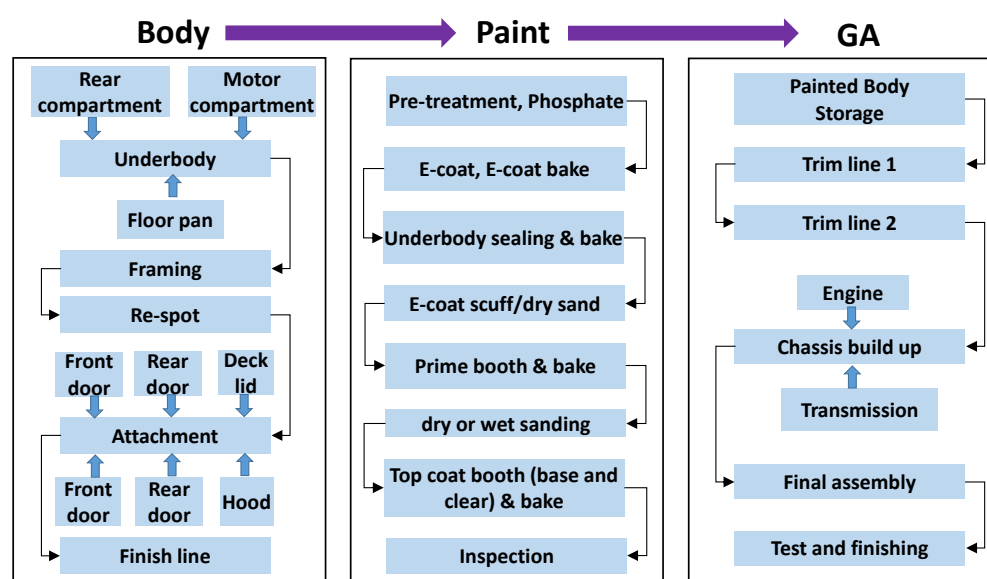
Figure 1. Main manufacturing processes in vehicle assembly plant.

In general, assessing a vehicle plant's productivity should require a multivariate model, including many explanatory variables, to account for the complex car manufacturing system. A multivariate model is a valuable tool in dealing with the interaction of multiple variables; however, with many intercorrelated explanatory variables, it may have the problem of suppressor variable (A suppressor variable may inflate the predictive validity of another variable (or set of variables) by its inclusion in a regression equation. Multicollinearity is the usual suspect of suppressor variables. Suppressors are present when too many variables are used in one model and some of input variables are positively correlated. Its effect is to flip the coefficient sign of suppressor variables because a suppressor variable suppresses itself, but boosts the other positively correlated variables. Suppressors tend to remain even after passing a variance inflation factor test.) presence, which has the potential to significantly limit findings that can be reported because the obvious solution for suppressor variables is omitting the problem variables, resulting in useful datasets being passed over. Note that the gain effort of competitors' data in the context of automotive industry benchmarking is highly expensive and time-consuming. The collected but neglected useful datasets could have been used to conduct analysis to find an effective productivity improvement path. Therefore, the data omission due to the inability in modeling complex manufacturing systems is rather frustrating.

This study will propose a systematic approach to assessing a plant's productivity, where the problem of suppressor variable presence caused by multicollinearity is relieved. The overall approach is to execute the following steps: (1) decompose a vehicle assembly plant into three shops–Bodyshop, Paintshop and GA; (2) build SFA models for these shops; and (3) reintegrate the general plant system by the appropriate combination of the shops' inefficiency distribution, which is part of the shop-level SFA model.

In this procedure, it is required to derive a probability density function from three different inefficiency distributions to aggregate shop-level assessment results to evaluate the plant-level productivity. We represent the density function as an integral of a common function and provide a computational error as well.

The objective of this paper is to provide the following contributions to researchers and practitioners from the proposed method. First, we provide insights into the use of SFA for accurate and objective assessment of plant productivity and the identification of productivity improvement opportunities. Second, we provide a step-by-step procedure for deriving a composite pdf and cdf from three half-normal inefficiency distributions for HPU in SFA. In particular, our study shows how to mitigate the problem of suppressor variable presence in assessing the productivity of highly complex vehicle assembly plants in such

a way that the overall process is divided into shops (i.e., Bodyshop, Paintshop, and GA). Third, this paper shows an illustrative example that demonstrates the application of the proposed approach for assessing hypothetical vehicle assembly plants where the total labor hours are split into Bodyshop, Paintshop and GA, exclusively and collectively. At last, this study offers a solution process to clarify the reasons for underperforming plants in terms of labor productivity and identify the course of actions to cure the issues with some managerial insights emphasizing the balanced approach incorporating people, process and technology.

The remainder of this paper is organized as follows. In Section 2, we review the existing literature. Section 3 deals with the theoretical development of the SFA model focused on a mathematical solution for the composition of multiple stochastic frontier distributions. An application of the developed SFA model to the automotive industry is discussed in Section 4 with some remarks relevant to practitioners. Some concluding remarks are drawn in the last section.

2. Literature Review

This work is motivated by two streams of extant research: (1) productivity analysis in manufacturing systems; and (2) efficiency assessment via SFA models.

In the first stream of studies, the productivity of manufacturing systems related to quality, flexibility and inventory is analyzed. A methodology that measures the efficiency of vehicle assembly plants in terms of productivity and quality was described by Krafcik (1988) [6]. This study cites an earlier empirical study that assesses productivity to identify superior manufacturing practices because a comprehensive comparative analysis was performed based on real visits to 38 automotive assembly plants in 13 countries around the world. The automotive industry has attracted significant attention from researchers in management science and decision optimization due to its significant impact on the real world economy and its data availability. The mechanism for the achievement of high productivity via efficient labor utilization was studied by Lieberman et al. [7]. Regarding the impact of flexibility on productivity, a study was undertaken to analyze how the increased complexity of parts involved in car manufacturing can result in a negative effect on productivity [1]. In addition, a similar study showed that an increased number of vehicle options has a negative impact on both productivity and quality [8]. By a comparison of US and Japanese automobile producers, it was shown that there is a strong relationship between higher productivity and inventory reduction [9]. Gopal et al. (2013) showed that launch events can disrupt manufacturing operations, resulting in productivity losses [2]. In their study, ordinary least squares regression was applied, but various sample selection methods were utilized to test their hypothesis. For example, a series of matched sample methodologies or an instrument variable was used to ascertain whether a plant hosting a launch suffers a decline in productivity, and the Heckman sample selection methodology [10] was used to examine how experience and flexibility can mitigate productivity losses. Alden et al. (2006) reported how analytics can contribute to the improvement of plant productivity, with assembly plants of General Motors taken as examples [11].

The studies in the second stream focus on the one-sided distributional form assumption of the efficiency component in SFA models. As the inefficiency component is widely assumed to be a half-normal distribution in many applications, we consider integrating different half-normal distributions here. We remark that a truncated normal distribution or an exponential distribution is also assumed for the inefficiency component. In particular, for the truncated normal distribution [12–14], the first parameter can capture firm characteristics in representing heterogeneous efficiency effects, and the half-normal distribution can be transformed from the truncated normal distribution. Meesters (2014) also showed that the exponential distribution is a special limiting case of the truncated normal distribution, as is the half-normal distribution [15].

Although practical experience with frontier analysis using SFA models has shown that such standard models are sufficiently accurate (see also [16]), some researchers have devel-

oped complicated but more realistic distribution models by using the Gamma distribution. Green (1990) proposed the use of the gamma distribution to estimate the inefficiency term, although the likelihood function for the convolution of the two error components of the SFA model is not available in closed form because the likelihood function involves an integral that cannot be evaluated analytically [17]. However, Beckers and Hammond (1987) showed that the likelihood can be computed using special functions in this context [18]. Subsequently, Greene (2003) proposed a simulation-based method to address this issue [19]. Tsionas (2012) showed that the fast Fourier transform of the characteristic function can be used to address the troublesome term in the likelihood function [20]. It is worth noting that Kumbhakar et al. (2013) introduced the zero inefficiency stochastic frontier model which can accommodate the presence of both efficient and inefficient units [21] and Makieła (2017) included more than one random inefficiency terms [22].

For the purpose of this study, it is necessary to compute the distribution of the sum of inefficiency random variables for standard one-sided distributions. Indeed, in probability theory and statistics, the computation of the distribution of the sum of independent of random variables has been widely studied. For example, Moschopoulos (1985) showed that the distribution of the sum of independent gamma variates with different parameters can be represented as a gamma series [23] and Bibingera (2013) provided a closed form of the distribution of the sum of independent exponentially distributed random variables [24,25]. In addition, Krenek et al. (2016) studied the sum of distributions for the case of truncated distributions [26].

In particular, this work focuses on the derivation as well as numerical computation of the probability density function from three independent half-normal distributions. In Bayesian SFA, numerical tools such as Gibbs sampling can be used to approximate the multivariate probability distribution. On the other hand, it is required to obtain a closed form of the probability density function for non-Bayesian SFA. We notice that the half-normal distribution is a limit of a skewed normal distribution [27] and the density function of the sum of independent skew normal random variables is given in a closed form [28]. However, the expression involves a special function, the Kampé de Fériet function, which is not easy to evaluate. In contrast to Nadarajah and Li's work [28], we present the required pdf and cdf in the form of the integral of the error function and estimate the computational errors.

SFA is the most commonly used efficient frontier measures in the productivity analysis in econometrics [29,30]. However, in context of the automotive industry, few researchers have studied the application of SFA. Boyd (2008, 2014) [31,32] and Oh and Hildreth (2014, 2016) [33,34] worked to address the topic of energy intensity and its influencing factors. Besides, a few literatures dealing with productivity and Research and Development (R&D) efficiency in automotive industry with SFA were found [35,36].

Added to SFA, DEA [5] is also an extremely useful tool developed in operations research community based on non-parametric modeling concept. There are plenty of literatures reporting DEA application to benchmarking studies (e.g., bank branch productivity study [37], warehouse performance study [38]). However, this study uses SFA because this study needs parametric modeling approach to integrate shop-level inefficiency distributions.

3. Composition of Multiple Stochastic Frontier Distributions

The goal of plant productivity analysis is to measure the overall plant productivity. However, due to the complexity of the manufacturing system, the multivariate model representing a whole plant may include too many explanatory variables intercorrelated with each other, resulting in the presence of suppressor variables. For example, let Z_1 and Z_2 denote the number of car platforms and the number of car models, respectively. Both Z_1 and Z_2 are used for the same purpose of indicating the complexity level of a given plant. Therefore, they are positively intercorrelated and have a negative relation with HPU. When both variables are introduced together into one multivariate model, the coefficient sign

of one of the two variables is likely to flip over owing to the effect of suppression such that one variable takes a plus sign while the other variable takes a minus sign. In fact, Z_1 mainly affects the Bodyshop process only, but the impact of Z_2 is limited to the GA area. In other words, although Z_1 and Z_2 can cause the suppressor variable problem if combined, by separating them into different SFA models (Z_1 for Bodyshop and Z_2 for GA), the problem can be avoided. This study proposes to build separate SFA models by shop and then compose their inefficiency distributions to estimate the overall plant productivity. In this section, a mathematical solution for the composition of multiple stochastic frontier distributions is built.

To this end, it is required to integrate inefficiency distributions of Bodyshop, Paintshop, and GA. As discussed earlier in explaining Figure 1, the three processes are independent of each other and therefore, the competitiveness gap of those processes can be assumed to be independent. We denote $u_i^{Bodyshop}$, $u_i^{Paintshop}$, and u_i^{GA} as the competitiveness gap of Bodyshop, Paintshop, and GA, respectively. Then, the overall inefficiency for plant i is given by

$$u_i = u_i^{Bodyshop} + u_i^{Paintshop} + u_i^{GA}. \tag{4}$$

As the percentile ranking of i-plant's productivity is calculated from (3), it is required to compute the associated probability density function. It is well-known that the probability density for the sum of independent random variables is given by the convolution of each density function. Thus, the distribution of the overall inefficiency ϕ can be represented as the convolution of the distributions $\phi_k (k = 1, 2, 3)$ for Bodyshop, Paintshop, and GA, respectively,

$$\phi = (\phi_1 * (\phi_2 * \phi_3)), \tag{5}$$

where

$$(\phi_j * \phi_k)(x) = \int_{-\infty}^{\infty} \phi_j(t)\phi_k(x - t)dt.$$

Here, we compute ϕ under the assumption that the inefficiency components are distributed half-normally. The use of half-normal distribution to represent the inefficiency components is in fact pretty reasonable in the context of the automotive industry. When Kernel distribution estimate [39] is applied to predict the shape of HPU competitiveness distribution in the automotive industry, it turned out that the distribution shape was far skewed to the left. The skewed shape to the left means that the industrial competition for labor productivity is highly intensive and many companies are already reaching out to their best practice, in which case the half-normal distribution can be introduced to account for the shape of HPU competitiveness distribution at its high rate observed in the far left side. To be sure, this study carried out LR (likelihood-ratio) tests [40] for $u_i^{Bodyshop}$, $u_i^{Paintshop}$, and u_i^{GA} where two models were compared: (1) Ordinary Least Square (OLS) and (2) SFA with the half-normal distribution used for the inefficiency component. All three tests told that the LR statistic exceeds 99%, indicating that the choice of half normal distribution to represent inefficiency in Body shop, Paint shop and GA is a correct decision.

Thus, for some $\sigma_k > 0$,

$$\phi_k(x) = 2g_k(x)\chi_{[0,\infty)}(x)$$
$$= \frac{\sqrt{2}}{\sigma_k\sqrt{\pi}}e^{-x^2/2\sigma_k^2}\chi_{[0,\infty)}(x), \quad k = 1, 2, 3, \tag{6}$$

where g_k is a Gaussian distribution function with mean 0 and variance σ_k^2; and χ is the characteristic function. Next, we state a technical lemma:

Lemma 1. *Let g_2, g_3 be the probability density functions for the normal distributions with mean 0, and variances σ_2^2 and σ_3^2, respectively. Then*

$$\int_0^t g_2(s)g_3(t-s)ds = \frac{1}{2\sqrt{2\pi}\sigma}e^{-t^2/2\sigma^2}\left[\mathrm{erf}\left(\frac{t}{\sqrt{2}\sigma}\frac{\sigma_3}{\sigma_2}\right) + \mathrm{erf}\left(\frac{t}{\sqrt{2}\sigma}\frac{\sigma_2}{\sigma_3}\right)\right],$$

where

$$\sigma = \sqrt{\sigma_2^2 + \sigma_3^2},$$

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-w^2}dw. \tag{7}$$

We provide a proof of Lemma 1 in the Appendix A.

If only two independent variables are considered, the distribution of their sum can be easily computed by the convolution integral (5). By considering the supports of ϕ_2 and ϕ_3, we have

$$(\phi_2 * \phi_3)(t) = \int_{-\infty}^{\infty} \phi_2(t-s)\phi_3(s)ds$$

$$= \int_0^t \phi_2(t-s)\phi_3(s)ds.$$

It follows from (6) and Lemma 1 that

$$(\phi_2 * \phi_3)(t) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma}e^{-t^2/2\sigma^2}\left[\mathrm{erf}\left(\frac{t}{\sqrt{2}\sigma}\frac{\sigma_3}{\sigma_2}\right) + \mathrm{erf}\left(\frac{t}{\sqrt{2}\sigma}\frac{\sigma_2}{\sigma_3}\right)\right].$$

However, we are interested in the probability distribution function for the sum of three independent random variables. Unfortunately, we cannot compute it from convolution theorem (5) as in the case of the sum of two independent variables.

Let $x_k (k = 1, 2, 3)$ be normally distributed random variables with mean 0 and variances $\sigma_k^2 (k = 1, 2, 3)$. Then

$$u^{Bodyshop} = |x_1|, \ u^{Paintshop} = |x_2|, \ u^{GA} = |x_3|,$$

and the density function ϕ for $u = u^{Bodyshop} + u^{Paintshop} + u^{GA}$ satisfies

$$P(|x_1| + |x_2| + |x_3| < x) = \int_0^x \phi(u)du. \tag{8}$$

On the other hand, the region $|x_1| + |x_2| + |x_3| < x$ in the three-dimensional $x_1 x_2 x_3$ space is bounded by the eight plains $x_1 + x_2 \pm x_3 = x$, $-x_1 + x_2 \pm x_3 = x$, $-x_1 - x_2 \pm x_3 = x$, and $x_1 - x_2 \pm x_3 = x$. Therefore,

$$P(|x_1| + |x_2| + |x_3| < x) = \int_0^x \int_0^{x-x_1} \int_{(x_1+x_2)-x}^{x-(x_1+x_2)} g_1(x_1)g_2(x_2)g_3(x_3)dx_3dx_2dx_1$$

$$+ \int_{-x}^0 \int_0^{x+x_1} \int_{(-x_1+x_2)-x}^{x-(-x_1+x_2)} g_1(x_1)g_2(x_2)g_3(x_3)dx_3dx_2dx_1$$

$$+ \int_{-x}^0 \int_{-x-x_1}^0 \int_{(-x_1-x_2)-x}^{x-(-x_1-x_2)} g_1(x_1)g_2(x_2)g_3(x_3)dx_3dx_2dx_1$$

$$+ \int_0^x \int_{-x+x_1}^0 \int_{(x_1-x_2)-x}^{x-(x_1-x_2)} g_1(x_1)g_2(x_2)g_3(x_3)dx_3dx_2dx_1. \tag{9}$$

Together with (8), $\phi(x)$ is obtained from Lemma 1 and the derivatives of the integrals in (9) with respect to x;

$$\phi(x) = \frac{2}{\pi\sigma\sigma_1} \int_0^x e^{-x_1^2/2\sigma_1^2 - (x-x_1)^2/2\sigma^2} \left[\mathrm{erf}\left(\frac{x - x_1}{\sqrt{2}\sigma}\frac{\sigma_3}{\sigma_2}\right) + \mathrm{erf}\left(\frac{x - x_1}{\sqrt{2}\sigma}\frac{\sigma_2}{\sigma_3}\right)\right] dx_1. \quad (10)$$

The detailed derivation of (10) is given in Appendix B. We remark that $\phi(x)$ can be rewritten in terms of $P(\sigma_i)$ for any permutation operator P due to the symmetry of $\sigma_1, \sigma_2, \sigma_3$.

Now, we compute the cumulative distribution function, that is,

$$\Phi(u) = \int_0^u \phi(x)dx. \quad (11)$$

To reduce the mathematical expression of (10), let us define $\psi(w)$ as

$$\psi(w) := e^{-w^2/2\sigma^2}\left[\mathrm{erf}\left(\frac{w}{\sqrt{2}\sigma}\frac{\sigma_3}{\sigma_2}\right) + \mathrm{erf}\left(\frac{w}{\sqrt{2}\sigma}\frac{\sigma_2}{\sigma_3}\right)\right].$$

Using the change of variables with $\psi(w)$, we rewrite (10) as

$$\phi(x) = \frac{2}{\pi\sigma\sigma_1} \int_0^x e^{-w^2/2\sigma_1^2}\psi(x - w)dw$$
$$= \frac{2}{\pi\sigma\sigma_1} \int_0^x e^{-(x-w)^2/2\sigma_1^2}\psi(w)dw.$$

Substitution of ϕ into (11) together with changing the order of integration yields

$$\Phi(u) = \frac{2}{\pi\sigma\sigma_1} \int_0^u \int_0^x e^{-(x-w)^2/2\sigma_1^2}\psi(w)dwdx$$
$$= \frac{2}{\pi\sigma\sigma_1} \int_0^u \int_w^u e^{-(x-w)^2/2\sigma_1^2}\psi(w)dxdw$$
$$= \sqrt{\frac{2}{\pi}}\frac{1}{\sigma} \int_0^u \mathrm{erf}\left(\frac{u - w}{\sqrt{2}\sigma_1}\right)\psi(w)dw. \quad (12)$$

We summarize our findings in Theorem 1.

Theorem 1. *Assume that $u^{Bodyshop}$, $u^{Paintshop}$, and u^{GA} are independent random variables that are half-normally distributed with parameters σ_1^2, σ_2^2, and σ_z^3, respectively. Then the probability density function and the cumulative distribution function for $u = u_i^{Bodyshop} + u_i^{Paintshop} + u_i^{GA}$ are given by (10) and (12), respectively.*

It should be noted that one can compute ϕ in a different form, using the result of Krenek et al. [26]. However, it is not suitable for our numerical implementation as it is created via induction.

4. Illustrative Example

This section discusses the application of pdf (Equation (10)) and cdf (Equation (12)) developed in the previous section. The example discussed in this section deals with HPU data from the automotive industry. Since (10) and (12) are expressed in integrals of common functions, they can be easily implemented in Microsoft Excel and therefore all implemented works in this section was done in Microsoft Excel.

The data source for HPU data can be found in the Harbour Report (Oliver Wyman consulting firm, US), which is an annual survey of all automotive manufacturing plants in the world. Currently, most automotive manufacturers voluntarily provide data for the survey. To prove the data integrity, Oliver Wyman consultants visit each plant and audit the data. They publish an annual report that documents and compares the performance of

plants based on HPU comparisons by market segment. In addition to HPU, the Harbour Report includes plant-specific operational data, including capacity, production volume, complexity, automation, souring, contents, and labor. However, due to the non-disclosure agreement signed by auto companies participating in the Harbour Report, the data in the report cannot be publicly shared (Before 2007, the Harbour report data was in the public domain. Gopal et al. (2013) [2] reported that when they used publicly available data from the years 1999–2007 from North American automotive plants, the average annual production was 186,000 per plant and the average HPU was 27.1.).

Therefore, for the illustration purpose, the authors generated five similar plants that are competing in the luxury car market segment. The datasets for plants were generated to resemble real-world data as close as possible. The dataset is a set of values of explainable variables listed in Table A1. From the datasets, SFA models for Bodyshop (13), Paintshop (14) and GA (15) were constructed. The Bodyshop SFA model includes production volumes, the number of platforms, body-on-frame, total welds, and welding automation. Turning to Paintshop, the SFA model takes production volumes, the number of models, sealer automation, paint content automation, and paint content outsourcing as explanatory variables. GA's SFA model uses production volumes, the number of models, body-on-frame, part numbers, GA outsourcing, logistic outsourcing, and the parts received as explanatory variables. Then, u_i and v_i for each plant was calculated.

$$HPU_i^{Bodyshop} = f^{Bodyshop}(PV, NoP, BoF, TW, WA; \beta^{Bodyshop}) + u_i^{Bodyshop} + v_i^{Bodyshop} \qquad (13)$$

$$HPU_i^{Paintshop} = f^{Paintshop}(PV, NoM, SA, PCA, PCO; \beta^{Paintshop}) + u_i^{Paintshop} + v_i^{Paintshop} \qquad (14)$$

$$HPU_i^{GA} = f^{GA}(PV, NoM, BoF, PN, GAO, LO, PR; \beta^{GA}) + u_i^{GA} + v_i^{GA} \qquad (15)$$

It is worth noting that the global market is divided into several markets based on car size and price (The car industry generally uses the standard for the global market segments set by IHS Markit, UK). Among the market segments, E1 and E2 segments represent large & luxury class and high luxury class, respectively. In detail, Segment E1 counts car models with sizes of 4700–5100 mm and prices between \$50K and \$85K. For example, BMW 5 Series, Mercedes-Benz E-Class, Audi A6, Chrysler 300 and Toyota Crown belong to the E1 segment. The E2 segment contains car models having sizes greater than 4500 mm and prices of \$85K–\$170K. For example, Lincoln Town Car, Mercedes-Benz S-Class, BMW 7 Series, Chevrolet Corvette, and Lexus LS match the standards of the E2 segment. This study observed real plants producing either E1 or E2, or both car segments, and generated five synthetic plants. Overall, the plants producing models belonging to E1 and E2 segments are characterized by higher HPU than low class segment cars because of special efforts to maintain high quality and the high rate of in-house workload with small production volume, different from regular mass production models.

The authors believe that the use of synthetic datasets will not be detrimental to the overall purpose of this study because they are generated to close to the real-world data. Each plant's HPU gaps for each theoretical best and the performance gaps are calculated in Table 1. Recall that $u_i^{Bodyshop}$, $u_i^{Paintshop}$, and u_i^{GA} denote the competitiveness gap of Bodyshop, Paintshop and GA, respectively, where $u_i^{Bodyshop} \sim N^+(0, \sigma_1^2)$, $u_i^{Paintshop} \sim N^+(0, \sigma_2^2)$, and $u_i^{GA} \sim N^+(0, \sigma_3^2)$. When maximum likelihood estimation was applied to estimate the parameters in (1), we obtained $\sigma_1^2 = 40.895$, $\sigma_2^2 = 16.463$ and $\sigma_3^2 = 116.304$.

Table 1. Competitiveness gap and performance comparison.

Plant	Competitiveness Gap				Performance			
	Bodyshop	Paintshop	GA	Plant	Bodyshop	Paintshop	GA	Plant
A	4.25	4.27	1.7	10.22	51%	29%	87%	79%
B	3.02	4.3	5.69	13.02	64%	29%	60%	65%
C	5.86	4.18	4.79	14.83	36%	30%	66%	56%
D	5.35	3.16	12.37	21.34	40%	37%	25%	27%
E	3.55	15.17	21.16	39.88	58%	0%	5%	1%

The curves in Figure 2 represent

- inefficiency cdf for Bodyshop ($= \int_u^\infty N^+(0, \sigma_1^2) dx$)
- inefficiency cdf for Paintshop ($= \int_u^\infty N^+(0, \sigma_2^2) dx$)
- inefficiency cdf for GA ($= \int_u^\infty N^+(0, \sigma_3^2) dx$)
- inefficiency cdf for overall plant (= $\int_u^\infty \phi(x) dx = 1 - \Phi(u)$)

Note that the cdf curve of the overall plant is obtained from Theorem 1, but using a numerical approach (because (12) cannot be solved analytically). This study uses the trapezoidal rule to calculate (12) numerically. First, the uniform grid for numerical calculation of $\Phi(u)$ is defined as follows:

$$0 = w_0, w_1, w_2, \cdots, w_M = u, \quad \Delta = w_{j+1} - w_j.$$

Then, Theorem 1 and the trapezoidal rule yield

$$\Phi(u) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \int_0^u \mathrm{erf}\left(\frac{u-w}{\sqrt{2}\sigma_1}\right) \psi(w) dw$$

$$\approx \sqrt{\frac{2}{\pi}} \frac{\Delta}{\sigma} \sum_{j=1}^{M-1} \mathrm{erf}\left(\frac{u-w_j}{\sqrt{2}\sigma_1}\right) \psi(w_j)$$

$$= \sqrt{\frac{2}{\pi}} \frac{\Delta}{\sigma} \sum_{j=1}^{M-1} \mathrm{erf}\left(\frac{w_{M-j}}{\sqrt{2}\sigma_1}\right) \psi(w_j),$$

where we adopt $\mathrm{erf}(0) = 0$.

Although the trapezoidal rule always finds an approximation for $\Phi(u)$, we estimate the error of the numerical integration compared with the actual $\Phi(u)$ in order to provide the accuracy for our simulation. Fix $u \in (0, \infty)$. Then it is known [41] that the error for the trapezoidal rule is given by

$$\epsilon(\Phi) := \Phi(u) - \sqrt{\frac{2}{\pi}} \frac{\Delta}{\sigma} \sum_{j=1}^{M-1} \mathrm{erf}\left(\frac{w_{M-j}}{\sqrt{2}\sigma_1}\right) \psi(w_j)$$

$$= -\frac{\Delta^2}{12} u \phi''(\xi) \quad \text{for some } 0 < \xi < u,$$

or asymptotically,

$$\epsilon(\Phi) = -\frac{\Delta^2}{12} \left[\phi'(u) - \phi'(0)\right] + O(\Delta^3) \quad \text{as } \Delta \to 0.$$

Direct calculation shows that

$$\phi'(0) = \frac{2}{\pi\sigma^2}\left(\frac{\sigma_3}{\sigma_2} + \frac{\sigma_2}{\sigma_3}\right)\text{erf}\left(\frac{u}{\sqrt{2}\sigma_1}\right)$$

$$\leq \frac{2}{\pi\sigma}\left(\frac{1}{\sigma_2} + \frac{1}{\sigma_3}\right)$$

$$\phi'(u) = -\frac{2}{\pi\sigma\sigma_1}\psi(u).$$

As obviously $|\psi(u)| \leq 2$ for any u, it follows that for a sufficiently small Δ,

$$|\epsilon(\Phi)| \leq \frac{\Delta^2}{12\sigma}\left(\frac{2}{\sigma_1} + \frac{1}{\sigma_2} + \frac{1}{\sigma_3}\right). \tag{16}$$

In our simulation, we take $\Delta = 0.01$ and thus

$$|\epsilon(\Phi)| \leq 10^{-6},$$

which gives a tolerance of our simulation results.
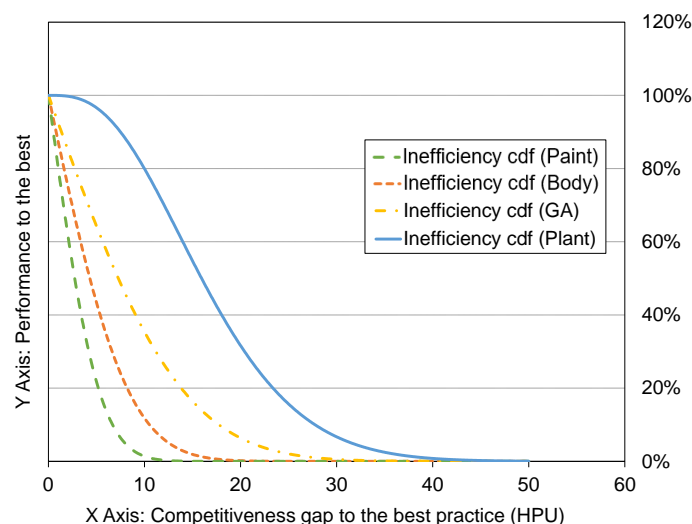
Figure 2. Cumulative distribution function of inefficiency for shops and overall plant.

Figures 3–5 show the performance of competing plants with respect to their competitiveness gaps. Figure 3 shows the order of performance ranking in Bodyshop, where plant B is ranked first, followed by plants E, A, D, and C. In Paintshop, Figure 4 indicates that plant D is ranked first, followed by plants C, A, B and E. In GA, plant A outperforms other plants, followed by plants C, B, D, and E, as shown in Figure 5. In terms of overall performance, as shown in Figure 6, plant A has the best performance, followed by plants B, C, D, and E. Note that plant A, which is the best in GA performance, becomes the best in overall plant-wise performance. This result seems reasonable in the sense that assembling parts in GA takes up the largest portion of the plant-wise HPU; furthermore, cars belonging to the E1 and E2 market segments require more labor hours in GA because of higher parts attachment and various factory fit options compared to ordinary car models. Therefore, plant A becomes the best in overall plant-wise performance competition, although it has mediocre performance in Bodyshop and Paintshop. Based on the competitiveness gap disclosed from the performance assessment, each plant can set up their feasible performance improvement target. For example, plant B needs to catch up plant A in Figure 6, where plant B can set up its target performance to be 79% to go after plant A, because plant A already shows that 79% achievement in the respective segment is possible.
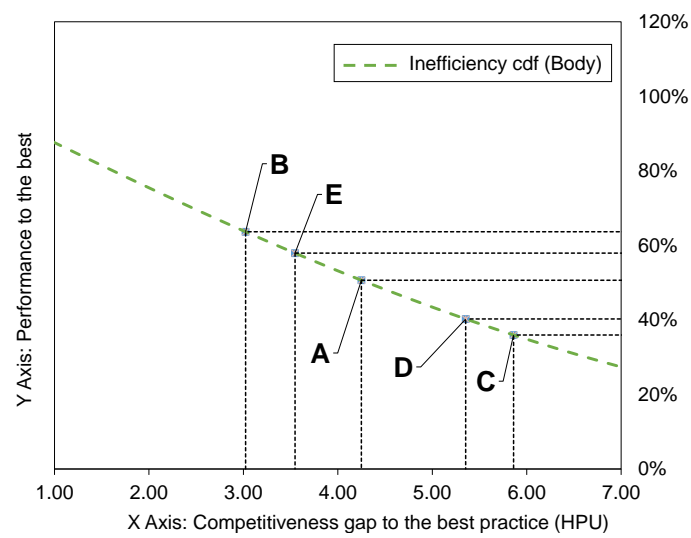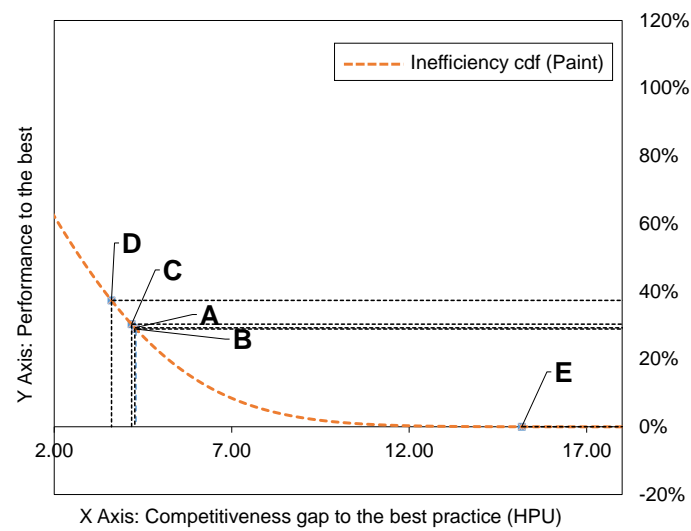
Figure 3. Assessment for Bodyshop.

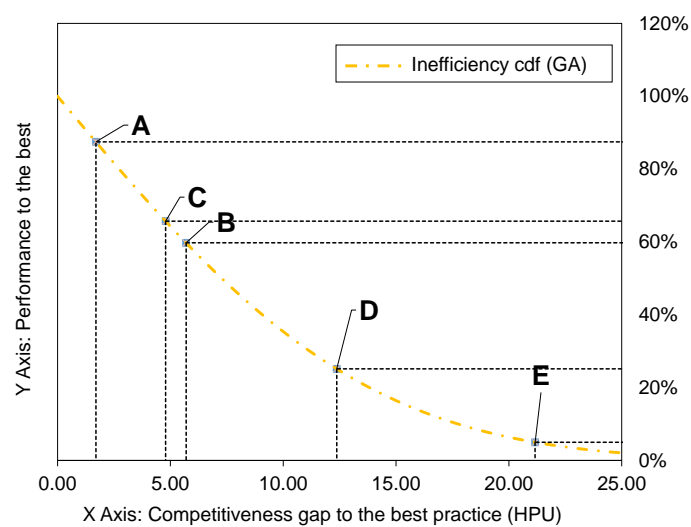**Figure 4.** Assessment for Paintshop.

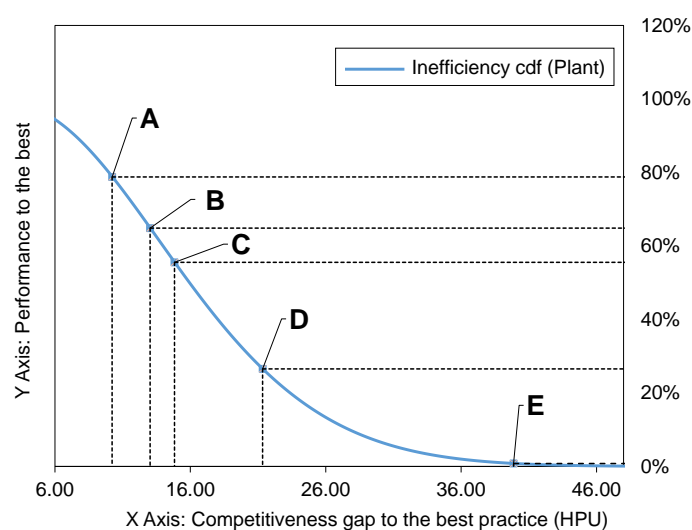**Figure 5.** Assessment for GA.

Figure 6. Assessment for plant.

Figures 7 and 8 show the pdfs and the reflected cdfs for our integrated model and the standard half-normal model for total HPU. We obtained $\sigma_u^2 = 334.202$ from the maximum likelihood estimation with the half-normal model.

Normally, the logical next step after a series of data-driven studies, including benchmarking, performance assessment, and feasible target setting, is to clarify the reasons why the outperforming (underperforming) plants show the best (worst) results, and what course of actions the underperforming plants should take to enhance their performance. While the data-driven studies can be done by techniques offered in this paper, it is people's role to clarify the reasons and identify the course of actions. To help activate and facilitate the people's role, this study offers a systematic solution process as shown in Figure 9 where a description of each step is as follows.
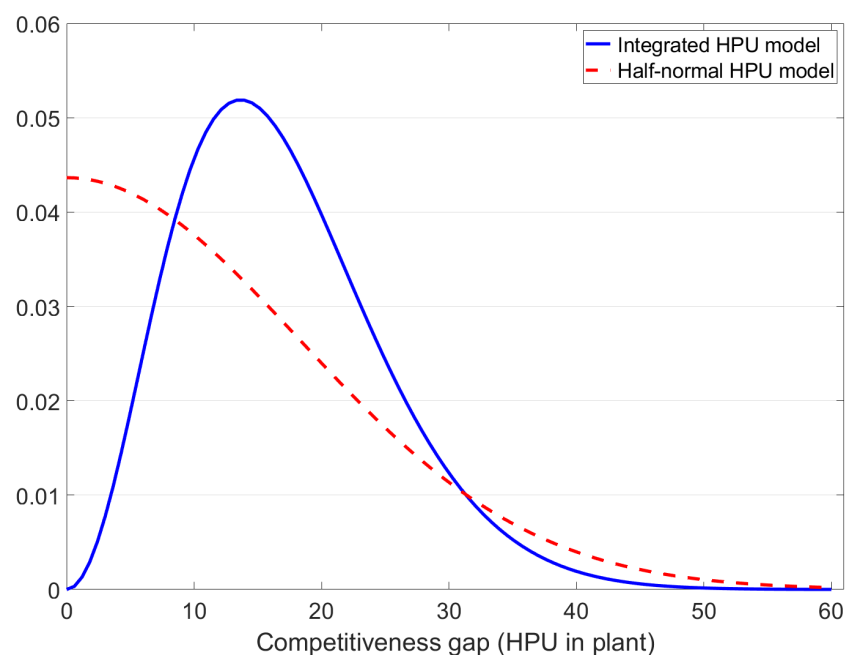

Figure 7. Probability density functions of inefficiency for integrated model and half-normal model.
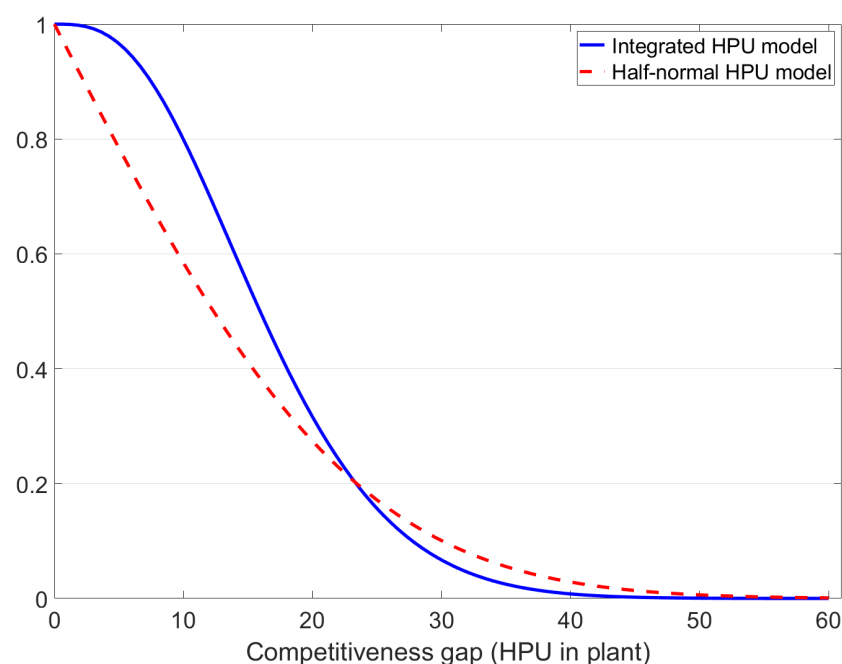
Figure 8. Cumulative distribution functions of inefficiency for integrated model and half-normal model.
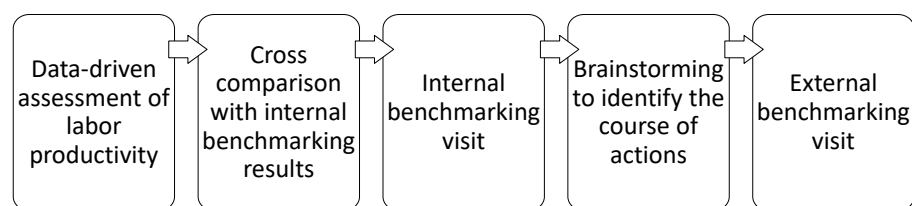

Figure 9. Solution process to close the competitiveness gap.

The first step is data-driven assessment of labor productivity. This step is about performance assessment, identification of best practice plant, and feasible target setting, which this paper has delivered thus far.

The second step is a cross comparison with internal benchmarking results. Typically, a manufacturing company has used its standard time-and-motion study tool to estimate workload, based on which the company practices an internal benchmarking study. At this step, the external benchmarking study executed in the first step can be cross compared with the internal benchmarking study results to figure out the similarity or difference between the two benchmarking studies, being able to confirm the areas of opportunity for further improvement. One advantage of this step comes from the plenty of internal data and knowledge available within the company and so it is possible to further split the labor hours into the core areas to fine tune a further segmentation of the functional area for allocating labor hours. The core areas include manufacturing (i.e., all preparation and processing tasks associated with joining, stamping, machining, loading, fabricating, painting, and fastening), maintenance (i.e., planned/preventive/breakdown maintenance, Die/jig maintenance, conveyor system maintenance, etc.), support (i.e., labor required to support production, engineering/IT, various reporting, and etc.), logistics (parts receiving, checking, repacking/sequencing, line side delivery, material handling, etc.), quality (i.e., inspection, audits/checks, rework, test, measurement, reporting, etc.), and central site services (i.e., medical, training, finance, management, house keeping, etc.)

The third step is internal benchmarking visit. In case of a company which has a global manufacturing footprint, it is highly possible to select a peer group consisting of comparable plants with similar manufacturing operation parameters inside the company.

In such a case, underperforming plants in the peer group may visit the best practice plant (identified from the first step) to learn how well the target plant performs and why the target plant is successful, and to adapt specific best practices.

The fourth step is brainstorming to identify the course of actions. Even after executing the third step, an underperforming plant may not discover enough actions for improvement to fill the competitiveness gap identified in the first step. In the case, turning to brainstorming ideas would be a solution. The typical brainstorming means the efforts of a group of people to generate new ideas and solutions around a specific problem of interest. Under the brainstorming scheme, subject matter experts, plant managers, and engineers hold a meeting where individuals can freely suggest as many spontaneous new ideas for labor productivity improvement as possible with no fear of being criticised.

The final step is external benchmarking visit. When the fourth step does not discover a complete solution set for labor productivity improvement due to lack of knowledge and experience, visiting a competitor's plant directly may be a last resort but can be an effective solution. In general, with a non-disclosure agreements concluded, two competing companies can agree to mutually exchange information beneficial to both companies and visit each other to identify leading edge practices. Since the best practice plant was already identified at the first step, it is clear which competitor's plant to aim to engage.

Further, this study also provide some managerial insights to practitioners that may increase the adoption and usage of the proposed methods in this paper as follows. Although every global manufacturer is driving its own version of a manufacturing production system in its local and regional facilities, it is possible to share common insights and ideas for clarification. While there are multiple reasons for underperformance, but the two reasons rated most important are overcapacity and the labor/operational competitiveness gap to best practices in HPU. The overcapacity issues are likely to be an expected challenge to face during the process of implementing the company's long-term product portfolio strategy (e.g., car makers in North America keep traditional passenger car portfolio even though the sales of passenger cars in North America have been declining for the past years and continue on a downward trend). Since it is too early to judge the success or failure of the strategy, a directional change in the strategy is not expected. Indeed, a strategy change may produce a whole series of changes in objectives through the hierarchical arrangement with cost and confusion resulting from any major reconstruction. Meanwhile, the competitiveness gap in labor and operational efficiency is considered to be something that a company can address in the time and space of the present-day world if right decisions are made for assessment of current competitiveness, target setting, and the identification of areas of opportunities or a course of action to take. It is worth noting that at present, the improvement of labor and operational efficiency is tight, but even more important than ten years ago, as the theme of "doing more with less" continues to dominate the automotive manufacturing culture. Nowadays, nearly half of the North American and Europe plants operate on a three-shift or three-crew model, drastically increasing capacity utilization and capital efficiency with far fewer factories, as existed before the downturn in 2008. With this three-shift pattern, even if the next downturn arrives, the plants will not be closed, but just reduced to two shifts. However, this shift pattern means that factories are being asked to add more and more vehicles, which are growing in terms of features and complexity. In addition, much higher content with upgraded standards for safety and emissions and shifting consumer demands due to the change of traditional personal transportation methods are impacting both production complexity and vehicle costs, which add extreme challenges to the production environment.

As the three-shift pattern with high complexity has become the new standard in the industry, marginal productivity growth opportunities and an increase in competition for better operational efficiency have been observed, resulting in marginal productivity growth between competitors. In this environment, the improvement of labor and operational efficiency is becoming more important because performance improvements only driven by technology and automation investments are often masked by increasing product complex-

ity, washing away much of the productivity gains. In general, higher operational efficiency is displayed by more flexible manufacturing. For example, in 2016, Honda transferred the Acura MDX from its Alabama plant to the East Liberty plant in Ohio to allow for more truck production in the Alabama plant, which also increased the utilization of car production in the Ohio plant in a very short period without a decline in production.

Regarding the identification of a course of action for labor and operational efficiency, one surprising discovery is that many of the newest factories with less automation equipment built in China, India, and Eastern Europe have shown incredible performance from young and eager workforces. This discovery is different from the general belief that the robots and automation inspired by Industry 4.0 are taking over the role of human workers, outperforming the labor-intensive traditional manufacturing system. Contrary to expectations, the best plants are not the automated plants, but the plants that display a good balance of people and technology, working in harmony. In fact, significant levels of automation may increase downtime which hinders production flows, driving more highly paid, skilled labor to keep the factory running, while avoiding throughput losses and quality problems. Rather, high-performance plants are commonly demonstrating less capital investment to full automation, but strong workforces at all levels, from the plant manager to the plant floor. These plants integrate automation technology more selectively, yet foster improvements with relatively low-budget technologies, such as small cobots to assist with unsafe or ergonomically difficult tasks, automated guided vehicles incorporated to bring materials, simple means of error proofing, visual controls to highlight potential problems, and improved part presentation (i.e., the use of specific part kits).

5. Conclusions

This study provided insight into the use of SFA for accurate and objective assessment of plant productivity and the identification of areas for productivity improvement. A composite pdf and cdf was derived from three half-normal inefficiency distributions for HPU by dividing the integration domain into eight non-overlapping subdomains. Furthermore, we estimated the error in computation of cdf that provides the accuracy of our simulation. It was shown that the new pdf and cdf can be used to assess the productivity of highly complex vehicle assembly plants such that the overall process is divided into shops (i.e., Bodyshop, Paintshop, and GA). The shop models are then hierarchically combined into the overall plant performance model, resulting in more accurate and objective assessment which mitigates the problem of suppressor variable presence and ensures no loss from useful datasets. This technique can be used to measure the gap between a plant's current productivity and set a feasible target. Granting all this, however, it is still the people's role to clarify the reasons for the competitiveness gap and identify productivity improvement opportunities. To help activate and facilitate the people's role, hence, this study offered a systematic solution process and also provide managerial insights emphasizing the balanced approach incorporating people, process and technology. Although robotization and automation inspired by Industry 4.0 are believed to revolutionize the automotive industry and dramatically increase productivity, the reality is that they are not a panacea. Instead, a balance of people, flexible process and technology is the deciding factor for success in improving productivity.

In future research, the extension of the proposed model's inefficiency component to allow other one-sided distributions than the half-normal distribution would be explored. Since the shop-level inefficiency can be modeled with an appropriate one-sided distribution accounting for the characteristics of the shop, it is necessary to develop a general framework to drive composite pdf and cdf, both numerically and analytically, from distinct one-sided distributions such as exponential, truncated normal, and gamma distribution; by doing so we can algebraically capture the different maturity stage of industrial competition on labor productivity. The authors also plan to reconsider the model formulation with focus on more accurate variable selection by following the method set forth by Koop et al. (1997) [42] because of the fact that the scale of the inefficiency term is data-dependent.

Appendix A. Proof of Lemma 1

In this section, we provide a proof of Lemma 1. From the definition of g_2 and g_3, we have

$$
\int_0^t g_2(s)g_3(t-s)ds = \int_0^t \left[\frac{1}{\sqrt{2\pi}\sigma_2}e^{-s^2/2\sigma_2^2}\right]\left[\frac{1}{\sqrt{2\pi}\sigma_3}e^{-(t-s)^2/2\sigma_3^2}\right]ds
$$
$$
= \frac{1}{2\pi\sigma_2\sigma_3}\int_0^t e^{-(t-s)^2/2\sigma_2^2-s^2/2\sigma_3^2}ds.
$$

It follows from completing the square that

$$
\frac{1}{2\pi\sigma_2\sigma_3}\int_0^t e^{-(t-s)^2/2\sigma_2^2-s^2/2\sigma_3^2}ds = \frac{1}{2\pi\sigma_2\sigma_3}e^{-t^2/2\sigma^2}\int_0^t e^{-\sigma^2\left(s-\frac{\sigma_2^2}{\sigma^2}t\right)^2/(2\sigma_2^2\sigma_3^2)}ds.
$$

Recall that $\sigma^2 = \sigma_2^2 + \sigma_3^2$. Let $w := \sigma\left(s - \frac{\sigma_2^2}{\sigma^2}t\right)/(\sqrt{2}\sigma_2\sigma_3)$. Then we deduce

$$
\frac{1}{2\pi\sigma_2\sigma_3}e^{-t^2/2\sigma^2}\frac{\sqrt{2}\sigma_2\sigma_3}{\sigma}\int_{-\frac{t}{\sqrt{2}\sigma}\frac{\sigma_2}{\sigma_3}}^{\frac{t}{\sqrt{2}\sigma}\frac{\sigma_3}{\sigma_2}} e^{-w^2}dw = \frac{1}{2\sqrt{2\pi}\sigma}e^{-t^2/2\sigma^2}\left[\mathrm{erf}\left(\frac{t}{\sqrt{2}\sigma}\frac{\sigma_3}{\sigma_2}\right)+\mathrm{erf}\left(\frac{t}{\sqrt{2}\sigma}\frac{\sigma_2}{\sigma_3}\right)\right].
$$

Appendix B. Derivation of Equation (10)

Here we derive the Equation (10). Differentiate the first integral in (9) with respect to x to get

$$
\frac{d}{dx}\int_0^x \int_0^{x-x_1} \int_{(x_1+x_2)-x}^{x-(x_1+x_2)} g_1(x_1)g_2(x_2)g_3(x_3)dx_3dx_2\,dx_1
$$
$$
= \int_0^x \int_0^{x-x_1} g_1(x_1)g_2(x_2)[g_3(x-(x_1+x_2))+g_3((x_1+x_2)-x)]dx_2\,dx_1
$$
$$
= 2\int_0^x \int_0^{x-x_1} g_1(x_1)g_2(x_2)g_3(x-(x_1+x_2))dx_2dx_1.
$$

Here, we use the formula

$$
\frac{d}{dx}\int_0^x h(w,x)dw = h(x,x) + \int_0^x \frac{\partial}{\partial x}h(w,x)dw
$$

successively and use the symmetric property of g_3. The same method for other integrals and the change of variables yield

$$
\phi(x) = 8\int_0^x \int_0^{x-x_1} g_1(x_1)g_2(x_2)g_3(x-(x_1+x_2))dx_2dx_1.
$$

Next, we apply Lemma 1 and obtain

$$\int_0^{x-x_1} g_2(x_2)g_3((x-x_1)-x_2))dx_2$$
$$= \frac{1}{2\sqrt{2\pi}\sigma}e^{-(x-x_1)^2/2\sigma^2}\left[\mathrm{erf}\left(\frac{x-x_1}{\sqrt{2}\sigma}\frac{\sigma_3}{\sigma_2}\right)+\mathrm{erf}\left(\frac{x-x_1}{\sqrt{2}\sigma}\frac{\sigma_2}{\sigma_3}\right)\right].$$

Together with (8), it follows

$$\phi(x) = \frac{2}{\pi\sigma\sigma_1}\int_0^x e^{-x_1^2/2\sigma_1^2-(x-x_1)^2/2\sigma^2}\left[\mathrm{erf}\left(\frac{x-x_1}{\sqrt{2}\sigma}\frac{\sigma_3}{\sigma_2}\right)+\mathrm{erf}\left(\frac{x-x_1}{\sqrt{2}\sigma}\frac{\sigma_2}{\sigma_3}\right)\right]dx_1,$$

as desired.

Appendix C. The Explanatory Variables Used in SFA Models

The definitions of explanatory variables used in SFA models are found in Table A1.

Table A1. Terms and definitions.

Production volumes (PV)	Total vehicle production of a plant during a year.
The number of platforms (NoP)	The number of platforms that a plant produces.
Body-on-frame (BoF)	It is a binary variable to indicate whether a plant produces Body-on-Frame cars like pick-up truck.
Total welds (TW)	Spot weld equivalents performed in-house by a plant. Spot weld equivalents is the production volume weighted average of the spot weld equivalents across all products at a plant. The spot weld equivalent is a means of normalizing all weld types (e.g., laser, MIG, Gas, and etc.) to a common "spot weld equivalent" using different factors.
Welding automation (WA)	The Bodyshop automation level(%) in a plant which is calculated by (automated stations + 0.5×semi-automated stations)/(automated stations + semi-automated stations + manual stations). It only accounts for mainline stations, not including sub assemblies.
The number of models (NoM)	The number of vehicle models that a plant produces.
Sealer automation (SA)	The automation level(%) of sealer operation in Paintshop in a plant which is the production volume weighted average of the percentage of sealer applied automatically across all products at a plant.
Paint content automation (PCA)	The automation level(%) of painting operation in Paintshop in a plant which is the production volume weighted average of the percentage of paint applied automatically across all products at the plant.
Paint content outsourcing (PCO)	The outsourcing level(%) of Paintshop in a plant which is calculated by 100%-(the percentage of in-plant paint content).
Part numbers (PN)	The production volume weighted average of parts attached to a vehicle across all products in GA of a plant.
GA outsourcing (GAO)	The outsourcing level(%) of GA in a plant which is calculated by 100%-(the percentage of in-plant assembly content).
Logistic outsourcing (LO)	The outsourcing level(%) of logistic operation of a plant which is calculated by 100%-(the percentage of in-plant logistic service content).
Parts received (PR)	The number of part numbers received by a plant, including fasteners, delivered by suppliers and assembled in GA. The part number is a unique identifier assigned to parts handled. Be sure that the part number should not be mingled with the number of parts.

References

1. MacDuffie, J.P.; Sethuraman, K.; Fisher, M.L. Product variety and manufacturing performance: Evidence from the international automotive assembly plant study. *Manag. Sci.* **1996**, *42*, 350–369. [CrossRef]
2. Gopal, A.; Goyal, M.; Netessine, S.; Reindorp, M. The impact of new product introduction on plant productivity in the north american automotive industry. *Manag. Sci.* **2013**, *59*, 2217–2236. [CrossRef]
3. Aigner, D.; Lovell, C.; Schmidt, P. Formulation and estimation of stochastic frontier production function models. *J. Econom.* **1977**, *6*, 21–37. [CrossRef]
4. Meeusen, W.; van Den Broeck, J. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *Int. Econ. Rev.* **1977**, *18*, 435–444. [CrossRef]
5. Charnes, A.; Cooper, W.; Rhodes, E. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
6. Krafcik, J.F. Comparative Analysis of Performance Indicators at World Auto Assembly Plants. Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1988.
7. Lieberman, M.B.; Lau, L.J.; Williams, M.D. Firm-level productivity and management influence: A comparison of u.s. and japanese automobile producers. *Manag. Sci.* **1990**, *36*, 1193–1215. [CrossRef]
8. Fisher, M.L.; Ittner, C.D. The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Manag. Sci.* **1999**, *45*, 771–786. [CrossRef]
9. Lieberman, M.B.; Demeester, L. Inventory reduction and productivity growth: Linkages in the japanese automotive industry. *Manag. Sci.* **1999**, *45*, 455–620. [CrossRef]
10. Heckman, J.J. Sample selection bias as a specification error. *Econometrica* **1979**, *47*, 153–161. [CrossRef]
11. Alden, J.M.; Burns, L.D.; Costy, T.; Hutton, R.D.; Jackson, C.A.; Kohls, D.S.K.K.A.; Owen, J.H.; Turnquist, M.A.; Veen, D.J.V. General motors increases its production throughput. *Interfaces* **2006**, *36*, 6–25. [CrossRef]
12. Kumbhakar, S.C.; Ghosh, S.; McGuckin, J.T. A generalized production frontier approach for estimating determinants of determinants of inefficiency in U.S. Dairy Farms. *J. Bus. Econ. Stat.* **1991**, *9*, 279–286.
13. Huang, C.J.; Liu, J.T. Estimation of a non-neutral stochastic frontier production function. *J. Product. Anal.* **1994**, *5*, 171–180. [CrossRef]
14. Battese, G.; Coelli, T.J. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empir. Econ.* **1995**, *20*, 325–332. [CrossRef]
15. Meesters, A. A note on the assumed distributions in stochastic frontier models. *J. Product. Anal.* **2014**, *42*, 171–173. [CrossRef]
16. van den Broeck, J.; Koop, G.; Osiewalski, J.; Steel, M. Stochastic frontier models: A Bayesian perspective. *J. Econom.* **1994**, *61*, 273–303. [CrossRef]
17. Greene, W.H. A Gamma-distributed stochastic frontier model. *J. Econom.* **1990**, *46*, 141–163. [CrossRef]
18. Beckers, D.E.; Hammond, C.J. A tractable likelihood function for the normal-gamma stochastic frontier model. *Econ. Lett.* **1987**, *24*, 33–38. [CrossRef]
19. Greene, W.H. Simulated likelihood estimation of the normal-gamma stochastic frontier function. *J. Product. Anal.* **2003**, *19*, 179–190. [CrossRef]
20. Tsionas, E.G. Maximum likelihood estimation of stochastic frontier models by the Fourier transform. *J. Econom.* **2012**, *170*, 234–248. [CrossRef]
21. Kumbhakar, S.C.; Parmeter, C.F.; Tsionas, E.G. A zero inefficiency stochastic frontier model. *J. Econom.* **2013**, *172*, 66–76. [CrossRef]
22. Makieła, K. Bayesian Inference and Gibbs Sampling in Generalized True Random-Effects Models. *Cent. Eur. J. Econ. Model. Econom.* **2017**, *9*, 69–95.
23. Moschopoulos, P.G. The distribution of the sum of independent gamma random variables. *Ann. Inst. Stat. Math.* **1985**, *37*, 541–544. [CrossRef]
24. Bibinger, M. Notes on the sum and maximum of independent exponentially distributed random variables with different scale parameters. *arXiv* **2013**, arXiv:1307.3945.
25. Feller, W. *An Introduction to Probability Theory and Its Applications. Volume II*; John Wiley & Sons Inc.: New York, NY, USA, 1971.
26. Krenek, R.; Cha, J.; Cho, B.R. Development of the convolutions of truncated normal random variables with three different quality characteristics in engineering applications. *Comput. Ind. Eng.* **2016**, *94*, 125–137. [CrossRef]
27. Genton, M.G. *Skew-Elliptical Distributions and Their Applications: A Journey beyond Normality*; CRC Press: Boca Raton, FL, USA, 2004.
28. Nadarajah, S.; Li, R. The exact density of the sum of independent skew normal random variables. *J. Comput. Appl. Math.* **2017**, *311*, 1–10. [CrossRef]
29. Majumdar, S.K. Incentive regulation and productive efficiency in the US telecommunications industry. *J. Bus.* **1997**, *70*, 547–576. [CrossRef]
30. Xie, C.; Li, L. The empirical test on investment efficiency and influence of equity incentive in supply-side structural reform: Based on the two-tier stochastic frontier approach. *Appl. Stoch. Model. Bus. Ind.* **2018**, *34*, 730–745. [CrossRef]
31. Boyd, G.A. Estimating plant level energy efficiency with a stochastic frontier. *Energy J.* **2008**, *29*, 23–43. [CrossRef]
32. Boyd, G.A. Estimating the changes in the distribution of energy efficiency in the U.S. automobile assembly industry. *Energy Econ.* **2014**, *42*, 81–87. [CrossRef]

33. Oh, S.C.; Hildreth, A.J. Estimating the technical improvement of energy efficiency in the automotive industry–stochastic and deterministic frontier benchmarking approaches. *Energies* **2014**, *7*, 1–27. [CrossRef]
34. Oh, S.C.; Hildreth, A.J. *Analytics for Smart Energy Management*; Springer: Berlin/Heidelberg, Germany, 2016.
35. Zhu, C.L.; Shi, P.; An, L.R. Human capital, its structure and R&D innovation efficiency based on the test of SFA model. *J. Ind. Eng. Eng. Manag.* **2012**, *4* , 58–64.
36. Díaz González, E.; López Zepeda, J.C.; Garduño Rivera, R. Technical efficiency in the Mexican automotive industry, 1988–2008. *Probl. Del Desarro.* **2019**, *50*, 141–175.
37. Oran, M.; Kerrani, O.; Yolalan, R. An empirical study on analyzing the productivity of bank branches. *IIE Trans.* **1992**, *24*, 166–176. [CrossRef]
38. Johnson, A.; McGinnis, L. Performance measurement in the warehousing industry. *IIE Trans.* **2010**, *43*, 220–230. [CrossRef]
39. Bowman, A.W.; Azzalini, A. *Applied Smoothing Techniques for Data Analysis*; Oxford University Press Inc.: New York, NY, USA, 1997.
40. King, G. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*; Cambridge University Press: New York, NY, USA, 1989.
41. Quarteroni, A.; Sacco, R.; Saleri, F. *Numerical Mathematics*; Springer: Berlin/Heidelberg, Germany, 2006.
42. Koop, G.; Osiewalski, J.; Steel, M.F. Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *J. Econom.* **1997**, *76*, 77–105. [CrossRef]