Macroeconomic Dynamics (2025), **29**, e95, pp. 1–51
doi:10.1017/S1365100525000148

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Distance functions and the analysis of inefficiency[†]

Muna Esheba and Apostolos Serletis

Department of Economics, University of Calgary, Calgary, AB, Canada
Corresponding author: Muna Esheba; Email: mesheba@ucalgary.ca

Abstract

Efficiency is a crucial factor in productivity growth and the optimal allocation of resources in the economy; therefore, measuring inefficiency is particularly important. This paper provides a comprehensive review of the latest developments in distance functions and the measurement of inefficiency within the stochastic frontier framework. Recent advances in several related areas are reviewed and evaluated, including various approaches to measuring inefficiency using distance functions, advancements in modeling inefficiency within the stochastic frontier framework, and the most common estimation techniques. A practical guide is provided on when these methods can be applied and how to implement them. The radial, hyperbolic, and directional measures of inefficiency are discussed and assessed. The development of modeling inefficiency concerning its temporal behavior, classification, and determinants is also examined. To ensure the use of appropriate estimation techniques, recent advancements in the most common estimation techniques are reviewed. This paper also addresses the importance of maintaining the theoretical regularity applied by neoclassical microeconomic theory when it is violated, as well as the econometric regularity when variables are non-stationary. Without regularity, inefficiency results can be extremely misleading. The paper discusses significant challenges related to estimation issues that must be managed in future applications. These challenges include the inaccurate choice of functional form, ignoring the possibility of heterogeneity and heteroskedasticity, and suffering from the endogeneity problem. The paper also examines various approaches to addressing these issues, as well as potentially productive areas for future research.

Keywords: Inefficiency; Distance functions; Stochastic frontier analysis

JEL Classification: B41; C51; D21

1. Introduction

Classical exogenous and contemporary endogenous theories of economic growth illustrate how enhancements in productivity and the efficient allocation of resources positively influence gross domestic product and, ultimately, economic growth. Efficiency is a crucial factor in productivity growth and the optimal allocation of resources in the economy; thus, measuring inefficiency is particularly important. The efficiency of a production unit is determined by comparing it to the most efficient production frontier. This comparison involves comparing observed input to the minimum potential input required to produce the output, comparing observed output to the maximum potential output obtainable from the input, or a combination of both. Optimality is defined in terms of production frontiers and value duals, such as cost, revenue, and profit frontiers.

Battese (1992) conducted a survey on production frontiers and technical efficiency, emphasizing econometric models and their empirical applications in agricultural economics. Greene (1993) offered an extensive review of the econometric approach to both technical and allocative

[†]This paper builds upon material from Chapter 1 of Muna Esheba's doctoral dissertation at the University of Calgary.

efficiency. Darku, et al. (2013) provided a thorough historical review that analyzed various studies on agricultural efficiency, assessing their methodologies and key findings. Nonetheless, recent years have seen significant theoretical and empirical advancements in the efficiency literature.

This paper provides a comprehensive review of the latest developments in distance functions and the measurement of inefficiency within the stochastic frontier framework. Distance functions are advantageous as they measure how close a producer is to the efficient production frontier or to the optimal benchmark, such as cost minimization, revenue, or profit maximization, thereby serving as a direct measure of inefficiency. The primary benefit of the stochastic frontier approach is its ability to separate the error term from inefficiency, thereby providing more precise measures of inefficiency. Recent advancements in several related areas are reviewed and evaluated, including various approaches to measuring inefficiency using distance functions, the development of modeling inefficiency in the stochastic frontier framework, and the most common estimation techniques.

The paper examines the radial measure of inefficiency, as defined by standard distance functions; the hyperbolic measure, as provided by the hyperbolic distance function; and the directional measure, as defined by directional distance functions.

The development of modeling inefficiency concerning its temporal behavior, classification, and determinants is also discussed. Initially, in stochastic frontier models, inefficiency is assumed to be time-invariant in both cross-sectional and panel data models. This assumption is later relaxed with the introduction of time-variant inefficiency models, which allow inefficiency to vary over time and among individual producers. Time-invariant and time-variant inefficiency models are developed to address both inefficiency components. More recently, models incorporating four random components have been proposed to address both inefficiencies and heterogeneous technology. Dynamic inefficiency models have been introduced to capture the dynamic nature of inefficiency, where inefficiency evolves through an autoregressive process in which past inefficiency values influence the current value.

In contrast to models that allow for the existence of extremely inefficient producers who cannot survive in highly competitive markets, threshold inefficiency models truncate the distribution of inefficiency by placing a threshold parameter of the minimum efficiency required for survival on inefficiency. These models, therefore, define an upper bound for the distribution of inefficiency, in addition to the zero lower bound. While threshold inefficiency models focus on the possibility of inefficient producers being out of the markets, the zero inefficiency models highlight the possibility of producers being fully efficient. Zero inefficiency models can incorporate both fully efficient and inefficient producers within a probabilistic framework.

Heterogeneous inefficiency models are proposed to capture heterogeneity in the inefficiency component by incorporating characteristics specific to each producer. These characteristics can be integrated into the inefficiency component itself, or into the mean, variance, or both parameters of the inefficiency distribution.

To ensure the use of appropriate estimation techniques, recent advancements in the most common estimation techniques are reviewed. This paper also addresses the importance of maintaining the theoretical regularity applied by neoclassical microeconomic theory when it is violated, as well as the econometric regularity when variables are non-stationary. Without regularity, inefficiency results can be extremely misleading. The paper further discusses techniques for imposing theoretical regularity and explores integration and cointegration methods that can be used to address the non-stationarity of the residuals.

This paper addresses significant challenges related to estimation issues that must be managed in future applications. These challenges include the inaccurate choice of functional form, ignoring the possibility of heterogeneity and heteroskedasticity, and suffering from the endogeneity problem.

The estimates of inefficiency can be distorted by an inaccurate choice of the functional form for production technology. This paper discusses several criteria for selecting a specific functional form for the production technology, based on theoretical properties such as the shape of the isoquants, separability, flexibility, and regular regions, as well as application properties like homogeneity and translation properties. Additionally, this paper addresses empirical techniques that can be used to assess the ability of different functional forms to approximate the unknown underlying function.

The selection of an appropriate functional form is insufficient without accommodating heterogeneous technologies that may exist among producers or heterogeneity in the inefficiency term. Ignoring heterogeneity can lead to incorrect conclusions regarding inefficiency measures because heterogeneity not captured by producer-specific characteristics is wrongly attributed to inefficiency. This paper addresses the importance of accommodating heterogeneity and discusses different approaches to account for both heterogeneous technologies and heterogeneity in the inefficiency term while estimating inefficiency.

Another potential issue in estimating inefficiency using distance functions is that inputs and outputs may be endogenous, which may lead to biased and inconsistent estimates of the parameters of the production technology and the associated measures of inefficiency. This paper discusses various approaches to addressing this issue.

The remainder of the paper is organized as follows: The next section presents the theoretical background on radial, hyperbolic, and directional measures of inefficiency using distance functions. Section 3 reviews the development of modeling inefficiency within the stochastic frontier framework. Section 4 provides a brief review of recent advances in the most common estimation techniques. Section 5 discusses estimation issues, and the last section concludes the paper with a brief discussion of the important issues that should be addressed in future applications when estimating inefficiency, as well as potentially productive areas for future research.

2. Distance functions and the measurement of inefficiency

There are various methods to measure inefficiency using distance functions. Inefficiency can be assessed radially with standard distance functions, hyperbolically with hyperbolic distance functions, or directionally with directional distance functions. It can also be evaluated by utilizing the duality between distance functions and cost, revenue, and profit functions. The selection of a method may depend on several criteria: the objectives of the producers, exogeneity assumptions, data availability, and the complexity of the estimation procedures.

To briefly review some of the literature on radial, hyperbolic, and directional measures of inefficiency using distance functions, consider a producer employing a vector of n inputs $x = (x_1, \ldots, x_n) \in \mathbb{R}^n_+$ available at fixed prices $w = (w_1, \ldots, w_n) \in \mathbb{R}^n_{++}$ to produce a vector of m outputs $y = (y_1, \ldots, y_m) \in \mathbb{R}^m_+$ that can be sold at fixed prices $p = (p_1, \ldots, p_m) \in \mathbb{R}^m_{++}$. Let $L(y)$ be the set of all input vectors x which can produce the output vector y

$$L(y) = \{x = (x_1, \ldots, x_n) \in R^N_+ : x \text{ can produce } y\}$$

and let $P(x)$ be the feasible set of outputs y that can be produced from the input vector x

$$P(x) = \{y = (y_1, \ldots, y_m) \in R^m_+ : y \text{ is producible from } x\}$$

The production technology T for a producer is defined as the set of all feasible input-output vectors

$$T = \{(x, y) : x \in \mathbb{R}^n_+, y \in \mathbb{R}^m_+, x \text{ can produce } y\}$$

Note that $(x, y) \in T \Leftrightarrow x \in L(y) \Leftrightarrow y \in P(x)$.

2.1 The Input and Output-Oriented Radial Measures

The radial measures of technical inefficiency are provided by standard distance functions. Distance functions were initially defined on the input or output production possibility sets by

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

Debreu (1951) and Shephard (1953, 1970). The input distance function examines how much the input vector can be proportionally contracted while the output vector remains fixed. Conversely, the output distance function examines how much the output vector can be proportionally expanded while the input vector remains fixed.

Input (output) distance functions do not consider the opposite output (input) orientation. This limitation can be challenging if the adjustability of both inputs and outputs is required. Therefore, the choice between input and output distance functions as a representation of production technology should be based on the producers' objectives and the assumption of exogeneity for inputs and outputs. If the goal of producers is to minimize costs, which involves choosing the optimal quantities of inputs to produce a specific output vector, the input distance function can be utilized to estimate inefficiency in the cost-minimization problem. Conversely, if the goal of producers is to maximize revenues, which involves producing the optimal quantities of outputs from a given input vector, the output distance function can be employed to estimate inefficiency in the revenue-maximization problem. Additionally, if producers have significant control over decisions regarding input usage (output production), selecting the input (output) distance function is more appropriate.

2.1.1 The input distance function

Following Shephard (1953), the input distance function (IDF) can be defined in relation to the input set $L(y)$ or the production technology T as follows:

$$D_I(y,x) = \max_{\vartheta_I}\left\{\vartheta_I\colon \frac{x}{\vartheta_I} \in L(y)\right\} = \max_{\vartheta_I}\left\{\vartheta_I\colon \left(\frac{x}{\vartheta_I}, y\right) \in T\right\}$$

where $1/\vartheta_I$ represents the proportional contraction of inputs required to reach the inner boundary of the input set, or the production frontier, with the outputs held constant. The function $D_I(y,x)$ is defined as the ratio of the observed input to the minimum input required to produce the given output. Therefore, for any input vector x, the expression $x/D_I(y,x)$ represents the minimum input vector on the ray from the origin through x that can produce y, as illustrated in Figure 1. Efficient producers, who produce on the boundary of the input set, or the production frontier, have $D_I(y,x) = 1$. Inefficiency is indicated by $D_I(y,x) > 1$.

The Debreu-Farrell input-oriented measure of technical efficiency is defined as follows:

$$TE_I(y,x) = \min_{\vartheta_{FI}}\left\{\vartheta_{FI}\colon \vartheta_{FI}x \in L(y)\right\} = \min_{\vartheta_{FI}}\left\{\vartheta_{FI}\colon (\vartheta_{FI}x, y) \in T\right\}$$

The Debreu-Farrell input-oriented measure of technical efficiency is defined as the reciprocal of the *IDF*: $TE_I(y,x) = [D_I(y,x)]^{-1}$. The expression $TE_I(y,x) \le 1$ represents a radial reduction of inputs required to be considered as being efficient. Technical inefficiency is defined as follows:

$$TI_I(y,x) = 1 - TE_I(y,x) = 1 - \frac{1}{D_I(y,x)}$$

where $0 \le TI_I(y,x) \le 1$. The input distance function has the following properties [see Färe and Primont (1995), and Färe and Grosskopf (2004) for more details]

i) representation, $D_I(y,x) \ge 1$ iff $x \in L(y)$ or $(x,y) \in T$
ii) non-increasing and quasi-concave in outputs, and
iii) non-decreasing, concave, and linearly homogeneous in inputs, $D_I(y,\lambda x) = \lambda D_I(y,x)$, $\lambda > 0$.

The input distance function has been applied across various sectors to assess efficiency. In the manufacturing sector, it has been used to evaluate the efficiency of production processes,

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

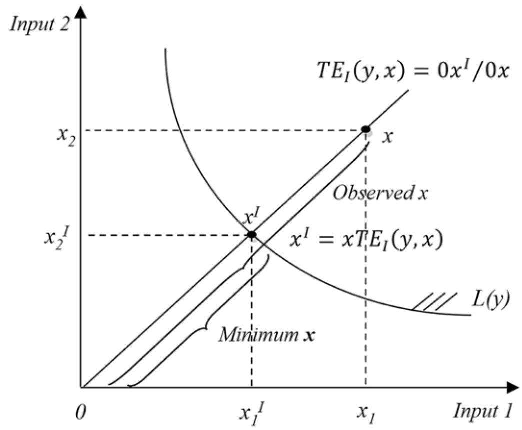

Figure 1. The Debreu-Farrell input-oriented measure of technical efficiency.

as shown by Hattori (2002) and Atkinson et al. (2003). In the banking sector, it has been utilized to measure the efficiency of banks by examining the use of inputs, such as capital, labor, and other resources, in providing financial services and identifying strategies for performance enhancement, as discussed by Sturm and Williams (2008). Additionally, the IDF has been applied to evaluate the efficiency of agricultural production by analyzing how inputs are transformed into outputs and identifying areas for improvement, as demonstrated by Tsionas et al. (2015). Furthermore, the IDF has been utilized to address undesirable outputs. This is modeled by holding desirable outputs y constant and treating undesirable outputs, denoted as b, as inputs x; $D_I\left(y, x, b\right) = \max_{\vartheta_I}\left\{\vartheta_I\colon \left(\frac{x}{\vartheta_I}, \frac{b}{\vartheta_I}\right) \in L\left(y\right)\right\} = \max_{\vartheta_I}\left\{\vartheta_I\colon \left(\frac{x}{\vartheta_I}, \frac{b}{\vartheta_I}, y\right) \in T\right\}$. See, for example, Atkinson and Dorfman (2005). This approach credits producers for proportionally reducing both inputs and undesirable outputs to reach the production frontier. However, if inputs are freely disposable, undesirable outputs are as well. This treatment of undesirable outputs has been criticized due to the implied strong disposability of undesirable outputs, as noted by Färe et al. (2005). Similarly, treating undesirable outputs as inputs allows for substitutability or complementarity among them. Furthermore, these studies overlook the fact that the production of undesirable outputs is influenced by the production of desirable outputs, represented as $b = f(y)$, rather than the opposite, $y = f(b)$. Thus, treating undesirable outputs as inputs is inappropriate because it imposes incorrect theoretical restrictions on the production technology. Therefore, the IDF can be used only to decrease inputs while keeping both outputs constant. Assaf et al. (2013) utilized the IDF and treat undesirable output as a technology shifter.

2.1.2 The output distance function

Instead of looking at how much the input vector x may be proportionally contracted with the output vector y held fixed, the output distance function (ODF) introduced by Shephard (1970), considers by how much the output vector may be proportionally expanded with the input vector held fixed. It is defined on the output set $P\left(x\right)$ or the production technology T as follows:

$$D_O\left(x, y\right) = \min_{\vartheta_O}\left\{\vartheta_O\colon \frac{y}{\vartheta_O} \in P\left(x\right)\right\} = \min_{\vartheta_O}\left\{\vartheta_O\colon \left(x, \frac{y}{\vartheta_O}\right) \in T\right\}$$

where $1/\vartheta_O$ represents the proportional expansion of outputs required to reach the upper boundary of the output set or the production frontier, holding the inputs constant. The function $D_O\left(x, y\right)$ is defined as the ratio of the observed output to maximum potential output obtainable from the given input. Therefore, for any output vector y, the expression $y/D_O\left(x, y\right)$ represents the

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

largest output vector on the ray from the origin through y that can be produced by x, as illustrated in Figure 2. If y is on the boundary of the output set or the production frontier, then $D_O(x, y) = 1$, which implies that the producer operates at full technical efficiency. If y is within the boundary of the output set or the production frontier, then $D_O(x, y) < 1$, indicating that the producer operates with technical inefficiency.

The Debreu-Farrell output-oriented measure of technical efficiency is defined as follows:

$$TE_O(x, y) = \max_{\vartheta_{FO}} \left\{ \vartheta_{FO}\colon \vartheta_{FO} y \in P(x) \right\} = \max_{\vartheta_{FO}} \left\{ \vartheta_{FO}\colon (x, \vartheta_{FO} y) \in T \right\}$$

The Debreu-Farrell output-oriented measure of technical efficiency is defined as the reciprocal of the ODF: $TE_O(x, y) = [D_O(x, y)]^{-1}$. The expression $TE_O(x, y) \geq 1$ represents a radial expansion of outputs required to achieve efficiency; the greater this measure, the lower the efficiency. Technical inefficiency is defined as follows:

$$TI_O(x, y) = TE_O(x, y) - 1 = \frac{1}{D_O(x, y)} - 1$$

where $TI_O(x, y) \geq 0$. The output distance function has the following properties [see Färe and Grosskopf (1994) for more details]

i) representation, $D_O(x, y) \leq 1$ iff $y \in P(x)$ or $(x, y) \in T$
ii) non-increasing and quasi-convex in inputs, and
iii) non-decreasing, convex and linearly homogeneous in outputs, $D_O(x, \lambda y) = \lambda D_O(x, y)$, $\lambda > 0$.

The output distance function has been applied across various sectors to accommodate multiple-output technologies and assess efficiency. In the banking sector, it has been used to evaluate the efficiency of banks in providing diverse services, such as loans and other financial services, and to identify banks that may not perform efficiently, as demonstrated by Cuesta and Orea (2002) and Almanidis et al. (2019). In the public sector, the output distance function has been applied to measure the efficiency of various public services, such as education, and to determine how efficiently financial and human resources are used to deliver various educational activities, as discussed by Letti et al. (2022). Additionally, it has been utilized to measure the efficiency of public hospitals in providing medical services. By focusing on the outputs, these functions assist in evaluating and improving healthcare delivery, as demonstrated by Devitt et al. (2024). Furthermore, the ODF has been applied to address undesirable outputs. This is modeled by holding inputs x constant and treating undesirable outputs, b, as desirable outputs, y; $D_O(x, y) = \min_{\vartheta_O} \left\{ \vartheta_O\colon \left(\frac{y}{\vartheta_O}, \frac{b}{\vartheta_O} \right) \in P(x) \right\} = \min_{\vartheta_O} \left\{ \vartheta_O\colon \left(x, \frac{y}{\vartheta_O}, \frac{b}{\vartheta_O} \right) \in T \right\}$. See, for example, Färe et al. (1993) and Färe et al. (1989). This approach credits producers for proportionally expanding both desirable and undesirable outputs to reach the production frontier. However, this is applicable only if the adjustability of both types of outputs is required. Producers have no control over reducing undesirable outputs without also reducing desirable outputs, and producing more desirable outputs requires producing more undesirable outputs, such as generating pollution as byproducts of producing desirable outputs. Furthermore, undesirable outputs are inevitably produced unless the entire production process is terminated.

The standard input and output distance functions adjust both desirable and undesirable outputs proportionally at the same rate. This approach may not align with the objectives of policy-makers who aim to simultaneously reduce undesirable outputs and increase desirable ones. Furthermore, these standard distance functions treat technical inefficiency as environmental inefficiency. Future studies that compare and differentiate these inefficiencies should contribute to a deeper understanding of their distinctions.

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

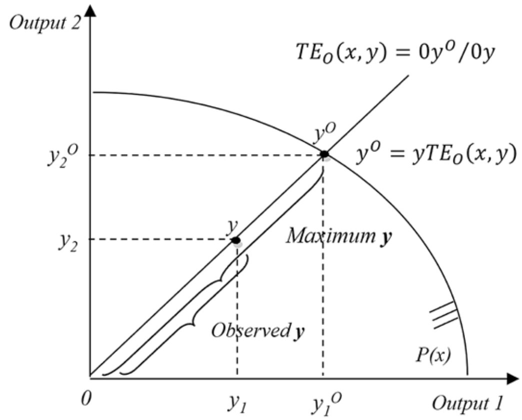

Figure 2. The Debreu-Farrell output-oriented measure of technical efficiency.

2.2 The hyperbolic measure

The hyperbolic measures of technical inefficiency are determined by hyperbolic distance functions. According to Färe et al. (1985), the hyperbolic distance function (HDF) is defined in relation to the production technology T as follows:

$$D_H(x,y) = \min_{\vartheta_H} \left\{ \vartheta_H \colon \left(\vartheta_H x, \frac{y}{\vartheta_H} \right) \in T \right\}$$

where $1 \geq \vartheta_H > 0$ represents the proportional contraction of inputs and expansion of outputs required to reach the production frontier. It is important to note that reducing ϑ_H implies expanding $1/\vartheta_H$. This is illustrated in figure 3 where the hyperbolic curve intersects with the production frontier at point $H = \left(\vartheta_H x, \frac{y}{\vartheta_H}\right)$. Efficient producers who produce on the boundary of the production frontier, have $D_H(x,y) = 1$. Inefficiency is indicated by $D_H(x,y) < 1$. The hyperbolic measure of technical efficiency, as proposed by Färe et al. (1985), is defined as follows:

$$TE_H(x,y) = \max_{\vartheta_{FH}} \left\{ \vartheta_{FH} \colon \left(\frac{x}{\vartheta_{FH}}, \vartheta_{FH} y \right) \in T \right\}$$

Note that the hyperbolic measure of technical efficiency is the reciprocal of the HDF, $TE_H(x,y) = \left[D_H(x,y)\right]^{-1}$. To measure profitability (or return-to-dollar) efficiency, Zofio and Prieto (2006) defined the hyperbolic measure of technical efficiency as $TE_H(x,y) = \min_{\vartheta_H} \left\{ \vartheta_H \colon \left(\vartheta_H x, \frac{y}{\vartheta_H} \right) \in T \right\} = \vartheta_H$. For technology frontiers with variable returns to scale, Nahm and Vu (2013) demonstrated that the hyperbolic measure of technical efficiency is the square of the HDF, $TE_H(x,y) = \left[D_H(x,y)\right]^2$. It is assumed that $TE_H(x,y) \geq 1$ under the condition of weak disposability of inputs and outputs. Technical inefficiency is defined as follows:

$$TI_H(x,y) = TE_H(x,y) - 1 = \frac{1}{D_H(x,y)} - 1$$

where $TI_H(x,y) \geq 0$. Färe et al. (2002) demonstrated that under constant returns to scale, the HDF is related to the standard input and output distance functions as $D_H(x,y) = \left[D_I(y,x)\right]^{-1/2} = \left[D_O(x,y)\right]^{1/2}$. Another type of relationship was developed by Simar and Vanhems (2012) and Daraio and Simar (2014) between the HDF and the directional technology distance function, expressed as $\ln D_H(x^*, y^*) = \vec{D}_T(x, y; g_x, g_y)$ where $x^* = \exp(x/g_x)$ and

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

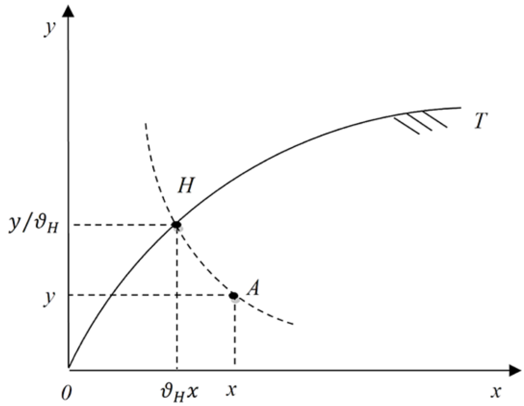

Figure 3. The hyperbolic measures of technical efficiency.

$y^* = \exp\left(y/g_y\right)$. The HDF has the following properties [see Färe et al. (1985, 1994) for more details]

i) representation, $D_H\left(x, y\right) \leq 1$ iff $\left(x, y\right) \in T$
ii) non-decreasing in outputs and non-increasing in inputs
iii) homogeneity, $D_H\left(\lambda^{-1}x, \lambda y\right) = \lambda D_H\left(x, y\right), \lambda > 0$
iv) almost homogeneous of degrees k_1, k_2, and k_3 if $D_H\left(\lambda^{k_1}x, \lambda^{k_2}y\right) = \lambda^{k_3}D_H\left(x, y\right)$, and
v) homogeneous of degree zero in inputs and outputs under constant returns to scale.

In contrast to standard distance functions, the hyperbolic distance function simultaneously contracts inputs and expands outputs proportionally, without imposing the restriction of holding either inputs or outputs constant. This function can be utilized to estimate inefficiency, assuming that producers have the capability to adjust both inputs and outputs to improve efficiency. In the banking sector, the HDF has been applied to assess the efficiency of banks by analyzing their delivery of financial services and cost reduction, as illustrated by Cuesta and Zofio (2005) and Chaffai (2020). Furthermore, the HDF can be utilized for the simultaneous contraction of inputs x and undesirable outputs b, and the expansion of desirable outputs y; $D_H\left(x, y, b\right) = \min_{\vartheta_H}\left\{\vartheta_H\colon \left(\vartheta_H x, \frac{y}{\vartheta_H}, \vartheta_H b\right) \in T\right\}$. See, for example, Cuesta et al. (2009), Fang and Yang (2014), and Adenuga et al. (2019).

2.3 The directional measure

Directional distance functions (DDF), unlike standard or hyperbolic distance functions, represent an additive measure of technical inefficiency in a specified direction, g, rather than a proportional or multiplicative one. The additive nature of directional distance functions allows for the inclusion of non-positive inputs or outputs. A significant consideration when using DDF is selecting an appropriate direction in which inefficient producers are projected onto the production frontier.

2.3.1 The directional technology distance function

The directional technology distance function (DTDF) generalizes Shephard's input and output distance functions, providing a tool to address efficiency issues in an integrated manner. Chambers et al. (1998) introduced it as a variant of the Luenberger (1995) shortage function. This

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

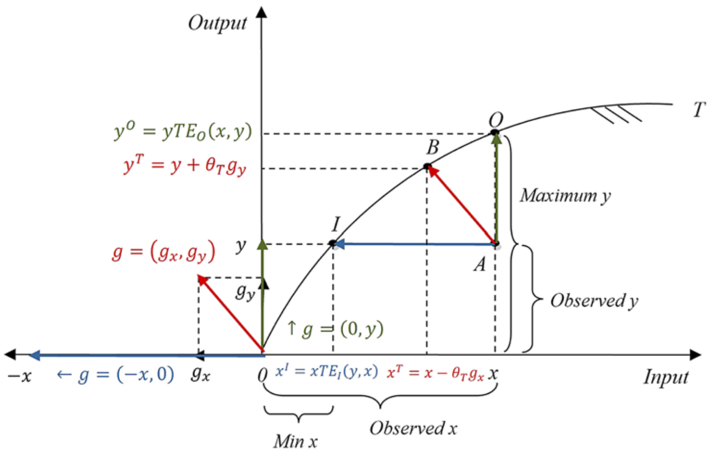

Figure 4. The directional measure of technical inefficiency.

function allows for the simultaneous expansion of y and contraction of x according to a specified direction vector $g = (g_x, g_y)$, where $g_x \in R_+^N$ and $g_y \in R_+^M$. This means that inputs are contracted in the direction of g_x, and outputs are expanded in the direction of g_y. In particular, the directional technology distance function is defined as follows:

$$\vec{D}_T\left(x, y; g_x, g_y\right) = \max_{\theta_T} \left\{\theta_T \colon \left(x - \theta_T g_x, y + \theta_T g_y\right) \in T\right\} \tag{1}$$

Efficient producers operating on the frontier of T have $\vec{D}_T\left(x, y; g_x, g_y\right) = 0$, indicating that no further expansion of outputs or contraction of inputs is feasible. Inefficiency is indicated by $\vec{D}_T\left(x, y; g_x, g_y\right) > 0$, with higher values denoting greater inefficiency when producers operate below the frontier of T. Eliminating technical inefficiency for those operating at point A would move them to point $B = \left(x^T, y^T\right) = \left(x - \theta_T g_x, y + \theta_T g_y\right)$ on the frontier of T, as illustrated in Figure 4. The DTDF serves as a technology-oriented measure of technical inefficiency.

$$TI_T = \vec{D}_T\left(x, y; g_x, g_y\right)$$

As noted by Chambers et al. (1998), the directional technology distance function has the following properties:

i) representation, $\vec{D}_T\left(x, y; g_x, g_y\right) \geq 0$ iff $(x, y) \in T$
ii) translation, $\vec{D}_T\left(x - \alpha g_x, y + \alpha g_y; g_x, g_y\right) = \vec{D}_T\left(x, y; g_x, g_y\right) - \alpha$, for $\alpha \in R$
iii) non-decreasing in x and non-increasing in y if inputs and outputs are freely disposable
iv) concave in (x, y)
v) homogeneous of degree -1 in g, That is, $\vec{D}_T\left(x, y; \lambda g_x, \lambda g_y\right) = \lambda^{-1}\vec{D}_T\left(x, y; g_x, g_y\right)$, for $\lambda > 0$, and
vi) homogeneous of degree $+1$ in x and y if the technology exhibits constant returns to scale, $\vec{D}_T\left(\lambda x, \lambda y; g_x, g_y\right) = \lambda \vec{D}_T\left(x, y; g_x, g_y\right)$, for $\lambda > 0$.

The directional technology distance function can be utilized to represent technologies that are not separable in desirable and undesirable outputs. This is useful in these production contexts because it allows for non-radial or hyperbolic expansions of desirable outputs and contractions of undesirable outputs, as demonstrated by Malikov et al. (2016). The DTDF has been applied across various sectors to assess efficiency. In the banking sector, these functions have been utilized to determine bank inputs and outputs and to evaluate the performance and efficiency of

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

banks by assessing how efficiently they convert inputs-such as capital, labor, and other resources-into outputs, like loans and other financial services, while minimizing undesirable outputs, such as non-performing loans. For examples, see Koutsomanoli-Filippaki et al. (2012), Guarda et al. (2013), and Huang and Chung (2017). In the public sector, the DTDF has been employed to measure the efficiency of public hospitals in delivering multiple health services by utilizing their resources, such as medical staff, equipment, and other resources, and minimizing undesirable outcomes. For examples, see Vardanyan et al. (2022).

2.3.2 The directional input distance function

The inefficiency measures derived from the directional distance function depend on the directional vector, $g = (g_x, g_y)$. By setting $g_y = 0$, the directional vector becomes $g = (g_x, 0)$, which allows for input contraction with outputs held constant, as illustrated in Figure 4. In this context, equation (1) transforms into the directional input distance function (DIDF), permitting only input contraction; $\vec{D}_T(x, y; g_x, 0) = \vec{D}_I(y, x; g_x)$

$$\vec{D}_I(y, x; g_x) = \max_{\theta_I}\{\theta_I: (x - \theta_I g_x) \in L(y)\} = \max_{\theta_I}\{\theta_I: (x - \theta_I g_x, y) \in T\}$$

Moreover, according to Chambers et al. (1996, 1998) and Färe and Grosskopf (2000), if the directional input vector, g_x, equals the observed input vector, x, (that is, $g_x = -x$), then

$$\vec{D}_I(y, x; g_x) = \vec{D}_I(y, x; -x) = 1 - \frac{1}{D_I(y, x)}$$

In this context, a relationship exists between the DIDF, $\vec{D}_I(y, x; -x)$, and the Shephard input distance function, $D_I(y, x)$. As illustrated in Figure 4, producers operating at point A can maintain constant output while reducing input in the direction of $g_x = -x$ to reach point I. The directional input distance function serves as an input-oriented measure of technical inefficiency.

$$TI_I = \vec{D}_I(y, x; g_x)$$

The directional input distance function satisfies the following properties, as outlined by Chambers et al. (1996):

i) representation, $\vec{D}_I(y, x; g_x) \geq 0$ iff $x \in L(y)$ or $(x, y) \in T$
ii) translation, $\vec{D}_I(y, x - \alpha g_x; g_x) = \vec{D}_I(y, x; g_x) - \alpha$, for $\alpha \in R$
iii) concavity in inputs
iv) positive monotonicity in inputs. That is, $x' > x$ implies $\vec{D}_I(y, x'; g_x) \geq \vec{D}_I(y, x; g_x)$
v) negative monotonicity in outputs. That is, $y' > y$ implies $\vec{D}_I(y', x; g_x) \leq \vec{D}_I(g, x; g_x)$, and
vi) homogeneity of degree -1 in g_x. That is, $\vec{D}_I(y, x; \lambda g_x) = \lambda^{-1}\vec{D}_I(y, x; g_x)$, for $\lambda > 0$.

The directional input distance function has been applied across various sectors to evaluate efficiency. In the public sector, researchers have utilized it to compare the performance of public hospitals with that of hospitals in other countries. This comparison assesses how efficiently hospitals utilize their inputs compared to their peers, identifying best practices and areas for improvement, as demonstrated by Dervaux et al. (2004). Furthermore, the DIDF has been employed to evaluate the efficiency of dairy farming by analyzing the use of various resources, such as labor, equipment, and other inputs, to determine how efficiently these resources are utilized to produce dairy products, as shown by Serra et al. (2011).

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

2.3.3 The directional output distance function

By setting $g_x = 0$, the directional vector becomes $g = \left(0, g_y\right)$, which allows for output expansion while inputs remain constant, as illustrated in Figure 4. In this context, equation (1) simplifies to the directional output distance function (DODF), allowing solely for the expansion of output; $\vec{D}_T\left(x, y; 0, g_y\right) = \vec{D}_O\left(x, y; g_y\right)$

$$\vec{D}_O\left(x, y; g_y\right) = \max_{\theta_o}\left\{\theta_O : \left(y + \theta_O g_y\right) \in P\left(x\right)\right\} = \max_{\theta_o}\left\{\theta_O : \left(x, y + \theta_O g_y\right) \in T\right\}$$

Moreover, as noted by Chambers et al. (1998) and Färe and Grosskopf (2000), if the directional output vector, g_y, equals the observed output vector, y (that is, $g_y = y$), then

$$\vec{D}_O\left(x, y; g_y\right) = \vec{D}_O\left(x, y; y\right) = \frac{1}{D_O\left(x, y\right)} - 1$$

In this context, a relationship exists between the DODF, $\vec{D}_O\left(x, y; y\right)$, and the Shephard output distance function, $D_O\left(x, y\right)$. As illustrated in Figure 4, producers operating at point A can maintain constant input while expanding output in the direction of $g_y = y$ to reach point O. The directional output distance function serves as an output-oriented measure of technical inefficiency.

$$TI_O = \vec{D}_O\left(x, y; g_y\right)$$

The directional output distance function satisfies the following properties, as outlined by Färe et al. (2005):

i) representation, $\vec{D}_O\left(x, y; g_y\right) \geq 0$ iff $y \in P\left(x\right)$ or $\left(x, y\right) \in T$

ii) translation, $\vec{D}_O\left(x, y + \alpha g_y; g_y\right) = \vec{D}_O\left(x, y; g_y\right) - \alpha$, for $\alpha \in R$

iii) concavity in outputs

iv) negative monotonicity in outputs. That is, $y' > y$ implies $\vec{D}_O\left(x, y'; g_y\right) \leq \vec{D}_O\left(x, y; g_y\right)$, and

v) homogeneity of degree minus -1 in g_y. That is, $\vec{D}_O\left(x, y; \lambda g_y\right) = \lambda^{-1}\vec{D}_O\left(x, y; g_y\right)$, for $\lambda > 0$.

The directional output distance function has been applied across various agricultural and industrial sectors to accommodate multiple-output technologies, assess efficiency, and mitigate undesirable outputs such as pollution and emissions. This function is employed to address these undesirable outputs, b, with the DODF defined as follows: $\vec{D}_O\left(x, y, b; g_y, g_b\right) = \max_{\theta_o}\left\{\theta_O : \left(y + \theta_O g_y, b - \theta_O g_b\right) \in P\left(x\right)\right\} = \max_{\theta_o}\left\{\theta_O : \left(x, y + \theta_O g_y, b - \theta_O g_b\right) \in T\right\}$; see, for example, Färe et al. (2006), Watanabe and Tanaka (2007), Feng et al. (2018), and Yang et al. (2021).

2.3.4 The directional vector

The measure of inefficiency derived from directional distance functions depends on the choice of the direction vector g, which projects the data onto the frontier of T. Inefficiency is measured by selecting either an exogenous or an endogenous direction vector. The exogenous vector is pre-specified, whereas the endogenous vector determines the direction based on specific internal behavior.

2.3.4.1 Exogenous directional vector. For an exogenous or a pre-specified direction vector, two widely used directions are the unit value direction $g = (-1, 1)$ and the observed input-output direction $g = \left(-x, y\right)$. The unit value direction $g = (-1, 1)$ implies that the amount by which a producer can decrease inputs and increase outputs will be $\vec{D}_T\left(x, y; -1, 1\right) \times 1$ units of x and y; see, for example, Färe et al. (2005). The advantage of choosing this directional vector lies in its simplicity, its ability to be aggregated at the industry level, its normalizing nature, and its convenience in explaining the results of measurement. Specifically, an inefficiency measure based on

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

such a directional vector provides a single number that indicates, regardless of the units of measurement, how many units of each input must be reduced and how many units of each output must be increased to any particular point in the technology set to reach the production frontier. As noted by Färe and Grosskopf (2004), the inefficiency of the industry equals the sum of the directional distance functions for all producers when a common directional vector is chosen for all producers.

Another commonly used pre-specified direction is the observed input-output direction $g = (-x, y)$. This type of directional vector measures the simultaneous maximum proportional expansion of outputs and contraction of inputs that are feasible given the technology. It assumes that a producer can reduce inefficiency by decreasing inputs and increasing outputs in proportion to the initial combination of actual inputs and outputs; see, for example, Färe et al. (2004).

The pre-specified direction vector has been extended in multiple directions. Koutsomanoli-Filippaki et al. (2012) employed the observed input-output averages direction $g = (\overline{x}, \overline{y})$. However, these producer-specific direction vectors cannot be aggregated to the industry level. Tzeremes (2015) utilized a range directional vector $g = (g_x, g_y) = (R, 0)$, where the range of possible input reduction of a specific producer is defined as the input minus the minimum inputs observed: $R_{ik'} = x_{ik'} - \min_k \{x_{ik}\}$ given a set of producers $k = \{1, \ldots..K\}$.

The primary concern with the predetermined direction vector is that the parameter estimates of the production technology, as well as the associated measures of inefficiency, are affected by the choice of direction vectors, as noted by Atkinson and Tsionas (2016) and Esheba and Serletis (2023).

2.3.4.2 Endogenous directional vector. An endogenous direction vector is selected to guide any producer towards the benchmark of minimizing costs or maximizing revenue/profit. This choice depends on the availability of price information and the fulfillment of necessary behavioral assumptions, specifically the behaviors of minimizing costs or maximizing revenue/profit.

Using the direction vector $g = (g^C_{x_1}, g^C_{x_2})$, the directional measure of input inefficiency is determined by projecting any inefficient producer onto the cost-minimizing bundle C, where producers achieve both technical and allocative efficiency. Allocative inefficiency arises from the failure to choose the cost-minimizing input vector given the relative input market prices. For an endogenous direction vector that projects to the cost-minimizing benchmark, see, for example, Malikov et al. (2016). Using the direction vector $g = \left(g^R_{y_1}, g^R_{y_2}\right)$, the directional measure of output inefficiency is determined by projecting any inefficient producer onto the revenue-maximizing bundle R, where producers achieve both technical and allocative efficiency. Allocative inefficiency arises from the failure to choose the revenue-maximizing output vector given the relative output market prices. Esheba (2018) and Esheba and Serletis (2023) provided a set of directions consistent with cost minimization, as well as revenue and profit maximization. Using the direction vector $g = \left(g^\pi_x, g^\pi_y\right)$, the directional measure of overall inefficiency is determined by projecting any inefficient producer to the profit-maximizing bundle π, where producers achieve both technical and allocative efficiency. Allocative inefficiency arises from the failure to choose the profit-maximizing input-output vector given the relative input and output market prices. All profit inefficiency for producers operating below the profit frontier can be regarded as measures of overall technical inefficiency, as noted by Zofio et al. (2013). Feng et al. (2018) discussed an endogenous direction vector projecting to the profit-maximizing benchmark, and Atkinson and Tsionas (2016) provided a set of directions consistent with cost minimization and profit maximization.

Färe et al. (2013) developed a method for selecting direction vectors that are endogenously determined based on exogenous normalization constraints using Data Envelopment Analysis, without the need for price data. Hampf and Kruger (2015) utilized optimization methods to endogenously determined optimal directions for a non-parametric efficiency analysis. However,

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

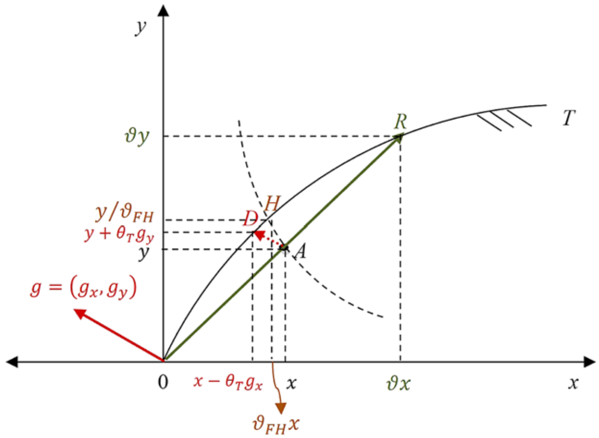

Figure 5. Radial, hyperbolic and directional measures of technical inefficiency.

further research is necessary to compare alternative choices of directional vectors and to establish a framework for determining an optimal set of directions using parametric measurement methods.

To summarize, Figure 5 illustrates the projections of the observed input-output vector at point A using various types of distance functions. The standard distance function projects A proportionally onto R. The radial measure of efficiency is given by $TE_R(x,y) = \max_{\vartheta}\{\vartheta: (\vartheta x, \vartheta y) \in T\}$. However, the radial measure provided by the standard distance function can yield high inefficiency measures even when the observed input-output vector is very close to the frontier, as noted by Hudgins and Primont (2007). The hyperbolic distance function projects A hyperbolically onto H, where the intersection between the hyperbolic curve and the frontier of T is the point $H = (\vartheta_{FH}x, y/\vartheta_{FH})$. The hyperbolic measure of efficiency is expressed as $TE_H(x,y) = \max_{\vartheta_{FH}}\{\vartheta_{FH}: (x/\vartheta_{FH}, \vartheta_{FH}y) \in T\}$. Implementing the hyperbolic measure using the hyperbolic distance function can be complex due to the non-linear optimization involved. The directional technology distance function is particularly well-suited for simultaneously contracting x and expanding y to project A onto D using the direction vector g. The directional measure of inefficiency is technology-oriented and expressed as $\vec{D}_T(x, y; g_x, g_y) = \max_{\theta_T}\{\theta_T: (x - \theta_T g_x, y + \theta_T g_y) \in T\}$. The important properties of the alternative distance functions that can be used for measuring inefficiency and the relationships among them are summarized in Table 1.

3. Modeling inefficiency

The deterministic frontier approach assumes that all deviations from the efficient frontier are under the control of producers and are considered inefficiencies. In contrast, the stochastic frontier approach introduces a random error term that accounts for exogenous stochastic factors beyond the control of producers, in addition to the inefficiency term, in the specification of the frontier model. The primary advantage of the stochastic frontier approach is its ability to separate the error term from inefficiency, thereby providing more precise measures of inefficiency. The stochastic frontier model is presented as follows:

$$Y = f(X; \beta)exp(v - u)$$

where $(v - u)$ is a composed error term, u represents the inefficiency term and v represents random errors associated with random factors that can positively or negatively affect production. Technical efficiency is defined as the ratio of the observed production values to the corresponding

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

Table 1. A summary of the important properties of alternative distance functions

Property Function	Feasibility	Monotonicity	Homogeneity/ Translation	Inputs	Outputs	Inefficiency	Relationships
IDF $D_I(y,x)$ $=\max_{\vartheta_I}:$ $\left(\frac{x}{\vartheta_I}, y\right)$ $\in T$	$D_I(y,x) \geq 1$ iff $(x,y) \in T$ if $(x,y)$is on the frontier of T, then $D_I(y,x)=1$	$\nabla_x D_I(.) \geq 0$ $\nabla_y D_I(.) \leq 0$	homogeneity $D_I(y,\lambda x)$ $=\lambda D_I(y,x)$ $\lambda > 0$	concave	quasi- concave	$TI_I(y,x)$ $=1-TE_I(y,x)$ $=1-\frac{1}{D_I(y,x)}$	under CRS; $D_I(y,x)=$ $1/D_o(x,y)$ duality; $1/D_I(y,x) \geq$ $C(y,w)/wx$
ODF $D_o(x,y)$ $=\min_{\vartheta_o}:$ $\left(x, \frac{y}{\vartheta_O}\right)$ $\in T$	$D_o(x,y) \leq 1$ iff $(x,y) \in T$ if $(x,y)$is on the frontier of T, then $D_o(x,y)=1$	$\nabla_x D_o(.) \leq 0$ $\nabla_y D_o(.) \geq 0$	homogeneity $D_o(x,\lambda y)$ $=\lambda D_o(x,y)$ $\lambda > 0$	quasi- convex	convex	$TI_o(x,y)$ $=TE_o(x,y)-1$ $=\frac{1}{D_o(x,y)}-1$	under CRS; $D_o(x,y)=$ $1/D_I(y,x)$ duality; $1/D_o(x,y) \leq$ $R(x,p)/py$
HDF $D_H(x,y)$ $=\min_{\vartheta_H}:$ $\left(\vartheta_H x, \frac{y}{\vartheta_H}\right)$ $\in T$	$D_H(x,y) \leq 1$ iff $(x,y) \in T$ if $(x,y)$is on the frontier of T, then $D_H(x,y)=1$	$\nabla_x D_H(.) \leq 0$ $\nabla_y D_H(.) \geq 0$	almost homogeneous $D_H(\lambda^{-1}x, \lambda y)$ $=\lambda D_H(x,y)$ under CRS; $D_H(\lambda x, \lambda y)$ $=D_H(x,y)$	convex	convex	$TI_H(x,y)$ $=TE_H(x,y)-1$ $=\frac{1}{D_H(x,y)}-1$	under CRS; $D_H(x,y)$ $=\frac{1}{\sqrt{D_I(y,x)}}$ $=\sqrt{D_O(x,y)}$ duality; $[1/D_H(x,y)]^2$ $\geq py/wx$
DTDF $\vec{D}_T(.)$ $=\max_{\theta_T}:$ $(x-\theta_T g_x,$ $y+\theta_T g_y)$ $\in T$	$\vec{D}_T(.) \geq 0$ iff $(x,y) \in T$ if $(x,y)$is on the frontier of T, then $\vec{D}_T(.)=0$	$\nabla_x \vec{D}_T(.) \geq 0$ $\nabla_y \vec{D}_T(.) \leq 0$	translation $\vec{D}_T(x-\alpha g_x,$ $y+\alpha g_y; g)=$ $\vec{D}_T(x,y;g)-\alpha$ homogeneity $\vec{D}_T(x,y;\lambda g)=$ $\lambda^{-1}\vec{D}_T(x,y;g)$	concave	concave	$TI_T(x,y)$ $=\vec{D}_T(.)$	$\vec{D}_T(x,y;0,g_y)$ $=\vec{D}_o(x,y;g_y)$ $\vec{D}_T(x,y;g_x,0)$ $=\vec{D}_I(y,x;g_x)$ duality; $\vec{D}_T(x,y;g) \leq$ $\frac{\pi(p,w)-(py-wx)}{pg_y+wg_x}$
DIDF $\vec{D}_I(y,x;g_x)$ $=\max_{\theta_I}:$ $(y,x-\theta_I g_x)$ $\in T$	$\vec{D}_I(.) \geq 0$ iff $(x,y) \in T$ if $(x,y)$is on the frontier of T, then $\vec{D}_I(.)=0$	$\nabla_x \vec{D}_I(.) \geq 0$ $\nabla_y \vec{D}_I(.) \leq 0$	translation $\vec{D}_I(y,x-\alpha g_x)=$ $\vec{D}_I(y,x;g_x)-\alpha$ homogeneity $\vec{D}_I(y,x;\lambda g_x)=$ $\lambda^{-1}\vec{D}_I(y,x;g_x)$	concave	quasi- concave	$TI_I(y,x)$ $=\vec{D}_I(.)$	$\vec{D}_I(y,x;-x)=$ $\vec{D}_T(x,y;-x,0)$ $=1-1/D_I(y,x)$ duality; $\vec{D}_I(y,x;g_x) \leq$ $\frac{wx-C(y,w)}{wg_x}$
DODF $\vec{D}_o(x,y;g_y)$ $=\max_{\theta_O}:$ $(x,y+\alpha g_y)$ $\in T$	$\vec{D}_o(.) \geq 0$ iff $(x,y) \in T$ if $(x,y)$is on the frontier of T, then $\vec{D}_o(.)=0$	$\nabla_x \vec{D}_o(.) \geq 0$ $\nabla_y \vec{D}_o(.) \leq 0$	translation $\vec{D}_O(x,y+\alpha g_y)=$ $\vec{D}_O(x,y;g_y)-\alpha$ homogeneity $\vec{D}_O(x,y;\lambda g_y)=$ $\lambda^{-1}\vec{D}_O(x,y;g_y)$	quasi- concave	concave	$TI_o(x,y)$ $=\vec{D}_o(.)$	$\vec{D}_o(x,y;y)$ $=\vec{D}_T(x,y;0,y)$ $=[1/D_o(x,y)]-1$ duality; $\vec{D}_O(x,y;g_y) \leq$ $\frac{R(x,p)-py}{pg_y}$

estimated frontier values.

$$TE = \frac{Y}{Y*} = \frac{f(X;\beta)exp(v-u)}{f(X;\beta)exp(v)} = \exp(-u)$$

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

where $u \geq 0$ represents a measure of technical inefficiency because $u = -lnTE \approx 1 - TE \approx 1 - exp(-u)$. Several techniques have been proposed in the literature to separate the composed error term $\varepsilon = \nu - u$. Refer to Section 4 for more details on estimation techniques.

This section examines the development of modeling inefficiency concerning its temporal behavior, classification, and determinants. Initially, in stochastic frontier models, inefficiency is assumed to be time-invariant in both cross-sectional and panel data models. This assumption is later relaxed with the introduction of time-variant inefficiency models, which allow inefficiency to vary over time and among individual producers. Time-invariant and time-variant inefficiency models are developed to address both inefficiency components. More recently, models incorporating four random components have been proposed to address both inefficiencies and heterogeneous technology. Dynamic inefficiency models have been introduced to capture the dynamic nature of inefficiency, where inefficiency evolves through an autoregressive process in which past inefficiency values influence the current value. Threshold inefficiency models examine the possibility that inefficient producers might be excluded from markets. In contrast, zero inefficiency models highlight the possibility for producers to be fully efficient. These models can incorporate both fully efficient and inefficient producers within a probabilistic framework. Heterogeneous inefficiency models are proposed to capture heterogeneity in the inefficiency component by incorporating characteristics specific to each producer. These characteristics can be integrated into the inefficiency component itself, or into the mean, variance, or both parameters of the inefficiency distribution.

3.1 Time-invariant inefficiency models

Time-invariant inefficiency models consider inefficiency as unchanging over time. In the inefficiency literature, it is sometimes referred to as long-term or persistent inefficiency. This inefficiency can be modeled using either cross-sectional or panel data.

3.1.1 The cross-sectional models

The early literature on stochastic frontier models utilized cross-sectional models, where specific distributions for inefficiency and error terms were assumed to estimate the frontier function. These distribution assumptions were also necessary to separate inefficiency from the error term.

The stochastic production frontier was independently proposed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977). Battese and Cora (1977) introduced the first application. It is presented as follows:

$$Y_i = \alpha + f(X_i; \beta) + \nu_i - u_i$$

In this context, inputs, outputs, stochastic factors, and inefficiency vary only across producers. A key issue with the cross-sectional model is its reliance on the strong assumption that inefficiencies are independent of the regressors. If this assumption is violated, it results in inconsistent estimates of the model's parameters and the measures of inefficiency.

3.1.2 The panel data models

The use of panel data addresses the limitations of cross-sectional models and offers several advantages in models with time-invariant inefficiency. It provides consistent estimates of inefficiency by incorporating more temporal observations for the same producer, assuming the time series is sufficiently large. Additionally, it does not require inefficiency to be independent of the regressors, which is beneficial for including time-invariant regressors in the model. Furthermore, there is no need to make specific distributional assumption regarding inefficiency, and all parameters of the model can be estimated using standard estimation procedures for panel data, such as fixed and random-effects. The production frontier panel data model can be represented as follows:

$$Y_{it} = \alpha + f(X_{it}; \beta) + \nu_{it} - u_i$$

In this context, inputs, outputs, and stochastic factors vary over time and among producers, but inefficiency varies only among producers, this assumption may be unrealistic. However, it might hold true if the time dimension of the panel is particularly short or if inefficiency is associated with management and there is no change in management during that period. If the time dimension is long, it seems unrealistic to assume constant inefficiency over time or for inefficient producers to continue operating in the market.

3.1.2.1 The fixed-effects models. If inefficiency is considered systematic and, therefore, u_i is treated as a producer-specific constant or an unknown fixed parameter to be estimated, a fixed-effect model can be implemented. No distribution assumption is required for u_i, which is assumed to be correlated with the regressors X_{it} or the random errors ν_{it}. Schmidt and Sickles (1984) defined a fixed-effects model of inefficiency as follows:

$$Y_{it} = \alpha_i + f(X_{it}; \beta) + \nu_{it}$$

Since u_i is treated as fixed, it becomes the producer-specific intercept $\alpha_i = \alpha - u_i$. These are regarded as fixed numbers that can be estimated as parameters or eliminated through suitable transformation if the number of producers is too large. Schmidt and Sickles (1984) considered various procedures for estimating the fixed-effects model: the within estimator, the Generalized Least Squares (GLS) estimator, and the Maximum Likelihood Estimation (MLE). The within estimator does not assume independence between u_i and the regressors, whereas the GLS estimator assumes that u_i is uncorrelated with the regressors. The MLE assumes both distributional and independence assumptions. Sickles (2005) presented a diverse array of methods for identifying producer-specific inefficiency using panel estimators. Koop et al. (1997) described procedures for Bayesian estimation of fixed-effects inefficiency models.

Using the within estimator, the fixed-effects estimate $\widehat{\beta}$, also known as the within estimate, can be determined by either regressing $\widetilde{Y}_{it} = \left(y_{it} - \overline{y}_i\right)$ on $\widetilde{X}_{it} = (x_{it} - \overline{x}_i)$, where $\overline{y}_i = \sum_{i=1}^{T} y_{it}/T$ and $\overline{x}_i = \sum_{i=1}^{T} x_{it}/T$, thereby eliminating α_i, or, equivalently, by regressing Y_{it} on X_{it} along with a set of specific dummy variables for producers using the Ordinary Least Squares (OLS). Consequently, $\widehat{\alpha}_i$ is obtained by averaging its residuals over time as $\widehat{\alpha}_i = \overline{y}_i - \overline{x}_i\widehat{\beta}$, or, equivalently, $\widehat{\alpha}_i$ represents the estimated coefficients of the dummy variables. Inefficiency is assessed by comparing the estimated intercept of each producer to the maximum estimated value.

$$\widehat{u}_i = \max_j \left\{\widehat{\alpha}_j\right\} - \widehat{\alpha}_i$$

Producer-specific efficiency can be obtained from $TE_i = \exp\left(-\widehat{u}_i\right)$. However, this approach considers the producer with the highest intercept as fully efficient, making inefficiency for other producers relative to that producer. Feng and Horrace (2012) estimated inefficiency relative to the least efficient producer instead of the most efficient one by comparing the estimated intercept of each producer to the minimum estimated value. They argued that these inefficiency estimates have smaller bias than those using the maximum estimated value when many producers operate close to the efficient frontier. However, in both cases, inefficiency is estimated as relative rather than absolute inefficiency. Furthermore, the intercept $\widehat{\alpha}_i$ captures all time-invariant unobserved heterogeneity, not only those related to inefficiency. Additionally, as pointed out by Kim and Schmidt (2000), Wang and Schmidt (2009), and Satchachai and Schmidt (2010), the estimation of inefficiency based on the fixed-effects estimator can be upwardly biased when the number of time series is small and the number of cross-sectional observations is large. The max operator induces an upward bias in $\widehat{\alpha} = \max_j \left\{\widehat{\alpha}_j\right\}$, which in turn induces an upward bias in the inefficiency estimates $\widehat{u}_i$.

Wikstrom (2016) introduced a modified fixed-effects estimator that does not suffer from bias in large cross-sectional observations. This is accomplished by utilizing the second central moment of the inefficiency distribution to correct the intercept value derived from the fixed-effects estimator. Wikstrom presented a consistent estimator of α, assuming a half-normal distribution and an exponential distribution for u_i as $\widehat{\alpha} = \widehat{\mu}_\alpha + \widehat{\mu}_u$ where $\widehat{\mu}_\alpha = \sum_{i=1}^{N} \widehat{\alpha}_i / N$, $\widehat{\mu}_u = \widehat{\sigma}_\alpha^2 \, (2/\pi - 2)^{1/2}$ assuming a half-normal distribution for u_i and $\widehat{\mu}_u = \left(\widehat{\sigma}_\alpha^2\right)^{1/2}$ assuming an exponential distribution for u_i, $\widehat{\sigma}_\alpha^2 = \left(\sum_{i=1}^{N} (\widehat{\alpha}_i - \widehat{\mu}_\alpha)^2 / N\right) - \left(\widehat{\sigma}_v^2 / T\right)$. The modified fixed-effects estimator of u_i is defined as follows:

$$\widehat{u}_i = \widehat{\alpha} - \widehat{\alpha}_i$$

The fixed-effects estimator has the advantage of not requiring a distributional assumption on inefficiency and allows inefficiency to be correlated with any other variables. However, the assumption of time invariance in inefficiency is very restrictive and may not be reasonable for relatively long panels. Additionally, fixed-effect time-invariant models are based on the assumption that all time-invariant effects are parts of inefficiency. Consequently, inefficiency measures include any other source of time-invariant unobserved heterogeneity, not only those related to inefficiency, making it challenging to distinguish unobserved heterogeneity from inefficiency, as noted by Greene (2004a). Furthermore, time-invariant regressors cannot be used in the specification of the frontier model, as this would lead to perfect multicollinearity between α_i and the time-invariant regressors.

3.1.2.2 The random-effects models. When the assumption of no correlation between the regressors and inefficiency is correct, random-effects models provide more efficient estimates than fixed effects models. Random-effects time-invariant inefficiency models were introduced by Pitt and Lee (1981), Kumbhakar (1987), and Battese and Coelli (1988), where inefficiency is treated as time-invariant. Similar to fixed effects models, inefficiency measures can be estimated using the GLS technique commonly applied to standard random-effects panel data models. Inefficiency measures can also be estimated by $E(u_i \mid \varepsilon_{it})$, where $\varepsilon_{it} = v_{it} - u_i$, using maximum likelihood estimation or the posterior mean $E(u_i \mid Y)$ using Bayesian estimation. See section 4 for more details on estimation techniques.

Fixed-effect models allow for a correlation between inefficiency and regressors, whereas random-effect models require independence among them and do not allow for endogenous regressors in the model. This assumption may be unrealistic because inefficiency could be related to the usage and quality of inputs when modeling production. Additionally, random-effect time-invariant models are based on the assumption that all time-invariant effects are parts of inefficiency. An advantage of random-effects models is that time-invariant regressors can be included in the model without causing collinearity issues.

3.2 Time-variant inefficiency models

To accommodate efficiency improvements and allow inefficiency to change over time, time-variant inefficiency models are employed. In the inefficiency literature, time-variant inefficiency is sometimes referred to as short-term or transient inefficiency. Estimates of inefficiency in these models depend on model specifications, distributional assumptions, and the temporal behavior of inefficiency. Time-variant inefficiency models facilitate the identification of both time-variant inefficiency and producer effects simultaneously, accounting for heterogeneous technologies. Additionally, they allow for the simultaneous identification of both time-variant inefficiency and technical change.

3.2.1 The fixed-effects models

Cornwell et al. (1990) modified the assumption of time invariance in the Schmidt and Sickles (1984) model by replacing α_i with a quadratic function of time. This adjustment allows inefficiency to vary over time and among individual producers. The model is represented as follows:

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

$$Y_{it} = \alpha_{it} + f(X_{it}; \beta) + v_{it}$$

where $\alpha_{it} = \theta_{0i} + \theta_{1i}t + \theta_{2i}t^2$. The model is estimated by regressing the residuals for each producer $\left(\widehat{\varepsilon}_{it} = Y_{it} - X'_{it}\widehat{\beta}\right)$ on a constant, time, and time squared. The fitted values from this regression provide an estimate of α_{it}. Inefficiency measures are computed relative to the most efficient producer across all time periods or, alternatively, to the most efficient producer in a given year. This modification allows the most efficient producer to change from year to year.

$$\widehat{u}_{it} = \max_j \left\{\widehat{\alpha}_{jt}\right\} - \widehat{\alpha}_{it}$$

The advantages of this model include its independence of distribution assumption on inefficiency and its allowance of inefficiency to vary among producers and over time. However, it is quite restrictive in describing the temporal behavior of inefficiency, as this is assumed to be deterministic. Additionally, this model cannot distinguish inefficiency from technical change, which represents a shift in the production frontier, because time is a factor in the inefficiency function.

Lee and Schmidt (1993) described inefficiency as the product of individual producer inefficiency and time effects; $\alpha_{it} = \theta_t \alpha_i$, where $\theta_t = \sum_t \delta_t$ with δ_t being a dummy variable for each period t and $\widehat{u}_{it} = \max_j \left\{\widehat{\theta_t \widehat{\alpha}_i}\right\} - \widehat{\theta}_t \widehat{\alpha}_i$. This specification differs from the time-invariant fixed-effect model by allowing inefficiency to vary over time. However, the producer effect, denoted as α_i, cannot be identified separately from θ unless a specification for the inefficiency component is considered without the intercept. Furthermore, the temporal behavior of inefficiency is assumed to be the same for all producers.

The primary challenge with these fixed-effects, time-variant inefficiency models is that they require the estimation of numerous parameters, which can be constrained by very short panels. Furthermore, inefficiency varies over time in both models through the use of a time trend or time dummies, which hinders the control of technical change.

Greene (2005a, 2005b) introduced what is known as the true fixed-effects model. It is represented as follows:

$$Y_{it} = \alpha_i + f(X_{it}; \beta) + v_{it} - u_{it}$$

In this context, α_i represents unobserved time-invariant heterogeneity and is treated as a random variable that is correlated with X_{it}, but does not capture inefficiency and can be estimated as a parameter. A true fixed-effects model can be estimated by adding dummy variables for each producer to the model. The disadvantage of this model is that it induces the incidental parameters problem, which can lead to inconsistency because the number of parameters to be estimated depends on the number of producers, as discussed by Neyman and Scott (1948). Recent studies have considered addressing the problem of incidental parameters in the true fixed-effects model by using within transformation to eliminate the producer effects for unobserved heterogeneity,[1] see, for example, Wang and Ho (2010) and Chen, et al. (2014).

3.2.2 The random-effects models

In these models, time-variant inefficiency can either be independently and identically distributed (iid) among producers and over time, or it can be modeled as the product of a deterministic function of time, $g(t; \gamma)$, and a non-negative time-invariant random variable u_i, such that $u_{it} = g(t; \gamma)u_i$, where γ is a parameter to be estimated. Thus, $g(t; \gamma)$ allows the data to determine the temporal behavior of inefficiency rather than imposing it a priori. Inefficiency is then estimated from $\widehat{u}_{it} = \widehat{g}(t; \gamma)E(u_i \mid \varepsilon_i)$, or, alternatively, $\widehat{u}_i = E(u_i \mid \varepsilon_{it})$; see Kumbhakar and Lovell (2000).

Kumbhakar (1990) assumed that $u_{it} = u_i \left(1 + \exp\left(\gamma_1 t + \gamma_2 t^2\right)\right)^{-1}$. His model allows inefficiency to either increase or decrease monotonically, depending on the values of γ_1 and γ_2. Battese and Coelli (1992) and Battese and Tessema (1993) assumed that $u_{it} = \exp\left(-\gamma\,(t - T)\right) u_i$, where T represents the final time period. Their model implies that inefficiency changes over time in a monotonic manner, increasing or decreasing exponentially for all producers based on the sign of γ. Specifically, inefficiency increases at a decreasing rate when γ is positive and decreases at an increasing rate when γ is negative. A time-invariant model is obtained when γ is equal to zero. Kumbhakar and Wang (2005) assumed that $u_{it} = \exp\left(-\gamma\,\left(t - \underline{t}\right)\right) u_i$. In this context, inefficiency evolves over time according to $\exp\left(-\gamma\,\left(t - \underline{t}\right)\right)$, where $\underline{t}$ denotes the initial time period, ensuring $u_{it} = u_i$ at time $\underline{t}$. A promising area for further research is developing a model that incorporates both the initial and final time periods, as it considers both market entry and exit in defining the reference point.

Lee and Schmidt (1993) assumed that $u_{it} = \gamma_t u_i$, where γ_t represents the parameters associated with the time dummy variables that need to be estimated. While Battese and Coelli (1992) and Battese and Tessema (1993) assumed a quite restrictive temporal behavior of inefficiency which is assumed to be the same for all producers, Cuesta (2000) modified the Battese and Coelli (1992) model to allow for greater flexibility in how inefficiency changes over time by assuming $u_{it} = \exp\left(-\gamma_i\,(t - T)\right) u_i$, $u_{it} = \exp\left(g_i\,(t, T, z_{it})\right) u_i$. This specification allows inefficiency to evolve over time at varying rates among producers.

Cuesta and Orea (2002) and Feng and Serletis (2009) extended the Battese and Coelli (1992) model by assuming $u_{it} = \exp\left(-\gamma_1\,(t - T) - \gamma_2\,(t - T)^2\right) u_i$. This specification relaxes the monotonicity of the time path of inefficiency using a two-parameter approach. Consequently, the model allows for producer effects to be either convex or concave and to vary over time, increasing in some periods and decreasing in others.

The primary advantage of random-effects models over fixed-effects time-variant inefficiency models is that they allow for the inclusion of time-invariant regressors. However, random-effects models require independence between inefficiency and regressors in the model, a condition not required in fixed-effects models.

The Hausman and Taylor (1981) model integrates fixed and random-effects models, allowing inefficiency to be uncorrelated with certain, but not all, regressors. It also enables the inclusion of time-invariant regressors in the model. In this framework, producer inefficiency can be consistently estimated and separated from the producer effects or the intercept, provided that the cross-sectional and temporal observations are sufficiently large.

To separate producer heterogeneity, or producer effects, and inefficiency, where inefficiency can vary over time and may be either iid or a function of exogenous variables, Greene (2005a, 2005b) introduced a time-invariant random effect to account for unobserved producer heterogeneity and proposed what is known as the true random-effects model.

$$Y_{it} = (\alpha + w_i) + f(X_{it}; \beta) + v_{it} - u_{it}$$

where $\alpha_i = \alpha + w_i$ represents unobserved time-invariant heterogeneity and is treated as a random variable that is uncorrelated with X_{it}. It is important to note that Kumbhakar and Wang (2005) introduced these producer-specific intercepts α_i to account for heterogeneous technologies. If α_i is treated as a random variable that is correlated with X_{it} but does not capture inefficiency, then the model becomes the true fixed-effects model.

3.3 Time-invariant and time-variant inefficiency models

Previous models for panel data have focused either on time-invariant inefficiency or time-variant inefficiency, but none of these models considers both simultaneously. Mundlak (1961) noted that time-invariant inefficiency reflects the effects of inputs such as management; therefore, it is important to estimate it, particularly in short panels. However, for large panels or when there are changes

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

in management, it is also important to estimate time-variant inefficiency. Colombi et al. (2014) argued that time-variant inefficiency arises due to the failure to allocate resources properly in the short run. Tsionas and Kumbhakar (2014) noted that estimating a model with only one inefficiency component, regardless of controlling for producer effects for unobserved heterogeneity, is likely to yield incorrect estimates of inefficiency. Kumbhakar and Heshmati (1995) proposed a model in which inefficiency is assumed to have both time-invariant and time-variant components as follows:

$$Y_{it} = \alpha + f(X_{it};\, \beta) + v_{it} - u_{i0} - u_{it}$$

In this context, u_{i0} represents time-invariant inefficiency, u_{it} represents time-variant inefficiency, and $u_{i0} + u_{it}$ constitutes total inefficiency. The error components are assumed to be independent of each other and also independent of X_{it}. The model can be estimated using MLE, Bayesian estimation, or a three-step procedure as follows: First, Employing a standard random-effects model for panel data provides consistent estimates of the model parameters and predicted values of u_{i0} and u_{it}. Second, The time-invariant inefficiency is estimated as $\widehat{u}_{i0} = \max_j \left\{\widehat{u}_{j0}\right\} - \widehat{u}_{i0}$. Finally, the time-variant inefficiency is estimated by maximizing the log-likelihood function for pooled data, expressed as $[r_{it} = \alpha + v_{it} - u_{it}]$ where $r_{it} = Y_{it} - f(X_{it};\, \beta) + u_{i0}$. Estimates of u_{it}, conditional on the estimated $(\varepsilon_{it} = v_{it} - u_{it})$, are obtained from $\widehat{u}_{it} = E(u_{it} \mid \varepsilon_{it})$, following the method of Jondrow et al. (1982). Total efficiency is then defined as the product of time-invariant and time-variant efficiencies.

$$\text{Total efficiency}_{it} = \exp\left[-\widehat{u}_{i0}\right] \times \exp\left[-\widehat{u}_{it}\right]$$

3.4 Four random components inefficiency models

Time-invariant and time-variant inefficiency models have not explicitly accounted for producer effects related to unobserved heterogeneity, nor have they separated these effects from time-invariant inefficiency. In response, Kumbhakar et al. (2014), Colombi et al. (2014), Tsionas and Kumbhakar (2014), and Fillipini and Greene (2016) expanded upon the true random-effects model proposed by Greene (2005a, 2005b). They introduced a time-invariant inefficiency component and proposed four random components inefficiency models to address both inefficiencies and heterogeneity. These models decompose the time-invariant producer effect into a producer effect and a time-invariant inefficiency effect.

$$Y_{it} = \alpha + f(X_{it};\, \beta) + \omega_i + v_{it} - u_{i0} - u_{it}$$

In this context, the error term comprises four random components: ω_i represents random producer effects for unobserved heterogeneity, u_{i0} denotes time-invariant inefficiency, u_{it} represents time-variant inefficiency, and v_{it} indicates random errors. Kumbhakar et al. (2014) estimated the model using a three-step procedure based solely on OLS. First, the model is rearranged as $Y_{it} = \alpha + f(X_{it};\, \beta) + \xi_i + \varepsilon_{it}$ where $\xi_i = \omega_i - u_{i0}$ and $\varepsilon_{it} = v_{it} - u_{it}$. Here, ξ_i can be viewed as the producer-specific component. Using a standard random-effect model for panel data provides consistent estimates of the model's parameters and predicted values of $\widehat{\xi_i}$ and $\widehat{\varepsilon}_{it}$. Second, estimates of u_{it} conditional on the estimated $(\varepsilon_{it} = v_{it} - u_{it})$ are obtained from $\widehat{u}_{it} = E(u_{it} \mid \varepsilon_{it})$ following Jondrow et al. (1982). Third, a similar procedure to that in stage two is used to estimate the time-invariant inefficiency component u_{i0}. It should be noted that Kumbhakar et al. (2014) used the procedure of Jondrow et al. (1982), which implicitly assumes that the marginal distribution of inefficiency given the observations is truncated-normal. However, as shown by Cartinhour (1990) and Horrace (2005), the marginal distributions of a multivariate truncated-normal distribution are not truncated-normal distributions.

Colombi et al. (2014) adopted a single-step approach using MLE, drawing on results from the closed skew-normal distribution, as opposed to the three-step procedure utilized by Kumbhakar

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

et al. (2014). The MLE is asymptotically more efficient than the three-step procedure because it estimates all parameters simultaneously. However, Tsionas and Kumbhakar (2014) observed that the MLE method employed by Colombi et al. (2014) becomes computationally challenging when T is large, due to the likelihood function's reliance on a $(T+1)$-dimensional integral of the normal distribution. Tsionas and Kumbhakar (2014) employed Bayesian estimation to estimate the model using a large panel of banks in the United States. Conversely, Fillipini and Greene (2016) utilized the Simulated Maximum Likelihood Estimation, as introduced by Greene (2005a, 2005b), to estimate the parameters and various random components of the model. By applying the moment-generating function for the closed skew-normal distribution, as developed by Colombi et al. (2014), they estimated the efficiency values.

3.5 Dynamic inefficiency models

The temporal behavior of inefficiency in dynamic inefficiency models is characterized by its evolution through an autoregressive process, where past values of inefficiency determine the current value. However, these models are infrequently utilized in the literature to measure inefficiency. Ahn and Sickles (2000) assumed that inefficiency follows a first-order autoregressive process, $AR(1)$, where the current inefficiency, u_{it}, is influenced by two components: the unadjusted portion of the inefficiency from the previous period, $(1-\rho_i)\,u_{i,t-1}$, where $0<\rho_i\leq 1$ represents the adjustment speed, and the new, unexpected inefficiency, e_{it}.

$$u_{it}=(1-\rho_i)\,u_{i,t-1}+e_{it}$$

The econometric method employed, specifically the Generalized Method of Moments (GMM), are suitable for stationary or trend-stationary data but not for data exhibiting stochastic trends. Tsionas (2006) applied Bayesian estimation to a panel of large U.S. commercial banks, considering inefficiency as a function of explanatory variables that reflect producer-specific characteristics to account for heterogeneity in inefficiency. Specific assumptions were also made regarding the initial value u_{i1}.

$$\ln u_{it}=z_{it}\delta+\rho\ln u_{i,t-1}+e_{it} \qquad \text{for } t=2,\ldots\ldots,T$$

$$\ln u_{i1}=z_{i1}\delta/\,(1-\rho)+e_{i1} \qquad \text{for } t=1$$

In this context, $e_{it}\sim N(0,\sigma_e^2)$, $e_{i1}\sim N(0,\sigma_e^2/\left(1-\rho^2\right))$. Deprins and Simar (1989) utilized the specification of log-normality for inefficiency. However, assuming a log-normal distribution for inefficiency cannot accommodate a situation where most producers are fully efficient. For the $\ln u_{it}$ process to be stationary, the restriction $|\rho|<1$ should be imposed. Tsionas (2006) found that the posterior mean was $\rho=0.91$, indicating that the autoregressive process is nearly static.

The evolving environment, which encompasses government policies and regulations, market conditions, and economic shocks, can cause producers to react differently at various times. This indicates potential instability over time in the relationships among macroeconomic variables. Rather than presuming that the parameters are stationary and constant, a promising area for future research is the examination of time-varying parameters when analyzing macroeconomic time series.

3.6 Threshold inefficiency models

In contrast to models that allow for the existence of extremely inefficient producers who cannot survive in highly competitive markets, threshold inefficiency models truncate the distribution of inefficiency by placing a threshold parameter of the minimum efficiency required for survival. These models, therefore, define an upper bound for the distribution of inefficiency, in addition to the zero lower bound.

Lee (1996) introduced a tail-truncated half-normal distribution with a threshold parameter θ; $u_i \sim N^+(0, \sigma_u^2)$, $0 \leq u_i \leq \theta$. In contrast, Lee and Lee (2014) assumed a uniform distribution, $u_i \sim U(0, \theta)$. Almanidis, et al. (2014) extended Lee's (1996) model to a panel data model and assumed that u_{it} are drawn from a time-variant distribution with an upper bound θ_t, which is considered to be the sum of weighted polynomials: $\theta_t = \sum_{i=0}^{N} b_i \, (t/T)^i$, where $t = 1, \ldots, T$ and b_i are constants. These threshold inefficiency models are beneficial for empirical studies that aim to estimate the inefficiency threshold.

3.7 Zero inefficiency models

The threshold inefficiency models focus on the possibility of inefficient producers being excluded from the markets, whereas the zero inefficiency models highlight the possibility of producers being fully efficient.

Wheat et al. (2014) noted that the probability of any producer being fully efficient is zero in models that do not allow for the presence of fully efficient producers. However, Bos, et al. (2010), along with Bos, et al. (2010), used latent class models and identified small groups of producers that are fully efficient. Kumbhakar, et al. (2013) observed that if the data represent a mixture of both fully efficient and inefficient producers, then models imposing inefficient behavior on all producers result in biased estimates of inefficiency. They introduced the zero inefficiency model, which can accommodate the presence of both fully efficient and inefficient producers within a probabilistic framework.

Assuming that some producers operate with full efficiency, where $u_i = 0$ for certain producers, and others operate with inefficiency, where $u_i > 0$, the zero-inefficiency model is represented as follows:

$$Y_i = \begin{matrix} f(X_i; \beta) + v_i & \text{with probability } p \\ f(X_i; \beta) + v_i - u_i & \text{with probability } \left(1 - p\right) \end{matrix}$$

In this context, p denotes the probability of a producer being fully efficient or the proportion of producers who are fully efficient, while $\left(1 - p\right)$ represents the proportion of producers who are inefficient. Kumbhakar, et al. (2013) specified the estimates of inefficiency as $\widetilde{u}_i = \left(1 - \widetilde{p}_i\right) \widehat{u}_i$, where $\widehat{u}_i$ is the zero-inefficiency estimator of inefficiency with $p = 0$, and $\widetilde{p}_i$ is the estimate of the probability of being fully efficient.

Kumbhakar, et al. (2013) and Rho and Schmidt (2015) proposed modeling the probability or proportion of producers achieving full efficiency as a parametric function of a set of explanatory variables that determine full efficiency through a logit or probit function. However, Tran and Tsionas (2016b) argued that misspecification of the parametric functional form of this probability affects the identification of fully efficient producers and the estimates of inefficiency. They employed a non-parametric approach for the probability of producers achieving full efficiency, utilizing an unknown smooth function of explanatory variables that influence the likelihood of a producer reaching full efficiency.

The zero inefficiency model addresses two classes a priori: fully efficient and inefficient producers. Therefore, it does not face the challenge of determining the number of classes, as is the case with latent class models. However, Rho and Schmidt (2015) discussed the presence of the incorrect skewness issue identified by Waldman (1982) and the identification challenges within zero inefficiency models. They argue that when all producers are fully efficient, it is unclear whether efficiency results from p being close to one or from σ_u^2 being close to zero, which has significant implications for conducting inference. Another concern with zero inefficiency models is that the consistency of the estimates depends on the exogeneity of the explanatory variables. Tran and Tsionas (2016a) investigated the endogeneity issues in zero inefficiency models through a simultaneous equation setting, allowing for one or more regressors to be endogenous.

3.8 Heterogeneous inefficiency models

Heterogeneous inefficiency models are proposed to capture heterogeneity in the inefficiency component by incorporating characteristics specific to each producer. These characteristics, represented as Z, can be integrated into the inefficiency component itself, or into the mean, variance, or both parameters of the inefficiency distribution. These models are also valuable for understanding the relationship between inefficiency and its exogenous determinants.

Heterogeneous inefficiency models can be estimated using either a two-step procedure, where inefficiency and explanatory variables Z are estimated sequentially, or a one-step procedure, where the explanatory variables are estimated simultaneously with the other parameters of the model. However, the two-step procedure has been criticized for potentially misspecifying the first-step model, suffering from omitted variable bias if X and Z are correlated, and for its bias from ignoring the impact of Z on inefficiency. For further discussion, see Caudill and Ford (1993), Battese and Coelli (1995), and Wang and Schmidt (2002).

3.8.1 Determinants of inefficiency models

In these models, inefficiency is represented as a function of explanatory variables Z that reflect characteristics specific to each producer and explain the differences in inefficiency among them. For further discussion, see Deprins and Simar (1989), Kumbhakar et al. (1991), and Huang and Liu (1994) who introduced interaction terms between producer-specific characteristics and regressors, $z_i x_i$.

$$u_i = g\left(z_i, z_i x_i; \delta\right) + e_i$$

where δ represents unknown parameters to be estimated, and e_i is a random variable defined by the truncation of a normal distribution. If there are no interaction terms $z_i x_i$, the model reduces to those of Deprins and Simar (1989) and Kumbhakar et al. (1991). Tsionas (2006) extended the Kumbhakar et al. (1991) model to a panel data model that allows for dynamic inefficiency; $\ln u_{it} = z_{it}\delta + \rho \ln u_{i,t-1} + e_{it}$. Srairi (2010) further extended the Kumbhakar et al. (1991) model to panel data where $u_{it} = g\left(z_{it}; \delta\right) + e_{it}$ to examine bank-specific variables that may explain the sources of inefficiency.

Determinants of inefficiency models face the challenge of ensuring non-negative inefficiency values. Kumbhakar et al. (1991) proposed a solution to address this issue: $u_i = \left|N\left(z_i\delta, \sigma_u^2\right)\right|$. Reifschneider and Stevenson (1991) assumed that $u_i = u_i^* + \exp\left(z_i\delta\right)$, where both $u_i^* \sim N^+(0, \sigma_u^2)$ and $\exp\left(z_i\delta\right)$ are positive. However, it is not necessary for both components to be positive to achieve a positive u_i. Huang and Liu (1994) extended the assumption of Reifschneider and Stevenson (1991) by assuming only that $u_i^* \geq -\exp\left(z_i\delta\right)$.

3.8.2 Determinants of inefficiency distribution models

In these models, producer-specific characteristics can be incorporated into the mean, variance, or both parameters of the inefficiency distribution. Battese and Coelli (1995) and Wang and Ho (2010) proposed that the mean of the inefficiency distribution be modeled as a function of explanatory variables that represent producer-specific characteristics.

$$u_{it} = g\left(z_{it}; \delta\right) u_i,\ u_{it} \sim N^+(\mu_{it}, \sigma_u^2),\ \mu_{it} = z_{it}\delta_m$$

Including producer-specific characteristics in the variance of the inefficiency distribution is motivated by the potential presence of heteroscedasticity in inefficiency. Reifschneider and Stevenson (1991), Caudill and Ford (1993), and Caudill et al. (1995) assumed that the inefficiency term, u, is heteroskedastic and included the standard deviation in exponential form to ensure a positive estimate of the variance parameter for all parameters involved, Z and γ_u.

$$\sigma_{ui} = \exp\left(z_{ui}\gamma_u\right),\ \sigma_{vi} = \exp\left(\gamma_v\right)$$

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

It is also possible to assume that both u and v are heteroskedastic, which is referred to in the literature as the doubly heteroskedastic model. The variance parameters of the u and v distributions are modeled as functions of the explanatory variables z_{ui} and z_{vi}, which may or may not be equivalent; see, for example, Hadri (1999) and Hadri et al. (2003).

$$\sigma_{ui} = \exp\left(z_{ui}\gamma_u\right), \sigma_{vi} = \exp\left(z_{vi}\gamma_v\right)$$

Including producer-specific characteristics in both the mean and the variance of the inefficiency distribution allows for non-monotonic inefficiency effects across producers. For examples, see Wang (2002) and Wang and Schmidt (2002).

$$u_{it} \sim N^{+}(\mu_{it}, \sigma^2_{uit}), \mu_{it} = z_{it}\delta_m, \sigma^2_{uit} = \exp\left(z_{it}\gamma_u\right)$$

Kumbhakar and Wang (2005) proposed that the variance parameter of the v distribution could be modeled as a function of the explanatory variables z_{vi}, alongside the mean and variance of the inefficiency distribution.

$$u_i \sim N^{+}(\mu_i, \sigma^2_{ui}), \mu_i = z_i\delta_m, \sigma^2_{ui} = \exp\left(z_{ui}\gamma_u\right), v_{it} \sim N(0, \sigma^2_{vi}), \sigma^2_{vi} = \exp\left(z_{vi}\gamma_v\right)$$

A key question to consider regarding heterogeneous inefficiencies is whether heterogeneity in inefficiency exists or if producer-specific inefficiency depends on a set of exogenous factors. As Kim and Schmidt (2008) suggested, one can test for the presence of these factors by regressing Y on X and Z, and then testing the significance of the parameters of these factors using the F-test.

A summary of the main characteristics of inefficiency models commonly used in the literature is presented in Table 2. A time-variant and time-invariant inefficiency model is derived by omitting ω_i from the four random components inefficiency model. A time-variant inefficiency model is derived by omitting u_{i0}, while a time-invariant inefficiency model is derived by omitting u_{it} from the time-variant and time-invariant inefficiency model.

4. Estimation techniques

A variety of econometric estimation techniques, incorporating recent advancements, have been proposed in the literature to estimate inefficiency within the stochastic frontier approach. Fixed-effects and random-effects estimators are briefly discussed in the previous section. Sickles (2005) summarized different panel frontier estimators of inefficiency that have been utilized in the literature. For developments in econometric estimation techniques, see, for example, Bauer (1990), Greene (1993), and Parmeter and Kumbhakar (2014). Given the extensive nature of this literature, this section provides a concise review of the most common estimation techniques: maximum likelihood estimation and Bayesian estimation.

4.1 Maximum likelihood estimation

Estimating inefficiency in the stochastic frontier approach using maximum likelihood estimation (MLE) requires a distribution assumption for the inefficiency term as well as the random error to disentangle one from the other. Various distributions have been assumed in the literature for the inefficiency term, with the most commonly used being the half-normal, exponential, gamma, truncated-normal, and skew-normal distributions. Greene (1993) employed different distribution assumptions and demonstrated that inefficiency measures are similar across these distributions. Berger and DeYoung (1997) found that assuming a truncated-normal distribution for the inefficiency term provides similar but statistically significant estimates compared to the half-normal assumption. However, Baccouche and Kouki (2003) found that inefficiency measures depend heavily on the distribution assumptions.

The MLE method involves specifying the model through the joint probability density function, denoted as $f(Y, \theta)$. When assuming independence, the joint density of Y is expressed as the

Table 2. A summary of the main characteristics of inefficiency models

	Inefficiency		ω_i	Heterogeneous inefficiency			
Inefficiency Models	u_i	u_{it}		u	Mean μ	Variance σ^2	Models are proposed by
Time-invariant Cross-section	u_i	NA	NA	NA	$u_i \sim N^+(0, \sigma_u^2)$	σ_u^2	Aigner et al. (1977), Meeusen & Van den Broeck (1977)
Time-invariant Fixed-effects	$\alpha_i = \alpha - u_i$	NA	NA	NA	No distribution assumption	No distribution assumption	Schmidt and Sickles (1984)
Time-invariant Random-effects	u_i	NA	NA	NA	$u_i \sim N^+(0, \sigma_u^2)$	σ_u^2	Pitt and Lee (1981), Battese and Coelli (1988)
Time-variant Fixed-effects	NA	$\alpha_{it} = g(t)$	NA	NA	No distribution assumption	No distribution assumption	Cornwell et al. (1990), Lee and Schmidt (1993)
Time-variant Random-effects	NA	$u_{it} = g(t)u_i$	NA	NA	$u_i \sim N^+(0, \sigma_u^2)$	σ_u^2	Kumbhakar (1990), Battese and Coelli (1992)
True fixed-effect	NA	u_{it}	α_i	NA	$u_{it} \sim N^+(0, \sigma_u^2)$	σ_u^2	Greene (2005a, b)
True random-effect	NA	u_{it}	ω_i	NA	$u_{it} \sim N^+(0, \sigma_u^2)$	σ_u^2	Greene (2005a, b)
Time-variant and time-invariant	u_{i0}	u_{it}	NA	NA	$u_{it} \sim N^+(0, \sigma_u^2)$ $u_{i0} \sim N^+(0, \sigma_{u_0}^2)$	σ_u^2 $\sigma_{u_0}^2$	Kumbhakar & Heshmati (1995)
Four random components	u_{i0}	u_{it}	ω_i	NA	$u_{it} \sim N^+(0, \sigma_u^2)$ $u_{i0} \sim N^+(0, \sigma_{u_0}^2)$	σ_u^2 $\sigma_{u_0}^2$	Colombi et al. (2014), Kumbhakar et al. (2014)
Dynamic	NA	$u_{it} = h(u_{i,t-1})$	NA	$u_{it} = g\,(z; \delta)$	$u_{i,t-1} \sim N^+(\mu_{i,t-1}, \sigma_u^2)$	σ_u^2	Ahn and Sickles (2000), Tsionas (2006)
Threshold	u_i	NA	NA	NA	$u_i \sim N^+(0, \sigma_u^2)$ $0 \leq u_i \leq \theta$	σ_u^2	Lee (1996)
Zero Inefficiency	$u_i = 0$ with p $u_i > 0$ with $(1-p)$	NA	NA	NA	$u_i \sim N^+(0, \sigma_u^2)$	σ_u^2	Kumbhakar et al. (2013), Rho and Schmidt (2015)
Heterogeneous *Z*on *u*	u_i	NA	NA	$u_i = g\,(z; \delta)$	$u_i \sim N^+(\mu, \sigma_u^2)$	σ_u^2	Deprins and Simar (1989), Huang and Liu (1994)
Heterogeneous *Z*on μ	NA	u_{it}	NA	$u_{it} = g\,(z; \delta)$	$u_{it} \sim N^+(\mu_{it}, \sigma_u^2)$ $\mu_{it} = z_{it}\delta_m$	σ_u^2	Battese and Coelli (1995)
Heterogeneous *Z*on σ_u	u_i	NA	NA	NA	$u_i \sim N^+(0, \sigma_{ui}^2)$	$\sigma_{ui} = \exp\,(z_{ui}\gamma_u)$	Reifschneider & Stevenson (1991)
Heterogeneous *Z*on σ_u and σ_v	u_i	NA	NA	NA	$u_i \sim N^+(0, \sigma_{ui}^2)$	$\sigma_{ui} = \exp\,(z_{ui}\gamma_u)$ $\sigma_{vi} = \exp\,(z_{vi}\gamma_v)$	Hadri (1999), Hadri et al. (2003)
Heterogeneous *Z*on μ and σ_u	NA	u_{it}	NA	NA	$u_{it} \sim N^+(\mu_{it}, \sigma_{uit}^2)$ $\mu_{it} = z_{it}\delta_m$	$\sigma_{uit}^2 = \exp\,(z_{it}\gamma_u)$	Wang (2002), Wang and Schmidt (2002)
Heterogeneous *Z*on μ, σ_u and σ_v	NA	$u_{it} = g(t)u_i$	α_i	NA	$u_i \sim N^+(\mu_i, \sigma_{ui}^2)$ $\mu_i = z_i\delta_m$	$\sigma_{ui}^2 = \exp\,(z_{ui}\gamma_u)$ $\sigma_{vi}^2 = \exp\,(z_{vi}\gamma_v)$	Kumbhakar & Wang (2005)

Note: ω_i denotes heterogeneous technologies and NA indicates that the component is not included in the inefficiency model.

product of the densities of the individual observations, $f_i(Y_i, \theta)$.

$$f(Y, \theta) = \prod_{i=1}^{N} f_i(Y_i, \theta)$$

Due to the potential for the product to be either extremely large or extremely small, it is more practical to work with the logarithm of the likelihood function.

$$L(Y, \theta) = \log f(Y, \theta) = \sum_{i=1}^{N} \log f_i(Y_i, \theta) = \sum_{i=1}^{N} \log l_i(Y_i, \theta)$$

In this context, $L(Y, \theta)$ represents the likelihood of the parameters θ given the observed data Y. It is important to note that $L(Y, \theta)$ provides the same parameter estimates because it is a monotonic transformation of $f(Y, \theta)$. The MLE of the model's parameters is obtained by maximizing the likelihood function with respect to the parameters. The estimated parameters are then used to obtain the estimate of inefficiency by employing one of the inefficiency estimators.

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

4.1.1 Normal-half-normal models

The likelihood function for the normal-half-normal cross-sectional models was derived by Aigner et al. (1977). In these models, it is assumed that inefficiency follows a half-normal distribution $u_i \sim N^+(0, \sigma_u^2)\,; f(u) = \left(\sigma_u \sqrt{2\pi}\right)^{-1} \exp\left(-u^2/2\sigma_u^2\right)$, the random errors follow a normal distribution $v_i \sim N(0, \sigma_v^2)$, and u_i and v_i being independently and identically distributed. The likelihood function is defined as the product of the densities of the composed error term $\prod_{i=1}^{N} f_\varepsilon(\varepsilon_i)$, where $f_\varepsilon(\varepsilon_i)$ represents the density of the composed error term $\varepsilon_i = v_i - u_i$. Inefficiency can be estimated using Jondrow et al. (1982) estimator for the half-normally distributed inefficiency in cross-section models.

$$\widehat{u}_i = E(u_i \mid \varepsilon_i) = \sigma_S \left[\frac{\phi\,(\psi_i)}{1 - \Phi\,(\psi_i)} - \psi_i \right]$$

where ϕ (.) represents the density of the standard normal distribution, Φ (.) represents the cumulative density function, $\sigma_S = \sigma\lambda/\left(1 + \lambda^2\right) = \left(\sigma_u^2 \sigma_v^2/\sigma^2\right)^{1/2} = \sigma_u \sigma_v/\sigma,\ \sigma = \left(\sigma_u^2 + \sigma_v^2\right)^{1/2},\ \psi_i = \lambda\varepsilon_i/\sigma$, and $\lambda = \sigma_u/\sigma_v$. To implement this estimator, it must be evaluated at the estimated parameters $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}_u^2, \widehat{\sigma}_v^2\right)$ and the implied values of $\widehat{\lambda}$, $\widehat{\sigma}^2$, and $\widehat{\varepsilon}_i = Y_i - \widehat{\alpha} - X_i'\widehat{\beta}$. However, Wang and Schmidt (2009) demonstrated that the distribution of $\widehat{E}(u_i \mid \varepsilon_i)$ differs from the distribution of u_i unless $\sigma_v \to 0$. As σ_v^2 increases, it converges to $E(u_i)$, indicating that ε_i is no longer useful in predicting inefficiency through the conditional mean of the Jondrow et al. (1982) estimator. Battese and Coelli (1988) proposed an alternative efficiency estimator given by

$$E(TE_i \mid \varepsilon_i) = E\left(\exp\left(-u_i\right) \mid \varepsilon_i\right) = \frac{1 - \Phi\,(\sigma_S - \psi_i)}{1 - \Phi\,(\psi_i)} \exp\left(\sigma_S \psi_i + \left((\sigma_S)^2/2\right)\right)$$

Fried et al. (2008) argued that the efficiency estimator of Battese and Coelli (1988) is preferable to $1 - E(u_i \mid \varepsilon_i)$ used in the Jondrow et al. (1982) estimator because the latter is merely a first-order approximation to the more general infinite power series approximation, $\exp(-\,(u_i \mid \varepsilon_i)) = 1 - u_i + u_i^2/2 - u_i^3/3 \ldots\ldots$ However, Fried et al. (2008) and Kumbhakar et al. (2014) argued that Jondrow et al. (1982) and Battese and Coelli (1988) estimators are not consistent in cross-sectional models. Although these estimators are unbiased, they do not provide consistent estimates of efficiency, as $p \lim E(u_i \mid \varepsilon_i) - u_i \neq 0$ or $E(u_i \mid \varepsilon_i)$ never approaches u_i as the number of producers or cross-sectional units approaches infinity.

Greene (1990) argued that the half-normal assumption for the distribution of inefficiency is relatively inflexible and implicitly assumes that most producers are nearly fully efficient. Furthermore, the distribution of the composed error term, ε_i, is no longer normal, as observed by Horrace (2005). In fact, it might be positively skewed in the wrong direction, leading to full efficiency measures for all producers.[2] Waldman (1982) demonstrated that if ε_i is incorrectly skewed in the positive direction, then maximum likelihood estimates are equivalent to OLS estimates for $\left(\alpha, \beta, \sigma_u^2, \sigma_v^2\right)$ and zero for λ. The wrong skewness direction of the OLS composed error term, and consequently a zero maximum likelihood estimate of σ_u^2, is expected given the dependence of the maximum likelihood estimator for σ_u^2 on the skewness of the OLS composed error term in the normal-half-normal model.[3] Feng et al. (2015) suggested using constrained optimization methods to impose the restriction that $\sigma_u^2 > 0$ in the normal-half-normal model. Hafner et al. (2018) generalized the inefficiency distribution to allow for the existence of the incorrect skewness, thereby obtaining well defined inefficiency measures.

Pitt and Lee (1981), Kumbhakar (1987), and Battese and Coelli (1988) extended the normal-half-normal model proposed by Aigner et al. (1977) to the panel data time-invariant inefficiency model. Inefficiency is estimated using an extension of the Jondrow et al. (1982) estimator for the panel data model.

$$\widehat{u}_i = E(u_i \mid \varepsilon_{it}) = \varphi_i^N + \sigma_P \left[\frac{\phi\left(\varphi_i^N/\sigma_P\right)}{1 - \Phi\left(\varphi_i^N/\sigma_P\right)} \right] \tag{2}$$

where $\varphi_i^N = \left(-\sigma_u^2 \sum_{t=1}^T \varepsilon_{it}\right) / \left(\sigma_v^2 + T\sigma_u^2\right)$, and $\sigma_P = \left(\sigma_v^2 \sigma_u^2\right) / \left(\sigma_v^2 + T\sigma_u^2\right)$. Kumbhakar (1987) noted that these estimates are asymptotically consistent. Lee (1996) introduced a tail-truncated half-normal distribution to incorporate a bound for inefficiency, as the extremely inefficient producers cannot survive in highly competitive markets. He introduced the threshold parameter of the minimum efficiency for survival, denoted as θ, where $u_i \sim N^+(0, \sigma_u^2)$, $0 \leq u_i \leq \theta$. Thus, the variance of inefficiency depends on two parameters: σ_u^2 and θ.

4.1.2 Normal-exponential models

Aigner et al. (1977) and Meeusen and van den Broeck (1977) proposed a likelihood function under the assumption that u_i follows an exponential distribution; $f(u) = \theta \exp(-\theta u)$, where $\theta > 0$ and $\theta = \sigma_u^{-1}$. Additionally, v_i is assumed to follow a normal distribution, $v_i \sim N(0, \sigma_v^2)$. Inefficiency can be estimated using the Jondrow et al. (1982) estimator for exponentially distributed inefficiency in cross-sectional models.

$$\widehat{u}_i = E(u_i \mid \varepsilon_i) = \sigma_v \left[\frac{\phi\left(\varphi_i^E\right)}{\Phi\left(\varphi_i^E\right)} + \varphi_i^E \right]$$

where $\varphi_i^E = \left(\varepsilon_i \sigma_u - \sigma_v^2\right) / (\sigma_u \sigma_v)$. Kim and Schmidt (2000) extended the normal-exponential model proposed by Aigner et al. (1977) to a panel data time-invariant inefficiency model. Inefficiency is estimated using an extension of the Jondrow et al. (1982) estimator for the panel data model by replacing ε_i by $\overline{\varepsilon}_i$ and σ_v^2 by σ_v^2/T. A simulation study conducted by Horrace and Parmeter (2018) indicates that a Laplace model, where v_i follows a Laplace distribution and u_i follows a truncated Laplace distribution, performs relatively well compared to the normal-exponential model when v_i is misspecified.

4.1.3 Normal-Gamma models

Greene (1980a, 1980b), Stevenson (1980), and Greene (1990) assumed a gamma distribution for the inefficiency term, where $f(u) = \left[\theta^P / \Gamma(P)\right] \exp(-\theta u)\, u^{P-1}$, $P > 0$, $\theta = \sigma_u^{-1}$, $\Gamma(P) = \int_0^\infty t^{P-1} e^{-t} dt$, and $v_i \sim N(0, \sigma_v^2)$. Stevenson (1980) considered only the Erlang form (integer values of P; 1.0 and 2.0), which produces a tractable formulation for $f_\varepsilon(\varepsilon_i)$ but significantly restricts the model. Beckers and Hammond (1987) derived the log-likelihood function for $f_\varepsilon(\varepsilon_i)$ without limiting P to integer values; however, the resulting functional form was intractable. When $P = 1$, the normal-gamma model reverts to the normal-exponential model. The inefficiency estimator for the gamma model is

$$\widehat{u}_i = E(u_{it} \mid \varepsilon_{it}) = \frac{q\,(P, \varepsilon_{it})}{q\,(P-1, \varepsilon_{it})}$$

The normal-gamma distribution provides a more flexible parameterization of the distribution. However, the computational complexity of the maximum likelihood estimator limits its application in empirical studies. Various efforts, including those by Ritter and Simar (1997) and Greene (2003), have been made to simplify the computation using simulation methods.

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

4.1.4 Normal-truncated-normal models

Stevenson (1980) argued that the assumption of a zero mean in the Aigner et al. (1977) model is an unnecessary constraint. He suggested that inefficiency follows a non-negative truncated distribution, represented as $u_i \sim N(\mu, \sigma_u^2); f(u) = \left(\Phi\left(\mu/\sigma_u\right)\sigma_u\sqrt{2\pi}\right)^{-1}\exp\left(-\left(u-\mu\right)^2/2\sigma_u^2\right)$. Greene (1993) demonstrated that the conditional expectation of inefficiency for the truncated-normal distribution, where μ can vary from zero in either direction, is obtained by replacing ψ_i with $\psi_i^T = \lambda\varepsilon_i/\sigma + \mu/\lambda\sigma$.

Pitt and Lee (1981) expanded the normal-truncated-normal model to accommodate panel data with time-invariant inefficiency. Battese and Coelli (1988) and Battese et al. (1989) further developed the Jondrow et al. (1982) estimator for application in the panel data model.

$$\widehat{u}_i = E(u_i \mid \varepsilon_{it}) = \varphi_i^T + \sigma_P\left[\frac{\phi\left(\varphi_i^T/\sigma_P\right)}{1-\Phi\left(\varphi_i^T/\sigma_P\right)}\right]$$

where $\varphi_i^T = \left(\mu\sigma_v^2 - \sigma_u^2\sum_{t=1}^T\varepsilon_{it}\right)/\left(\sigma_v^2 + T\sigma_u^2\right)$. By setting $\mu = 0$, it reverts to the estimator for the normal-half-normal model in equation (2). Battese and Coelli (1988, 1992) extended the panel data to $E\left(\exp\left(-u_i\right) \mid \varepsilon_i\right)$ as

$$E\left(\exp\left(-u_i\right) \mid \varepsilon_{it}\right) = \left[\frac{\Phi\left[\left(\varphi_i^T/\sigma_P\right) - \sigma_P\right]}{\Phi\left(\varphi_i^T/\sigma_P\right)}\right]\exp\left(-\varphi_i^T + \left(\sigma_P/2\right)\right)$$

Almanidis et al. (2014) specified inefficiency as a doubly truncated-normal distribution. In addition to the zero lower bound, they specified an upper bound for inefficiency to exclude extremely inefficient producers. The upper bound, θ_t, is assumed to be the sum of weighted polynomials, $\theta_t = \sum_{i=0}^N b_i\left(t/T\right)^i$, where $t = 1, \ldots.., T$, and b_i are constants. Their specification provides a closed-form solution for $f_\varepsilon(\varepsilon_i)$ and the log-likelihood. Furthermore, this specification results in non-zero estimates of σ_u^2 in the presence of wrong skewness of the composed error term.

The truncated-normal distribution is applicable when producers are assumed to be inefficient, as it has a mode at zero only if $\mu \leq 0$. Furthermore, it provides a method for introducing heterogeneity into the distribution of inefficiency by incorporating characteristics specific to producers into the mean, variance, or both parameters of the inefficiency distribution.

4.1.5 Skew-normal models

Recent studies have focussed on the distribution of the composed error term $f_\varepsilon(\varepsilon_i)$ rather than on the distribution of inefficiency. In models with four random components, these components $(\omega_i + v_{it} - u_{i0} - u_{it})$ can be treated as two terms because they can be expressed as the sum of the time-invariant terms $(\xi_i = \omega_i - u_{i0})$ and the time-variant terms $(\varepsilon_{it} = v_{it} - u_{it})$. The time-invariant terms combine the producer-specific effects for unobserved heterogeneity (ω_i) and time-invariant inefficiency (u_{i0}), whereas the time-variant terms combine the random errors (v_{it}) and time-variant inefficiency (u_{it}). These two terms are assumed to result from the difference between a normal random variable and an independent normal random variable that is left-truncated at zero. Consequently, each of the two terms follows its own skew-normal distribution rather than a normal distribution.[4]

The full unconditional log-likelihood function for this model, based on the joint distribution of $(\varepsilon_{it}, \xi_i)$, was derived by Colombi et al. (2014). They estimated the four random components as $E\left(\exp\left(\omega_i\right) \mid y_i\right)$, and $E\left(\exp\left(t' u_i\right) \mid y_i\right)$, where the first element of $E\left(\exp\left(t' u_i\right) \mid y_i\right)$ represents the conditional expected value of the time-invariant inefficiency u_{i0} for each producer i. However, the computational complexity of the maximum likelihood estimator arises from the $(T+1)$ dimentional multivariate normal integrals.[5] Tsionas and Kumbhakar (2014) observed

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

that the maximum likelihood estimator developed by Colombi et al. (2014) becomes computationally challenging when T is large. However, for models with time-invariant inefficiency, whether or not they include a producer-specific component to account for heterogeneous technologies, the integral is one-dimensional. For models with time-variant inefficiency that do not include a producer-specific component for heterogeneous technologies, it is a product of T one-dimensional integrals. Consequently, the computational challenge is primarily associated with time-variant inefficiency models that include a producer-specific component for heterogeneous technologies, or time-invariant and time-variant inefficiency models, where time-variant inefficiency coexists with time-invariant inefficiency. Fillipini and Greene (2016) utilized Butler and Moffitt's (1982) formulation to propose a simplified density of y_i conditional on u_{i0} and ω_i, which is the product over time of T univariate closed skew-normal densities.

The efficiency estimates are derived using the results from Colombi et al. (2014), which are based on the moment-generating function for the closed skew-normal distribution. Inefficiency is calculated using the expression $-\log E\left(\exp\left(t^{'}u_i\right) \mid e_i\right)$, computed element by element. Fillipini and Greene (2016) followed the methodologies of Kumbhakar and Heshmati (1995) and Kumbhakar et al. (2014) to measure total efficiency as follows:

$$\text{Total efficiency}_{it} = E\left(\exp\left(-u_{i0}\right) \mid e_i\right) \times E\left(\exp\left(-u_{it}\right) \mid e_i\right)$$

Concerning the significance of distribution assumptions, if the goal is to estimate inefficiency specific to individual producers, then the assumption about the distribution of inefficiency becomes crucial. However, if the objective is to compare the rankings of producers, employing models without distribution assumptions or adhering to the suggestion by Ritter and Simar (1997) to utilize a simple one-parameter distribution for inefficiency may be adequate.

4.1.6 Confidence intervals

Distributions imposed on v and u create distributions for $(u \mid \varepsilon)$ and $\left(\exp\left(-u\right) \mid \varepsilon\right)$, which can be used to construct confidence intervals for inefficiency. Several studies have demonstrated that it is possible to obtain confidence intervals for any of the inefficiency estimators. Hjalmarsson et al. (1996) developed confidence intervals for the Jondrow et al. (1982) estimator, and Bera and Sharma (1999) did so for the Battesse and Coelli (1988) estimator. Horrace and Schmidt (1996) derived lower and upper bounds on $\left(\exp\left(-u\right) \mid \varepsilon\right)$ based on the lower and upper bounds of $(u \mid \varepsilon)$. However, Wheat et al. (2014) argued that the form of confidence intervals derived by Horrace and Schmidt (1996) is not of minimum width because $f\left(u \mid \varepsilon\right)$ is truncated-normal at zero and thus asymmetric. They proposed a minimum width prediction interval for u given ε. Parmeter and Kumbhakar (2014) noted that the narrower interval proposed by Wheat et al. (2014) is preferable to the intervals of Horrace and Schmidt (1996) if the aim is to accurately predict producer-specific inefficiency.

4.2 Bayesian estimation

Bayesian estimation of inefficiency was initially introduced in the literature for cross-sectional models by Van den Broeck et al. (1994) and Koop et al. (1994, 1995). Koop et al. (1997), Fernandez et al. (1997), and Osiewalski and Steel (1998) expanded the application of Bayesian estimation to panel data models. Koop et al. (1997) outlined procedures for Bayesian estimation in both fixed-effects and random-effects models. Fernandez et al. (2000, 2002) further extended the use of Bayesian estimation to situations where some outputs produced might be undesirable, differentiating between technical and environmental inefficiency.

The Bayesian approach treats the model parameters as random variables that are conditional on the data, rather than as known or fixed values estimated solely from the data. Instead of using

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

the distribution of u conditional on ε, $E(u \mid \varepsilon)$, inefficiency inference is derived from the conditional posterior distribution, $p(u \mid \theta_{-u}, Y)$, based on its marginal posterior, where θ_{-u} includes all parameters except u. Bayesian inference, including point and interval estimation, hypothesis evaluation, and prediction, is obtained from the posterior distribution.

Bayesian estimation involves specifying prior distributions for the unknown parameters, $p(\theta)$; deriving the likelihood function, $L(Y \mid \theta)$; and deriving the joint posterior distributions for all the parameters by using Bayes' theorem to combine the likelihood function with the joint prior distributions. Iterative Markov Chain Monte Carlo (MCMC) methods are then employed to obtain the marginal posteriors for the estimated parameters.

4.2.1 The prior distribution

The prior distributions of the parameters to be estimated, including inefficiency, represent information that is not contained in the data and are expressed as a probability distribution, $p\,(\theta)$. These distributions are categorized as either informative priors, which are based on previous findings and theoretical predictions, or uninformative priors, which are based on a lack of prior knowledge available for estimation.

Specifying a uniform or flat prior distribution allows the prior to play a minimal role in estimating the posterior distribution by relying on the data through the likelihood function. This approach is equivalent to specifying a prior distribution with a large variance, which makes the prior distribution of the parameter values nearly flat. However, non-informative prior distributions are often improper. Fernandez et al. (1997) demonstrated that choosing an uninformative prior on the scale parameter leads to an improper prior.

Informative priors convey information and summarize existing knowledge about parameters. Since the normal distribution allows for negative numbers, it is not suitable as a prior distribution for inefficiency or scale parameters. Van den Broeck et al. (1994) found that the exponential distribution is more robust to prior assumptions than other distributions. Alvarez et al. (2014) compared the inverse Wishart, scaled inverse Wishart, and hierarchical inverse Wishart as potential priors for the scale parameter in multivariate models. They found that all priors perform well except the inverse Wishart prior, which is biased toward large values when the true variance is small relative to the prior mean. Esheba and Serletis (2023) utilized the Wishart distribution for the scale parameter and found it to be biased toward large values, resulting in large values for the scale parameter and consequently large values for the inefficiency measures. They also noted that the MCMC algorithm for a system of equations terminated after a small number of iterations due to the large values involved. In general, the prior distribution for the scale parameter, which plays a crucial role in the estimation of inefficiency, is essential in any multivariate model and becomes more challenging as the dimension increases due to the quadratic growth in the number of parameters and the need to ensure the matrix remains non-negative definite.

Informative priors can be utilized to impose constraints derived from economic theory, such as monotonicity and curvature constraints, as noted in Terrell (1996). They can also be employed for linear constraints among the elements of the parameters, as demonstrated by Geweke (1993), or for constraints on inefficiency, where $u \geq 0$, as shown by Feng et al. (2018) and Esheba and Serletis (2023). It is important to note that selecting a prior distribution that is conjugate to the likelihood results in a posterior that retains the same form as the prior.

4.2.2 The posterior distribution

Updating the prior information of the parameters is accomplished by combining the prior distribution, $p(\theta)$, with the likelihood function, $L(Y, \theta)$. This process results in the posterior distribution, which serves as the basis for Bayesian estimation and is defined by Bayes' theorem as follows:

$$p(\theta \mid Y) \propto L(Y, \theta)p\,(\theta)$$

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

In this context, $p(\theta \mid Y)$ represents the posterior distribution, which is proportional to the likelihood function multiplied by the prior. The posterior mean $E(\theta \mid Y)$ serves as the optimal Bayesian estimator of θ. However, when the model involves multidimensional parameters to be estimated, the posterior distribution becomes a joint posterior distribution. The marginal posterior distribution for a given parameter θ_i is defined by integrating the joint posterior density of θ with respect to all elements of θ other than θ_i. This process can be too complex for direct analytical integration or may not be analytically tractable. Implementing the Bayesian approach often requires the use of an iterative MCMC algorithm. Two common algorithms are Gibbs sampling, introduced by Geman and Geman (1984), and the MetropolisHastings algorithm, introduced by Metropolis et al. (1953) and further developed by Hastings (1970).

When the joint posterior distribution is challenging to work with, Gibbs sampling, which involves drawing from the conditional posterior distributions, can be employed to approximate joint and marginal distributions[6] This method is advantageous in situations where the conditional posterior distributions have relatively simpler forms than the joint distribution, facilitating simulation. To assess whether the draws from the conditional posterior distributions have converged to the marginal posterior distribution, Geweke (1992) proposed a convergence test. If there is insufficient evidence for convergence, the number of draws must be increased.

The Gibbs sampling algorithm relies on conditional distributions. However, in some situations, conditional posterior distributions do not belong to any known family distributions or are not available in closed form, making simulation from them challenging. In such situations, the Metropolis-Hastings algorithm, which is more general than Gibbs sampling, serves as an alternative MCMC algorithm that can be used to approximate the posterior distribution. The Metropolis-Hastings algorithm requires the specification of a proposal density, $q(\theta^* \mid \theta^S)$, which is easier to simulate from than the target density, $p(\theta \mid Y)$.

The initial MCMC iterations are discarded as a burn-in, and estimates of the parameters are obtained by averaging over the remaining iterations. It should be noted that the Gibbs sampler is considered a specific case of the Metropolis-Hastings algorithm, where the candidate density $q(\theta^* \mid \theta^S)$ coincides with the target density, resulting in an acceptance probability of 1 for each draw.

4.3 Theoretical regularity

As required by microeconomic theory, production technology must satisfy the theoretical regularity conditions of monotonicity and curvature. Barnett (2002) noted that these regularity conditions can be violated unless they are imposed. However, Lau (1986) demonstrated that imposing these regularity conditions globally can compromise the flexibility of flexible functional forms.[7] Ryan and Wales (2000) demonstrated that imposing curvature locally at a single point can be sufficient to achieve global regularity while preserving the flexibility of flexible functional forms. Terrell (1996) indicated that imposing regularity conditions over small regions of data can preserve the flexibility of flexible functional forms. Wolff (2016) imposed regularity conditions locally, globally, and regionally on a flexible input demand system using the same data set. He found that regional estimators outperform global and local estimators in terms of the model's fit to the sample data and preserving the flexibility of flexible functional forms.

Barnett (2002) and Barnett and Pasupathy (2003) noted that monotonicity conditions have often been disregarded in stochastic frontier estimation. However, monotonicity is crucial in assessing inefficiency, as it ensures that additional units of input do not reduce output. Violating monotonicity conditions can lead to highly misleading results, such as incorrectly identifying a producer as efficient when it is not. Consider an example of two producers, Producer A and Producer B, with a non-monotone technology frontier, as illustrated in Figure 6. Under this non-monotone technology frontier, Producer A is deemed efficient, while Producer B is considered inefficient because it operates below the production frontier. However, Producer B produces the

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

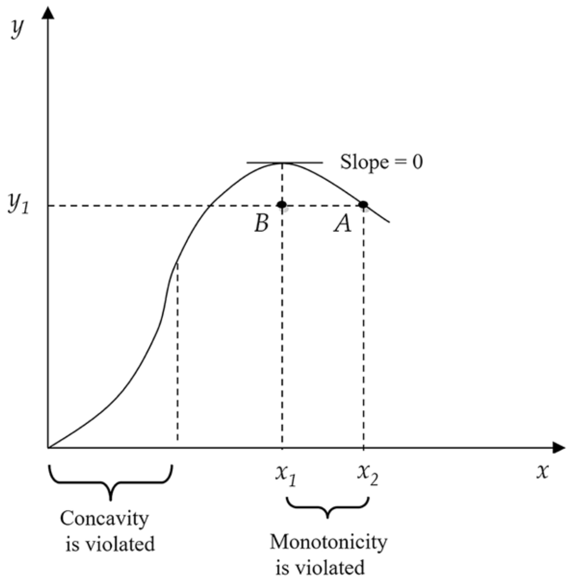

Figure 6. Violation of monotonicity and curvature conditions.

same output (y_1) as Producer A but uses less input $(x_1 < x_2)$. Consequently, the technical inefficiency measures of these two producers based on this non-monotone technology frontier are reversed; in this situation, Producer A is inefficient relative to Producer B. Imposing monotonicity conditions prevents the production technology from exhibiting negative marginal productivities, which are implied by a downward-sloping production frontier, such as at point A.

Curvature conditions are primarily required by microeconomic theory for the duality theorem to hold. Unless exploiting the duality theory and using the dual system specifications, measuring inefficiency does not require curvature conditions; instead, monotonicity conditions must be satisfied. In general, the regularity conditions can be verified as follows:

- Monotonicity is verified by analyzing the first-order derivatives of the estimated production technology with respect to the input and output. Refer to Table 1 for the properties of various distance functions.
- Concavity (convexity) is determined using the unbordered Hessian matrix, which must be negative (positive) semi-definite. This determination can be verified by examining whether the values of the Cholesky factors are non-positive (negative), as discussed by Lau (1978a).

If the regularity conditions are not attained, the model can be estimated by imposing these conditions, thereby treating them as maintained hypotheses. This process may require the use of Bayesian estimation to enforce the necessary inequality restrictions for regularity conditions. Regularity conditions can be imposed either locally at a single point in the regressor space, globally across the entire domain, or regionally on a connected subset of the domain. Techniques for imposing regularity conditions locally have been developed by Ryan and Wales (1998), globally by Lau (1978a) and Diewert and Wales (1987), and regionally by Gallant and Golub (1984), Terrell (1996), Wolff et al. (2010), and Wolff (2016).

4.4 *Econometric regularity*

The non-stationarity of residuals in production technology is an important issue when estimating inefficiency, as inefficiency measures are derived from these estimated residuals. However,

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

non-stationarity is often disregarded in inefficiency studies, primarily because standard methods for addressing non-stationarity in linear models do not apply to non-stationarity in non-linear models. Nonetheless, ignoring the potential non-stationarity of the residuals can lead to misleading inefficiency results.[8] Barnett (1977) demonstrated that consistency and asymptotic efficiency require stationarity assumptions as part of the econometric regularity conditions.

The non-stationarity of the residuals in production technology can result from the non-stationarity of either the dependent or explanatory variables, or from the omission of non-stationary variables. When all variables are integrated of order one, denoted as $I(1)$ in the terminology of Engle and Granger (1987), the production technology represents a cointegrating relationship, and the OLS provides super-consistent estimates.[9] If all variables are non-stationary, the production technology represents a spurious relationship, leading to significantly misleading inefficiency measurements and rendering the results meaningless. Additional complications arise if the production technology is unbalanced, with different variables having different orders of integration or some variables being stationary while others are non-stationary. Feng and Serletis (2008) presented an empirical comparison and evaluation of the effectiveness of four flexible cost functions: the locally flexible generalized Leontief (GL), translog, and normalized quadratic (NQ), as well as one globally flexible cost function, the asymptotically ideal model (AIM). They found that the GL and translog models fail both economic and econometric regularity, and the NQ and AIM models fail econometric regularity, indicating that these models are non-stationary.

Serially correlated residuals are commonly modeled in the literature by assuming a first-order autoregressive process, $AR(1)$, in the error terms as $\varepsilon_t = \rho\varepsilon_{t-1} + e_t$, where ρ is an unknown parameter and e_t is a non-autocorrelated random error term. The $AR(1)$ process is stationary when $|\rho| < 1$ and becomes a non-stationary random walk process when $\rho = 1$. Consequently, tests for stationarity can be conducted by examining whether ρ is equal to one or significantly less than one. These tests are referred to as unit root tests for stationarity. Additionally, the following tests are utilized to assess the presence of a unit root and non-stationarity in the residuals of the production technology: the augmented Dickey-Fuller test proposed by Dickey and Fuller (1981), the non-parametric test of Phillips (1987), the numerical Bayesian test by Dorfman (1995), the test proposed by Harris and Tzavalis (1999) for dynamic panels, and the Fisher test by Maddala and Wu (1999).

If stationarity is not achieved, cointegration techniques can be employed to address the non-stationarity of the residuals.[10] If all variables are non-stationary, these variables must be cointegrated in levels, given that inefficiency models are linear. Ng (1995) and Attfield (1997) argued that standard estimation techniques are inadequate for obtaining accurately estimated standard errors in cointegrated panels. Tsionas and Christopoulos (2001) applied panel cointegration techniques to estimate inefficiency using Fully Modified Ordinary Least Squares (FM-OLS), as proposed by Phillips and Hansen (1990), Phillips (1995), Phillips and Moon (1999), and Pedroni (2001) for cointegrated panels.[11] They compared their results with those obtained by estimating inefficiency using standard estimation techniques and found significant quantitative differences. However, these cointegration techniques are applicable to linear models. Park and Hahn (1999) considered models that are linearized in the non-stationary variables. Lewbel and Ng (2005) proposed a reformulation of the translog model, which can be modified into a linear form to address non-stationarity.

If cointegration between the $I(1)$ variables is not found, a suitable solution is to convert the non-stationary series to stationary series by taking first differences if they are difference stationary, by de-trending, or alternatively by including a trend variable in the model if they are trend stationary. However, Serletis and Shahmoradi (2007) argued that correcting serially correlated residuals increases the number of curvature violations and induces spurious violations of monotonicity.

Several attempts have been proposed in the literature to develop estimation techniques for non-stationary models. Chang et al. (2001) extended earlier work by Phillips and Hansen (1990) and developed an estimator for non-linear, non-stationary models. Their estimator is consistent under

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

fairly general conditions, but the convergence rate critically depends on the type of functional form. Han and Phillips (2010) proposed a consistent GMM estimation method for estimating autoregressive roots near unity with both time series and panel data. However, their estimator has little bias even in very small samples. Therefore, with non-linear, non-stationary inefficiency models, further research is needed to modify linear model cointegration techniques and to advance existing non-linear cointegration techniques.

5. Estimation issues

The estimates of inefficiency can be distorted by an inaccurate choice of the functional form for production technology, ignoring the possibility of heterogeneity and heteroskedasticity, and suffering from the endogeneity problem.

5.1 Functional forms

The estimates of inefficiency can be distorted by an inaccurate selection of the functional form for production technology. Berger and Mester (1997) argued that achieving a close fit of the actual data for the estimated production frontier is crucial for estimating technical inefficiency, as technical inefficiency is assessed based on deviations from this production frontier. Giannakas et al. (2003b) demonstrated that an inaccurate choice of the functional form results in biased estimates of inefficiency, confidence intervals, and production elasticities.

Although the true functional form is unknown, several properties of the production technology are known from economic theory. Various empirical methods can be employed to evaluate how well different functional forms approximate the unknown underlying function. A functional form may be considered appropriate due to its theoretical properties, the feasibility and ease of application and empirical estimation, or a combination of these criteria. However, many studies do not explicitly state the rationale for selecting a specific functional form for production technology.

The selection of a specific functional form for production technology can be based on theoretical properties such as the shape of isoquants, separability, flexibility, and regular regions. Greene (1993) noted that the choice of functional form for production technology has important implications with respect to the shape of the isoquants. Färe and Vardanyan (2016) compared the quadratic and translog functional forms in terms of their ability to approximate convex frontiers of the input set and found that both functional forms provide a reliable approximation when a true frontier is assumed to be convex. Their findings support those of Färe et al. (2010) and Chambers et al. (2013), who found that the translog functional form tends to yield convex frontier estimates even when the true frontier is concave. Therefore, the translog functional form, which can approximate convex frontiers of the input set, should perform relatively well when modeling input isoquants, such as input distant functions. On the other hand, if the true production frontier is concave, simulation studies by Färe et al. (2010) and Chambers et al. (2013) suggest that the concave frontier of the output set is better parameterized using a quadratic functional form than a translog functional form. Chambers et al. (2013) further found that the translog specification of a concave frontier can yield imprecise estimates of the technology. Consequently, the quadratic functional form, which can approximate concave frontiers of the output set, should perform relatively well when modeling output isoquants, such as standard or directional output distance functions. The separability properties are important for consistent aggregation. Thompson (1988) noted that both the translog and the quadratic are separable functional forms.

The selection of a specific functional form for particular studies can be based on choosing between functional forms that globally satisfy the theoretical regularity conditions of economic theory and those that possess flexibility. Flexible functional forms are characterized by their second-order approximation property and are sufficiently flexible to ensure that the production elasticities and substitution elasticity are not restricted by the choice of the functional form.[12]

However, the selection of functional forms for estimating inefficiency should prioritize regular functional forms that are consistent with economic theory, rather than focusing on flexibility. Greene (1980b) argued that flexible functional forms might suffer from multicollinearity due to the large number of parameters that need to be estimated, and single-equation estimates are likely to be imprecise.

Different distance functions have various application properties that influence the choice of functional forms, such as homogeneity and translation properties.[13] For instance, the selection of a functional form for standard distance functions should be based on the satisfaction of the homogeneity property. Griffin et al. (1987) noted that commonly used functional forms that are not linearly homogeneous include logarithmic and augmented Fourier forms. Some functional forms can be linearly homogeneous by incorporating the appropriate restrictions, such as the quadratic, Cobb-Douglas, transcendental, constant elasticity of substitution, and the translog. In contrast, selecting a functional form for performing directional distance functions should be based on satisfying the translation property and ensuring homogeneity of degree (-1) in the direction vector. Chambers (1998) suggested two functional forms that satisfy the translation property: the logarithmic transcendental and the quadratic. Further research is needed to identify alternative functional forms that satisfy both the translation property and the homogeneity of degree (-1) in the direction vector.

5.1.1 The translog functional form

The translog functional form is a generalization of the Cobb-Douglas functional form and was introduced by Christensen et al. (1973). It is a locally flexible functional form that provides a second-order local approximation. Caves and Christensen (1980), Guilkey and Lovell (1980), Barnett and Lee (1985), and Barnett et al. (1985) argued that most locally flexible functional forms are not globally regular and have very small regions where theoretical regularity conditions are satisfied. The translog functional form is defined over N inputs and M outputs as follows:

$$\ln\left(D\left(x,y\right)\right)=\alpha_0+\sum_{n=1}^{N}\alpha_n\ln x_n+\sum_{m=1}^{M}\beta_m\ln y_m+\frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\alpha_{nn'}\ln x_n\ln x_{n'}$$
$$+\frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M}\beta_{mm'}\ln y_m\ln y_{m'}+\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_{nm}\ln x_n\ln y_m$$

Symmetry requires that $\alpha_{nn'}=\alpha_{n'n}$ $(n\neq n')$, and $\beta_{mm'}=\beta_{m'm}$ $(m\neq m')$. The translog functional form includes numerous parameters that require estimation, totaling $\left(k^2+3k+2\right)/2$ parameters, including the intercept. It is linear in the parameters, which can be constrained to satisfy the homogeneity property of standard and hyperbolic distance functions; however, it cannot be constrained to satisfy the translation property of directional distance functions.

The constraints necessary for achieving homogeneity of degree one in inputs are: $\sum_{n=1}^{N}\alpha_n=1$, $\sum_{n'=1}^{N}\alpha_{nn'}=0$, and $\sum_{n=1}^{N}\gamma_{nm}=0$. One approach to imposing these restrictions on the input distance function is to normalize the function by one of the inputs. This is achieved by setting the parameter of the homogeneity property to $\lambda=1/x_N$, resulting in $D_I\left(y,x/x_N\right)=D_I\left(y,x\right)/x_N$. For further details, refer to Sturm and Williams (2008). The constraints necessary for achieving homogeneity of degree one in outputs are: $\sum_{m=1}^{M}\beta_m=1$, $\sum_{m'=1}^{M}\beta_{mm'}=0$, and $\sum_{m=1}^{M}\gamma_{nm}=0$. One method to impose these restrictions on the output distance function is to normalize the function by one of the outputs. This is achieved by setting the parameter of the homogeneity property to $\lambda=1/y_M$, resulting in $D_O\left(x,y/y_M\right)=D_O\left(x,y\right)/y_M$. For further details, refer to O'Donnell and Coelli (2005). The restrictions required for almost homogeneity of degrees -1, 1, and 1 are: $\sum_{m=1}^{M}\beta_m-\sum_{n=1}^{N}\alpha_n=1$, $\sum_{m=1}^{M}\gamma_{nm}-\sum_{n'=1}^{N}\alpha_{nn'}=0$, and $\sum_{m'=1}^{M}\beta_{mm'}-\sum_{n=1}^{N}\gamma_{nm}=0$.

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

One approach to imposing these restrictions on the hyperbolic distance function is to normalize the function by one of the inputs. This is achieved by setting the parameter of the homogeneity property to $\lambda = 1/x_N$, resulting in $D_H\left(x/x_N, yx_N\right) = D_H\left(x, y\right)/x_N$. Alternatively, normalization can be accomplished by one of the outputs by setting $\lambda = 1/y_M$, resulting in $D_H\left(xy_M, y/y_M\right) = D_H\left(x, y\right)/y_M$. For further details, refer to Cuesta and Zofio (2005). The first-order and second-order partial derivatives are expressed as follows:

$$\frac{\partial \ln\left(D\left(x,y\right)\right)}{\partial \ln x_n} = \alpha_n + \sum_{n'=1}^{N} \alpha_{nn'} \ln x_{n'} + \sum_{m=1}^{M} \gamma_{nm} \ln y_m, \qquad \frac{\partial^2 \ln\left(D\left(x,y\right)\right)}{\partial \ln x_n \ln x_{n'}} = \alpha_{nn'}$$

$$\frac{\partial \ln\left(D\left(x,y\right)\right)}{\partial \ln y_m} = \beta_m + \sum_{m'=1}^{M} \beta_{mm'} \ln y_{m'} + \sum_{n=1}^{N} \gamma_{nm} \ln x_n, \qquad \frac{\partial^2 \ln\left(D\left(x,y\right)\right)}{\partial \ln y_m \ln y_{m'}} = \beta_{mm'}$$

The translog functional form, unlike the Cobb-Douglas functional form, is neither monotonic nor globally convex. Caves and Christensen (1980) noted that the translog functional form exhibits satisfactory local properties when the technology is nearly homothetic and the substitution between factors of production is high. Guilkey et al. (1983) demonstrated that the translog functional form is globally regular if and only if the technology is Cobb-Douglas. Färe and Vardanyan (2016) found that the translog functional form often violates theoretical regularity conditions and requires the imposition of appropriate regularity conditions, which significantly compromise its flexibility. Their findings align with the simulation results of Wales (1977) and Guilkey et al. (1983), who compared the performance of various functional forms, including the translog.

5.1.2 The quadratic functional form

Chambers (1998) suggested the use of a quadratic functional form for directional distance functions, as its parameters can be constrained to satisfy the translation property. Lau (1978b) introduced this quadratic functional form, which is expressed as follows:

$$D\left(x,y\right) = \alpha_0 + \sum_{n=1}^{N} \alpha_n x_n + \sum_{m=1}^{M} \beta_m y_m + \frac{1}{2} \sum_{n=1}^{N} \sum_{n'=1}^{N} \alpha_{nn'} x_n x_{n'} + \frac{1}{2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \beta_{mm'} y_m y_{m'} + \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_{nm} x_n y_m \qquad (3)$$

Symmetry requires that $\alpha_{nn'} = \alpha_{n'n}$ $(n \neq n')$, and $\beta_{mm'} = \beta_{m'm}$ $(m \neq m')$. The quadratic functional form includes numerous parameters that require estimation, totaling $\left(k^2 + 3k + 2\right)/2$ parameters, including the intercept. It is linear in the parameters, which can be restricted to satisfy the translation property of the directional distance functions. The set of linear restrictions required for the translation property is as follows: $\sum_{m=1}^{M} \beta_m g_{y_m} - \sum_{n=1}^{N} \alpha_n g_{x_n} = -1$; $\sum_{m=1}^{M} \beta_{mm'} g_{y_m} = 0$; $\sum_{n=1}^{N} \alpha_{nn'} g_{x_n} = 0$; $\sum_{m=1}^{M} \gamma_{nm} g_{y_m} = 0$; and $\sum_{n=1}^{N} \gamma_{nm} g_{x_n} = 0$. One approach to imposing these restrictions is to apply them directly in Equation (3) to derive a restricted version, as illustrated by Atkinson and Tsionas (2016). Alternatively, these restrictions can be imposed by setting the parameter of the translation property, α, equal to a selected input, $\alpha = x_N$, or the negative of a selected output, $\alpha = -y_M$. The corresponding direction vector is then normalized so that $g_{x_N} = 1$ or $g_{y_M} = 1$. In the situation where $\alpha = x_N$ is chosen, the expression $\vec{D}_T\left(\widetilde{x} - x_N \widetilde{g}_x, y + x_N g_y; g_x, g_y\right) = \vec{D}_T\left(x, y; g_x, g_y\right) - x_N$ holds, where $\widetilde{x} = (x_1, \ldots.., x_{N-1})$ and $\widetilde{g}_x = (g_{x_1}, \ldots., g_{x_{N-1}})$. The input x_N is absent from $\vec{D}_T\left(\widetilde{x} - x_N \widetilde{g}_x, y + x_N g_y; g_x, g_y\right)$ because $x_N - x_N(1) = 0$. In the situation where $\alpha = -y_M$ is chosen, the expression $\vec{D}_T\left(x + y_M g_x, \widetilde{y} - y_M \widetilde{g}_y; g_x, g_y\right) = \vec{D}_T\left(x, y; g_x, g_y\right) + y_M$ holds, where $\widetilde{y} = (y_1, \ldots.., y_{M-1})$ and $\widetilde{g}_y = (g_{y_1}, \ldots., g_{y_{M-1}})$. The output y_M is absent from

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

$\vec{D}_T\left(x + y_M g_x, \widetilde{y} - y_M \widetilde{g}_y; g_x, g_y\right)$ because $y_M - y_M(1) = 0$; see, for example, Malikov et al. (2016). The first-order and second-order partial derivatives are expressed as follows:

$$\frac{\partial D\left(x,y\right)}{\partial x_n} = \alpha_n + \sum_{n'=1}^{N} \alpha_{nn'} x_{n'} + \sum_{m=1}^{M} \gamma_{nm} y_m, \qquad \frac{\partial^2 D\left(x,y\right)}{\partial x_n x_{n'}} = \alpha_{nn'}$$

$$\frac{\partial D\left(x,y\right)}{\partial y_m} = \beta_m + \sum_{m'=1}^{M} \beta_{mm'} y_{m'} + \sum_{n=1}^{N} \gamma_{nm} x_n, \qquad \frac{\partial^2 D\left(x,y\right)}{\partial y_m y_{m'}} = \beta_{mm'}$$

Thompson (1988) noted that the quadratic functional form can satisfy global curvature restrictions without additional constraints in estimation. This finding is supported by the results of Färe and Vardanyan (2016), who demonstrated that the quadratic functional form satisfies global regularity without curvature restrictions and maintains its flexibility. A simulation study by Chambers et al. (2013) suggests that the quadratic functional form outperforms the translog in large samples with a relatively high degree of curvature. Diewert and Fox (2008) noted that curvature restrictions can be globally imposed on the quadratic functional form without losing flexibility. However, monotonicity cannot be imposed simultaneously with curvature without compromising the flexibility of the functional form. As Barnett (2002) noted, imposing global curvature on the quadratic functional form may induce spurious violations of monotonicity.

5.1.3 The logarithmic-transcendental functional form

Chambers (1998) suggested the use of a logarithmic-transcendental functional form for directional distance functions due to its inherent satisfaction of the translation property. However, it has been largely ignored in the inefficiency literature, as researchers have favored the quadratic functional form for its linearity in parameters. The logarithmic-transcendental or the transcendental-exponential functional form is a flexible functional form that provides a second-order approximation and can be expressed as follows:

$$\begin{aligned} \exp\left(D\left(x,y\right)\right) = \alpha_0 &+ \frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N} \alpha_{nn'} \exp\left(\frac{x_n}{2}\right)\exp\left(\frac{x_{n'}}{2}\right) \\ &+ \frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M} \beta_{mm'} \exp\left(-\frac{y_m}{2}\right)\exp\left(-\frac{y_{m'}}{2}\right) + \sum_{n=1}^{N}\sum_{m=1}^{M} \gamma_{nm} \exp\left(\frac{x_n}{2}\right)\exp\left(-\frac{y_m}{2}\right) \end{aligned} \tag{4}$$

The logarithmic-transcendental functional form requires fewer parameters to be estimated than the translog and quadratic functional forms, with a total of $\left(k^2 + k + 2\right)/2$ parameters, including the intercept. Symmetry requires that $\alpha_{nn'} = \alpha_{n'n}$ $(n \neq n')$, and $\beta_{mm'} = \beta_{m'm}$ $(m \neq m')$. The first-order partial derivatives are expressed as follows:

$$\frac{\partial \exp\left(D\left(x,y\right)\right)}{\partial \exp\left(\frac{x_n}{2}\right)} = \sum_{n'=1}^{N} \alpha_{nn'} \exp\left(\frac{x_{n'}}{2}\right) + \sum_{m=1}^{M} \gamma_{nm} \exp\left(-\frac{y_m}{2}\right)$$

$$\frac{\partial \exp\left(D\left(x,y\right)\right)}{\partial \exp\left(-\frac{y_m}{2}\right)} = \sum_{m'=1}^{M} \beta_{mm'} \exp\left(-\frac{y_{m'}}{2}\right) + \sum_{n=1}^{N} \gamma_{nm} \exp\left(\frac{x_n}{2}\right)$$

The second-order partial derivatives are expressed as follows:

$$\frac{\partial^2 \exp\left(D\left(x,y\right)\right)}{\partial \exp\left(\frac{x_n}{2}\right)\exp\left(\frac{x_{n'}}{2}\right)} = \alpha_{nn'}$$

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

$$\frac{\partial^2 \exp\left(D\left(x,y\right)\right)}{\partial \exp\left(-\frac{y_m}{2}\right)\exp\left(-\frac{y_{m'}}{2}\right)} = \beta_{mm'}$$

Empirical techniques can also be used to assess the ability of different functional forms to approximate the unknown underlying function. Several methods have been proposed in the literature, including Monte Carlo simulations, parametric modeling, and constructive techniques. Monte Carlo simulations assess the approximation capabilities of different functional forms relative to the underlying technology. For applications of this technique, see, for example, Guilkey and Lovell (1980), Giannakas et al. (2003b), Färe et al. (2010), Chambers et al. (2013), and Färe and Vardanyan (2016). Parametric modeling examines the plausibility of various functional forms in fitting actual data. For applications of this technique, see, for example, Griffin et al. (1987), Giannakas, et al. (2003a), and Feng and Serletis (2008). The primary challenge with parametric modeling is that the true functional form for production technology is unknown. Evaluating the performance of different functional forms in fitting actual data is beneficial if the focus is on analyzing the data itself, rather than the functional forms. As noted by Giannakas, et al. (2003a), the appropriate functional form in this context is specific to each situation. Constructive techniques provide a method to determine preferable functional forms by deriving and graphically displaying their regular regions. For applications of this technique, see, for example, Caves and Christensen (1980), and Barnett et al. (1985, 1987).

5.2 Heterogeneity issue

The selection of an appropriate functional form for production technology is insufficient without considering heterogeneity within the production model. Heterogeneity may manifest in the technology by shifting the production frontier, in the inefficiency term by altering the inefficiency distribution, or in both. To account for heterogeneous technologies, it is essential to include producer-specific characteristics directly in the functional form of the technology. Since inefficiency heterogeneity affects the location and scale parameters of the inefficiency distribution, it can be addressed by incorporating producer-specific characteristics either in the inefficiency term or in the parameters of the inefficiency distribution. Greene (2005a) argued that these characteristics are important sources of heterogeneity that have been largely ignored in the inefficiency literature.

5.2.1 Heterogeneity in the production technology

Ignoring the diversity of technologies among producers can lead to incorrect conclusions regarding inefficiency measures, as demonstrated by Casu and Molyneux (2003) and Bos and Schmiedel (2007). Brown and Glennon (2000) noted that assuming a uniform production technology for all producers is a highly restrictive assumption. According to Tsionas (2002), assuming a common technology for all producers might result in a producer being ranked as inefficient, even though the producer employs a different technology and fully utilizes its own resources. Mester (1997), Greene (1993, 2005b) and Caiazza et al. (2016) confirmed that heterogeneity leads to biased estimates obtained from the stochastic frontier approach.

To illustrate the importance of accommodating heterogeneity in the production frontier when estimating inefficiency, consider an example involving two producers, labeled A and B, with production frontiers denoted as A and B, respectively, as shown in Figure 7. If a common frontier C is assumed for these two producers, the directional measure of overall technical inefficiency is given by $TI_T^{pooled}(x_A, y_A) = \|AC^T\| \,/\, \|0g\|$ for producer A and $TI_T^{pooled}(x_B, y_B) = \|BC^T\| \,/\, \|0g\|$ for producer B. Under this assumption, producer A appears less efficient than producer B because $\|AC^T\| > \|BC^T\|$. However, when considering each producer operating on their own frontier, the measures become $TI_T^{own}(x_A, y_A) = \|AA^T\| \,/\, \|0g\|$ and $TI_T^{own}(x_B, y_B) = \|BB^T\| \,/\, \|0g\|$.

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

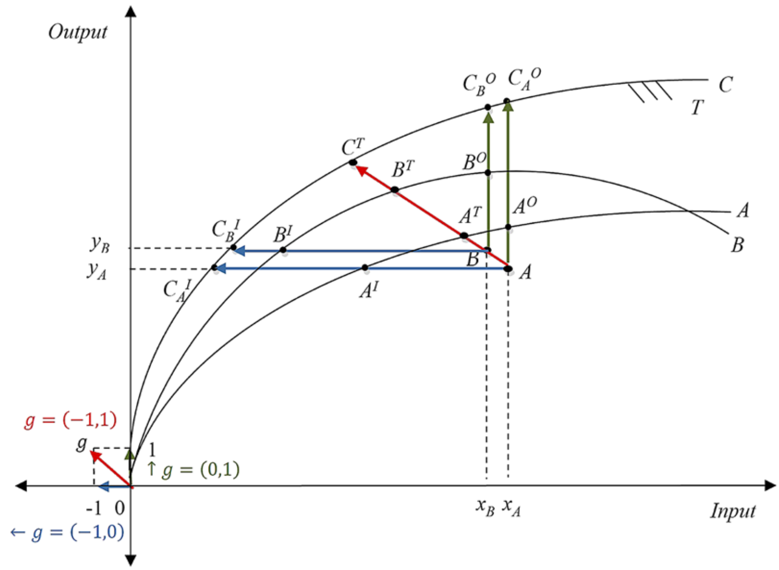

Figure 7. Heterogeneous technologies and inefficiency.

Consequently, the efficiency ranking of these two producers reverses, as in this case, producer B appears less efficient than producer A because $\|BB^T\| > \|AA^T\|$. Similarly, the directional measure of output technical inefficiency is expressed as $TI_O^{pooled}(x_A, y_A) = \|AC_A^O\|$ for producer A and $TI_O^{pooled}(x_B, y_B) = \|BC_B^O\|$ for producer B, assuming a common frontier C. Under this assumption, producer A is less efficient than producer B. However, when each producer operates on their individual frontier, the measures become $TI_O^{own}(x_A, y_A) = \|AA^O\|$ and $TI_O^{own}(x_B, y_B) = \|BB^O\|$. Consequently, the efficiency ranking of these two producers is reversed. In this situation, producer B is less efficient than producer A because $\|BB^O\| > \|AA^O\|$.

The directional measure of input technical inefficiency is expressed as $TI_I^{pooled}(x_A, y_A) = \|AC_A^I\|$ for producer A and $TI_I^{pooled}(x_B, y_B) = \|BC_B^I\|$ for producer B, assuming a common frontier C. Under this assumption, producer A is less efficient than producer B. However, when each producer operates on their individual frontier, the measures become $TI_I^{own}(x_A, y_A) = \|AA^I\|$ and $TI_I^{own}(x_B, y_B) = \|BB^I\|$. Consequently, the efficiency ranking of these two producers is reversed. In this situation, producer B is less efficient than producer A because $\|BB^I\| > \|AA^I\|$.

There are several approaches to addressing heterogeneous technologies. One approach involves introducing a producer-specific intercept into the model, as demonstrated by Greene (2005a, 2005b). Cornwell et al. (1990) and Swamy and Tavlas (1995) assumed that both the intercept and the slope parameters are random. Akhavein et al. (1997), Tsionas (2002), and Feng et al. (2018) proposed a stochastic frontier model with random coefficients. Alternatively, varying coefficient models, in which the coefficients are expressed as functions of other variables, can be employed, as shown by Hastie and Tibshirani (1993) and Tran (2014).

Another approach involves categorizing producers into groups based on factors such as size, ownership, organizational structure, or geographic regions, and then estimating a model for each group. This method is illustrated by Mester (1996) and Altunbas et al. (2001), among others. However, a limitation of this approach is that the model for each group is estimated independently, without incorporating information from producers in other groups, as noted by Greene (1993, 2004b), Orea and Kumbhakar (2004), and Parmeter and Kumbhakar (2014). Alternatively,

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

threshold models categorize producers into technology groups based on the value of the threshold variable. The model for each group is estimated using information provided by producers in other groups. For examples of a single threshold model, see Hansen (1999, 2000) and Yélou et al. (2010); for multiple threshold models, refer to Almanidis (2013) and Tsionas et al. (2019). However, Almanidis (2013) noted that the joint estimation of the threshold parameters requires a grid search over an enormous number of points, which increases with the number of break points. The solution is to use sequential estimation of the threshold parameters. This method, however, yields asymptotically efficient estimates only for the last threshold parameter in the process. Bai (1997) introduced a refinement for estimating threshold parameters. This method involves re-estimating the threshold parameters in reverse order while keeping the estimates of the previous thresholds constant. The refined estimator has been demonstrated to be asymptotically efficient.

5.2.2 Heterogeneity in the inefficiency term

Ignoring the presence of heterogeneity in the inefficiency term can result in inaccurate measures of inefficiency, as heterogeneity not accounted for by producer-specific characteristics is incorrectly attributed to inefficiency. This heterogeneity can be addressed by incorporating producer-specific characteristics into the mean, variance, or both parameters of the inefficiency distribution. For more details on these models, refer to Section 3.

To summarize, exogenous factors that influence a producer's output but are not considered inefficiencies, because they are beyond the producer's control, are intended to capture technological differences and diversity. These factors should be specified in the production frontier itself. When exogenous factors that can be managed by the producer are more related to inefficiency, inefficiency heterogeneity directly impacts inefficiency and is often included in the location or scale parameters of the inefficiency distribution.

5.3 Heteroscedasticity issue

Several inefficiency models are based on the assumption that the random errors v and the inefficiency term u are homoscedastic, meaning both σ_u^2 and σ_v^2 remain constant. However, this may not be the case in practice, as they can be heteroscedastic. Heteroscedasticity refers to models in which σ_u^2 and σ_v^2 are not constant but are instead functions of explanatory variables that reflect producer-specific characteristics. Kumbhakar and Lovell (2000) and Wang and Schmidt (2002) concluded that ignoring the heteroscedasticity of v results in consistent estimates of the parameters of the production technology but leads to biased estimates of the intercept and inefficiency. In contrast, ignoring the heteroscedasticity of u causes biased estimates of both the parameters of the production technology and the estimates of inefficiency. To address heteroscedasticity, the scale parameter of the distribution of the random error and inefficiency can be modeled as functions of explanatory variables that reflect characteristics specific to each producer. For more details on these models, refer to Section 3.

5.4 Endogeneity issue

A potential issue in estimating inefficiency using distance functions is that inputs and outputs may be endogenous. This means they could be correlated with the random errors, inefficiency, or both, leading to biased and inconsistent estimates of the parameters of the production technology and the associated measures of inefficiency. For further discussion, refer to Atkinson and Primont (2002), Atkinson et al. (2003), and O'Donnell (2014).

There are two approaches to addressing this issue: one involves the use of instruments for inputs and outputs, and the other employs a systems approach. The use of instruments involves

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

selecting instrumental variables that are uncorrelated with the composed error term, estimating the stochastic frontier model with exogenous and endogenous variables, and utilizing the reduced form equation for the endogenous variables, which includes the exogenous variables and the instruments. Tran and Tsionas (2013) proposed a simple GMM procedure for estimating stochastic frontier models in the presence of endogenous variables. Assaf et al. (2013) compared the use of instruments with GMM and Bayesian estimation and found that the Bayesian estimates are more precise compared to GMM. Tran and Tsionas (2015) considered an alternative procedure that does not involve the use of instruments and is based on the copula function to directly model and capture the dependency between the endogenous variables and the composed error term.

Alternatively, the endogeneity issue can be addressed by employing a system approach. To meet the rank condition for identifying the system, it is necessary to include a total number of potentially endogenous variables as independent equations within the system, including the production technology. The selection of the system should be based on the behavioral assumptions of producers, duality theory, and the endogeneity of inputs and outputs. If only inputs (outputs) are endogenous, choosing the first-order conditions of cost minimization (revenue maximization), along with the IDF (ODF) or DIDF (DODF) may be preferable. Coelli (2000) showed that OLS provides consistent estimates of an IDF (ODF) under the assumption of cost-minimizing (revenue-maximizing) behavior when estimating distance functions in a system of equations, suggesting that instrumental variables may not be necessary. Refer to Tsionas et al. (2015) as an example of a system based on the IDF and the first-order conditions for cost minimization and Esheba and Serletis (2023) for an example of a system based on the DIDF (DODF) and the first-order conditions for cost minimization (revenue maximization). However, if both inputs and outputs are endogenous, it may be preferable to select the first-order conditions of profit maximization in conjunction with the HDF or DTDF. Atkinson and Tsionas (2016) and Esheba and Serletis (2023) provided examples of systems based on the DTDF and the assumption of profit-maximizing behavior. Tsionas et al. (2022) estimated a stochastic ray production frontier along with additional equations derived from profit maximization to address the issue of endogeneity. However, Malikov et al. (2016) considered the DTDF and the first-order conditions for cost minimization, without addressing the endogeneity of outputs. Feng et al. (2018) utilized the DODF and the first-order conditions for profit maximization, considering inputs as fixed in the DODF and endogenous in profit maximization.

The systems approach is not only considered a method to address the endogeneity issue but also offers several advantages. Berndt and Christensen (1973) argued that using the systems approach overcomes the multicollinearity issue that a single equation may suffer due to the large number of parameters that need to be estimated. When evaluating different functional forms using both a single equation and a system of equations, Guilkey et al. (1983) found that the functional form considered in the system of equations outperforms the single equation in terms of bias. Furthermore, the system approach incorporates a significant amount of information through the first-order conditions, resulting in more meaningful outcomes.

6. Conclusion

Efficiency is a crucial factor in productivity growth and the optimal allocation of resources in the economy; therefore, measuring inefficiency is particularly important. This paper provides a comprehensive review of the latest developments in distance functions and the measurement of inefficiency within the stochastic frontier framework. It examines the radial measure of inefficiency, as defined by standard distance functions; the hyperbolic measure, as provided by the hyperbolic distance function; and the directional measure, as defined by directional distance functions. The radial measure can result in high inefficiency measures even when the observed

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

input-output vector is very close to the frontier. However, implementing the hyperbolic measure can be complex due to the non-linear optimization involved. The directional measure is technology-oriented and simultaneously contracts inputs and expands outputs using either an exogenous or an endogenous directional vector to reach the efficient frontier. Additionally, the paper discusses the development of modeling inefficiency concerning its temporal behavior, classification, and determinants.

To ensure the use of appropriate estimation techniques, recent advancements in the most common estimation techniques are reviewed. This paper also addresses the importance of maintaining the theoretical regularity applied by neoclassical microeconomic theory when it is violated, as well as the econometric regularity when variables are non-stationary. If regularity conditions are not attained, the model can be estimated subject to imposed regularity conditions, which may require the use of Bayesian estimation. If stationarity is not achieved, cointegration techniques can be utilized to address the non-stationarity of the residuals. However, with nonlinear non-stationary inefficiency models, further research is necessary to modify linear model cointegration techniques and to develop existing nonlinear cointegration techniques.

Regarding estimation issues, inefficiency estimates can be distorted by an inaccurate choice of functional form for the production technology, ignoring the possibility of heterogeneity and heteroskedasticity, and suffering from the endogeneity problem. It is crucial for future applications to estimate inefficiency and address these issues using one of the various procedures discussed in the paper. The paper outlines several criteria for selecting a specific functional form for the production technology, based on theoretical properties such as the shape of the isoquants, separability, flexibility, and regular regions, as well as application properties like homogeneity and translation properties. Additionally, this paper addresses empirical techniques that can be used to assess the ability of different functional forms to approximate the unknown underlying function.

The selection of an appropriate functional form is insufficient without accommodating heterogeneous technologies that may exist among producers or heterogeneity in the inefficiency term. Ignoring heterogeneity can lead to incorrect conclusions regarding inefficiency measures because heterogeneity not captured by producer-specific characteristics is wrongly attributed to inefficiency. This paper addresses the importance of accommodating heterogeneity and discusses different approaches to account for both heterogeneous technologies and heterogeneity in the inefficiency term while estimating inefficiency. In general, exogenous factors that influence a producer's output but are not considered inefficiencies, because they are beyond the producer's control, are intended to capture technological differences and diversity. These factors should be specified in the production frontier itself. When exogenous factors that can be managed by the producer are more related to inefficiency, inefficiency heterogeneity directly impacts inefficiency and is often included in the location or scale parameters of the inefficiency distribution. Including producer-specific characteristics in the scale parameter of the inefficiency distribution also accounts for heteroscedasticity.

Another potential issue in estimating inefficiency using distance functions is that inputs and outputs may be endogenous, which may lead to biased and inconsistent estimates of the parameters of the production technology and the associated measures of inefficiency. This paper discusses various approaches to addressing this issue and identifies potentially productive areas for future research.

Notes

1 Performing the within transformation on Greene (2005a) true fixed-effects model yields $\widetilde{Y}_{it} = \widetilde{X}_{it}\beta + \widetilde{v}_{it} - \widetilde{u}_{it}$ where $\widetilde{Y}_{it} = y_{it} - \overline{y}_i$ are the deviations from the producer means, $\overline{y}_i = \sum_t y_{it}/T$. Similarly for $\widetilde{X}_{it}$,$\widetilde{v}_{it}$,and $\widetilde{u}_{it}$. The transformation from Y_{it} to $\widetilde{Y}_{it}$ is called the within transformation. Note that this transformation removes time-invariant heterogeneity α_i since $\widetilde{\alpha}_i = 0$. See Hsiao (2003) for a detailed discussion regarding the advantages of using within transformation.

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

2 Azzalini (1985) defined a continuous random variable ε to have a skew-normal distribution if it has density function $f(\varepsilon) = 2\phi(\varepsilon)\Phi(a\varepsilon)$, where a is a fixed arbitrary number. The distribution is right skewed if $a > 0$ and is left skewed if $a < 0$.
3 Proper specification testing can be undertaken to check the sign of the skewness of the OLS residuals. See, for example, Kuosmanen and Fosgerau (2009).
4 See Gonzalez-Farias, et al. (2004), and Arellano-Valle and Azzalini (2006) for the probability density function of the skew-normal distribution.
5 See Genz and Bretz (2009) for a detailed review on computation methods of multi-normal integrals.
6 See, for example, Gelfand and Smith (1990), Casella and George (1992), Smith and Roberts (1993), Roberts and Smith (1994), Koop (1994), McCulloch and Rossi (1994), Dorfman (1997), and Geweke (1999) for further details on Gibbs sampling method.
7 For example, imposing both monotonicity and curvature conditions globally on a translog functional form transforms it into the Cobb-Douglas functional form.
8 See Stock (1994) and Watson (1994) for a review of the econometric issues associated with non-stationary variables.
9 Series that can be made stationary by taking the first difference, represented as $[\triangle Y_t = Y_t - Y_{t-1}]$, are referred to as integrated of order one, denoted as $I(1)$. Stationary series are considered integrated of order zero, denoted as $I(0)$. Generally, the order of integration of a series is the minimum number of times it must be differenced to achieve stationary.
10 When Y_t and X_t are non-stationary $I(1)$ variables, their difference, or any linear combination of them, is also $I(1)$. In this case, Y_t and X_t are said to be cointegrated.
11 Another estimation technique that can be used for cointegrated panels with higher-order integrated systems is the Dynamic Ordinary Least Squares proposed by Stock and Watson (1993).
12 For different definitions of the flexibility property, see, for example, Diewert (1971), Gallant (1981), and Barnett (1983). Diewert (1971) formalized the notion of flexibility in functional forms by defining a second-order approximation to an arbitrary function. Gallant (1981) proposed the sobolev norm as a measure of global flexibility. See, for example, Griffin et al. (1987) for a comprehensive review of the flexibility property.
13 Table 1 presents the properties of alternative distance functions.

References

Adenuga, A., J. Davis, G. Hutchinson, T. Donnellan and M. Patton (2019) Environmental efficiency and pollution costs of nitrogen surplus in dairy farms: a parametric hyperbolic technology distance function approach. *Environmental and Resource Economics* 74(3), 1273–1298.

Ahn, S.C. and R. Sickles (2000) Estimation of long-run inefficiency levels: a dynamic frontier approach. *Econometric Reviews* 19(4), 461–492.

Aigner, D.J., C.A. Lovell and P. Schmidt (1977) Formulation and estimation of Stochastic Frontier production functions. *Journal of Econometrics* 6(1), 21–37.

Akhavein, J.D., P.A. Swamy and S.B. Taubman (1997) A general method of deriving the efficiencies of banks from a profit function. *Journal of Productivity Analysis* 8(1), 71–93.

Almanidis, P. (2013) Accounting for heterogeneous technologies in the banking industry: a time-varying stochastic frontier model with threshold effects. *Journal of Productivity Analysis* 39(2), 191–205.

Almanidis, P., M. Karakaplan and L. Kutlu (2019) A dynamic Stochastic Frontier model with threshold effects: U.S. bank size and efficiency. *Journal of Productivity Analysis* 52(1–3), 69–84.

Almanidis, P., J. Qian and R. Sickles (2014) Stochastic frontier models with bounded inefficiency. In Almanidis, P., J. Qian and R. Sickles. (eds.), *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*, pp. 47–81. New York: Springer New York.

Altunbas, Y., L. Evans and P. Molyneux (2001) Bank ownership and efficiency. *Journal of Money, Credit and Banking* 33(4), 926–954.

Alvarez, I., J. Niemi and M. Simpson (2014). Bayesian Inference for a covariance matrix. In Alvarez, I., J. Niemi and M. Simpson. (eds.), Proceedings of the 26th Annual Conference on Applied Statistics in Agriculture, pp. 71–82. New Prairie Press.

Arellano-Valle, R.B. and A. Azzalini (2006) On the unification of families of skew-normal distribution. *Scandinavian Journal of Statistics* 33(3), 561–574.

Assaf, A.G., R. Matousek and E.G. Tsionas (2013) Turkish Bank efficiency: Bayesian estimation with undesirable outputs. *Journal of Banking and Finance* 37(2), 506–517.

Atkinson, S., C. Cornwell and O. Honerkamp (2003) Measuring and decomposing productivity change: stochastic distance function estimation versus data envelopment analysis. *Journal of Business and Economic Statistics* 21(2), 284–294.

Atkinson, S., R. Färe and D. Primont (2003) Stochastic estimation of firm inefficiency using distance functions. *Southern Economic Journal* 69(3), 596–611.

Atkinson, S.E. and J.H. Dorfman (2005) Bayesian measurement of productivity and efficiency in the presence of undesirable outputs: crediting electric utilities for reducing air pollution. *Journal of Econometric* 126(2), 445–468.

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

Atkinson, S.E. and D. Primont (2002) Stochastic estimation of firm technology, inefficiency, and productivity growth using shadow cost and distance functions. *Journal of Econometric* 108(2), 203–225.

Atkinson, S.E. and E.G. Tsionas (2016) Directional distance functions: optimal endogenous directions. *Journal of Econometric* 190(2), 301–314.

Attfield, C.L.F. (1997) Estimating a cointegrating demand system. *European Economic Review* 41(1), 61–73.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12(2), 171–178.

Baccouche, R. and M. Kouki (2003) Stochastic production frontier and technical inefficiency: a sensitivity analysis. *Econometric Reviews* 22(1), 79–91.

Bai, J. (1997) Estimating multiple breaks one at a time. *Economic Theory* 13(3), 315–352.

Barnett, W.A. (1977) Pollak and Wachter on the household production function approach. *Journal of Political Economy* 85(5), 1073–1082.

Barnett, W.A. (1983) Definitions of 'second order approximation' and of 'flexible functional form. *Economics Letters* 12(1), 31–35.

Barnett, W.A. (2002) Tastes and technology: curvature is not sufficient for regularity. *Journal of Econometrics* 108(1), 199–202.

Barnett, W.A. and Y.W. Lee (1985) The global properties of the minflex Laurent, generalized Leontief, and translog flexible functional forms. *Econometrica* 53(6), 1421–1437.

Barnett, W.A., Y.W. Lee and M.D. Wolfe (1985) The three-dimensional global properties of the minflex Laurent, generalized Leontief, and translog flexible functional forms. *Journal of Econometrics* 30(1-2), 3–31.

Barnett, W.A., Y.W. Lee and M.D. Wolfe (1987) The global properties of the two minflex Laurent flexible functional forms. *Journal of Econometrics* 36(3), 281–298.

Barnett, W.A. and M. Pasupathy (2003) Regularity of the generalized quadratic production model: a counterexample. *Econometric Reviews* 22(2), 135–154.

Battese, G.E. (1992) Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics. *Agricultural Economics* 7(3–4), 185–208.

Battese, G.E. and T.J. Coelli (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics* 38(3), 387–399.

Battese, G.E. and T.J. Coelli (1992) Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis* 3(1–2), 153–169.

Battese, G.E. and T.J. Coelli (1995) A model for technical inefficiency effects in a Stochastic Frontier production function for panel data. *Empirical Economics* 20(2), 325–332.

Battese, G.E., T.J. Coelli and T. Colby (1989) Estimation of Frontier production functions and the efficiencies of Indian farms using panel data from ICRISTAT's village level studies. *Journal of Quantitative Economics* 5, 327–348.

Battese, G.E. and G.S. Cora (1977) Estimation of a production frontier model: with application to the Pastoral zone of Eastern Australia. *Australian Journal of Agricultural Economics* 21(3), 169–179.

Battese, G.E. and G.A. Tessema (1993) Estimation of Stochastic Frontier production functions with time-varying parameters and technical efficiencies using panel data from Indian villages. *Agricultural Economics* 9(4), 313–333.

Bauer, P. (1990) Recent developments in the econometric estimation of Frontiers. *Journal of Econometrics* 46(1–2), 39–56.

Beckers, D.E. and C.J. Hammond (1987) A tractable likelihood function for the normal-Gamma stochastic frontier model. *Economics Letters* 24(1), 33–38.

Bera, A. K. and S.C. Sharma (1999) Estimating production uncertainty in Stochastic Frontier production function models. *Journal of Productivity Analysis* 12(3), 187–210.

Berger, A. and L. Mester (1997) Inside the black box: what explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance* 21(7), 895–947.

Berger, A.N. and R. DeYoung (1997) Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance* 21(6), 849–870.

Berndt, E.R. and L.R. Christensen (1973) The translog function and the substitution of equipment, structures and labor in U.S. manufacturing 1929–68. *Journal of Econometrics* 1(1), 81–113.

Bos, J. and H. Schmiedel (2007) Is there a single frontier in a single European banking market? *Journal of Banking and Finance* 31(7), 2081–2102.

Bos, J.W.B., C. Economidou and M. Koetter (2010) Technology clubs, R&D and growth patterns: evidence from EU manufacturing. *European Economic Review* 54(1), 60–79.

Bos, J.W.B., C. Economidou, M. Koetter and JW. Kolari (2010) Do all countries grow alike? *Journal of Development Economics* 91(1), 113–127.

Brown, J.A. and C. Glennon (2000) Cost structures of banks grouped by strategic conduct. *Applied Economics* 32(12), 1591–1605.

Butler, J. and R. Moffitt (1982) A computationally efficient quadrature procedure for the one factor multinomial probit model. *Econometrica* 50(3), 761–764.

Caiazza, S., A.F. Pozzolo and G. Trovato (2016) Bank efficiency measures, M&A decision and heterogeneity. *Journal of Productivity Analysis* 46(1), 25–41.

Cartinhour, J. (1990) One-dimensional marginal density functions of a truncated multivariate normal density function. *Journal of Communications in Statistics—Theory and Methods* 19(1), 197–203.

Casella, G. and E. George (1992) Explaining the Gibbs sampler. *The American Statistician* 46(3), 167–174.

Casu, B. and P. Molyneux (2003) A comparative study of efficiency in european banking. *Applied Economics* 35(17), 1865–1876.

Caudill, S.B. and J.M. Ford (1993) Biases in frontier estimation due to heteroskedasticity. *Economics Letters* 41(1), 17–20.

Caudill, S.B., J.M. Ford and D.M. Gropper (1995) Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity. *Journal of Business and Economic Statistics* 13(1), 105–111.

Caves, D.W. and L.R. Christensen (1980) Global properties of flexible functional forms. *American Economic Review* 70, 422–432.

Chaffai, M. (2020) Hyperbolic distance function, technical efficiency and stability to shocks: a comparison between Islamic Banks and conventional banks in MENA region. *Global Finance Journal* 46, 100485.

Chambers, R., R. Färe, S. Grosskopf and M. Vardanyan (2013) Generalized quadratic revenue functions. *Journal of Econometrics* 173(1), 11–21.

Chambers, R.G. (1998) Input and Output Indicators. In Chambers, R.G. (ed.), *Index Numbers: Essays in Honour of Sten Malmquist*, pp. 241–271. Dordrecht: Springer.

Chambers, R.G., Y. Chung and R. Färe (1996) Benefit and distance functions. *Journal of Economic Theory* 70(2), 407–419.

Chambers, R.G., Y. Chung and R. Färe (1998) Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications* 98(2), 351–364.

Chang, Y., J. Park and P. Phillips (2001) Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal* 4(1), 1–36.

Chen, Y.-Y., P. Schmidt and H.-J. Wang (2014) Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* 181(2), 65–76.

Christensen, L.R., D.W. Jorgenson and L. J. Lau (1973) Transcendental logarithmic production frontiers. *The Review of Economics and Statistics* 55(1), 28–45.

Coelli, T.J. (2000) On the econometric estimation of the distance function representation of a production technology. *CORE Discussion Paper No.2000/42*. Center for Operations Research and Econometrics, Université catholique de Louvain. http:hdl.handle.net/2078.1/4131

Colombi, R., S. Kumbhakar, G.M. Martini and G. Vittadini (2014) Closed-skew Normality in Stochastic frontiers with individual effects and long/Short run efficiency. *Journal of Productivity Analysis* 42(2), 123–136.

Cornwell, C., P. Schmidt and R. Sickles (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics* 46(1–2), 185–200.

Cuesta, R. and L. Orea (2002) Mergers and technical efficiency in Spanish savings banks: a stochastic distance function approach. *Journal of Banking and Finance* 26(12), 2231–2247.

Cuesta, R.A. (2000) A production model with firm-specific temporal variation in technical inefficiency: with application to Spanish dairy farms. *Journal of Productivity Analysis* 13(2), 139–158.

Cuesta, R.A., C.A.K. Lovell and J. L. Zofio (2009) Environmental efficiency measurement with translog distance functions: a parametric approach. *Ecological Economics* 68(8–9), 2232–2242.

Cuesta, R.A. and J.L. Zofio (2005) Hyperbolic efficiency and parametric distance functions: with application to Spanish savings banks. *Journal of Productivity Analysis* 24(1), 31–48.

Daraio, C. and L. Simar (2014) Directional distances and their robust versions: computational and testing issues. *European Journal of Operational Research* 237(1), 358–369.

Darku, A.B., S. Malla and K.C. Tran (2013) *Historical Review of Agricultural Efficiency Studies*, CAIRN Research Network.

Debreu, G. (1951) The coefficient of resource utilization. *Econometrica* 19(3), 273–292.

Deprins, D. and L. Simar (1989) Estimating technical inefficiencies with correction for environmental conditions. *Annals of Public and Cooperative Economics* 60(1), 81–102.

Dervaux, B., G. Ferrier, H. Leleu and V. Valdmanis (2004) Comparing French and US hospital technologies: a directional input distance function approach. *Applied Economics* 36(10), 1065–1081.

Devitt, N., M. Zieba and D. Dineen (2024) Technical efficiency in Irish public hospitals: a multi-output distance function SFA approach. *International Journal of Computational Economics and Econometrics* 14(2), 122–150.

Dickey, D.A. and W.A. Fuller (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49(4), 1057–1072.

Diewert, W.E. (1971) An application of the Shephard duality theorem: a generalized Leontief production function. *Journal of Political Economy* 79(3), 481–507.

Diewert, W.E. and K.J. Fox (2008) On the estimation of returns to scale, technical progress and monopolistic markups. *Journal of Econometrics* 145(1–2), 174–193.

Diewert, W.E. and T.J. Wales (1987) Flexible functional forms and global curvature conditions. *Econometrica* 55(1), 43–68.

Dorfman, J.H. (1995) A numerical Bayesian test for cointegration of AR processes. *Journal of Econometrics* 66(1-2), 289–324.

Dorfman, J.H. (1997) *Bayesian Economics Through Numerical Methods: A Guide to Econometrics and Decision-Making with Prior Information*, Springer Science & Business Media.

Engle, R.F. and C.W.J. Granger (1987) Cointegration and error correction: representation, estimation and testing. *Econometrica* 55(2), 251–276.

Esheba, M. (2018) Essays on Distance Functions and Inefficiency Measurement, [Doctoral dissertation, University of Calgary, Calgary, Canada]. Retrieved from, doi: 10.11575/PRISM/32405, https://prism.ucalgary.ca

Esheba, M. and A. Serletis (2023) Interactive effects between input and output technical inefficiencies. *Macroeconomic Dynamics* 27(4), 1138–1165.

Fang, X. and F. Yang (2014) Assessing chinese commercial bank technical efficiency with a parametric hyperbolic distance function. *American Journal of Operations Research* 4(03), 124–131.

Färe, R. and S. Grosskopf (1994) Theory and calculation of productivity indexes. In Eichhorn, W. (ed.), *Models and Measurement of Welfare and Inequality*, pp. 921–940. Berlin: Springer. https://doi.org/10.1007/978-3-642-79037-9_49.

Färe, R. and S. Grosskopf (2000) Notes on some inequalities in economics. *Economic Theory* 15(1), 227–233.

Färe, R. and S. Grosskopf (2004) *New directions: Efficiency and Productivity*. Boston: Kluwer Academic Publishers.

Färe, R., S. Grosskopf and C.A.K. Lovell (1985) *The Measurement of Efficiency of Production*. Boston: Kluwer Academic Publishers.

Färe, R., S. Grosskopf and C.A.K. Lovell (1994) *Production Frontiers*. New York: Cambridge University Press.

Färe, R., S. Grosskopf, C.A.K. Lovell and S. Yaisawarng (1993) Derivation of shadow prices for undesirable outputs: a distance function approach. *The Review of Economics and Statistics* 75(2), 374–380.

Färe, R., S. Grosskopf, CA.K. Lovell and C. Pasurka (1989) Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. *The Review of Economics and Statistics* 71(1), 90–98.

Färe, R., S. Grosskopf, D. Noh and W. Weber (2005) Characteristics of a polluting technology: theory and practice. *Journal of Econometrics* 126(2), 469–492.

Färe, R., S. Grosskopf and W. Weber (2004) The effect of risk-based capital requirements on profit efficiency in banking. *Applied Economics* 36(15), 1731–1743.

Färe, R., S. Grosskopf and W. Weber (2006) Shadow prices and pollution costs in U.S. agriculture. *Ecological Economics* 56(1), 89–103.

Färe, R., S. Grosskopf and G. Whittaker (2013) Directional output distance functions: endogenous directions based on exogenous normalization constraints. *Journal of Productivity Analysis* 40(3), 267–269.

Färe, R., S. Grosskopf and O. Zaim (2002) Hyperbolic efficiency and return to the dollar. *European Journal of Operational Research* 136(3), 671–679.

Färe, R., C. Martins-Filho and M. Vardanyan (2010) On functional form representation of multi-output production technologies. *Journal of Productivity Analysis* 33(2), 81–96.

Färe, R. and D. Primont (1995) *Multi-output Production and Duality: Theory and Applications*. Boston: Kluwer Academic Publishers.

Färe, R. and M. Vardanyan (2016) A note on parameterizing input distance functions: does the choice of a functional form matter? *Journal of Productivity Analysis* 45(2), 121–130.

Feng, G. and A. Serletis (2008) Productivity trends in the US manufacturing: evidence from the NQ and AIM cost functions. *Journal of Econometrics* 142(1), 281–311.

Feng, G. and A. Serletis (2009) Efficiency and productivity of the US banking industry, 1998–2005: evidence from fourier cost functions satisfying global regularity conditions. *Journal of Applied Econometrics* 24(1), 105–138.

Feng, G., C. Wang and A. Serletis (2018) Shadow prices of CO_2 emissions at US electric utilities: a random-coefficient, random-directional-vector directional output distance function approach. *Empirical Economics* 54(1), 231–258.

Feng, Q. and W.C. Horrace (2012) Alternative technical efficiency measures: skew, bias and scale. *Journal of Applied Econometrics* 27(2), 253–268.

Feng, Q., W.C. Horrace and G.L. Wu (2015) Wrong skewness and finite sample correction. *Working Paper No. 154*, Center for Policy Research, Syracuse University. https://surface.syr.edu/cpr/389

Fernandez, C., G. Koop and M.F.J. Steel (2000) A Bayesian analysis of multiple output stochastic frontiers. *Journal of Econometrics* 98(1), 47–79.

Fernandez, C., G. Koop and M.F.J. Steel (2002) Multiple-output production with undesirable outputs: an application to nitrogen surplus in agriculture. *Journal of the American Statistical Association* 97(458), 432–442.

Fernandez, C., J. Osiewalski and M. Steel (1997) On the use of data in Stochastic Frontier models with improper priors. *Journal of Econometrics* 79(1), 169–193.

Filippini, M. and W. Greene (2016) Persistent and transient productive inefficiency: a maximum simulated likelihood approach. *Journal of Productivity Analysis* 45(2), 187–196.

Fried, H.O., C.A.K. Lovell and S.S. Schmidt (2008) Efficiency and Productivity. In Fried, H.O., C.A.K. Lovell and S.S. Schmidt. (eds.), *The Measurement of Productive Efficiency and Productivity Growth*, pp. 3–91. Oxford: Oxford University Press.

Gallant, A.R. (1981) On the bias in flexible functional forms and an essentially unbiased form. *Journal of Econometrics* 15(2), 211–245.

Gallant, A.R. and G.H. Golub (1984) Imposing curvature restrictions on flexible functional forms. *Journal of Econometrics* 26(3), 295–322.

Gelfand, A. and A. Smith (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85(410), 398–409.

Geman, S. and D. Geman (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(7), 609–628.

Genz, A. and F. Bretz (2009) *Computation of Multivariate Normal and t Probabilities*. Dordrecht: Springer.

Geweke, J. (1992) Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Geweke, J. (ed.), Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, Dedicated to the memory of Morris H. DeGroot, 1931-1989, pp. 169–194, Oxford: Oxford University Press.

Geweke, J. (1993) Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics* 8(S1), S19–S40.

Geweke, J. (1999) Using simulation methods for bayesian econometric models: inference, development and communication. *Econometric Reviews* 18(1), 1–73.

Giannakas, K., K.C. Tran and V. Tzouvelekas (2003a) On the choice of functional form in stochastic frontier modeling. *Empirical Economics* 28(1), 75–100.

Giannakas, K., K.C. Tran and V. Tzouvelekas (2003b) Predicting technical effciency in Stochastic production frontier models in the presence of misspecification: a Monte-Carlo analysis. *Applied Economics* 35(2), 153–161.

Gonzalez-Farias, G., J.A. Dominguez-Molina and A.K. Gupta (2004) The closed skew-normal distribution. In Gonzalez-Farias, G., J.A. Dominguez-Molina and A.K. Gupta. (eds.), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, pp. 25–42. New York: Chapman and Hall/CRC.

Greene, W.H. (1980a) Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics* 13(1), 27–56.

Greene, W.H. (1980b) On the estimation of a flexible frontier production model. *Journal of Econometrics* 13(1), 101–115.

Greene, W.H. (1990) A Gamma-distributed Stochastic Frontier model. *Journal of Econometrics* 46(1–2), 141–163.

Greene, W.H. (1993) The Econometric Approach to Efficiency Analysis. In Greene, W.H. (ed.), *The Measurement of Productive Efficiency: Techniques and Applications*, pp. 68–119. Oxford: Oxford University Press.

Greene, W.H. (2003) Simulated likelihood estimation of the normal-Gamma Stochastic Frontier function. *Journal of Productivity Analysis* 19(2/3), 179–190.

Greene, W.H. (2004a) Fixed and random effects in Stochastic Frontier models. *Journal of Productivity Analysis* 23(1), 7–32.

Greene, W.H. (2004b) Distinguishing between heterogeneity and inefficiency: Stochastic Frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics* 13(10), 959–980.

Greene, W.H. (2005a) Fixed and random effects in Stochastic Frontier models. *Journal of Productivity Analysis* 23(1), 7–32.

Greene, W.H. (2005b) Reconsidering heterogeneity in panel data estimators of the Stochastic Frontier model. *Journal of Econometrics* 126(2), 269–303.

Griffin, R.C., J.M. Montgomery and M.E. Rister (1987) Selecting functional form in production function analysis. *Western Journal of Agricultural Economics* 12(2), 216–227.

Guarda, P., A. Rouabah and M. Vardanyan (2013) Identifying bank outputs and inputs with a directional technology distance function. *Journal of Productivity Analysis* 40(2), 185–195.

Guilkey, D.K. and C.A.K. Lovell (1980) On the flexibility of the translog approximation. *International Economic Review* 21(1), 137–147.

Guilkey, D.K., C.A.K. Lovell and R. Sickles (1983) A comparison of the performance of three flexible functional forms. *International Economic Review* 24(3), 591–616.

Hadri, K. (1999) Estimation of a doubly heteroskedastic stochastic frontier cost function. *Journal of Business and Economic Statistics* 17(3), 359–363.

Hadri, K., C. Guermat and J. Whittaker (2003) Estimation of technical inefficiency effects using panel data and doubly heteroskedastic stochastic production frontiers. *Empirical Economics* 28(1), 203–222.

Hafner, C.M., H. Manner and L. Simar (2018) The "wrong skewness" problem in Stochastic Frontier models: a new approach. *Econometric Reviews* 37(4), 380–400.

Hampf, B. and J.J. Kruger (2015) Optimal directions for directional distance functions: an exploration of potential reductions of greenhouse gases. *American Journal of Agricultural Economics* 97(3), 920–938.

Han, C. and P. Phillips (2010) GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory* 26(1), 119–151.

Hansen, B.E. (1999) Threshold effects in non-dynamic panels: estimation, testing, and inference. *Journal of Econometrics* 93(2), 345–368.

Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* 68(3), 575–603.

Harris, R.D.F. and E. Tzavalis (1999) Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91(2), 201–226.

Hastie, T. and R. Tibshirani (1993) Varying-coefficient models. *Journal of the Royal Statistical Society* 55(4), 757–796.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrica* 57(1), 97–109.
Hattori, T. (2002) Relative performance of U.S. and Japanese electricity distribution: an application of Stochastic Frontier analysis. *Journal of Productivity Analysis* 18(3), 269–284.
Hausman, J.A. and W.E. Taylor (1981) Panel data and unobservable individual effects. *Econometrica* 49(6), 1377–1399.
Hjalmarsson, L., S.C. Kumbhakar and A. Heshmati (1996) DEA, DFA and SFA: a comparison. *Journal of Productivity Analysis* 7(2-3), 303–327.
Horrace, W.C. (2005) Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis* 94(2), 209–221.
Horrace, W.C. and C.F. Parmeter (2018) A Laplace Stochastic Frontier model. *Econometric Reviews* 37(3), 260–280.
Horrace, W.C. and P. Schmidt (1996) Confidence statements for efficiency estimates from Stochastic Frontier models. *Journal of Productivity Analysis* 7(2–3), 257–282.
Hsiao, C. (2003) *Analysis of Panel Data*. Cambridge: Cambridge University Press.
Huang, C.J. and J.T. Liu (1994) Estimation of a non-neutral Stochastic Frontier production function. *Journal of Productivity Analysis* 5(2), 171–180.
Huang, T.-H. and M.-T. Chung (2017) Do undesirables matter on the examination of banking efficiency using Stochastic directional distance functions. *The Quarterly Review of Economics and Finance* 65, 194–211.
Hudgins, L.B. and D. Primont (2007) Derivative properties of directional technology distance functions. In Hudgins, L.B. and D. Primont (eds.), *Aggregation, Efficiency, and Measurement*, pp. 31–43. New York: Springer New York.
Jondrow, J., C.A.K. Lovell, I.S. Materov and P. Schmidt (1982) On the estimation of technical inefficiency in the Stochastic Frontier production function model. *Journal of Econometrics* 19(2–3), 233–238.
Kim, M. and P. Schmidt (2008) Valid tests of whether technical inefficiency depends on firm characteristics. *Journal of Econometrics* 144(2), 409–427.
Kim, Y. and P. Schmidt (2000) A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in Stochastic Frontier models with panel data. *Journal of Productivity Analysis* 14(2), 91–118.
Koop, G. (1994) Recent progress in applied Bayesian econometrics. *Journal of Economic Surveys* 8(1), 1–34.
Koop, G., J. Osiewalski and M.F. Steel (1994) Bayesian efficiency analysis with a flexible form: the AIM cost function. *Journal of Business and Economic Statistics* 12(3), 339–346.
Koop, G., J. Osiewalski and M.F. Steel (1997) Bayesian efficiency analysis through individual effects: hospital cost frontiers. *Journal of Econometrics* 76(1–2), 77–105.
Koop, G., M.F. Steel and J. Osiewalski (1995) Posterior analysis of Stochastic Frontier models using Gibbs sampling. *Computational Statistics* 10, 353–373.
Koutsomanoli-Filippaki, A., D. Margaritis and C. Staikouras (2012) Profit efficiency in the European Union banking industry: a directional technology distance function approach. *Journal of Productivity Analysis* 37(3), 277–293.
Kumbhakar, S.C. (1987) The specification of technical and allocative inefficiency in stochastic production and profit frontiers. *Journal of Econometrics* 34(3), 335–348.
Kumbhakar, S.C. (1990) Production Frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics* 46(1–2), 201–211.
Kumbhakar, S.C., S. Ghosh and J. McGuckin (1991) A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. *Journal of Business and Economics Statistics* 9(3), 279–286.
Kumbhakar, S.C. and A. Heshmati (1995) Efficiency measurement in Swedish dairy farms: an application of rotating panel data, 1976–88. *American Journal of Agricultural Economics* 77(3), 660–674.
Kumbhakar, S.C., G. Lien and J.B. Hardaker (2014) Technical efficiency in competing panel data models: a study of Norwegian grain farming. *Journal of Productivity Analysis* 41(2), 321–337.
Kumbhakar, S.C. and C.A.K. Lovell (2000) *Stochastic Frontier analysis*. Cambridge: Cambridge University Press.
Kumbhakar, S.C., C.F. Parmeter and E.G. Tsionas (2013) A zero inefficiency Stochastic frontier model. *Journal of Econometrics* 172(1), 66–76.
Kumbhakar, S.C. and H.-J. Wang (2005) Estimation of growth convergence using a Stochastic production function approach. *Economics Letters* 88(3), 300–305.
Kuosmanen, T. and M. Fosgerau (2009) Neoclassical versus frontier production models? Testing for the skewness of regression residuals. *The Scandinavian Journal of Economics* 111(2), 351–367.
Lau, L.J. (1978a) Testing and imposing monotonicity, convexity, and quasi-convexity constraints. In Lau, L.J. (ed.), *Production Economics: A Dual Approach to Theory and Applications*, pp. 409–453. Amsterdam: North-Holland Publishing Co.
Lau, L.J. (1978b) Application of Profit Functions. In Lau, L.J. (ed.), *Production Economics: A Dual Approach to Theory and Applications*, pp. 133–216. Amsterdam: North-Holland Publishing Co.
Lau, L.J. (1986) Functional forms in econometric model building. In Lau, L.J. (ed.), *Handbook of econometrics*, vol. 3, pp. 1515–1566. Amsterdam: Elsevier Science.
Lee, S. and Y.H. Lee (2014) Stochastic Frontier models with threshold efficiency. *Journal of Productivity Analysis* 42(1), 45–54.

Lee, Y. and P. Schmidt (1993) A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency. In Lee, Y. and P. Schmidt. (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, pp. 237–255. Oxford: Oxford University Press.

Lee, YH. (1996) Tail truncated Stochastic Frontier models. *Journal of Economic Theory and Econometrics* 2, 137–152.

Letti, A.G., M. V. L. Bittencourt and L. E. Vila (2022) Stochastic vs. Deterministic frontier distance output function: evidence from Brazilian higher education institutions. *Journal of Productivity Analysis* 58(1), 55–74.

Lewbel, A. and S. Ng (2005) Demand systems with nonstationary prices. *Review of Economics and Statistics* 87(3), 479–494.

Luenberger, D. (1995) *Microeconomic Theory*. New York: McGraw Hill.

Maddala, G.S. and S. Wu (1999) A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics* 61(S1), 631–652.

Malikov, E., S.C. Kumbhakar and E.G. Tsionas (2016) A cost system approach to the stochastic directional technology distance function with undesirable outputs: the case of U.S. banks in 2001–2010. *Journal of Applied Econometrics* 31(7), 1407–1429.

McCulloch, R. and P.E. Rossi (1994) An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* 64(1–2), 207–240.

Meeusen, W. and J. Van den Broeck (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18(2), 435–444.

Mester, L. (1996) A study of bank efficiency taking into account risk-preferences. *Journal of Banking and Finance* 20(6), 1025–1045.

Mester, L. (1997) Measuring efficiency at U.S. banks: accounting for heterogeneity is important. *European Journal of Operational Research* 98(2), 230–242.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21(6), 1087–1092.

Mundlak, Y. (1961) Empirical production function free of management bias. *Journal of Farm Economics* 43(1), 44–56.

Nahm, D. and H.T. Vu (2013) Measuring scale efficiency from a parametric hyperbolic distance function. *Journal of Productivity Analysis* 39(1), 83–88.

Neyman, J. and E.L. Scott (1948) Consistent estimation from partially consistent observations. *Econometrica* 16(1), 1–32.

Ng, S. (1995) Testing for homogeneity in demand systems when the regressors are nonstationary. *Journal of Applied Econometrics* 10(2), 147–163.

O'Donnell, C. J. (2014) Econometric estimation of distance functions and associated measures of productivity and efficiency change. *Journal of Productivity Analysis* 41(2), 187–200.

O'Donnell, C. J. and T. J. Coelli (2005) A Bayesian approach to imposing curvature on distance functions. *Journal of Econometrics* 126(2), 493–523.

Orea, L. and S.C. Kumbhakar (2004) Efficiency measurement using a latent class Stochastic Frontier model. *Empirical Economics* 29(1), 169–183.

Osiewalski, J. and M. Steel (1998) Numerical tools for the Bayesian analysis of Stochastic Frontier models. *Journal of Productivity Analysis* 10(1), 103–117.

Park, J.Y. and S.B. Hahn (1999) Cointegrating regressions with time varying coefficients. *Econometric Theory* 15(5), 664–703.

Parmeter, C. and S.C. Kumbhakar (2014) Efficiency analysis: a primer on recent advances. *Foundations and Trends in Econometrics* 7(3-4), 191–385.

Pedroni, P. (2001) Fully Modified OLS for Heterogeneous Cointegrated Panels. In Pedroni, P. (ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels (Advances in Econometrics*, vol. 15), pp. 93–130, Emerald Publishing Limited.

Phillips, P. and H. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5), 1057–1111.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55(2), 277–301.

Phillips, P.C.B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63(5), 1023–1078.

Phillips, P.C.B. and L.P. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies* 57(1), 99–125.

Pitt, M.M. and L.-F. Lee (1981) The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics* 9(1), 43–64.

Reifschneider, D. and R. Stevenson (1991) Systematic departures from the Frontier: a framework for the analysis of firm inefficiency. *International Economic Review* 32(3), 715–723.

Rho, S. and P. Schmidt (2015) Are all firms inefficient? *Journal of Productivity Analysis* 43(3), 327–349.

Ritter, C. and L. Simar (1997) Pitfalls of normal-Gamma Stochastic Frontier models. *Journal of Productivity Analysis* 8(2), 167–182.

Roberts, G.O. and A.F.M. Smith (1994) Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms. *Stochastic Processes and their Applications* 49(2), 207–216.

Ryan, D.L. and T.J. Wales (1998) A simple method for imposing local curvature in some flexible consumer demand systems. *Journal of Business and Economic Statistics* 16(3), 331–338.

Ryan, D.L. and T.J. Wales (2000) Imposing local concavity in the translog and generalized Leontief cost functions. *Economics Letters* 67(3), 253–260.
Satchachai, P. and P. Schmidt (2010) Estimates of technical inefficiency in Stochastic Frontier models with panel data: generalized panel jackknife estimation. *Journal of Productivity Analysis* 34(2), 83–97.
Schmidt, P. and R.C. Sickles (1984) Production Frontiers and panel data. *Journal of Business and Economic Statistics* 2(4), 367–374.
Serletis, A. and A. Shahmoradi (2007) Flexible functional forms, curvature conditions, and the demand for assets. *Macroeconomic Dynamics* 11(4), 455–486.
Serra, T., A.O. Lansink and S.E. Stefanou (2011) Measurement of dynamic efficiency: a directional distance function parametric approach. *American Journal of Agricultural Economics* 93(3), 756–767.
Shephard, R. (1953) *Cost and production functions*. Princeton: Princeton University Press.
Shephard, R. (1970) *Theory of cost and production functions*. Princeton: Princeton University Press.
Sickles, R.C. (2005) Panel estimators and the identification of firm specific efficiency levels in parametric, semiparametric and nonparametric settings. *Journal of Econometrics* 126(2), 305–324.
Simar, L. and A. Vanhems (2012) Probabilistic characterization of directional distances and their robust versions. *Journal of Econometrics* 166(2), 342–354.
Smith, A.F.M. and C.O. Roberts (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)* 55(1), 3–23.
Srairi, S.A. (2010) Cost and profit efficiency of conventional and Islamic Banks in GCC countries. *Journal of Productivity analysis* 34(1), 45–62.
Stevenson, RE. (1980) Likelihood functions for generalized Stochastic Frontier estimation. *Journal of Econometrics* 13(1), 57–66.
Stock, J.H. (1994) Unit roots, structural breaks and trends. In Stock, J.H. (ed.), *Handbook of Econometrics*, vol. 4, pp. 2739–2841. Amsterdam: Elsevier.
Stock, J.H. and M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61(4), 783–820.
Sturm, J.-E. and B. Williams (2008) Characteristics determining the efficiency of foreign banks in Australia. *Journal of Banking and Finance* 32(11), 2346–2360.
Swamy, P.A.V. B. and G.S. Tavlas (1995) Random coefficient models: theory and applications. *Journal of Economic Surveys* 9(2), 165–196.
Terrell, D. (1996) Incorporating monotonicity and concavity conditions in flexible functional forms. *Journal of Applied Econometrics* 11(2), 179–194.
Thompson, G.D. (1988) Choice of flexible functional forms: review and appraisal. *Western Journal of Agricultural Economics* 13(2), 169–183.
Tran, K.C. (2014) Nonparametric estimation of functional-coefficient dynamic panel data model with fixed effects. *Economics Bulletin* 34(3), 1751–1761.
Tran, K.C. and E.G. Tsionas (2013) GMM estimation of Stochastic Frontier model with endogenous regressors. *Economics Letters* 118(1), 233–236.
Tran, K.C. and E.G. Tsionas (2015) Endogeneity in Stochastic Frontier models: Copula approach without external instruments. *Economics Letters* 133, 85–88.
Tran, K.C. and E.G. Tsionas (2016a) On the estimation of zero-inefficiency Stochastic Frontier models with endogenous regressors. *Economics Letters* 147, 19–22.
Tran, K.C. and E.G. Tsionas (2016b) Zero-inefficiency Stochastic Frontier model with varying mixing proportion: semiparametric approach. *European Journal of Operational and Research* 249(3), 1113–1123.
Tsionas, E. and D. Christopoulos (2001) Efficiency measurement with nonstationary variables: an application of panel cointegration techniques. *Economics Bulletin* 3(14), 1–7.
Tsionas, E.G. (2002) Stochastic Frontier models with random coefficients. *Journal of Applied Econometrics* 17(2), 127–147.
Tsionas, E.G. (2006) Inference in dynamic Stochastic Frontier models. *Journal of Applied Econometrics* 21(5), 669–676.
Tsionas, E.G., S.C. Kumbhakar and E. Malikov (2015) Estimation of input distance functions: a system approach. *American Journal of Agricultural Economics* 97(5), 1478–1493.
Tsionas, E.G., K.C. Tran and P. Michaelides (2019) Bayesian inference in threshold Stochastic Frontier models. *Empirical Economics* 56(2), 399–422.
Tsionas, G. and S. Kumbhakar (2014) Firm heterogeneity, persistent and transient technical inefficiency. *Journal of Applied Econometrics* 29(1), 110–132.
Tsionas, M., M. Izzeldin, A. Henningsen and E. Paravalos (2022) Addressing endogeneity when estimating stochastic ray production frontiers: a Bayesian approach. *Empirical Economics* 62(3), 1345–1363.
Tzeremes, N.G. (2015) Efficiency dynamics in Indian banking: a conditional directional distance approach. *European Journal of Operational Research* 240(3), 807–818.

https://doi.org/10.1017/S1365100525000148 Published online by Cambridge University Press

Van den Broeck, J., G. Koop, J. Osiewalski and M. Steel (1994) Stochastic Frontier models: a Bayesian perspective. *Journal of Econometrics* 61(2), 273–303.

Vardanyan, M., V. Valdmanis, H. Leleu and G. Ferrier (2022) Estimating technology characteristics of the U.S. hospital industry using directional distance functions with optimal directions. *Omega* 113, 102716.

Waldman, D. (1982) A stationary point for the Stochastic Frontier likelihood. *Journal of Econometrics* 18(2), 275–279.

Wales, T.J. (1977) On the flexibility of flexible functional forms. *Journal of Econometrics* 5(2), 183–193.

Wang, H.-J. (2002) Heteroskedasticity and non-monotonic efficiency effects of a Stochastic Frontier model. *Journal of Productivity Analysis* 18(3), 241–253.

Wang, H.-J. and C.W. Ho (2010) Estimating fixed-effect panel data Stochastic Frontier models by model transformation. *Journal of Econometrics* 157(2), 286–296.

Wang, H.-J. and P. Schmidt (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis* 18(2), 129–144.

Wang, W.S. and P. Schmidt (2009) On the distribution of estimated technical efficiency in Stochastic Frontier models. *Journal of Econometrics* 148(1), 36–45.

Watanabe, M. and K. Tanaka (2007) Efficiency analysis of Chinese industry: a directional distance function approach. *Energy Policy* 35(12), 6323–6331.

Watson, M.W. (1994) Vector Autoregressions and Cointegration. In Stock, J.H. (ed.), *Handbook of Econometrics*, vol. 4, pp. 2843–2915. Amsterdam: Elsevier.

Wheat, P., B. Greene and A. Smith (2014) Understanding prediction intervals for firm specific inefficiency scores from parametric Stochastic Frontier models. *Journal of Productivity Analysis* 42(1), 55–65.

Wikstrom, D. (2016) Modified fixed effects estimation of technical inefficiency. *Journal of Productivity Analysis* 46(1), 83–86.

Wolff, H. (2016) Imposing and testing for shape restrictions in flexible parametric models. *Econometric Reviews* 35(6), 1013–1039.

Wolff, H., T. Heckelei and R.C. Mittelhammer (2010) Imposing curvature and monotonicity on flexible functional forms: an efficient regional approach. *Computational Economics* 37(1), 309–339.

Yang, J., J. Cheng, R. Zou and Z. Geng (2021) Industrial SO_2 technical efficiency, reduction potential and technology heterogeneities of China's prefecture-level cities: a multi-hierarchy meta-frontier parametric approach. *Energy Economics* 104, 105626.

Yélou, C., B. Larue and K.C. Tran (2010) Threshold effects in panel data Stochastic Frontier models of dairy production in Canada. *Economic Modelling* 27(3), 641–647.

Zofio, J.L., J.T. Pastor and J. Aparicio (2013) The directional profit efficiency measure: on why profit inefficiency is either technical or allocative. *Journal of Productivity Analysis* 40(3), 257–266.

Zofio, J.L. and A.M. Prieto (2006) Return to dollar, generalized distance function and the Fisher productivity index. *Spanish Economic Review* 8(2), 113–138.

Cite this article: Esheba M and Serletis A (2025). "Distance functions and the analysis of inefficiency[†]." *Macroeconomic Dynamics* **29**(e95), 1–51. https://doi.org/10.1017/S1365100525000148