Macroeconomic Dynamics (2023), **27**, pp. 1138–1165
doi:10.1017/S1365100522000153

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Interactive effects between input and output technical inefficiencies†

Muna Esheba and Apostolos Serletis

Department of Economics, University of Calgary, Calgary, AB, Canada

Abstract

This paper derives a new set of results that provide corrective measures of overall technical inefficiency that either have been ignored or wrongly assumed in the literature. Using directional distance functions, we argue that overall technical inefficiency is not only a function of input and output technical inefficiencies as previous studies claim but also of the interaction between them. The derivation of the interactive effects between input and output technical inefficiencies (IEIOs) solves the arbitrary decomposition of overall technical inefficiency into input and output components. We also show that the IEIO depends on the choice of the directional vector and whether quantities and prices are taken into consideration. Using exogenous and endogenous directional vectors, we prove these results theoretically and empirically using the US commercial banking data set. Using Bayesian estimation with the monotonicity conditions imposed at each observation, we estimate input and output technical inefficiencies separately using directional input and output distance functions with the three commonly used directional vectors; the unit value, the observed input—output, and the optimal directional vectors. The overall technical inefficiency is estimated using systems of equations to incorporate the interactive effect equation and to address the endogeneity of inputs and outputs. Consistent with the theoretical results, we find significant evidence of the IEIO which has a negative effect on the overall technical inefficiency. This result is robust to alternative directional vectors and model specifications, suggesting that the adjustability of both inputs and outputs is required for the improvement of the efficiency of the US commercial banks.

Keywords: Production; Technical Inefficiency Measurement; Directional Distance Function; Interactive Inefficiency Effect; Bayesian Estimation; Banking

1. Introduction

The economic costs associated with bank failures and performance problems are high for the real economy due to their role of financing the economy. It often results in significant macroeconomic costs and negative externalities on other financial institutions, economic activities, and economic stability and growth – see, for example, Lang and Stulz (1992), Ashcraft (2005), Kang et al. (2015), and Githinji-Muriithi (2017). Several empirical studies find significant positive relationships between bank efficiency and economic growth (e.g., Lucchetti et al. (2001), Bui (2020), and Ege and Nur (2020)), the transmission of monetary policy (e.g., Havranek et al. (2016)), and financial stability. Assaf et al. (2019) explore the impact of bank efficiency on financial stability by examining how bank efficiency during normal times affects survival and profitability during subsequent financial crises. They find that the main factor affecting bank performance during crises is bank efficiency in the normal times prior to the crises. They also find that bank

†We would like to thank Rolf Färe and two referees for comments that greatly improved the paper. Address correspondence to: Apostolos Serletis, Department of Economics, University of Calgary, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada, Email: serletis@ucalgary.ca.

efficiency during normal times helps reduce bank failure probabilities and enhance profitability during subsequent financial crises. Thus, it is critical to understand the major determinants of bank inefficiency to have the basis to propose the appropriate public policies and regulations that reduce the occurrence of costly bank problems early.

The extant literature on the determinants of bank inefficiency generally employs either an input or an output-oriented measurement technique. However, adopting an input (output) oriented measurement technique ignores the opposite output (input) orientation, and this restriction may substantially bias the measures of bank inefficiency. An efficiency survey by Berger et al. (1993) suggests comparing these input and output approaches with a complete approach to investigate the relationship between input and output inefficiencies. However, few studies examine total technical inefficiency and decompose it into input and output components – see Berger et al. (1993), Akhavein et al. (1997), Barros et al. (2012), and Fujii et al. (2014). Even though these studies disaggregate and quantify the impact of input and output on inefficiency, the arbitrary decomposition of total technical inefficiency into input and output inefficiency components results in concluding that total technical inefficiency equals the sum of input and output technical inefficiencies and shows no interactive effects between input and output technical inefficiencies (IEIOs).

In contrast to previous studies that employ either an input- or an output-oriented measurement technique, this paper extends the literature by focusing on overall technical inefficiency that includes both input and output inefficiencies as well as the interactions between them. This is important because input and output inefficiencies measure different concepts and may affect future bank outcomes through different channels in the economy. Input inefficiency arises from employing the wrong level of inputs and measures the maximum contraction of a bank's input to that of a best-practice or most efficient bank producing the same output. Low input inefficiency (high input efficiency) may reflect superior managerial quality that produces favorable performance. Output inefficiency arises from producing at the wrong level of outputs and measures the maximum expansion of a bank's output to that of a best-practice or most efficient bank using the same input. Low output inefficiency (high output efficiency) may be associated with high charter values that result in favorable performance. The interactions between input and output inefficiencies captures output (input) implications of any errors in the input usage (output production). For instance, if a certain input (output) is viewed as being relatively overused (underproduced), it is likely that outputs (inputs) that are intensive in using (producing) that input (output) will be overproduced (underused) relative to other outputs (inputs). Thus, the interactive effect contains more information than could be captured in an input (output)-oriented efficiency study which excludes output (input) effects of input usage (output production) errors.

We follow the suggestion in Berger et al. (1993) and compare these input and output approaches with a complete approach using directional input, output, and technology distance functions. Using exogenous and endogenous directional vectors, we derive the IEIO theoretically and empirically using a sample of 148 US commercial banks over the period 2001–2015. A potential issue when estimating technical inefficiency empirically using directional distance functions is that inputs and outputs may be endogenous, leading to biased and inconsistent estimates of the parameters of the production technology and the associated measures of inefficiency—see, for example, Atkinson and Primont (2002). The literature considers two approaches to deal with this issue; one approach relies on using instrumental variable estimation and the other relies on employing a system of equations approach. This paper follows the latter approach. Furthermore, the directional vectors of these models are allowed to be endogenous and vary across banks to account for heterogeneity across banks. The obtained estimates of the directional vectors can be interpreted as being optimal directional vectors—see Malikov et al. (2016). All these models are estimated using Bayesian estimation with the monotonicity conditions imposed at each observation to produce inference that is consistent with neoclassical microeconomic theory.

The theoretical and empirical results show that overall technical inefficiency is not only a function of input and output technical inefficiencies as previous studies claim but also of the interaction between them. Both input and output technical inefficiencies have significant positive effects on the overall technical inefficiency. However, the IEIO has a significant negative effect on the overall technical inefficiency. This result is robust to alternative directional vectors and model specifications. Therefore, the IEIO directly results in a decrease in overall technical inefficiency. This suggests that the overuse of inputs creates input technical inefficiency and has an effect on reducing (improving) output technical inefficiency (efficiency) and therefore improving overall technical efficiency. Intuitively, the overuse of inputs whether physical inputs involving overuse of labor or overuse of financial inputs involving overpayment of interest may encourage banks to produce more loans to pay salaries for its employees and interest rates on deposits. Similarly, the loss of production of outputs creates output technical inefficiency and has an effect on reducing (improving) input technical inefficiency (efficiency) and therefore improving overall technical efficiency. Intuitively, the loss of revenue due to the loss of production of loans may encourage banks to reduce the number of labor used in the production process or lower the interest rates paid on deposits. That is, a loss of output creates a loss of revenue and has an effect on the use of inputs, while an overuse of input creates additional costs and has an effect on the production of outputs.

From a positive perspective, our result can help understanding the determinants of overall bank inefficiency. Namely, the IEIO, which is a new determinant of this paper compared to the previous studies, plays a role in decreasing overall bank inefficiency and as such should be taken into account next to input and output inefficiencies. That is, our result indicates that measures that increase the IEIO can promote bank efficiency which in turn foster a better macroeconomic performance. Policymakers need to find ways to transform any errors in the input usage (output production) into their most productive use (an efficient use of input).

In addition to these implications, our result is important for researchers since it helps make our argument that input and output approaches are not an appropriate metric for measuring inefficiency in the presence of the interactive effect. Instead, using models that incorporate both input and output inefficiencies is superior to the standard input or output approach, since it allows to credit input and output while simultaneously crediting the interactive effect. Our results also show that models that ignore bank heterogeneity and endogeneity issues tend to underestimate bank inefficiency measures. This suggests the importance of managing the heterogeneity and endogeneity issues to obtain unbiased and consistent estimates of the parameters of the production technology and the associated measures of inefficiency.

This paper contributes to the literature in many ways. To the best of our knowledge, it is the first in the literature that examines the relationships among input, output, and overall technical inefficiencies theoretically and empirically using the three directional distance functions with the three commonly used directional vectors. It derives a new set of results that provide corrective measures of overall technical inefficiency that either have been ignored or wrongly assumed in the literature. It argues that overall technical inefficiency is not only a function of input and output technical inefficiencies as previous studies claim but also of the interaction between them. It provides theoretical and empirical evidence of the IEIO which solves the arbitrary decomposition of overall technical inefficiency into input and output components.

The rest of the paper is organized as follows. The next section presents some theoretical background on directional measures of technical inefficiency using directional distance functions. It also derives the IEIO theoretically using the directional distance functions assuming exogenous and endogenous directional vectors. Section 3 presents an empirical application. Section 4 discusses the Bayesian procedure for estimating the empirical models. Section 5 defines the data used in this paper. In Section 6, the methodology is applied to a sample of the US commercial banks, and the results are reported. Section 7 summarizes and concludes the paper.

2. Theoretical foundations

To briefly review some of the literature on the directional measures of technical inefficiency using directional distance functions, consider a bank employing a vector of n inputs $x = (x_1, \cdots, x_n) \in \mathbb{R}^n_+$ 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, and let $P(x)$ be the feasible set of outputs y that can be produced from the input vector x. The production technology T for a bank is defined as the set of all feasible input–output vectors. Note that $(x, y) \in T \Leftrightarrow x \in L(y) \Leftrightarrow y \in P(x)$.

2.1. The Directional Distance Functions

The directional technology distance function (DTDF) allows for simultaneous contraction of inputs and expansion of outputs in terms of a direction vector $g = (g_x, g_y)$, where $g_x \in R^N_+$ and $g_y \in R^M_+$ such that it contracts inputs in the direction g_x and expands outputs in the direction g_y. In particular, the DTDF is defined as:

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

Efficient banks who produce on the frontier of T have $\vec{D}_T(x, y; g_x, g_y) = 0$. Inefficiency is indicated by $\vec{D}_T(x, y; g_x, g_y) > 0$, with higher values indicating greater inefficiency when banks operate beneath the frontier of T. A technology-oriented measure of technical inefficiency is defined as $TI_T = \vec{D}_T(x, y; g_x, g_y)$.

As noted by Chambers et al. (1998), the DTDF is nonnegative, nondecreasing in x, nonincreasing in y, and concave in (x, y). Moreover, it satisfies the following translation property:

$$\vec{D}_T(x - \alpha g_x, y + \alpha g_y; g_x, g_y) = \vec{D}_T(x, y; g_x, g_y) - \alpha, \tag{2}$$

where $\alpha \in R$ is an arbitrary scaling factor. By setting $g_y = 0$, the directional vector becomes $g = (g_x, 0)$ and allows only for input contraction holding outputs fixed—see Figure 1 and 2. In this case, equation (1) becomes the directional input distance function (DIDF), $\vec{D}_T(x, y; g_x, 0) = \vec{D}_I(y, x; g_x)$. The DIDF serves as an input-oriented measure of technical inefficiency; $TI_I = \vec{D}_I(y, x; g_x)$:

$$\vec{D}_I(y, x; g_x) = \max_{\theta_I} \{\theta_I : (x - \theta_I g_x) \in L(y)\} = \max_{\theta_I} \{\theta_I : (x - \theta_I g_x, y) \in T\}. \tag{3}$$

By setting $g_x = 0$, the directional vector becomes $g = (0, g_y)$ and allows only for output expansion holding inputs fixed—see Figures 1 and 2. In this case, equation (1) becomes the directional output distance function (DODF), $\vec{D}_T(x, y; 0, g_y) = \vec{D}_O(x, y; g_y)$. The DODF serves as an output-oriented measure of technical inefficiency; $TI_O = \vec{D}_O(x, y; g_y)$:

$$\vec{D}_O(x, y; g_y) = \max_{\theta_o} \{\theta_O : (y + \theta_O g_y) \in P(x)\} = \max_{\theta_o} \{\theta_O : (x, y + \theta_O g_y) \in T\}. \tag{4}$$

2.2. Duality Relationships

The directional distance functions are primal representations of the technology. The dual representations of the technology are given by the profit $(\pi(p, w))$, cost $(C(y, w))$, and revenue

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

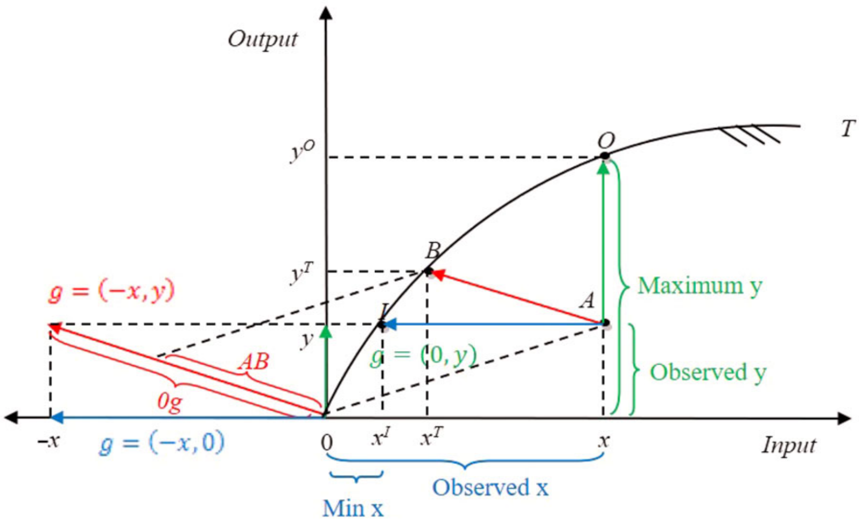

Figure 1. Inefficiency measures with the observed input–output directional vector.

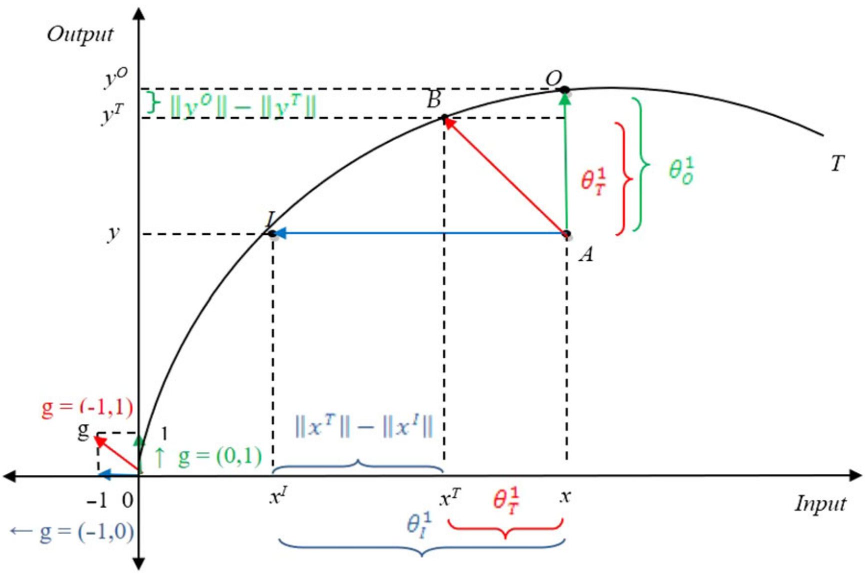

Figure 2. Inefficiency measures with the unit value directional vector.

$(R(x, p))$ functions. The relationship between the DTDF (DIDF or DODF) and the profit (cost or revenue) functions can be represented as [see Chambers et al. (1998)]:

$$\vec{D}_T(x, y; g_x, g_y) \leq \frac{\pi(p, w) - (py - wx)}{pg_y + wg_x}; \vec{D}_I(y, x; g_x) \leq \frac{wx - C(y, w)}{wg_x}; \vec{D}_O(x, y; g_y) \leq \frac{R(x, p) - py}{pg_y}.$$

The right-hand side can be interpreted as a measure of profit (cost or revenue) inefficiency comparing observed profit (cost or revenue) to maximum profit (minimum cost or maximum revenue) normalized by the value of the directional vector. The left-hand side captures overall (input or output) technical inefficiency, respectively. The inequality can be turned into equality by adding a residual term that captures allocative inefficiency, where allocative inefficiency is due to the failure of choosing the profit-maximizing (cost-minimizing or revenue-maximizing) input—output (input or output) vector given relative input and output market prices. Thus, technical inefficiency is due to the overuse of inputs or the loss of production of outputs or both, and allocative inefficiency results from employing inputs and outputs in the wrong proportions.

2.3. Exogenous Directional Vectors

Two widely used exogenous directions are the observed input—output direction and the unit value direction. The observed input—output direction $g = \left(-x, y\right)$ assumes that an inefficient bank can decrease inefficiency while decreasing input and increasing output in proportion to the initial combination of the actual input and output. Figure 1 illustrates how inefficiency can be measured using this type of directional vector. Banks who operate at point A are technically inefficient. The simultaneous maximum proportional contraction of input and expansion of output can be measured in terms of the lengths of x and y, with the use of the Pythagorean theorem, as:

$$\vec{D}_T\left(x, y; g_x, g_y\right) = \vec{D}_T\left(x, y; -x, y\right) = \frac{\|AB\|}{\|0g\|} = \frac{\sqrt{\left(\|x\| - \|x^T\|\right)^2 + \left(\|y^T\| - \|y\|\right)^2}}{\sqrt{\|x\|^2 + \|y\|^2}} = \theta_T.$$

The maximum proportional contraction of input holding output constant can be measured using the directional vector $g = (-x, 0)$. Technical inefficiency is considered to be input-oriented technical inefficiency and is defined by the difference of the lengths of x and x^I divided by the length of x:

$$\vec{D}_T\left(x, y; -x, 0\right) = \vec{D}_I\left(y, x; -x\right) = \frac{\|x\| - \|x^I\|}{\|x\|} = 1 - \frac{\|x^I\|}{\|x\|} = 1 - \frac{1}{D_I\left(y, x\right)} = \theta_I,$$

where $D_I\left(y, x\right)$ is the standard input distance function. The maximum proportional expansion of output holding input constant can be measured using the directional vector $g = \left(0, y\right)$. In this case, technical inefficiency is considered to be output-oriented technical inefficiency and is defined as:

$$\vec{D}_T\left(x, y; 0, y\right) = \vec{D}_O\left(x, y; y\right) = \frac{\|y^O\| - \|y\|}{\|y\|} = \frac{\|y^O\|}{\|y\|} - 1 = \frac{1}{D_O\left(x, y\right)} - 1 = \theta_O,$$

where $D_O\left(x, y\right)$ is the standard output distance function. A critical question that needs to be considered is whether $\theta_T = \theta_I + \theta_O$?

Proposition 1 *Let $\vec{D}_T\left(x, y; -x, 0\right) = \vec{D}_I\left(y, x; -x\right) = \theta_I$ be input-oriented technical inefficiency, $\vec{D}_T\left(x, y; 0, y\right) = \vec{D}_O\left(x, y; y\right) = \theta_O$ be output-oriented technical inefficiency, and $\vec{D}_T\left(x, y; -x, y\right) = \theta_T$ be overall technical inefficiency. Then, $\theta_T = \theta_I + \theta_O - \theta_{IO}$, where θ_{IO} is the IEIO.*

θ_I arises from employing the wrong level of inputs and measures the maximum contraction of a bank's input to that of a best-practice or most efficient bank producing the same output. θ_O arises from producing at the wrong level of outputs and measures the maximum expansion of a bank's output to that of a best-practice or most efficient bank using the same input. The interactions between input and output inefficiencies, θ_{IO}, captures output (input) implications of any errors in the input usage (output production).

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

Another widely used prespecified direction is the unit value direction $g = (-1, 1)$. This type of directional vector implies that the amount by which a bank could decrease input and increase output will be $\vec{D}_T(x, y; -1, 1) \times 1$ units of x and y. Figure 2 illustrates how inefficiency can be measured using the unit value direction. Banks who operate at point A are technically inefficient. The simultaneous maximum contraction of input and expansion of output can be measured as:

$$\vec{D}_T\left(x, y; -1, 1\right) = \frac{\|AB\|}{\|0g\|} = \frac{\sqrt{\left(\|x\| - \|x^T\|\right)^2 + \left(\|y^T\| - \|y\|\right)^2}}{\sqrt{2}} = \theta_T^1.$$

The maximum contraction of input holding output constant can be measured using the directional vector $g = (-1, 0)$. Technical inefficiency is considered to be input-oriented technical inefficiency and is defined as $\vec{D}_I(y, x; -1) = \|AI\| = \|x\| - \|x^I\| = \theta_I^1$.

The maximum expansion of output holding input constant can be measured using the directional vector $g = (0, 1)$. In this case, technical inefficiency is considered to be output-oriented technical inefficiency and is defined as $\vec{D}_O(x, y; 1) = \|AO\| = \|y^O\| - \|y\| = \theta_O^1$. Does $\theta_T^1 = \theta_I^1 + \theta_O^1$?

Proposition 2 *Let $\vec{D}_T(x, y; -1, 0) = \vec{D}_I(y, x; -1) = \theta_I^1$ be input-oriented technical inefficiency, $\vec{D}_T(x, y; 0, 1) = \vec{D}_O(x, y; 1) = \theta_O^1$ be output-oriented technical inefficiency, and $\vec{D}_T(x, y; -1, 1) = \theta_T^1$ be overall technical inefficiency. Then $\theta_T^1 = \theta_I^1 + \theta_O^1 - \theta_{IO}^1$, where θ_{IO}^1 is the IEIO.*

Corollary 1 *Let θ_T^1 be overall technical inefficiency, θ_I^1 be input-oriented technical inefficiency, and θ_O^1 be output-oriented technical inefficiency derived using the unit value directional vectors. Then the IEIO, θ_{IO}^1, is related to θ_I^1 and θ_O^1 as:*

$$\theta_{IO}^1 = \theta_I^1 + \left(\|y^O\| - \left\|y^T\right\|\right),$$

$$\theta_{IO}^1 = \theta_O^1 + \left(\left\|x^T\right\| - \|x^I\|\right).$$

While banks who operate on the production frontier have zero IEIO since $\theta_T^1 = \theta_I^1 = \theta_O^1 = 0$, banks who operate beneath the production frontier have negative IEIO since $\theta_T^1 < \theta_I^1 + \theta_O^1$. Furthermore, there is a relationship between the IEIO, θ_{IO}^1, and the input and output technical inefficiencies. The IEIO equals the input technical inefficiency θ_I^1 plus the loss of production of output $\|y^O\| - \|y^T\|$ (i.e., part of output technical inefficiency) that is forgone to reduce input and eliminate part of the input technical inefficiency $\|x\| - \|x^T\|$—see Figure 2. This suggests that the loss of output creates output technical inefficiency and has an effect on reducing input technical inefficiency. That is, a loss of bank output creates a loss of revenue and has an effect on the reduction of the use of inputs. The IEIO also equals the output technical inefficiency θ_O^1 plus the overuse of input $\|x^T\| - \|x^I\|$ (i.e., part of input technical inefficiency) that is used to produce more output and eliminate part of the output technical inefficiency $\|y^T\| - \|y\|$—see Figure 2. This suggests that the overuse of input creates input technical inefficiency and has an effect on reducing output technical inefficiency. That is, an overuse of a certain input creates additional resources that can be used to produce more outputs that are intensive in using that input.

2.4. Endogenous Directional Vectors

When information on input and output prices is available and banks are assumed to exhibit cost-minimizing (or revenue or profit-maximizing) behavior, technical inefficiency can be measured

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

by choosing an endogenous directional vector such that it projects any inefficient bank to the cost-minimizing (or revenue or profit-maximizing) benchmark.

Following Zofio et al. (2013), the directional vector $g = \left(g_x^{\pi}, g_y^{\pi}\right)$ is assumed to satisfy the price normalization constraint $pg_y^{\pi} + wg_x^{\pi} = 1$ and projects any inefficient bank towards the profit-maximizing bundle (x^{π}, y^{π}) where banks are both technically and allocatively efficient[1]. Thus, the directional vector can be defined as:

$$g = \left(g_x^{\pi}, g_y^{\pi}\right) = \left(\frac{x - x^{\pi}}{\pi\left(p, w\right) - \left(py - wx\right)}, \frac{y^{\pi} - y}{\pi\left(p, w\right) - \left(py - wx\right)}\right), \tag{5}$$

to ensure that $pg_y^{\pi} + wg_x^{\pi} = 1$. Then, the DTDF, $\vec{D}_T\left(x, y; g_x^{\pi}, g_y^{\pi}\right)$, equals the loss of profit due to technical inefficiency and gives a measure of overall technical inefficiency in monetary values.

Proposition 3 *Let $(x, y) \in T$, (p, w) be the vector of output and input prices, and $g = \left(g_x^{\pi}, g_y^{\pi}\right)$ be a vector such that it satisfies $pg_y^{\pi} + wg_x^{\pi} = 1$ and projects any inefficient bank to the profit-maximizing bundle (x^{π}, y^{π}), where banks are both technically and allocatively efficient. Then, $\vec{D}_T\left(x, y; g_x^{\pi}, g_y^{\pi}\right) = \theta_T^{\pi} = \pi\left(p, w\right) - \left(py - wx\right)$ and all profit inefficiency is technical, since allocative inefficiency equals zero.*

Similar cost (revenue) results can be obtained by choosing endogenous directional vectors that satisfy the price normalization constraint and projects any inefficient bank towards the cost-minimizing (revenue-maximizing) bundle.

Corollary 2 *Let $x \in L\left(y\right)$, w be the vector of input prices, and $g = \left(g_x^C, 0\right) = \left(\frac{x - x^C}{wx - C(y,w)}, 0\right)$ a vector that satisfies $wg_x^C = 1$ and projects any inefficient bank to the cost-minimizing bundle $\left(x^C, y\right)$ where banks are both technically and allocatively efficient. Then, $\vec{D}_I\left(y, x; g_x^C\right) = \theta_I^C = wx - C\left(y, w\right)$ and all cost inefficiency is technical, since allocative inefficiency equals zero.*

Corollary 3 *Let $y \in P\left(x\right)$, p be the vector of output prices, and $g = \left(0, g_y^R\right) = \left(0, \frac{y^R - y}{R(x,p) - py}\right)$ a vector that satisfies $pg_y^R = 1$ and projects any inefficient bank to the revenue-maximizing bundle $\left(x, y^R\right)$ where banks are both technically and allocatively efficient. Then, $\vec{D}_O\left(x, y; g_y^R\right) = \theta_O^R = R\left(x, p\right) - py$ and all revenue inefficiency is technical, since allocative inefficiency equals zero.*

Does $\theta_T^{\pi} = \theta_I^C + \theta_O^R$?

Proposition 4 *Let $\vec{D}_T\left(x, y; g_x^{\pi}, g_y^{\pi}\right) = \theta_T^{\pi}$ be overall technical inefficiency, $\vec{D}_I\left(y, x; g_x^C\right) = \theta_I^C$ be input-oriented technical inefficiency, and $\vec{D}_O\left(x, y; g_y^R\right) = \theta_O^R$ be output-oriented technical inefficiency. Then $\theta_T^{\pi} = \theta_I^C + \theta_O^R \pm \theta_{IO}^{CR}$, where θ_{IO}^{CR} is the IEIO.*

Consequently, including price information on the directional vector such as the directional vector that projects any inefficient bank to the profit-maximizing bundle, the IEIO may have positive or negative effects on the overall technical inefficiency, depending on the relationship between cost and profit inefficiency. Cost and profit inefficiency may be positively related if bank managers aim for both low cost and profit inefficiency. Cost and profit inefficiency may be negatively related if high-quality services require higher costs and result in higher measured cost inefficiency but get higher output prices that result in higher profits and lower measured profit inefficiency (see Rogers (1998) and Maudos et al. (2002)).

3. Empirical Application

Two widely used estimation techniques to measuring technical inefficiency are the parametric and nonparametric methods. This paper adopts the parametric stochastic frontier approach (SFA) to examine the efficiency of the US commercial banks. SFA is used, mainly because this approach can separate the effects of noises on the estimated inefficiency measures.

3.1. Model Specification

To obtain the estimates of the directional distance functions and therefore the measure of technical inefficiencies, this subsection provides the parametric specification of these functions. This involves choosing a functional form, imposing the parameter restrictions for the translation property, modeling the interactive effect, and specifying the directional vector g.

3.1.1. The quadratic functional form

The quadratic functional form is used to parameterize the functions in (1), (3), and (4), mainly because this functional form can easily impose the translation property. To avoid any estimation biases that may arise due to potential changes in bank performance due to technological change, technical change is incorporated by a trend variable, t, while nonneutral technical change is modeled by including terms capturing the interaction between trend and inputs and trend and outputs, as is common in the literature. Thus, the directional distance functions defined in (1), (3), and (4) can be rewritten as $\vec{D}_T(x, y, t; g_x, g_y)$, $\vec{D}_I(y, x, t; g_x)$, and $\vec{D}_O(x, y, t; g_y)$, respectively.

3.1.2. Imposing the restrictions

The translation property can be imposed by setting α equal to an arbitrarily chosen input or the negative of an arbitrarily chosen output which is specific to each bank, say $\alpha = -y_1$, and normalizing the corresponding directional vector $g_{y_1} = 1$—see, for example, Malikov et al. (2016)[2]. Using this transformation process and applying it to the empirical implementation that uses two inputs to produce two outputs, the translation property in equation (2) can be rewritten as:

$$\vec{D}_T(x + y_1 g_x, y_2 - y_1 g_{y_2}, t; g_x, g_y) = \vec{D}_T(x, y, t; g_x, g_y) + y_1. \tag{6}$$

Note that the output y_1 disappears from the left-hand side of (6) because of $y_1 - y_1(1) = 0$. Rearranging and adding a random error v_T to equation (6) yields the standard stochastic frontier model with two error terms, as follows:

$$y_1 = \vec{D}_T(x + y_1 g_x, y_2 - y_1 g_{y_2}, t; g_x, g_y) + v_T - u_T, \tag{7}$$

where v_T is a two-sided random error assumed to be identically and independently distributed (iid) with mean zero and variance $\sigma^2_{v_T} = \Sigma$, $v_T \sim N(0, \Sigma)$ and $\vec{D}_T(x, y, t; g_x, g_y) = u_T \geq 0$ is a one-sided error term which captures bank-specific overall technical inefficiency. Applying the quadratic functional form to the first term on the right-hand side of (7), (7) can be written as:

$$y_1 = \alpha_0 + \sum_{n=1}^{2} \alpha_n \widetilde{x}_n + \beta_2 \widetilde{y}_2 + \delta_t t + \frac{1}{2} \sum_{n=1}^{2} \sum_{n'=1}^{2} \alpha_{nn'} \widetilde{x}_n \widetilde{x}_{n'} + \frac{1}{2} \beta_{22} \left(\widetilde{y}_2\right)^2$$
$$+ \frac{1}{2} \delta_{tt} t^2 + \sum_{n=1}^{2} \gamma_{n2} \widetilde{x}_n \widetilde{y}_2 + \sum_{n=1}^{2} \delta_{tx_n} t \widetilde{x}_n + \delta_{ty_2} t \widetilde{y}_2 + v_T - u_T, \tag{8}$$

where $\widetilde{x}_n = x_n + y_1 g_{x_n}$ $(n = 1, 2)$, $\widetilde{y}_2 = y_2 - y_1 g_{y_2}$, and y_1 corresponds to the dependent variable. The parameters of (8) must satisfy the usual restrictions for symmetry $\alpha_{nn'} = \alpha_{n'n}$ $(n \neq n')$. Within a panel data framework, the DTDF model in (8) can be notationally simplified as:

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

$$y_{1},_{it} = \widetilde{R}_{T,it}(g)'\beta_T + v_{T,it} - u_{T,it}, \tag{9}$$

where $i = 1, \ldots, K$ indicates banks; $t = 1, \ldots, T$ indicates time; $\widetilde{R}_T(g)$, which is a function of g, is a vector of all the relevant variables on the right-hand side of (8) including a unity for the intercept term; and β_T is the corresponding vector of coefficients (including the intercept).

Similarly, the translation property of the DIDF can be imposed by setting α equal to an arbitrarily chosen input which is specific to each bank, say $\alpha = x_1$, and normalizing the corresponding directional vector $g_{x_1} = 1$. Using this transformation process and following the above methodology, the quadratic functional form can be written as:

$$-x_1 = \alpha_0 + \alpha_2\widetilde{x}_2 + \sum_{m=1}^{2} \beta_m y_m + \delta_t t + \frac{1}{2}\alpha_{22}\left(\widetilde{x}_2\right)^2 + \frac{1}{2}\sum_{m=1}^{2}\sum_{m'=1}^{2} \beta_{mm'} y_m y_{m'}$$
$$+ \frac{1}{2}\delta_{tt}t^2 + \sum_{m=1}^{2} \gamma_{2m}\widetilde{x}_2 y_m + \delta_{tx_2}t\widetilde{x}_2 + \sum_{m=1}^{2} \delta_{ty_m} t y_m + v_I - u_I, \tag{10}$$

where $\widetilde{x}_2 = x_2 - x_1 g_{x_2}$, x_1 corresponds to the dependent variable, $u_I \geq 0$ is a one-sided error term which captures bank-specific input technical inefficiency. The parameters of (10) must satisfy the usual restrictions for symmetry $\beta_{mm'} = \beta_{m'm}$ $(m \neq m')$. Within a panel data framework, the DIDF model in (10) can be notationally simplified as:

$$-x_{1},_{it} = \widetilde{R}_{I,it}(g)'\beta_I + v_{I,it} - u_{I,it} \tag{11}$$

When $g_x = 0$, the DTDF model in (8) reduces to the DODF that allows for only output expansion:

$$y_1 = \alpha_0 + \sum_{n=1}^{2} \alpha_n x_n + \beta_2\widetilde{y}_2 + \delta_t t + \frac{1}{2}\sum_{n=1}^{2}\sum_{n'=1}^{2} \alpha_{nn'} x_n x_{n'} + \frac{1}{2}\beta_{22}\left(\widetilde{y}_2\right)^2$$
$$+ \frac{1}{2}\delta_{tt}t^2 + \sum_{n=1}^{2} \gamma_{n2} x_n \widetilde{y}_2 + \sum_{n=1}^{2} \delta_{tx_n} t x_n + \delta_{ty_2} t\widetilde{y}_2 + v_O - u_O, \tag{12}$$

where $u_O \geq 0$ is a one-sided error term which captures output technical inefficiency. Within a panel data framework, the DODF model in (12) can be notationally simplified as:

$$y_{1},_{it} = \widetilde{R}_{O,it}(g)'\beta_O + v_{O,it} - u_{O,it}. \tag{13}$$

3.1.3. Modeling the interactive effect

Following Battese and Coelli (1995), overall technical inefficiency, $u_{T,it}$, can be modeled as a linear function of a vector of explanatory bank-specific variables Z_{it} that are expected to influence $u_{T,it}$. Bank-specific variables Z_{it} include input technical inefficiency, $u_{I,it}$, output technical inefficiency, $u_{O,it}$, and a term capturing the interactions between them, $u_{I,it} \times u_{O,it}$:

$$u_{T,it} = Z_{it}\delta + v_{u,it}, \tag{14}$$

where δ is an unknown vector of coefficients (including the intercept) to be estimated, and $v_{u,it}$ is an error term that is defined by the truncation of a normal distribution. It is possible to allow the explanatory variables Z_{it} to affect the production frontier by adding interaction terms between Z_{it} and the regressors, $Z_{it}X_{it}$, similar to Huang and Liu (1994). However, this is not the focus of this paper.

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

3.1.4. Specifying the directional vector

As discussed earlier in Section 2, there are two approaches in the literature. The first is to choose the directional vector a priori. The second approach is to let the data determine the directional vector. In this paper, both approaches are used and compared.

The prespecified directional vector. Two widely used prespecified directions are the unit value direction—see, for example, Park and Weber (2006) and Koutsomanoli-Filippaki et al. (2009), and the observed input–output direction—see, for example, Färe et al. (2004). Input-oriented measures of technical inefficiency $u_{I,it}$ using the unit value (or the observed input) direction can be obtained by setting $g_x = 1$ $\left(\text{or } g_{x_1} = 1, \text{and } g_{x_2} = x_2\right)$ in the case of (11) and estimate the single-equation DIDF subject to the usual symmetry restrictions and theoretical monotonicity restrictions. Similarly, output-oriented measures of technical inefficiency $u_{O,it}$ using the unit value (or the observed output) direction can be obtained by setting $g_y = 1$ $\left(\text{or } g_{y_1} = 1, \text{and } g_{y_2} = y_2\right)$ in the case of (13) and estimate the single-equation DODF subject to the usual symmetry restrictions and theoretical monotonicity restrictions. Overall or technology-oriented measures of technical inefficiency $u_{T,it}$ using the unit value (or the observed input–output) direction can be obtained by setting $g = (g_x, g_y) = (-1, 1)$ $\left(\text{or } g_{x_1} = x_1, g_{x_2} = x_2, g_{y_1} = 1, \text{and } g_{y_2} = y_2\right)$ in the case of (9) and estimate the system of equations that includes the DTDF and the interactive effect equation in (14) subject to the usual symmetry restrictions and theoretical monotonicity restrictions, while the translation property is already imposed by construction. More specifically, the system can be written as:

$$\begin{bmatrix} y_{1,it} \\ u_{T,it} \end{bmatrix} = \begin{bmatrix} \widetilde{R}_{T,it}(g)' & \\ & Z'_{it} \end{bmatrix} [\beta] - \begin{bmatrix} u_{T,it} \\ 0 \end{bmatrix} + \begin{bmatrix} v_{T,it} \\ v_{u,it} \end{bmatrix}$$

which can be written in a compressed form as:

$$Y_{it} = R_{it}(g)\beta - u_{T,it}\iota + v_{it}, \tag{15}$$

where $\iota = [1, 0]$ and $v_{it} = \left(v_{T,it}, v_{u,it}\right)' \sim N(0, \Sigma)$. Bank-specific variables Z is a vector of the relevant variables on the right-hand side of (14) that are obtained using the unit value (or the observed input and output) directional vector in equations (11) and (13). In this system of equations, bank-specific variables Z that determine overall technical inefficiency are estimated simultaneously with the variables that determine the frontier.

The optimal directional vector. The optimal directional vector is treated as unknown parameters in which the bank's movement toward the efficient frontier is to be estimated.

The cost-optimal directional vector. The bank cost-minimizing objective can be defined in terms of the DIDF in equation (3), to keep consistency with the endogeneity of inputs and exogeneity of outputs, as:

$$C(y, w) = \min_x \left\{w'x : \vec{D}_I\left(y, x; g_x\right) \geq 0\right\}. \tag{16}$$

Following Luenberger (1992), the problem in (16) can be represented by an unconstrained problem as:

$$C(y, w) = \min_x \left\{w'x - \vec{D}_I\left(y, x; g_x\right) \times w'g_x\right\}.$$

The corresponding first-order conditions are:

$$w_n = \nabla_{x_n} \vec{D}_I(.)\lambda_I \quad \textit{for } n = 1, 2, \tag{17}$$

where $\lambda_I = w'g_x = \sum_{n=1}^{N} w_n g_{x_n}$ is the sum of direction-weighted cost. The first-order conditions in (17) are the inverse demand functions. To meet the rank condition for the identification of the model, a total of at least the total number of potentially endogenous variables is needed as

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

independent equations in the system. As it is well known, the system (17) is singular and only $(n-1)$ equations in (17) can be used for the estimation—see Barten (1969) for more details. The DIDF in (11) plus the $(n-1)$ first-order conditions in (17) provide a system of n equations. Precisely, the system consists of the DIDF in (11) and the first-order condition for w_2. Note that $\nabla_{x_2}\vec{D}_I(y, \tilde{x}; g_x) = \nabla_{x_2}\vec{D}_I(y, x; g_x)$ by the translation property, then the first-order condition in (17) can be rewritten in terms of the parameters of (11) after adding an iid normal error term v_C as follows:

$$w_2 = \lambda_I \left(\alpha_2 + \alpha_{22}\tilde{x}_2 + \sum_{m=1}^{2} \gamma_{2m} y_m + \delta_{tx_2} t \right) + v_C. \tag{18}$$

Note that, solving the first-order condition for $\tilde{x}_2$ treats $\tilde{x}_2$ as an endogenous variable (as opposed to x_2). The equivalence of working with $\tilde{x}_2$ and working with x_2 holds because of $\partial\tilde{x}_2/\partial x_2 = 1$. By allowing g to differ across banks, (18) can be further written in a panel data framework as:

$$w_{2,it} = \tilde{R}_{C,it}\,(g_i)'\beta_C + v_{C,it}, \tag{19}$$

where $\tilde{R}_C$ is a vector of the relevant variables on the right-hand side of (18) and β_C is the corresponding vector of coefficients. β_C is a subset of β_I, which can be obtained using a selection matrix A_I that contains elements which are either 0 or 1, where $\beta_C = A_I\beta_I$.

The DIDF system (equations (11) and (19)) is a simultaneous equation model where the entire vector x is endogenous. The entire system can be written as:

$$\begin{bmatrix} -x_{1,it} \\ w_{2,it} \end{bmatrix} = \begin{bmatrix} \tilde{R}_{I,it}(g_i)' & \\ & \tilde{R}_{C,it}(g_i)' \end{bmatrix} [\beta] - \begin{bmatrix} u_{I,it} \\ 0 \end{bmatrix} + \begin{bmatrix} v_{I,it} \\ v_{C,it} \end{bmatrix}$$

which can be written in a compressed form as:

$$Y_{it} = R_{it}(g_i)\beta - u_{I,it}\iota + v_{it}, \tag{20}$$

where $\iota = [1, 0]$, $v_{it} = (v_{I,it}, v_{C,it})' \sim N(0, \Sigma)$, and $g_i = (g_{x_{2i}})'$ with $i = 1, \ldots, K$ indexing banks. The system is estimated subject to the symmetry restrictions and theoretical monotonicity restrictions. The directional vector g_i is treated as unknown parameters which are estimated jointly with the remaining parameters in the system. The obtained estimates of the directional vector can be interpreted as being cost-optimal due to the inclusion of the cost-minimizing first-order conditions in the system—see Malikov et al. (2016). That is, the estimated DIDF direction captures the bank movement to the point on a technological frontier where costs are minimized.

The revenue-optimal directional vector. The bank revenue-maximizing objective can be defined in terms of the DODF in equation (4), to keep consistency with the endogeneity of outputs and exogeneity of inputs, as:

$$R(x, p) = \max_{y} \left\{ p'y + \vec{D}_O(x, y; g_y) \times p'g_y \right\}$$

The first-order conditions for the revenue maximization problem are

$$p_m = -\nabla_{y_m}\vec{D}_O(.)\lambda_O \quad \textit{for } m = 1, 2 \tag{21}$$

where $\lambda_O = p'g_y = \sum_{m=1}^{M} p_m g_{y_m}$ is the sum of direction-weighted revenues. The first-order conditions in (21) are the inverse supply functions. A system of m equations consists of the DODF in equation (13) plus the first-order condition for p_2 in (21), which can be rewritten in terms of the parameters of (13) after adding an iid normal error term v_R as follows:

$$p_2 = -\lambda_O \left(\beta_2 + \beta_{22}\tilde{y}_2 + \sum_{n=1}^{2} \gamma_{n2} x_n + \delta_{ty_2} t \right) + v_R. \tag{22}$$

By allowing g to differ across banks, (22) can be further written in a panel data framework as:

$$p_{2,it} = \widetilde{R}_{R,it}(g_i)' \beta_R + v_{R,it}, \tag{23}$$

where $\widetilde{R}_R$ is a vector of the relevant variables on the right-hand side of (22) and β_R, which is a subset of β_O, is the corresponding vector of coefficients.

The DODF system (equations (13) and (23)) is a simultaneous equation model where the entire vector y is endogenous. The entire system can be written as:

$$\begin{bmatrix} y_{1,it} \\ p_{2,it} \end{bmatrix} = \begin{bmatrix} \widetilde{R}_{O,it}(g_i)' & \\ & \widetilde{R}_{R,it}(g_i)' \end{bmatrix} [\beta] - \begin{bmatrix} u_{O,it} \\ 0 \end{bmatrix} + \begin{bmatrix} v_{O,it} \\ v_{R,it} \end{bmatrix}$$

which can be written in a compressed form as:

$$Y_{it} = R_{it}(g_i)\beta - u_{O,\, it}\iota + v_{it}, \tag{24}$$

where $\iota = [1, 0]$, $v_{it} = \left(v_{O,it}, v_{R,it}\right)' \sim N(0, \Sigma)$, and $g_i = \left(g_{y_{2i}}\right)'$ for $i = 1, \ldots, K$. The obtained estimates of the directional vector g_i can be interpreted as being revenue-optimal due to the inclusion of the revenue-maximizing first-order conditions in the system.

The profit-optimal directional vector. Following Chambers et al. (1998), the bank profit-maximizing objective can be defined in terms of the DTDF in equation (1) as:

$$\pi(p, w) = \max_{x,y} \left\{ \left(p'y - w'x\right) + \vec{D}_T\left(x, y; g_x, g_y\right)\left(p'g_y + w'g_x\right) \right\}.$$

The first-order conditions for the profit maximization problem are

$$\begin{aligned} w_n &= \nabla_{x_n} \vec{D}_T(.)\lambda_T \qquad \text{for } n = 1, \ldots, N, \\ p_m &= -\nabla_{y_m} \vec{D}_T(.)\lambda_T \quad \text{for } m = 1, \ldots, M, \end{aligned} \tag{25}$$

where $\lambda_T = p'g_y + w'g_x = \sum_{m=1}^{M} p_m g_{y_m} + \sum_{n=1}^{N} w_n g_{x_n}$ is the sum of direction-weighted profits. The first-order conditions in (25) can be rewritten in terms of the parameters of (9) after adding iid normal error terms v_π as follows:

$$\begin{aligned} w_n &= \lambda_T \left(\alpha_n + \sum_{n'=1}^{2} \alpha_{nn'} \widetilde{x}_{n'} + \gamma_{n2}\widetilde{y}_2 + \delta_{tx_n} t \right) + v_{\pi w_n} \quad \text{for } (n = 1, 2) \\ p_2 &= -\lambda_T \left(\beta_2 + \beta_{22}\widetilde{y}_2 + \sum_{n=1}^{2} \gamma_{n2}\widetilde{x}_n + \delta_{ty_2} t \right) + v_{\pi p_2}. \end{aligned} \tag{26}$$

By allowing g to differ across banks, (26) can be further written in a panel data framework as:

$$\begin{aligned} w_{n,it} &= \widetilde{R}_{w_n,it}(g_i)' \beta_{w_n} + v_{\pi w_n,it} \qquad \text{for } (n = 1, 2) \\ p_{2,it} &= \widetilde{R}_{p_2,it}(g_i)' \beta_{p_2} + v_{\pi p_2,it}, \end{aligned} \tag{27}$$

where $\widetilde{R}_{w_n}$ and $\widetilde{R}_{p_2}$ are the vectors of the relevant variables on the right-hand sides of (26) and β_{w_n} and β_{p_2}, which are subsets of β_T, are the corresponding vectors of coefficients.

The DTDF in equation (9) plus the ($(n + m) - 1$) first-order conditions for profit maximization given in equation (27) provide a system of $(n + m)$ equations. This system of equations is a simultaneous equation model where the entire vector (x, y) is endogenous. Adding the interactive effect equation in (14), the entire system can be written as:

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

$$\begin{bmatrix} y_{1,it} \\ w_{1,it} \\ w_{2,it} \\ p_{2,it} \\ u_{T,it} \end{bmatrix} = \begin{bmatrix} \widetilde{R}_{T,it}(g_i)' & & & & \\ & \widetilde{R}_{w_1,it}(g_i)' & & & \\ & & \widetilde{R}_{w_2,it}(g_i)' & & \\ & & & \widetilde{R}_{p_2,it}(g_i)' & \\ & & & & Z'_{it} \end{bmatrix} [\beta] - \begin{bmatrix} u_{T,it} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} v_{T,it} \\ v_{\pi w_1,it} \\ v_{\pi w_2,it} \\ v_{\pi p_2,it} \\ v_{u,it} \end{bmatrix}$$

which can be written in a compressed form as:

$$Y_{it} = R_{it}(g_i)\beta - u_{T,it}\iota + v_{it}, \tag{28}$$

where $\iota=[1, 0, 0, 0, 0]$, $v_{it}=\left(v_{T,it}, v_{\pi w_1,it}, v_{\pi w_2,it}, v_{\pi p_2,it}, v_{u,it}\right)' \sim N(0, \Sigma)$, and $g_i=\left(g_{x_{1i}}, g_{x_2 i}, g_{y_{2i}}\right)'$ for $i = 1, \ldots, K$. Bank-specific variables Z is a vector of the relevant variables on the right-hand side of (14) that are obtained using the cost-optimal and revenue-optimal directional vectors defined in the system of equations in (20) and (24), respectively. The obtained estimates of the directional vector g_i can then be interpreted as being profit-optimal due to the inclusion of the profit-maximizing first-order conditions in the system.

Setting all the elements of g equal to ones; $g_{x_{n1}} = g_{x_{n2}} = \ldots = g_{x_{nk}} = 1$ for $(n = 1, 2)$, and $g_{y_{21}} = g_{y_{22}} = \ldots = g_{y_{2k}} = 1 \left(\text{or } g_{x_{1i}} = x_{1_i}, g_{x_{2i}} = x_{2_i}, \text{and } g_{y_{2_i}} = y_{2_i}\right)$ without additional first-order condition equations, the system in (28) reduces to (15) in the case of the unit value (or the observed input–output) directional vector.

4. Bayesian Estimation

Bayesian approach is used to estimate the DIDF, DODF, and DTDF models using the unit value and the observed input–output directional vectors defined by (11), (13), and (15), and the cost-optimal, revenue-optimal, and profit-optimal directional vectors defined by (20), (24), and (28), respectively. Bayesian estimation involves using a Markov Chain Monte Carlo (MCMC) sampling algorithm to generate sequences of samples from the joint posterior distribution of inefficiency and the unknown parameters of the model. This paper uses Metropolis–Hastings algorithm introduced by Metropolis et al. (1953) and Hastings (1970). Bayesian estimation and MCMC sampling algorithm have been widely documented in the stochastic frontier literature and thus are not discussed in this paper—see, for example, Koop and Steel (2003), and O'Donnell and Coelli (2005). Bayesian approach is used, mainly because this approach can easily impose monotonicity conditions implied by microeconomic theory and estimate directional vectors that vary across banks.

4.1. Prior Distributions

The use of Bayesian approach requires choosing prior distributions for the parameters β, Σ^{-1}, u_{it}, λ^{-1}, and g_i. For the ease of the comparison of the results among the three directional distance function models, the same prior distributions for the parameters are used. Following Gelfand et al. (1990), a normal prior distribution with zero mean and a large variance for β is used to ensure that the prior distribution for β is relatively uninformative:

$$p(\beta) \sim N(\beta_0, \Omega_\beta) I\left(\beta \in S_j\left(g_i\right)\right), \tag{29}$$

where β_0 is a vector of zeros and Ω_β is a diagonal matrix with 10^4 in diagonal elements. $I\left(\beta \in S_j\left(g_i\right)\right)$ is an indicator function which takes the value 1 if the constraints are satisfied and 0 otherwise, and $S_j\left(g_i\right)$, which depends on g_i, is the set of permissible parameter values when no theoretical regularity constraints $\left(j = 0\right)$ are imposed and when the theoretical regularity constraints

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

$(j = 1)$ must be satisfied. The indicator function restricts prior support to the region where the theoretical regularity constraints are satisfied.

For the covariance matrix Σ, the Wishart distribution is used first due to its conjugacy properties with the normal sampling model. However, it is found to be biased toward large values which result in large values for Σ and consequently large values for the inefficiency measures. The MCMC algorithm for a system of equations is also terminated after a small number of iterations due to the large values involved. Then, following O'Donnell and Coelli (2005), the following prior is used:

$$p\left(\Sigma^{-1}\right) \propto \Sigma \tag{30}$$

which implies that Σ^{-1} is fully determined by the likelihood function—see the conditional posterior density for Σ^{-1} in equation (36).

As noted by Van den Broeck et al. (1994), models based on exponential distribution are reasonably robust to changes in prior assumptions about the parameters. Therefore, an exponential distribution is used for the technical inefficiency u_{it} with an unknown parameter λ following Koop and Steel (2003); $u_{it} \sim i.i.d.\ \exp\left(\lambda^{-1}\right)$. Since the exponential distribution is a gamma distribution when its first parameter equals one, the prior for u_{it} can be written as:

$$p\left(u_{it} \mid \lambda^{-1}\right) = f_{\text{Gamma}}\left(u_{it} \mid 1, \lambda^{-1}\right). \tag{31}$$

According to Fernandez et al. (1997), to obtain a proper posterior, a proper prior for the parameter λ should be used. Therefore, the parameter λ is assumed to have independent exponential prior with mean equals to $-1/\ln \tau^*$ following Van den Broeck et al. (1994). The prior independence of λ leads to marginally prior independent of inefficiencies:

$$p(\lambda^{-1}) = f_{\text{Gamma}}\left(\lambda^{-1} \mid 1, -\ln \tau^*\right), \tag{32}$$

where τ^* is the prior estimate of the mean of the technical efficiency distribution—see, for example, Koop et al. (1997) and O'Donnell and Coelli (2005). Our best prior knowledge of the efficiency of US banks is the mean efficiency value of 0.4583 for DTDF with a prespecified directional vector, and 0.9431 for DTDF with a cost-optimal directional vector reported by Malikov et al. (2016) who apply a Bayesian DTDF-cost system approach to large US commercial banks for 2001–2010 period. The mean output technical efficiency value of 0.9279 is reported by Feng and Serletis (2010) who apply a Bayesian output distance function to US large banks for 2000–2005 period. To our knowledge, the input technical efficiency is not reported by any US banking study over a comparable period. However, the mean input technical efficiency value of 0.690 for all banks, 0.707 for small banks, and 0.735 for large banks are reported by Marsh et al. (2003) who apply a Bayesian input distance function to US commercial banks during 1990–2000. Since changing τ^* changes the prior moments, various values of τ^* within its possible range is experimented to assess the sensitivity of the results to changes to τ^*. The results are the same up to the number of digits presented in Section 6, implying that the results are very robust to large changes in τ^*.

To account for heterogeneity in the directional vectors for banks, prior distribution for g_i is specified as a normal prior distribution with mean $G_0 = 1$ and a large variance:

$$p(g_i) \sim N(G_0, \Omega_G), \tag{33}$$

where G_0 is a vector of ones, and Ω_G is a diagonal matrix with 10^4 in diagonal elements. The directional vector that projects any inefficient bank to the cost (revenue or profit) minimizing (maximizing) benchmark does not impose any sign restrictions on the adjustments of inputs and outputs. Therefore, these directional vectors may have negative components such that inputs are expanded, or outputs are contracted to reach the frontier at the cost (revenue or profit) minimizing (maximizing) benchmark—see, for example, Zofio et al. (2013) for a profit-optimal directional vector and Atkinson et al. (2018) for a cost and a profit-optimal directional vectors.

Using the priors in (29)–(33), and assuming that the prior distributions of the parameters are independent, the joint prior probability density function is therefore

$$f\left(\beta, \Sigma^{-1}, u_{it}, \lambda^{-1}, g_i\right) = p(\beta)p\left(\Sigma^{-1}\right) p\left(u_{it} \mid \lambda^{-1}\right) p(\lambda^{-1})p(g_i). \tag{34}$$

4.2. Full Conditional Posterior Distributions

Let $\Gamma = \left(\beta, \Sigma^{-1}, u_{it}, \lambda^{-1}, g_i\right)$ denotes all the parameters of the model, and Γ_{-a} denotes all parameters other than a. To derive the likelihood function, the Jacobian transformation matrix from the vector of random errors to the endogenous variables (inputs, outputs, or all the inputs and outputs) for the DIDF-cost, DODF-revenue, or DTDF-profit system is defined as $J_{it}\left(g_i, \beta\right) = \frac{\partial\left(v_{I,it} - u_{I,it}, v_{C,it}\right)}{\partial\left(x_{1,it}, x_{2,it}\right)}$, $J_{it}\left(g_i, \beta\right) = \frac{\partial\left(v_{O,it} - u_{O,it}, v_{R,it}\right)}{\partial\left(y_{1,it}, y_{2,it}\right)}$, or $J_{it}\left(g_i, \beta\right) = \frac{\partial\left(v_{T,it} - u_{T,it}, v_{\pi w_1,it}, v_{\pi w_2,it}, v_{\pi p_2,it}\right)}{\partial\left(y_{1,it}, x_{1,it}, x_{2,it}, y_{2,it}\right)}$, respectively. Applying the Jacobian transformation, the likelihood function of Y, given Γ is

$$L\left(Y \mid \Gamma\right) = \left[\prod_{i=1}^{K}\prod_{t=1}^{T} f_{\text{Normal}}\left(Y_{it} \mid R_{it}\left(g_{it}\right)\beta - u_{it}\iota, I \otimes \Sigma\right)\right] \prod_{i=1}^{K}\prod_{t=1}^{T} \left|\det\left(J_{it}\left(g_i, \beta\right)\right)\right|. \tag{35}$$

Using Bayes' theorem and combining the likelihood function in (35) and the joint prior distributions in (34), the full conditional posterior distributions for all the parameters are found to be

$$p\left(\Sigma^{-1} \mid Y, \Gamma_{-\Sigma^{-1}}\right) \propto f_{\text{Gamma}}\left(\Sigma^{-1} \mid \frac{KT}{2}, \frac{1}{2}\left(Q_{it} + u_{it}\iota\right)'\left(Q_{it} + u_{it}\iota\right)\right), \tag{36}$$

$$p\left(\beta \mid Y, \Gamma_{-\beta}\right) \propto f_{\text{Normal}}\left(\beta \mid Dd, D\right) \prod_{i=1}^{K}\prod_{t=1}^{T} \left|\det\left(J_{it}\left(g_i, \beta\right)\right)\right| I\left(\beta \in S_j\left(g_i\right)\right), \tag{37}$$

$$p\left(\lambda^{-1} \mid Y, \Gamma_{-\lambda^{-1}}\right) \propto f_{\text{Gamma}}\left(\lambda^{-1} \mid KT + 1, u'\iota_{KT} - \ln\tau^*\right), \tag{38}$$

$$p\left(u_{it} \mid Y, \Gamma_{-u_{it}}\right) \propto f_{\text{Normal}}\left(u_{it} \mid -\frac{\left(Q_{it}'\left(g_i\right)\Sigma^{-1}\iota + \lambda^{-1}\right)}{\iota'\Sigma^{-1}\iota}, \frac{1}{\iota'\Sigma^{-1}\iota}\right) I\left(u_{it} \geq 0\right), \tag{39}$$

$$p\left(g_i \mid Y, \Gamma_{-g_i}\right) \propto \left[\prod_{i=1}^{K} f_{\text{Normal}}\left(Y_i \mid R_i\left(g_i\right)\beta - u_i\iota, I \otimes \Sigma\right)\right] \prod_{i=1}^{K}\prod_{t=1}^{T} \left|\det\left(J_{it}\left(g_i, \beta\right)\right)\right|$$
$$\times \prod_{i=1}^{K} f_{\text{Normal}}\left(g_i \mid G_0, \Omega_G\right) \prod_{i=1}^{K} I\left(\beta \in S_j\left(g_i\right)\right), \tag{40}$$

where $D = \left(R_{it}\left(g_i\right)'\left(I \otimes \Sigma\right)^{-1} R_{it}\left(g_i\right) + \Omega_\beta^{-1}\right)^{-1}$, $d = R_{it}\left(g_i\right)'\left(I \otimes \Sigma\right)^{-1}\left(Y_{it} + u_{it}\iota\right) + \Omega_\beta^{-1}\beta_0$, and $Q_{it} = Y_{it} - R_{it}\left(g_i\right)\beta$. $I\left(u_{it} \geq 0\right)$ is an indicator function that takes the value 1 if the constraint $u_{it} \geq 0$ is satisfied and 0 otherwise.

Bayesian estimation for a single-equation stochastic directional distance function without additional first-order condition equations and with a prespecified directional vector $g = (-1, 1)$ or $g = \left(-x, y\right)$ can be implemented by setting $\prod_{i=1}^{K}\prod_{t=1}^{T}\left|\det\left(J_{it}\left(g_i, \beta\right)\right)\right| = 1$ and setting the relevant elements of g equal to ones or $g_{x_{1i}} = x_{1i}$, $g_{x_{2i}} = x_{2i}$, $g_{y_{2i}} = y_{2i}$, and normalizing the relevant directional vector. Bayesian estimation without theoretical regularity constraints can be implemented by setting $I\left(\beta \in S_j\left(g_i\right)\right)$ in (37) and (40) equal to 1 and then drawing sequentially from the full conditional posteriors in (36)–(40).

4.3. Estimating the Interactive Effect

In the Bayesian framework, Koop et al. (1997) propose a model where a time-invariant inefficiency is assumed to be exponentially distributed with producer-specific mean inefficiencies λ_i and

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

independent exponential priors; $u_i \sim i.i.d.\ \exp(\lambda_i^{-1})$ where $\lambda_i = \exp(Z_i'\delta)$. Following Koop et al. (1997), the inefficiency term $u_{T,it}$, can be specified to be time-variant inefficiency by including bank-specific time-varying covariates in the parameter of an exponential distribution as $u_{T,it} \sim i.i.d.\ \exp(\lambda_{it}^{-1})$, and $\lambda_{it} = \exp(Z_{it}'\delta)$, where δ is an unknown vector of coefficients (including the intercept) to be estimated. The prior for $u_{T,it}$ can be written as:

$$p\left(u_{T,it}, \delta \mid Z\right) = f_{\text{Gamma}}\left(u_{T,it} \mid 1, \exp\left(Z_{it}'\delta\right)\right). \tag{41}$$

Following Koop et al. (1997), the parameter vector δ is assumed to have a proper prior independent of the other parameters. A normal prior distribution with mean δ_0 and variance Ω_δ for δ is used:

$$p(\delta) \sim N(\delta_0, \Omega_\delta), \tag{42}$$

where δ_0 is a vector of zeros and Ω_δ is a diagonal matrix with 10^4 in diagonal elements. Note that by conditioning on Y and Z, bank-specific variables Z are allowed to be correlated with the variables describing the frontier Y. The full conditional posterior distributions for δ and $u_{T,it}$ are found to be

$$p\left(\delta \mid Y, \Gamma_{-\delta}\right) \propto f_{\text{Normal}}\left(\delta \mid \frac{\delta_0 \Omega_\delta^{-1}\iota - Z_{it}' u_{T,it}}{\iota'\Omega_\delta^{-1}\iota}, \frac{1}{\iota'\Omega_\delta^{-1}\iota}\right), \tag{43}$$

$$p\left(u_{T,it} \mid Y, \Gamma_{-u_{T,it}}\right) \propto f_{\text{Normal}}\left(u_{T,it} \mid -\frac{\left(Q_{it}'\left(g_i\right)\Sigma^{-1}\iota + \mu_{it}\right)}{\iota'\Sigma^{-1}\iota}, \frac{1}{\iota'\Sigma^{-1}\iota}\right) I\left(u_{T,it} \geq \mu_{it}\right), \tag{44}$$

where $Q_{it}(g_i) = Y_{it} - R_{it}(g_i)\beta$, and $\mu_{it} = Z_{it}'\delta$. Using MCMC methods, posteriors for input and output inefficiencies that derived separately are then used as proposal distributions in a computationally efficient second stage. See Lunn et al. (2013).

5. Data

The annual data on US commercial banks used in this paper is obtained from the Reports of Income and Condition (Call Reports) over the period from 2001 to 2015. Only continuously operating banks are examined to avoid the impact of entry through new charters and exit through failure or merger, and to focus on the performance of a core of healthy, surviving banks during the sample period. The data sample consists of a balanced panel of a total of 148 banks ($K = 148$) observed over 15 years, for a total of 2220 observations.

To select the relevant variables, the commonly accepted asset approach proposed by Sealey and Lindley (1977) is used. It defines loans and other assets as outputs, while deposits and other liabilities are treated as inputs. On the input side, two inputs are included; the quantity of labor, x_1, and the quantity of purchased funds and deposits, x_2. On the output side, two outputs are included; total loans y_1 which is composed of consumer loans, commercial and industrial loans, and real estate loans; and securities y_2 which includes all nonloan financial assets (i.e., all financial and physical assets minus the sum of total loans and physical capital (premises and other fixed assets)), so that all financial assets are included.

While nontraditional banking activities are becoming increasingly important in identifying bank outputs, the imperfect data and the wide range of activities such as securitization, brokerage services, management of financial assets for depositors and borrowers, and others make the measurement of nontraditional banking activities controversial. See Stiroh (2000) for a discussion of the different approaches to the measurement of nontraditional banking activities. To avoid the uncertainties associated with the measurement of nontraditional banking activities, it is not included as an additional output.

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

Table 1. Data summary statistics

Variable	Mean	Fifth percentile	Median	95th percentile	Standard deviation
Financial assets and liabilities					
x_1	333.6545	70.0000	281.0000	741.5000	239.2208
x_2	1211.1000	312.4149	1030.9000	2701.0000	810.5971
y_1	911.6341	173.4561	750.6042	2179.6000	686.9423
y_2	402.4122	68.2664	328.4993	977.2095	308.4258
Total assets	1350.7000	359.9879	1140.1000	3048.6000	918.6617
Bank-specific prices					
w_1	59.6961	37.4599	54.5960	91.3389	24.5955
w_2	0.0148	0.0014	0.0133	0.0341	0.0104
p_1	0.0611	0.0388	0.0591	0.0830	0.0227
p_2	0.0625	0.0217	0.0491	0.1435	0.0628
Market prices					
w_1	58.0550	50.7531	56.8821	63.9205	3.9677
w_2	0.0148	0.0031	0.0139	0.0318	0.0093
p_1	0.0612	0.0457	0.0611	0.0807	0.0096
p_2	0.0578	0.0473	0.0585	0.0728	0.0079

Note: All variables but labor and input and output prices are in thousands of real 2005 US dollars.

All the quantities of inputs and outputs are constructed by following the data construction method in Berger and Mester (2003). These quantities are deflated by the consumer price index (CPI) to the base year 2005, except for the quantity of labor. Given the numerical size of inputs and outputs reported in Table 1, we experienced convergence problems in estimation. Therefore, the data are normalized by dividing each input and output by its sample mean prior to the estimation following Färe et al. (2005).

For the input and output prices, the actual price paid by the bank for each input or the bank-specific price of input is obtained by dividing total expenses on each input by the corresponding input quantity. Similarly, the actual price received by the bank for each output or the bank-specific price of output is obtained by dividing total revenues from each output by the corresponding output quantity. Thus, for example, the bank-specific price of labor w_1 is obtained from expenses on salaries and benefits divided by the number of full-time employees x_1. The same approach is used to obtain w_2, p_1, and p_2.

Following Berger and Mester (2003), the market-average prices faced and determined exogenously rather than the actual prices paid or received by the bank are used. These market-average prices are more likely to be exogenous to the bank than the bank-specific prices. The bank market-average price at a given year is the weighted average of the other banks prices at that year excluding the bank-specific price, where the weights are each bank respective market share at that year. For example, the market-average price of labor that bank i faces in the labor market L at year t is obtained as:

$$w_{1it} = \sum\nolimits_{j=1,j\neq i}^{l} \left(\frac{x_{1jt}}{\sum\nolimits_{h=1,h\neq i}^{l} x_{1ht}} \right) w_{1jt}$$

where l is the number of banks operating in labor market L, x_{1jt} is the number of full-time employees of bank j at year t and w_{1jt} is the bank j specific price of labor at year t. The same approach is used to obtain w_{2it}, p_{1it}, and p_{2it} for $i = 1, \ldots, K$ over the years $t = 1, \ldots, T$. Data summary statistics are presented in Table 1.

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

6. Empirical Results

To investigate the relationships among input, output, and overall technical inefficiencies, several models are estimated. Specifically, input, output, and technology-oriented technical inefficiencies are estimated separately using the Bayesian procedure outlined in Section 4 and the DIDF, DODF, and DTDF, respectively. All of these inefficiencies are estimated with the three commonly used directions: the unit value, the observed input–output, and the optimal directional vectors. The unit value direction models are referred to as UDIDF, UDODF, and UDTDF, respectively. The observed input–output direction models are referred to as VDIDF, VDODF, and VDTDF, respectively. The optimal direction models are referred to as ODIDF, ODODF, and ODTDF, respectively. In total, nine models are estimated. For each of the nine models, a total of 450,000 observations are generated and then the first 150,000 observations are discarded as a burn-in. The simulation inefficiency factor (SIF) values for all the parameters of these models are estimated to check the mixing performance of the samplers following Kim et al. (1998). The SIF values for the three directional vectors models are reported in Table 4, suggesting that the samplers for these models have converged.

6.1. Imposing the Theoretical Regularity Conditions

As required by neoclassical microeconomic theory, the production technology has to satisfy the theoretical regularity conditions of monotonicity and curvature. Monotonicity requires that the directional distance function be nondecreasing in inputs and nonincreasing in outputs. Therefore, monotonicity conditions of the DTDF imply the following restrictions:

$$\begin{aligned}
\frac{\partial \vec{D}_T(.)}{\partial x_n} &= \alpha_n + \sum_{n'=1}^{2} \alpha_{nn'} x_{n'} + \left[\alpha_{nn} g_{x_n} + \alpha_{nn'} g_{x_{n'}} - \gamma_{n2} g_{y_2}\right] y_1 + \gamma_{n2} y_2 + \delta_{tx_n} t \geq 0 \quad (n = 1, 2)\,; \\
\frac{\partial \vec{D}_T(.)}{\partial y_1} &= \left[\alpha_1 g_{x_1} + \alpha_2 g_{x_2} - \beta_2 g_{y_2} - 1\right] \\
&\quad + \left[\alpha_{11} g_{x_1}^2 + \alpha_{22} g_{x_2}^2 + \beta_{22} g_{y_2}^2 + \alpha_{12} g_{x_1} g_{x_2} - \gamma_{12} g_{x_1} g_{y_2} - \gamma_{22} g_{x_2} g_{y_2}\right] y_1 \\
&\quad + \left[\gamma_{12} g_{x_1} + \gamma_{22} g_{x_2} - \beta_{22} g_{y_2}\right] y_2 \\
&\quad + \left[\alpha_{11} g_{x_1} + \alpha_{12} g_{x_2} - \gamma_{12} g_{y_2}\right] x_1 \\
&\quad + \left[\alpha_{21} g_{x_1} + \alpha_{22} g_{x_2} - \gamma_{22} g_{y_2}\right] x_2 \\
&\quad + \left[\delta_{tx_1} g_{x_1} + \delta_{tx_2} g_{x_2} - \delta_{ty_2} g_{y_2}\right] t \leq 0; \\
\frac{\partial \vec{D}_T(.)}{\partial y_2} &= \beta_2 + \left[\gamma_{12} g_{x_1} + \gamma_{22} g_{x_2} - \beta_{22} g_{y_2}\right] y_1 + \beta_{22} y_2 + \sum_{n=1}^{2} \gamma_{n2} x_n + \delta_{ty_2} t \leq 0.
\end{aligned} \tag{45}$$

When $g_{x_1} = g_{x_2} = 0$, monotonicity conditions of the DTDF in (45) reduces to the monotonicity conditions of the DODF. Monotonicity conditions of the DIDF imply the following restrictions:

$$\frac{\partial \vec{D}_I(.)}{\partial x_1} = \left[1 - \alpha_2 g_{x_2}\right] + \alpha_{22} g_{x_2}^2 x_1 - \alpha_{22} g_{x_2} x_2 - \sum_{m=1}^{2} \gamma_{2m} g_{x_2} y_m - \delta_{tx_2} g_{x_2} t \geq 0;$$

$$\frac{\partial \vec{D}_I(.)}{\partial x_2} = \alpha_2 - \alpha_{22} g_{x_2} x_1 + \alpha_{22} x_2 + \sum_{m=1}^{2} \gamma_{2m} y_m + \delta_{tx_2} t \geq 0;$$

$$\frac{\partial \vec{D}_I(.)}{\partial y_m} = \beta_m + \sum_{m'=1}^{2} \beta_{mm'} y_{m'} - \gamma_{2m} g_{x_2} x_1 + \gamma_{2m} x_2 + \delta_{tym} t \leq 0 \quad (m = 1, 2)\,.$$

Curvature restrictions can be imposed by ensuring that every principal minor of the Hessian matrix of odd order (even order) is nonpositive (nonnegative)—see, for example, Morey (1986). However, the US banking industry is highly regulated at both the federal and state levels, and different states change their regulatory restrictions at different times. This implies that the curvature condition would involve a bordered Hessian matrix that accounts for those regulatory restrictions. However, quantifying all those regulatory restrictions in the US banking industry is not an easy task. Furthermore, Barnett (2002) notes that the imposition of global curvature on the quadratic functional form may induce spurious violations of monotonicity. Therefore, directional distance functions are estimated subject to theoretical monotonicity only following Färe et al. (2005).

The unit value, the observed input−output, and the optimal direction models are first estimated without imposing the monotonicity conditions. However, the monotonicity conditions with respect to labor and all outputs are violated for all models at most observations. Thus, to produce inference that is consistent with neoclassical microeconomic theory, all models are re-estimated with the monotonicity conditions imposed at each observation, by following the Bayesian procedure discussed in O'Donnell and Coelli (2005).

6.2. Technical Inefficiency Measures

Observation-specific posterior estimates of technical inefficiency are obtained from the posterior conditional mean of u. The average technical inefficiency measures for each sample year from the three directional vectors regularity constrained models are summarized in Table 3.

Comparing technical inefficiency measures across the unit value, the observed input−output, and the optimal direction models, the prespecified unit value and observed input−output direction models which leave the endogeneity of inputs and outputs unaddressed produce lower estimates of technical inefficiency. This implies that models that ignore the endogeneity problem tend to underestimate bank inefficiency measures. This suggests the importance of managing the endogeneity issue to obtain unbiased and consistent estimates of the parameters of the production technology and the associated measures of inefficiency—see, for example, Atkinson and Primont (2002).

As can be seen in Table 3, the total average of input and output technical inefficiency measures are larger than the average overall technical inefficiency measures in the case of the unit value and the optimal direction models and smaller than the average overall technical inefficiency measures in the case of the observed input−output direction model[3]. These results are true for all years, except for years 2014 and 2015 in the case of the unit value direction model. A possible explanation of this is the problem of averaging inefficiency measures for all banks with different sizes for each sample year and recommends future research in comparing these inefficiencies for banks in small, medium, and large size classes.

Furthermore, output technical inefficiency is on average larger than input technical inefficiency in the three directional vectors models. This finding is consistent with Berger et al. (1993) and English et al. (1993), who find that output inefficiency measures are as large or larger than input inefficiency measures. That is, most of the technical inefficiency in the US banking is in the form of loss of production, rather than overuse of inputs. These results indicate that banks may have more control over inputs than over outputs.

6.3. Results on the Interactive Effects

Now, we explore the effect of the variables included in the interactive effect equation, input and output technical inefficiencies, and the term capturing the interactions between them, on overall technical inefficiency. These technical inefficiency measures are of interest for several reasons. First, input and output inefficiencies measure different concepts and may affect future bank outcomes through different channels in the economy. Second, they include the effects of the overuse of inputs and the loss of production of outputs that occur when banks are inefficient. Third, they include the effects of the interactions between input and output inefficiencies on the efficiency of the banks. For instance, if a certain input (output) is viewed as being relatively overused (underproduced), it is likely that outputs (inputs) that are intensive in using (producing) that input (output) will be overproduced (underused) relative to other outputs (inputs). The interactive effect term captures output (input) implications of any errors in the input usage (output production). Thus, the interactive effect contains more information than could be captured in an input (output)-oriented efficiency study which excludes output (input) effects of input usage (output production) errors.

To focus on the relationships among input, output, and overall technical inefficiencies obtained from the systems of equations, consisting of DTDF and the interactive effect equations without and with the profit-maximizing first-order conditions, the estimated parameters of the DIDF and DODF for the three directional vectors models are not discussed. The estimated parameters of the DTDF, their associated 95% Bayes intervals, and their SIF values from the unit value, the observed input−output, and the optimal directional vector regularity constrained models are summarized in Table 4.

As can be seen in Table 4, UDTDF, VDTDF, and ODTDF models show that both input and output technical inefficiencies have significant positive effects on the overall technical inefficiency. However, the IEIO, δ_{IO}, has a significant negative effect on the overall technical inefficiency. This result is robust to alternative directional vectors and model specifications. Banks with larger values of the IEIO tend to have a lower level of overall technical inefficiency which indicates that they are more efficient. This suggests that the overuse of inputs creates input technical inefficiency and has an effect on reducing (improving) output technical inefficiency (efficiency) and therefore improving overall technical efficiency. Intuitively, the overuse of inputs whether physical inputs involving overuse of labor or overuse of financial inputs involving overpayment of interest may encourage banks to produce more loans to pay salaries for its employees and interest rates on deposits. Similarly, the loss of production of outputs creates output technical inefficiency and has an effect on reducing (improving) input technical inefficiency (efficiency) and therefore improving overall technical efficiency. Intuitively, the loss of revenue due to the loss of production of loans may encourage banks to reduce the number of labor used in the production process or lower the interest rates paid on deposits.

The most clarifying insights come from comparing bank-specific input, output, and overall technical inefficiency measures over the years 2001−2015. Figure 3 show an example of the interactions between input and output technical inefficiencies obtained based on the unit value, the observed input−output, and the optimal directional vector models. It is apparent that the increase in the output technical inefficiency reflects on a reduction on the input technical inefficiency and vice versa. Therefore, the IEIO directly results in a decrease in overall technical inefficiency. That is, a loss of output creates a loss of revenue and has an effect on the use of inputs, while an overuse of input creates additional costs and has an effect on the production of outputs. This helps make our argument that input and output approaches are not an appropriate metric for measuring inefficiency in the presence of the interactive effect. Instead, using models that incorporate both input and output inefficiencies is superior to the standard input or output approach for measuring inefficiency, since it allows to credit input and output while simultaneously crediting the interactive effect.

Differences in the magnitude of the IEIO are observed when using exogenous and endogenous directional vectors. In particular, the IEIO obtained based on the unit value and the observed input−output directional vector models appear to be of low magnitudes. On the other hand, considering banks heterogeneity in the directional vector, the IEIO obtained based on the optimal directional vector model appears to be of higher magnitude. This implies that models that ignore banks heterogeneity can lead to wrong conclusions concerning the estimates obtained from the models. This finding is consistent with Mester (1997) and Greene (2005) who find that

Figure 3. Technical inefficiency measures based on the unit value, the observed input—output, and the optimal directional vectors.

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

Table 2. Estimates of optimal directional parameters

Direction Vector	ODIDF			ODODF			ODTDF		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
$g_{x_{1i}}$	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.9928	0.8514	1.1156
$g_{x_{2i}}$	1.0063	0.8059	1.1996	0.0000	0.0000	0.0000	1.0049	0.8340	1.1441
$g_{y_{1i}}$	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$g_{y_{2i}}$	0.0000	0.0000	0.0000	1.0049	0.5568	1.4208	0.9950	0.8022	1.1661

Table 3. Average technical inefficiency over time based on the unit value, observed input–output, and optimal directional regularity-constrained models

Year	UDIDF	UDODF	UDTDF	VDIDF	VDODF	VDTDF	ODIDF	ODODF	ODTDF
	Unit value direction			Observed input–output direction			Optimal direction		
2001	11.8294	38.0998	5.9640	1.8966	13.0282	4.3780	110.3720	44.5414	72.2921
2002	21.1690	54.6814	16.8900	3.3405	15.6542	16.0843	132.2072	74.2772	90.7688
2003	31.1416	70.2523	31.0910	5.6046	19.2972	30.5473	155.7102	107.1022	116.7214
2004	41.7216	85.5096	48.5342	8.6740	24.4736	47.7446	173.8357	143.4714	148.2288
2005	51.9758	101.4479	69.1852	12.4314	31.0880	67.7554	196.3768	184.6959	189.0018
2006	62.5347	116.3714	93.1282	16.9193	38.8666	90.6375	221.2248	228.6927	237.5864
2007	72.8167	130.8457	120.3147	22.0298	48.0702	116.4475	246.0232	275.8839	293.8951
2008	81.9493	150.5572	150.4222	27.4962	60.5054	145.0767	265.5990	331.7615	357.0892
2009	89.1105	166.3904	184.0269	33.5365	72.8146	176.8909	288.3704	387.7037	428.9286
2010	98.1810	184.6105	220.6380	40.4244	87.7846	211.8751	317.0758	450.7542	509.1689
2011	107.1871	195.5057	260.8635	48.2354	101.5196	250.1111	345.1988	510.5639	597.3874
2012	116.5829	211.9362	304.0676	56.7076	118.1862	291.5732	382.4024	580.8563	694.1378
2013	127.2603	224.2782	350.6689	66.1682	135.8242	336.4630	415.2908	650.9366	798.7750
2014	134.9243	248.7085	399.8923	75.5638	159.1207	384.5791	443.7975	735.9464	909.8987
2015	145.2475	255.3833	453.1177	86.3371	178.3353	436.7667	474.5047	806.1380	1029.2937
Average	79.5754	148.9719	180.5870	33.6910	73.6379	173.7954	277.8659	367.5550	431.5449

heterogeneity causes biased estimates obtained from the SFA. This suggests the importance of managing banks heterogeneity to obtain unbiased estimates of the parameters of the production technology and the associated measures of inefficiency. Table 2 presents the minimum, maximum, and mean of the estimates of the optimal directional parameters. Letting the data select the directional vectors produces estimates of the directional parameters with a range of variation across banks.

6.4. Technical Change

The parameter estimates of technical change δ_t, and δ_{tt} obtained based on the unit value, the observed input–output, and the optimal directional vector models appear to be of high magnitude—See Table 4. Specifically, it indicates significant technological advancement by the US banking industry over the period 2001–2015, which seems realistic given the recent advances in information technologies and its effects on the US banking industry.

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

Table 4. Parameter estimates for the regularity-constrained UDTDF, VDTDF, and ODTDF models

Parameter	UDTDF model			VDTDF model			ODTDF model		
	Estimate	95% Bayes interval	SIF	Estimate	95% Bayes interval	SIF	Estimate	95% Bayes interval	SIF
		The Frontier			The Frontier			The Frontier	
α_0	10.7097	(10.5691, 10.9404)	28.9949	8.3265	(8.3214, 8.3333)	53.1989	10.4904	(10.4588, 10.5220)	19.3941
α_1	0.3172	(0.3119, 0.3212)	26.2283	0.1966	(0.1921, 0.2004)	53.1405	0.2325	(0.2281, 0.2370)	17.0840
α_2	0.2938	(0.2892, 0.2986)	26.6859	0.3417	(0.3354, 0.3457)	53.1031	0.2977	(0.2935, 0.3020)	17.4929
β_2	−0.1710	(−0.1791, −0.1641)	23.4118	−0.1843	(−0.1881, −0.1823)	51.9249	−0.2349	(−0.2384, −0.2315)	17.1590
δ_t	11.3041	(11.1091, 11.5001)	30.0065	7.8266	(7.8234, 7.8304)	52.0048	9.8543	(9.8116, 9.8996)	20.0275
α_{11}	−0.0610	(−0.0643, −0.0580)	12.6863	−0.0348	(−0.0375, −0.0317)	52.3867	−0.0525	(−0.0554, −0.0498)	12.7861
α_{12}	0.0735	(0.0701, 0.0770)	14.2537	0.0693	(0.0633, 0.0770)	53.2463	0.0671	(0.0634, 0.0709)	13.7315
α_{22}	−0.0685	(−0.0711, −0.0645)	13.3931	−0.0640	(−0.0661, −0.0618)	51.5508	−0.0662	(−0.0688,−0.0638)	12.6097
β_{22}	−0.0141	(−0.0451, 0.0238)	5.3655	−0.0023	(−0.0064, 0.0016)	52.9642	−0.0138	(−0.0448, 0.0257)	4.1099
δ_{tt}	3.4889	(3.4595, 3.5139)	29.8515	3.1232	(3.1164, 3.1279)	51.2509	9.0387	(8.9692, 9.1077)	19.1570
γ_{12}	−0.0089	(−0.0346, 0.0245)	5.5158	−0.0390	(−0.0410, −0.0362)	51.3234	−0.0055	(−0.0314, 0.0271)	4.5671
γ_{22}	−0.0196	(−0.0437, 0.0221)	7.3338	0.0001	(−0.0019, 0.0019)	50.6460	−0.0215	(−0.0423, 0.0194)	6.7070
δ_{tx_1}	−0.0026	(−0.0169, 0.0135)	12.0490	0.0181	(0.0149, 0.0211)	51.8347	0.0017	(−0.0122, 0.0147)	7.2960
δ_{tx_2}	0.0026	(−0.0147, 0.0165)	12.4428	−0.0054	(−0.0092, −0.0019)	52.4963	0.0029	(−0.0140, 0.0166)	9.0611
δ_{ty_2}	0.0014	(−0.0116, 0.0143)	11.5602	−0.0090	(−0.0112, −0.0067)	51.2171	0.0049	(−0.0113, 0.0170)	8.8610
		The interactive effect equation			The interactive effect equation			The interactive effect equation	
δ_0	2.0163	(2.0121, 2.0205)	29.2707	2.4684	(2.4665, 2.4704)	49.9582	1.2458	(1.2276, 1.2639)	19.6232
δ_I	5.7224	(5.7002, 5.7444)	29.8449	0.5693	(0.5665, 0.5741)	53.0057	0.9452	(0.9294, 0.9613)	19.9741
δ_O	5.7957	(5.7527, 5.8305)	30.0774	3.5000	(3.4945, 3.5038)	53.1133	1.3628	(1.3350, 1.3905)	20.0692
δ_{IO}	−1.2044	(−1.2258, −1.1833)	30.0231	−0.8709	(−0.8740, −0.8676)	52.4857	−2.5344	(−2.5660, −2.5026)	20.0838

7. Conclusion

This paper derives a new set of results that provide corrective measures of overall technical inefficiency that either have been ignored or wrongly assumed in the literature. Using directional distance functions, we argue that overall technical inefficiency is not only a function of input and output technical inefficiencies as previous studies claim but also of the interaction between them. The derivation of the IEIO solves the arbitrary decomposition of overall technical inefficiency into input and output components. Ignoring the IEIO results in a decomposition of overall technical inefficiency into input and output components that are significantly different from the ones that incorporate it. We also show that the IEIO depends on the choice of the directional vector and whether quantities and prices are taken into consideration.

Using exogenous and endogenous directional vectors, we prove these results theoretically using the relationship between the directional distance functions and both the standard distance functions and their dual representations: cost, revenue, and profit functions. We also provide empirical support of the theoretical results using the US commercial banking data set over the period from 2001 to 2015. Using Bayesian estimation with the monotonicity conditions imposed at each observation, we estimate input and output technical inefficiencies separately using directional input and output distance functions with the three commonly used directional vectors: the unit value, the observed input—output, and the optimal directional vectors. The overall technical inefficiency is estimated using systems of equations to incorporate the interactive effect equation and to address the endogeneity of inputs and outputs. Furthermore, the directional vectors of these models are allowed to be endogenous and vary across banks to account for heterogeneity across banks.

Consistent with the theoretical results, we find significant evidence of the IEIO, where the increase in the output technical inefficiency reflects on a reduction on the input technical inefficiency and vice versa. The empirical results also show that both input and output technical inefficiencies have significant positive effects on the overall technical inefficiency. However, the IEIO has a significant negative effect on the overall technical inefficiency. This result is robust to alternative directional vectors and model specifications. These results are quite significant, since these inefficiency components have different implications for bank performance and may affect future outcomes through different channels, suggesting that the adjustability of both inputs and outputs is required for the improvement of the efficiency of the US commercial banks.

Notes

1 The normalizing constraint of the value of the directional vector is used by Luenberger (1992) in the context of consumer theory.

2 Alternatively, the translation property can be imposed by imposing a set of parameter restrictions that applied to the DTDF directly during the estimation and estimating the restricted version of the DTDF—see, for example, Atkinson and Tsionas (2016).

3 Note that by using the observed input—output direction $g = (-x, y)$, an inefficient bank can decrease inefficiency while decreasing input and increasing output in proportion to the initial combination of the actual input and output.

References

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

Ashcraft, A. (2005) Are banks really special? New evidence from the FDIC-Induced failure of healthy banks. *American Economic Review* **95**(5), 1712–1730.

Assaf, A. G., A. N. Berger, R. A. Roman and M. G. Tsionas (2019) Does efficiency help banks survive and thrive during financial crises? *Journal of Banking and Finance* **106**(1), 445–470.

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., D. Primont, D and M. G. Tsionas (2018) Statistical inference in efficient production with bad inputs and outputs using latent prices and optimal directions. *Journal of Econometrics* **204**(2), 131–146.

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

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

Barros, P. C., S. Managi and R. Matousek (2012) The technical efficiency of the Japanese banks: Non-radial directional performance measurement with undesirable output. *Omega* **40**(1), 1–8.

Barten, A. P. (1969) Maximum likelihood estimation of a complete system of demand equations. *European Economic Review* **1**(1), 7–73.

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.

Berger, A., D. Hancock and D. Humphrey (1993) Bank efficiency derived from the profit function. *Journal of Banking and Finance* **17**(2-3), 317–347.

Berger, A., W. Hunter and S. Timme (1993) The efficiency of financial institutions: A review and preview of research, past, present and future. *Journal of Banking and Finance* **17**(2-3), 221–249.

Berger, A. N. and L. J. Mester (2003) Explaining the dramatic changes in the performance of US banks: Technological change deregulation and dynamic changes in competition. *Journal of Financial Intermediation* **12**(1), 57–95.

Bui, T. N. (2020) How does the relationship between economic growth and bank efficiency? Evidence from Asean countries. *International Transaction Journal of Engineering Management, and Applied Sciences and Technologies* **11(7), 49–58.**

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.

Ege, I. and T. Nur (2020) The relationship between financial performance of banking sector and economic growth: A research on EU countries. *Gümüşhane Üniversitesi Sosyal Bilimler Enstitüsü Elektronik Dergisi* **11**(2), 508–518.

English, M., S. Grosskopf, K. Hayes and S. Yaisawarng (1993) Output allocative and technical efficiency of the financial services sector. *Journal of Banking and Finance* **17**(2-3), 349–366.

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.

Feng, G. and A. Serletis (2010) Efficiency, technical change, and returns to scale in large US banks: Panel data evidence from an output distance function satisfying theoretical regularity. *Journal of Banking and Finance* **34**(1), 127–138.

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.

Fujii, H., S. Managi and R. Matousek (2014) Indian bank efficiency and productivity changes with undesirable outputs: A disaggregated approach. *Journal of Banking and Finance* **38**, 41–50.

Gelfand, A., S. Hills, S. A.Racine-Poon and A. Smith (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972–985.

Githinji-Muriithi, M. (2017) Relationship between bank failure and economic activities: A review of literature. *International Journal of Management, Accounting and Economics* **4**, 1136–1151.

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

Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrica* **57**(1), 97–109.

Havranek, T., Z. Irsova and J. Lesanovska (2016) Bank efficiency and interest rate pass-through: Evidence from Czech loan products. *Economic Modelling* **54**(1), 153–169.

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.

Kang, A., R. Lowery and M. Wardlaw (2015) The costs of closing failed banks: A structural estimation of regulatory incentives. *The Review of Financial Studies* **28**(4), 1060–1102.

Kim, S., N. Shephard and S. Chib (1998) Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies* **65**(3), 361–393.

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. and M. F. Steel (2003) Bayesian analysis of stochastic frontier models. In: Koop, G. and M. F. Steel (eds.), *A Companion to Theoretical Econometrics*, Blackwell, MA, USA, 520–537.

Koutsomanoli-Filippaki, A., D. Margaritis and C. Staikouras (2009) Efficiency and productivity growth in the banking industry of central and eastern Europe. *Journal of Banking and Finance* **33**(3), 557–567.

Lang, L. and R. Stulz (1992) Contagion and competitive intra-industry effects of bankruptcy announcements. *Journal of Financial Economics* **32**(1), 45–60.

Lucchetti, R., L. Papi and A. Zazzaro (2001) Banks' inefficiency and economic growth: A micro-macro approach. *Scottish Journal of Political Economy* **48**(4), 400–424.

Luenberger, D. (1992) Benefit functions and duality. *Journal of Mathematical Economics* **21**(5), 461–481.

Lunn, D., J. Barrett, M. Sweeting and S. Thompson (2013) Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *Journal of the Royal Statistical Society Series C, Applied Statistics* **62**(4), 551–572.

Malikov, E., S. Kumbhakar and M. 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.

Marsh, T. L., A. M. Featherstone and T. A. Garrett (2003) Input inefficiency in commercial banks: A normalized quadratic input distance approach. *Federal Reserve Bank of St. Louis* Working Paper 2003-036A. http://research.stlouisfed.org/wp/2003/2003-036.pdf.

Maudos, J., J. M. Pastor, F. Perez and J. Quesada (2002) Cost and profit efficiency in European banks. *Journal of International Financial Markets, Institutions and Money* **12**(1), 33–58.

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.

Morey, E. R. (1986) An introduction to checking, testing, and imposing curvature properties: The true function and the estimated function. *Canadian Journal of Economics* **19**(2), 207–235.

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.

Park, K. and W. Weber (2006) A note on efficiency and productivity growth in the Korean banking industry, 1992-2002. *Journal of Banking and Finance* **30**(8), 2371–2386.

Rogers, K. E. (1998) Nontraditional activities and the efficiency of U.S. commercial banks. *Journal of Banking and Finance* **22**(4), 467–482.

Sealey, C. W. and J. T. Lindley (1977) Inputs, outputs, and a theory of production and cost at depository financial institutions. *Journal of Finance* **32**(4), 1251–1266.

Stiroh, K. J. (2000) How did bank holding companies prosper in the 1990s? *Journal of Banking and Finance* **24**(11), 1703–1745.

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

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.

Appendix

Proof of Proposition 1 *The sum of input- and output-oriented technical inefficiencies is defined as* $\theta_I + \theta_O$*. Since* $\theta_I = 1 - \left[1/D_I\left(y,x\right)\right]$ *and* $\theta_O = \left[1/D_O\left(x,y\right)\right] - 1$*, then* $\theta_I + \theta_O = \left[1/D_O\left(x,y\right)\right] - \left[1/D_I\left(y,x\right)\right]$*. Since* $D_I\left(y,x\right) \geq 1$*, then* $1/D_I\left(y,x\right) \leq 1$ *and since* $1/D_I\left(y,x\right) = 1 - \theta_I \leq 1$*, then* $0 \leq \theta_I \leq 1$ *which implies that the maximum proportional contraction of input would not exceed the initial input and the resulting input could still produce the output (i.e.,* $x \geq x - \theta_I x$*,* $x \geq 0$ *and* $y\left(x - \theta_I x\right) = y$*). Similarly, since* $D_O\left(x,y\right) \leq 1$*, then* $1/D_O\left(x,y\right) \geq 1$ *and since* $1/D_O\left(x,y\right) = 1 + \theta_O \geq 1$*, then* $\theta_O \geq 0$*. Since* $0 \leq \theta_I \leq 1$ *and* $\theta_O \geq 0$*, then* $\theta_I + \theta_O \geq 0$*.*

The overall technical inefficiency θ_T *is* $0 \leq \theta_T \leq 1$ *to ensure that the inequality* $x \geq x - \theta_T x$ *holds. Furthermore, the DTDF contracts input simultaneously with expanding output while the DIDF contracts input holding output fixed,* θ_T *is less than* θ_I *(i.e.,* $\theta_T < \theta_I$ *and* $y\left(x - \theta_T x\right) > y$*) which implies that more input is needed to produce the expanding output.*

Now, we have $0 \leq \theta_T \leq 1, 0 \leq \theta_I \leq 1, \theta_T < \theta_I, \theta_O \geq 0$*, and* $\theta_I + \theta_O \geq 0$*. As a result,* $\theta_T \leq \theta_I + \theta_O$*. Thus, the inequality can be turned into equality by subtracting a residual term that captures the IEIO,* θ_{IO}*, where* $\theta_T = \theta_I + \theta_O - \theta_{IO}$*, and the IEIO is defined as the gap in the inequality, namely,* $\theta_{IO} = \theta_I + \theta_O - \theta_T$*. The IEIO can be interpreted as the remainder of the overall technical inefficiency after the effects of both input and output technical inefficiencies have been subtracted out.* □

Proof of Proposition 2 *The sum of input- and output-oriented technical inefficiencies is* $\theta_I^1 + \theta_O^1$*. Since* $\theta_I^1 = \|x\| - \left\|x^I\right\|$*, then adding and subtracting* $\left\|x^T\right\|$ *yields* $\theta_I^1 = \left(\|x\| - \left\|x^T\right\|\right) + \left(\left\|x^T\right\| - \left\|x^I\right\|\right)$*. Since* $\theta_O^1 = \left\|y^O\right\| - \left\|y\right\|$*, then adding and subtracting* $\left\|y^T\right\|$ *yields* $\theta_O^1 = \left(\left\|y^O\right\| - \left\|y^T\right\|\right) + \left(\left\|y^T\right\| - \left\|y\right\|\right)$*. Thus,* $\left(\|x\| - \left\|x^T\right\|\right) + \left(\left\|x^T\right\| - \left\|x^I\right\|\right) + \left(\left\|y^O\right\| - \left\|y^T\right\|\right) + \left(\left\|y^T\right\| - \left\|y\right\|\right) = \theta_I^1 + \theta_O^1$*, and* $\theta_T^1 = \sqrt{\left(\|x\| - \left\|x^T\right\|\right)^2 + \left(\left\|y^T\right\| - \left\|y\right\|\right)^2}/\sqrt{2}$*. Since each point on*

https://doi.org/10.1017/S1365100522000153 Published online by Cambridge University Press

the 45-degree line equates the variable measured on the vertical axis with the variable measured on the horizontal axis, then $\left(\|x\| - \|x^T\|\right) = \left(\|y^T\| - \|y\|\right) = k$*. Substituting* $\left(\|x\| - \|x^T\|\right) = \left(\|y^T\| - \|y\|\right) = k$*, then* $\theta_T^1 = \sqrt{2k^2}/\sqrt{2} = k$*, and* $\theta_I^1 + \theta_O^1 = 2k + j$ *where* $j = \left(\|x^T\| - \|x^I\|\right) + \left(\|y^O\| - \|y^T\|\right)$*. As a result,* $\theta_T^1 \leq \theta_I^1 + \theta_O^1$*. Thus, the inequality can be turned into equality by subtracting a residual term that captures the IEIO,* θ_{IO}^1*, where* $\theta_T^1 = \theta_I^1 + \theta_O^1 - \theta_{IO}^1$ *and* $\theta_{IO}^1 = k + j$ *is the IEIO.* □

Proof of Proposition 3 *For every* $(x, y) \in T$*, the projected vector* (x^π, y^π) *based on the directional vector* $g = \left(g_x^\pi, g_y^\pi\right)$ *is* $\left(x - \theta_T^\pi g_x^\pi, y + \theta_T^\pi g_y^\pi\right) \in T$ *or, equivalently,* $(x, y) + \theta_T^\pi \left(g_x^\pi, g_y^\pi\right) \in T$, where $\theta_T^\pi = \vec{D}_T\left(x, y; g_x^\pi, g_y^\pi\right)$. Thus, $(py^\pi - wx^\pi) = (py - wx) + \theta_T^\pi \left(pg_y^\pi + wg_x^\pi\right)$. *Using the directional vector given in (5) yields, after some rearranging,* $\theta_T^\pi = \pi(p, w) - (py - wx)$. □

Proof of Proposition 4 *The sum of input- and output-oriented technical inefficiency can be defined as* $\theta_I^C + \theta_O^R$*. Since* $\theta_I^C = wx - C(y, w)$ *and* $\theta_O^R = R(x, p) - py$*, then* $\theta_I^C + \theta_O^R = R(x, p) - C(y, w) - (py - wx)$*. Since* $\theta_T^\pi = \pi(p, w) - (py - wx)$ *and* $\pi(p, w) \gtreqless C(y, w)$*, then* $\theta_T^\pi \gtreqless \theta_I^C + \theta_O^R$*. Thus, the inequality can be turned into equality by adding or subtracting a residual term that captures the IEIO,* θ_{IO}^{CR}*, where* $\theta_T^\pi = \theta_I^C + \theta_O^R \pm \theta_{IO}^{CR}$ *and the IEIO is defined as the gap in the inequality, namely,* $\theta_{IO}^{CR} = \theta_T^\pi \mp \left(\theta_I^C + \theta_O^R\right)$*.* □

Cite this article: Esheba M and Serletis A (2023). "Interactive effects between input and output technical inefficiencies." *Macroeconomic Dynamics* **27**, 1138–1165. https://doi.org/10.1017/S1365100522000153