A Proportional Scale-Invariant Measurement of Heterogeneous Directional Distances to the Technological Frontier*

Emir Malikov[1] Mike G. Tsionas[2] Diego A. Restrepo Tobón[3]

[1]Auburn University, Auburn, AL, United States
[2]Lancaster University Management School, Lancaster, UK
[3]EAFIT University, Medellín, Colombia

November 1, 2017

Abstract

We consider a stochastic multiplicative-directional-distance-function-based formulation of the production technology in the presence of bad outputs. In contrast to the popular additive directional distance function, our approach preserves the highly desired scale invariance property of the proportional radial distance function and can easily assume popular translog specification. Our model can also meaningfully accommodate technological heterogeneity across individual firms by letting the direction in which the distance to frontier is measured be an unknown function of the firm's idiosyncratic characteristics reflective of the heterogeneous economic environment in which it operates. To mitigate the uncertainty associated with an arbitrary choice of the direction, we rely on a data-driven selection method to determine these varying heterogeneous directional vectors for individual firms. We showcase practical advantages of the proposed model by estimating the production technology of U.S. commercial banks in the presence of undesirable non-performing loans during the 2001–2010 period.

Keywords: Bad Outputs, Commercial Banks, Efficiency, MCMC, Multiplicative Directional Distance Function, Nonperforming Loans, Productivity

JEL Classification: C11, D24, G21

*Email: emalikov@auburn.edu (Malikov), m.tsionas@lancaster.ac.uk (Tsionas), drestr16@eafit.edu.co (Restrepo).

1 Introduction

Modeling polluting production technologies in a way that explicitly accounts for the social and economic undesirability of the so-called "bad" outputs by-produced along with the firm's intended desirable outputs has led to the emergence of vast literature discussing the appropriate methodology for accommodating such bad outputs in the analysis of production technologies [for an excellent review, see Dakpo, Jeanneaux & Latruffe (2016)]. Popular approaches to the formulation of production processes in the presence of bad outputs include (i) additive directional distance functions that allow additive expansion of good and contraction of bad outputs in a pre-specified direction (e.g., Chung, Färe & Grosskopf, 1997; Färe, Grosskopf, Noh & Weber, 2005), (ii) "enhanced" hyperbolic distance functions that allow equiproportional expansion of good and contraction of bad outputs (e.g., Färe, Grosskopf, Lovell & Pasurka, 1989; Cuesta, Lovell & Zofio, 2009) and (iii) by-production systems of separable technologies for desirable production and undesirable pollution generation (e.g., Fernández, Koop & Steel, 2002, 2005; Murty, Russell & Levkoff, 2012; Malikov, Kumbhakar & Tsionas, 2015a; Kumbhakar & Tsionas, 2015). Among the three, the additive directional distance function has arguably become the most popular formulation of polluting production processes among practitioners (for an excellent survey, see Färe, Grosskopf & Margaritis, 2015).

However, despite the wide popularity of (additive) directional distance functions, it is still not uncommon to encounter studies employing the conventional radial distance function (Shephard, 1953, 1970), which measures equiproportional changes in inputs or desirable outputs, to formulate polluting technologies. In such studies, the conventional radial distance function is usually augmented to incorporate bad outputs in two ways: the function is either conditioned on the quantity of undesirable outputs treated as theoretically unregulated technology shifters (e.g., Atkinson & Dorfman, 2005; Assaf, Matousek & Tsionas, 2013) or expanded to effectively incorporate bad outputs in the role of inputs (e.g., Reinhard, Lovell & Thijssen, 1999, 2000; Hailu & Veeman, 2000, 2001). Arguably, these studies opt for the radial distance function in spite of its well-known criticisms (Färe & Grosskopf, 2003; Färe et al., 2005) because, unlike its alternative *additive* directional counterpart, the radial function provides a *proportional*, and hence scale-invariant, measurement of the distance to the technological frontier as well as can easily be made to take on, perhaps the most popular, translog parametric specification.

In this paper, we contribute to the literature at least on two fronts. First, we consider a "compromise" distance-function-based formulation of the polluting technology which can meaningfully accommodate bad outputs in the spirit of the additive directional distance function while still preserving the highly desired scale invariance property of the proportional radial distance function along with its ability to assume flexible functional forms. Specifically, we generalize the *multiplicative* directional technology distance function (MDTDF) of Peyrache & Coelli (2009) to the production environment with undesirable by-product outputs.[1] Like the more standard *additive* directional technology distance function (ADTDF) (Chambers, Chung & Färe, 1998; Malikov, Kumbhakar & Tsionas, 2015b), the MDTDF that we consider in this paper also seeks the simultaneous maximal expansion in good outputs and maximal contraction in inputs and bad outputs. The latter allows us to credit firms for the reduction in undesirables along with the expansion in desirables when computing their productivity and technological efficiency. However, the MDTDF has several important advantages thereby rendering it a more attractive formulation of the technology. It is advantageous over traditional (input- and/or output-oriented) radial distance functions, which measure

[1] Peyrache & Coelli (2009) assume all outputs are desirable. Mehdiloozad, Sahoo & Roshdi (2014) provide a different generalization of the MDTDF (without bads) that allows for slacks. For a recently proposed alternative multiplicative measure of efficiency for deterministic (modeled via data envelopment analysis) polluting technologies, also see Valadkhani, Roshdi & Smyth (2016).

*equi*proportional changes, because the proportional distance that MDTDF measures is technology-oriented and, more importantly, can vary across individual inputs and outputs as regulated by the argument-specific directions. Unlike for the ADTDF, this advantage however comes at *no* cost. Specifically, owing to its multiplicative nature, the MDTDF constitutes a *proportional* measure of efficiency in a given direction thus guaranteeing the scale invariance of the MDTDF-based measurement of technological efficiency, which the more popular ADTDF however cannot deliver. Furthermore, owing to its proportionality and the almost homogeneity property, the MDTDF presents the practitioners interested in the econometric estimation of polluting production technologies with the opportunity to adopt popular flexible parametric specifications (such as translog or Fourier) to approximate the unknown technology. The latter may be especially valuable given that most attempts to parametrically estimate a stochastic additive directional distance function in the literature have been limited to a somewhat less popular quadratic specification (in levels) primarily due to the ease with which the ADTDF's translation property can be imposed onto it (e.g., Färe et al., 2005; Feng & Serletis, 2014; Atkinson & Tsionas, 2015).

Our second contribution lies in extending the stochastic MDTDF model to meaningfully accommodate the potential technological heterogeneity across individual firms. We accomplish this by allowing the directional vector along which the distance toward technological frontier is being quantified to be an unknown function of firms' idiosyncratic characteristics. These variables are intended to contextualize both the external and internal economic environment in which firms operate and thereby help us identify firm-specific "target" points on the technological frontier relative to which firms' performance is to be measured. Thus, we rely on the data to determine the varying heterogeneous directional vector for each firm. In doing so, we are able to circumvent a well-known shortcoming of directional distance functions which produce technological measurements that are subject to uncertainty associated with the choice of the directional vector. Our approach to modeling data-driven directions also nests that of Tsionas, Malikov & Kumbhakar (2015) and Atkinson & Tsionas (2015) who also estimate directions (albeit, for an additive distance function) simultaneously with the technology. However, unlike these previous studies, we do not restrict the unknown directional vector to be fixed and thus common to all firms but rather allow it to vary across firms. Our approach therefore allows for a greater degree of flexibility.

To highlight practical usefulness of the heterogeneous MDTDF along with its potentially significant advantages over its alternatives, we apply our methodology to study the production technology of U.S. commercial banks in 2001–2010. By employing the directional technology distance function, we are able to explicitly accommodate the presence of non-performing loans, i.e., loans that are not paid back duly as obligated contractually, in the bank's balance sheet. The latter is an inherent characteristic of the bank's production due to its exposure to credit risk which materializes in the form of non-performing loans. Researchers have long ago acknowledged the importance of taking non-performing loans into account when estimating banking technology and growingly favor treating non-performing loans as an undesirable output (e.g., Park & Weber, 2006; Assaf et al., 2013; Guarda, Rouabah & Vardanyan, 2013; Malikov et al., 2015b) because not only does such an approach recognize that non-performing loans are a by-product of producing desirable outputs, such as earning loans and securities, but it also accommodates the undesirability of non-performing loans. That is, by acknowledging that nonperforming loans are undesirable, we can credit banks for the reduction in them along with the expansion in desirable assets. Furthermore, by employing a model of varying context-dependent distances to the technological frontier we can successfully capture the well-documented heterogeneity in banks' business strategies along various dimensions including the scope and competitiveness of the market, banks' size and product mix as well as risk taking. That is, banks are naturally expected to target different points on the frontier in order to fulfill their heterogeneous business strategies in pursuit of the maximal franchise value.

To empirically assess the sensitivity of various technological metrics to the choice of the directional distance function and the treatment of the directional vector along which the latter function is estimated, in addition to the proposed heterogeneous MDTDF model we also estimate its several popular alternatives. The knowledge of the degree of such sensitivity is pivotal from the policy perspective especially when the estimates of banking production technology are taken as the basis for the evaluation of the existing and/or devising of the new regulations for the industry. All models are estimated via MCMC methods subject to theoretical regularity conditions imposed at each observation. We document non-negligible distortions in the measurement of banks' technological efficiency, productivity growth and scale elasticity when the estimated banking technology is constrained to be common to all banks under a rather unrealistic assumption of the (known or unknown) directional vector towards the bank's production frontier being fixed. Thus, the model that accommodates the well-documented heterogeneity among banks by permitting the directional vector to be bank-specific by varying with banks' idiosyncratic characteristics thus capturing their heterogeneous business strategies is likely to provide a more robust estimation strategy. The data also seem to support the proportional measurement of technological inefficiency over the widely used additive alternative.

The rest of the paper proceeds as follows. Section 2 introduces the MDTDF along with its properties. We describe the details of our estimation strategy in Section 3. Section 4 outlines the conceptual framework for the analysis of the banking technology and describes the data used in the paper. The empirical results are reported in Section 5. Section 6 concludes.

2 *Multiplicative* Directional Distance to the Technological Frontier in the Presence of Bad Outputs

Consider the production process in which J inputs $\mathbf{x} \in \Re^J_+$ are being transformed into M desirable ("good") outputs $\mathbf{y} \in \Re^M_+$ and P undesirable ("bad") outputs $\mathbf{b} \in \Re^P_+$. The firm's production technology is given by

$$\mathbb{T} \overset{\text{def}}{=} \{(\mathbf{x}, \mathbf{y}, \mathbf{b}) : \ \mathbf{x} \text{ can produce } (\mathbf{y}, \mathbf{b})\}, \tag{2.1}$$

subject to usual axioms (see, e.g., Färe et al., 2005):

(1) closedness of $\mathbb{T}$;

(2) no free lunch: if $(\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$ and $\mathbf{x} = \mathbf{0}$, then $(\mathbf{y}, \mathbf{b}) = (\mathbf{0}, \mathbf{0})$;

(3) null-jointness of the output production: if $(\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$ and $\mathbf{b} = \mathbf{0}$, then $\mathbf{y} = \mathbf{0}$;

(4) free input disposability: if $(\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$ and $\mathbf{x}' \geq \mathbf{x}$, then $(\mathbf{x}', \mathbf{y}, \mathbf{b}) \in \mathbb{T}$;

(5) weak joint disposability of desirable and undesirable outputs: if $(\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$ and $0 \leq \kappa \leq 1$, then $(\mathbf{x}, \kappa\mathbf{y}, \kappa\mathbf{b}) \in \mathbb{T}$;

(6) free disposability of desirable outputs: if $(\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$ and $\mathbf{y}' \leq \mathbf{y}$, then $(\mathbf{x}, \mathbf{y}', \mathbf{b}) \in \mathbb{T}$;

(7) feasibility of inaction: $(\mathbf{0}, \mathbf{0}, \mathbf{0}) \in \mathbb{T}$;

(8) convexity of $\mathbb{T}$.

The analysis of $\mathbb{T}$ necessitates an adoption of its "functional" representation for which the predominant choice in the literature is an *additive* directional distance function by Chambers, Chung & Färe (1996), Chambers et al. (1998) and Chung et al. (1997). Building on Atkinson &

Tsionas (2015) and Malikov et al. (2015b), we can define the additive directional technology distance function (ADTDF) as the function measuring the maximal distance between the observed $(\mathbf{x},\mathbf{y},\mathbf{b})$ and the boundary of production technology $\mathbb{T}$ in a given direction $\boldsymbol{d} \equiv (\boldsymbol{d}_x, \boldsymbol{d}_y, \boldsymbol{d}_b) \in \Re^J_+ \times \Re^M_+ \times \Re^P_+$, i.e.,

$$\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) \stackrel{\text{def}}{=} \sup_{\beta} \left\{\beta \in \Re : \ (\mathbf{x} - \beta\boldsymbol{d}_x, \mathbf{y} + \beta\boldsymbol{d}_y, \mathbf{b} - \beta\boldsymbol{d}_b) \in \mathbb{T}\right\}. \tag{2.2}$$

The above ADTDF seeks the simultaneous maximal expansion in desirable outputs and maximal reduction in inputs and undesirable outputs and is a fusion of the additive directional output (Chung et al., 1997) and technology (Chambers et al., 1998; Hudgins & Primont, 2007) distance functions. It is advantageous over traditional radial distance functions by Shephard (1953, 1970) because the distance that it measures is technology-oriented and, more importantly, can vary across individual inputs and outputs as regulated by the argument-specific directions. However, this advantage comes at quite a significant cost. Specifically, unlike the traditional radial distance function, the ADTDF constitutes an additive, i.e., *not* proportional, measure of inefficiency in a given direction $\boldsymbol{d}$. The latter leads to two important implications for applied research. First, the ADTDF-based measurements of technological efficiency are *not* scale-invariant (i.e., they lack commensurability).[2] Second, virtually all attempts to parametrically estimate stochastic additive directional distance functions in the literature have been limited to a quadratic specification over the more popular translog due to the ease with which the (additive) "translation property" of the ADTDF can be imposed onto the former (e.g., Färe et al., 2005; Feng & Serletis, 2014). The multiplicative functional representation of the polluting technology $\mathbb{T}$, which we opt for in this paper, however does not suffer from the above two shortcomings.

In what follows, we introduce the *multiplicative* directional technology distance function (MDTDF) which generalizes the recent work by Peyrache & Coelli (2009) to the environment with undesirable outputs. Specifically, the maximal distance between the observed input-output vector $(\mathbf{x},\mathbf{y},\mathbf{b})$ and the firm's technological frontier in a given direction $\boldsymbol{g} \equiv \left(\boldsymbol{g}_x, \boldsymbol{g}_y, \boldsymbol{g}_b\right) \in \Re^J_+ \times \Re^M_+ \times \Re^P_+$ is said to be given by the value of the multiplicative directional technology distance function defined as

$$\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{g}) \stackrel{\text{def}}{=} \inf_{\lambda} \left\{\lambda \in \Re_+ : \ \left(\lambda^{\mathrm{diag}\{\boldsymbol{g}_x\}}\mathbf{x}, \lambda^{-\mathrm{diag}\{\boldsymbol{g}_y\}}\mathbf{y}, \lambda^{\mathrm{diag}\{\boldsymbol{g}_b\}}\mathbf{b}\right) \in \mathbb{T}\right\}. \tag{2.3}$$

Like the ADTDF in (2.2), the MDTDF above seeks the simultaneous maximal expansion in good outputs and maximal reduction in inputs and bad outputs. It is also advantageous over traditional (input- or output-oriented) radial distance functions because the distance that MDTDF measures is technology-oriented and can vary across individual inputs and outputs. However, unlike in the case of the ADTDF, this advantage does not come at a cost. Specifically, owing to its multiplicative nature, the MDTDF in (2.3) constitutes a *proportional* measure of efficiency in a given direction $\boldsymbol{g}$, thus guaranteeing the scale invariance of the MDTDF-based measurement of technological efficiency, which the more popular ADTDF is however unable to deliver.

The function $\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{g})$ satisfies the following theoretical properties, which all directly follow from the axioms describing technology $\mathbb{T}$ in (2.1):[3]

(1) $\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{g}) \in (0,1] \iff (\mathbf{x},\mathbf{y},\mathbf{b}) \in \mathbb{T}$, where the unit value of $\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{g})$ implies full technological efficiency; and $\vec{D}_{\mathrm{M}}(\mathbf{x},\kappa\mathbf{y},\kappa\mathbf{b};\boldsymbol{d}) \in (0,1]$ for $(\mathbf{x},\mathbf{y},\mathbf{b}) \in \mathbb{T}$ and $0 \leq \kappa \leq 1$;

[2]As emphasized by Peyrache & Coelli (2009, p.2) with the reference to Salnykov & Zelenyuk (2005), "scale invariance is a very desirable property, since it avoids the danger that different researchers (using the same methodology and datasets) may obtain different results because they happen to define input and/or output measures using different units".

[3]The proofs directly follow from those in Peyrache & Coelli (2009).

(2) almost homogeneity of degrees $\left(-\boldsymbol{g}_x, \boldsymbol{g}_y, -\boldsymbol{g}_b, 1\right)$ in $(\mathbf{x}, \mathbf{y}, \mathbf{b})$, i.e., for any $\gamma > 0$, we have

$$\vec{D}_{\mathrm{M}}\left(\gamma^{-\mathrm{diag}\{\boldsymbol{g}_x\}}\mathbf{x}, \gamma^{\mathrm{diag}\{\boldsymbol{g}_y\}}\mathbf{y}, \gamma^{-\mathrm{diag}\{\boldsymbol{g}_b\}}\mathbf{b}; \boldsymbol{g}\right) = \gamma \vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g}); \tag{2.4}$$

(3) homogeneity of degree -1 in $\boldsymbol{g}$: $\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \gamma\boldsymbol{g}) = \gamma^{-1}\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g})$ for any $\gamma > 0$;

(4) convexity of $\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g})$ in $(\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$;

(5) negative monotonicity in inputs: if $(\mathbf{x}', \mathbf{y}, \mathbf{b}) \leq (\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$, then $\vec{D}_{\mathrm{M}}(\mathbf{x}', \mathbf{y}, \mathbf{b}; \boldsymbol{g}) \geq \vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g})$;

(6) positive monotonicity in good outputs: if $(\mathbf{x}, \mathbf{y}', \mathbf{b}) \geq (\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$, then $\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}', \mathbf{b}; \boldsymbol{g}) \geq \vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g})$;

(7) negative monotonicity in bad outputs: if $(\mathbf{x}, \mathbf{y}, \mathbf{b}') \leq (\mathbf{x}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}$, then $\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}'; \boldsymbol{g}) \geq \vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g})$.

The properties of the MDTDF are quite similar to those of the ADTDF listed in Supplementary Appendix A. For instance, the MDTDF's monotonicity and curvature regularities are mere pairwise "counter-reflections" of those corresponding to the ADTDF, owing to the fact the MDTDF measures technological efficiency whereas the ADTDF gauges technological *in*efficiency. Among the theoretical properties, the almost homogeneity of the MDTDF in (2.4), which is a logical counterpart of the ADTDF's translation property, is probably the most useful for the purposes of stochastic estimation. Specifically, unlike the additive translation property of the ADTDF, the almost homogeneity property of MDTDF has a multiplicative formulation thereby enabling a researcher to employ popular flexible specifications in logs (such as the translog or Fourier) for $\vec{D}_{\mathrm{M}}(\cdot)$.

Note that, in the case of no by-production of bad outputs ($\mathbf{b} = \mathbf{0}$ and hence $\boldsymbol{g}_b = \mathbf{0}$), the MDTDF in (2.3) nests several special cases of traditional distance functions (also see Peyrache & Coelli, 2009). Specifically, by setting the directional vector $(\boldsymbol{g}_x, \boldsymbol{g}_y, \boldsymbol{g}_b) = (\mathbf{1}, \mathbf{0}, \mathbf{0})$, the MDTDF collapses to the inverse of the conventional (radial) input distance function, i.e., $\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{0}; (\mathbf{1}, \mathbf{0}, \mathbf{0})) = 1/D_i(\mathbf{x}, \mathbf{y})$. Similarly, when $(\boldsymbol{g}_x, \boldsymbol{g}_y, \boldsymbol{g}_b) = (\mathbf{0}, \mathbf{1}, \mathbf{0})$, the MDTDF becomes the radial output distance function: $\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{0}; (\mathbf{0}, \mathbf{1}, \mathbf{0})) = D_o(\mathbf{x}, \mathbf{y})$.

The MDTDF in (2.3) also nests hyperbolic distance functions. In a popular setup with no bad outputs, setting the directional vector $(\boldsymbol{g}_x, \boldsymbol{g}_y, \boldsymbol{g}_b) = (\mathbf{1}, \mathbf{1}, \mathbf{0})$ yields Färe, Grosskopf & Lovell's (1985) hyperbolic distance function:

$$\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{0}; (\mathbf{1}, \mathbf{1}, \mathbf{0})) = \inf_{\lambda}\left\{\lambda \in \Re_+ :\ \left(\lambda\mathbf{x}, \lambda^{-1}\mathbf{y}\right) \in \mathcal{T}\right\} = D_h(\mathbf{x}, \mathbf{y}), \tag{2.5}$$

where $\mathcal{T} \stackrel{\mathrm{def}}{=} \{(\mathbf{x}, \mathbf{y}) :\ \mathbf{x} \text{ can produce } \mathbf{y}\}$.

Alternatively, we may choose to allow for the presence of bad outputs and set $(\boldsymbol{g}_x, \boldsymbol{g}_y, \boldsymbol{g}_b) = (\mathbf{0}, \mathbf{1}, \mathbf{1})$. By doing so, we can obtain the "enhanced" hyperbolic distance function studied by Färe et al. (1989):

$$\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; (\mathbf{0}, \mathbf{1}, \mathbf{1})) = \inf_{\lambda}\left\{\lambda \in \Re_+ :\ \left(\lambda^{-1}\mathbf{y}, \lambda\mathbf{b}\right) \in \mathbb{P}(\mathbf{x})\right\} = D_h(\mathbf{x}, \mathbf{y}, \mathbf{b}), \tag{2.6}$$

where the output set $\mathbb{P}(\mathbf{x})$ is defined as $\mathbb{P}(\mathbf{x}) \stackrel{\mathrm{def}}{=} \{(\mathbf{y}, \mathbf{b}) :\ \mathbf{x} \text{ can produce } (\mathbf{y}, \mathbf{b})\}$.

Since the MDTDF in (2.3) explicitly accommodates the by-production of undesirable outputs, we are able to credit firms for the reduction in undesirable outputs along with the expansion in desirable outputs when computing their productivity growth. Specifically, letting the time enter the MDTDF directly and recognizing that $(\mathbf{x},\mathbf{y},\mathbf{b})$ are all changing over time, the total time-differentiation of $\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b},t;\boldsymbol{g})$ yields

$$\frac{\mathrm{d}\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b},t;\boldsymbol{g})}{\mathrm{d}t}=\sum_j\frac{\partial\vec{D}_{\mathrm{M}}(\cdot)}{\partial\log x_j}\dot{x_j}+\sum_m\frac{\partial\vec{D}_{\mathrm{M}}(\cdot)}{\partial\log y_m}\dot{y_m}+\sum_p\frac{\partial\vec{D}_{\mathrm{M}}(\cdot)}{\partial\log b_p}\dot{b_p}+\frac{\partial\vec{D}_{\mathrm{M}}(\cdot)}{\partial t},\tag{2.7}$$

dividing both sides of which by $\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b},t;\boldsymbol{g})$ yields

$$\begin{aligned}\mathrm{PG_M}&\stackrel{\text{def}}{=}\sum_j\frac{\partial\log\vec{D}_{\mathrm{M}}(\cdot)}{\partial\log x_j}\dot{x_j}+\sum_m\frac{\partial\log\vec{D}_{\mathrm{M}}(\cdot)}{\partial\log y_m}\dot{y_m}+\sum_p\frac{\partial\log\vec{D}_{\mathrm{M}}(\cdot)}{\partial\log b_p}\dot{b_p}\\&=-\frac{\partial\log\vec{D}_{\mathrm{M}}(\cdot)}{\partial t}+\frac{\mathrm{d}\log\vec{D}_{\mathrm{M}}(\cdot)}{\mathrm{d}t},\end{aligned}\tag{2.8}$$

where the left-hand side of the equality is a Divisia index of productivity growth $\mathrm{PG_M}$, and the right-hand side of the equality provides a decomposition of $\mathrm{PG_M}$ into technical change $\mathrm{TC_M}\equiv-\partial\log\vec{D}_{\mathrm{M}}(\cdot)/\partial t$ and efficiency change $\mathrm{EC_M}\equiv\mathrm{d}\log\vec{D}_{\mathrm{M}}(\cdot)/\mathrm{d}t$. The $\mathrm{PG_M}$ index is an average of percentage growth rates in inputs and good and bad outputs computed using the following weights: $\partial\log\vec{D}_{\mathrm{M}}(\cdot)/\partial\log x_j\le 0\ \forall\ j=1,\dots,J$, $\partial\log\vec{D}_{\mathrm{M}}(\cdot)/\partial\log y_m\ge 0\ \forall\ m=1,\dots,M$ and $\partial\log\vec{D}_{\mathrm{M}}(\cdot)/\partial\log b_p\le 0\ \forall\ p=1,\dots,P$, where the signs follow the monotonicity properties of the MDTDF and are in line with one's intuition.

3 Estimation of Stochastic Heterogeneous Directional Distances

In this paper, we consider a *stochastic* production process. Before we proceed, note that the direction in which any directional distance function (be it additive or multiplicative) is specified is uniquely defined by $(J+M+P-1)$ ratios of the elements in the $(J+M+P)$-dimensional directional vector. This is essentially implied by the homogeneity of the distance function in the directional vector. For instance, consider the case of one input ($J=1$), one good ($M=1$) and zero bad outputs ($P=0$). The directional vector for the MDTDF is then uniquely defined by the ratio of g_x to g_y. The intuition here is as follows: the direction is uniquely defined by $(J+M+P-1)$ *angles* between the vector $\boldsymbol{g}$ and its $(J+M+P-1)$ projections on all two-dimensional subspaces. The magnitudes of elements in $\boldsymbol{g}$ *per se* do not matter. Without the loss of generality, we can therefore normalize one of the elements in the directional vector to be equal to one and control the direction of the distance function using (magnitudes of) its remaining elements.[4]

We normalize the stochastic MDTDF in (2.3) in order to impose the almost homogeneity property onto it for a given direction $\boldsymbol{g}$ along the lines of the procedure used by Cuesta & Zofío (2005) for a hyperbolic distance function.[5] More specifically, we set $\gamma=b_p$ in (2.4) and normalize the corresponding direction $g_{b_p}=1$ (in the light of the argument above). The almost homogeneity property (2.4) can then be written in logs as follows:

$$\ln\vec{D}_{\mathrm{M}}\left(b_p^{-\mathrm{diag}\{\boldsymbol{g}_x\}}\mathbf{x},b_p^{\mathrm{diag}\{\boldsymbol{g}_y\}}\mathbf{y},b_p^{-\mathrm{diag}\{\widehat{\boldsymbol{g}}_b\}}\widehat{\mathbf{b}};\boldsymbol{g}\right)=\ln b_p+\ln\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{g});\tag{3.1}$$

where $\widehat{\mathbf{b}}\stackrel{\text{def}}{=}(b_1,\dots,b_{p-1},b_{p+1},\dots,b_P)$ and $\widehat{\boldsymbol{g}}_b\stackrel{\text{def}}{=}(g_{b_1},...,g_{b_{p-1}},g_{b_{p+1}},\dots,g_{b_P})$. Note that, due to the normalization $g_{b_p}=1$, the bad output b_p now enters the left-hand side of (3.1) in the capacity of "γ"

[4]Such a normalization is essentially equivalent to the imposition of the $\boldsymbol{g}$-homogeneity property onto the distance function via dividing the directional vector by one of its elements.

[5]Note that the property may alternatively be imposed using the set of parameter restrictions applied to the MDTDF directly during the estimation.

from (2.4) only. The "output" b_p has however disappeared from $\vec{D}_{\mathrm{M}}\left(b_p^{-\mathrm{diag}\{\boldsymbol{g}_x\}}\mathbf{x}, b_p^{\mathrm{diag}\{\boldsymbol{g}_y\}}\mathbf{y}, b_p^{-\mathrm{diag}\{\widehat{\boldsymbol{g}}_b\}}\widehat{\mathbf{b}}; \boldsymbol{g}\right)$ because $b_p^{-g_{b_p}} b_p = b_p^{-1} b_p = 1$ which vanishes after taking logs. After rearranging, from (3.1) we get

$$\begin{aligned}\ln b_p &= \ln \vec{D}_{\mathrm{M}}\left(b_p^{-\mathrm{diag}\{\boldsymbol{g}_x\}}\mathbf{x}, b_p^{\mathrm{diag}\{\boldsymbol{g}_y\}}\mathbf{y}, b_p^{-\mathrm{diag}\{\widehat{\boldsymbol{g}}_b\}}\widehat{\mathbf{b}}; \boldsymbol{g}\right) - \ln \vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g}) \\ &\overset{\mathrm{def}}{=} \ln \vec{D}_{\mathrm{M}}\left(b_p^{-\mathrm{diag}\{\boldsymbol{g}_x\}}\mathbf{x}, b_p^{\mathrm{diag}\{\boldsymbol{g}_y\}}\mathbf{y}, b_p^{-\mathrm{diag}\{\widehat{\boldsymbol{g}}_b\}}\widehat{\mathbf{b}}; \boldsymbol{g}\right) + u_{\mathrm{M}}, \end{aligned} \tag{3.2}$$

where $u_{\mathrm{M}} \overset{\mathrm{def}}{=} -\ln \vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g}) \geq 0$ which can be linked to (i) the good-output *in*efficiency defined as $[1/\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g}) - 1]$ via the following identity: $1/\vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g}) - 1 = \exp\{u_{\mathrm{M}}\}$ and (ii) the input/bad-output *in*efficiency defined as $[1 - \vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g})]$ via $1 - \vec{D}_{\mathrm{M}}(\mathbf{x}, \mathbf{y}, \mathbf{b}; \boldsymbol{g}) = 1 - \exp\{-u_{\mathrm{M}}\}$. The normalized MDTDF can then be estimated as a stochastic frontier model after appending it with the white noise term. Due to its multiplicative nature, the choice set of functional forms for the MDTDF includes many popular flexible specifications, including the translog and Fourier forms. In this paper, we opt for the translog:

$$\begin{aligned}\ln b_p = \theta_0 &+ \sum_j \alpha_j \ln \breve{x}_j + \sum_m \beta_m \ln \breve{y}_m + \sum_{k(\neq p)} \gamma_k \ln \breve{b}_k + \theta_t t + \\ &\frac{1}{2}\sum_j\sum_{j'} \alpha_{jj'} \ln \breve{x}_j \ln \breve{x}_{j'} + \frac{1}{2}\sum_m\sum_{m'} \beta_{mm'} \ln \breve{y}_m \ln \breve{y}_{m'} + \frac{1}{2}\sum_{k(\neq p)}\sum_{k'(\neq p)} \gamma_{kk'} \ln \breve{b}_k \ln \breve{b}_{k'} + \frac{1}{2}\theta_{tt}t^2 + \\ &\sum_j\sum_m \delta_{jm} \ln \breve{x}_j \ln \breve{y}_m + \sum_j\sum_{k(\neq p)} \phi_{jk} \ln \breve{x}_j \ln \breve{b}_k + \sum_m\sum_{k(\neq p)} \varphi_{mk} \ln \breve{y}_m \ln \breve{b}_k + \\ &\sum_j \theta_{x,j} t \ln \breve{x}_j + \sum_m \theta_{y,m} t \ln \breve{y}_m + \sum_{k(\neq p)} \theta_{b,k} t \ln \breve{b}_k + u_{\mathrm{M}}, \end{aligned} \tag{3.3}$$

where $\breve{x}_j \overset{\mathrm{def}}{=} b_p^{-g_{x_j}} x_j \ \forall\ j = 1, \ldots, J$, $\breve{y}_m \overset{\mathrm{def}}{=} b_p^{g_{y_m}} y_m \ \forall\ m = 1, \ldots, M$ and $\breve{b}_k \equiv b_p^{-g_{b_k}} b_k \ \forall\ k(\neq p) = 1, \ldots, P$.

3.1 Bank-Specific Varying Directional Vectors

In our discussion above, following a common tradition in the literature, we have treated the directional vector $\boldsymbol{g}$ (or its normalized counterpart $\widehat{\boldsymbol{g}}$) as fixed. However, it is crucial to recognize that there exists an infinite number of the MDTDFs as defined by the unique values of its directional vector. This is a well-known shortcoming of directional distance functions that are implicit functions of the pre-specified directions. Hence, the estimates of the MDTDF-based measure of efficiency, productivity as well as of any other technological metric are likely to change with the different choice of the directional vector.

To eliminate this uncertainty, we rely on the data to determine the direction in which the firm's distance to the technological frontier is to be estimated. More specifically, we treat elements of the normalized directional vector $\widehat{\boldsymbol{g}}$ as an unknown (non-negative) function of some R exogenous contextual variables $\mathbf{z}_{it}$ capturing firm's characteristics. We approximate this directional vector via (complete) S_n-order polynomial sieves, i.e.,

$$\widehat{g}_\kappa(\mathbf{z}_{it}) \approx \exp\left(\sum_{\sum_r s_r = S_n;\ s_r \geq 0} \lambda_{\kappa, s_1, \ldots, s_R} z_{1,it}^{s_1} \cdots z_{R,it}^{s_R}\right) \quad \forall \quad \kappa = 1, \ldots, (J + M + P - 1), \tag{3.4}$$

where S_n regulates the degree of approximation. By employing the exponential link function, we ensure the non-negativity of the directional vector.

Our approach yields "data-driven" estimates of *firm-year-specific* (nonparametric) directional vectors that can meaningfully vary with firm's characteristics captured by contextual variables $\mathbf{z}_{it}$. The directional vector function $\widehat{\boldsymbol{g}}(\cdot)$ is estimated jointly with the remaining parameters in the distance function. Note that our procedure yields the estimates of the firm-year-specific direction towards the technological frontier without requiring any behavioral assumptions. Our approach also nests that of Tsionas et al. (2015) who also estimate the directional vector along with parameters of the distance function. However, unlike us, they treat the directional vector as a vector of unknown *fixed* parameters, which amounts to a rather restrictive assumption of the common direction towards the frontier for all firms across all years. We thus allow for greater degree of flexibility.

In its design, whereby directions are allowed to vary with firm's characteristics, our modeling of firm-specific directional vectors is closely related to the approach pursued by Daraio & Simar (2014) who also propose a data-driven approach to the selection of "context-dependent" directions (albeit, for *additive* distance functions) which accounts for heterogeneity among firms. We differ from them at least in two key aspects: (i) we opt for an easy-to-implement global approximation of unknown directional vector functions as opposed to employing local approximation methods, and (ii) we estimate both the technological frontier and directional vectors in one stage. Our approach also does not require the transformation of the data to polar coordinates.

Lastly, we note that the data-driven selection of the directional vector for the directional distance function is also discussed, but in a deterministic framework, by Färe, Grosskopf & Whittaker (2013) and Hampf & Krüger (2014). Unlike our approach which treats the directional vector as a latent quantity along the lines of Atkinson & Tsionas (2015), these two studies formalize the directional vector as the firm's choice variable.

3.2 Theoretical Regularity Conditions

We estimate directional technology distance function in (3.3) [simultaneously with (3.4)] subject to the Slutsky symmetry as well as the theoretical monotonicity and curvature restrictions[6] in order to ensure that our results are economically meaningful, in line with arguments of Barnett, Geweke & Wolfe (1991), Barnett (2002), Feng & Serletis (2009) and Malikov et al. (2015b).

Since first-order gradients of the distance function contain variables from the data, the monotonicity restrictions are observation-specific and imposed at every data point. The convexity of the MTDF is imposed by ensuring that its corresponding Hessian matrix, which includes only parameters and hence is not observation-specific, is positive semidefinite. In total, we have $(J + M + P)NT + 1$ theoretical regularity conditions, where N and T are the numbers of cross-sectional units and time periods, respectively.

Monotonicity of the (nonlinear) translog MDTDF in (3.3) in inputs and desirable and undesirable outputs, respectively, requires that the log-derivatives satisfy the following conditions:

$$\frac{\partial \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln x_j} = \alpha_j + \sum_{j'} \alpha_{jj'} \ln \breve{x}_{j'} + \sum_{m} \delta_{jm} \ln \breve{y}_m + \sum_{k(\neq p)} \phi_{jk} \ln \breve{b}_k + \theta_{x,j} t \leq 0 \quad \forall \quad j$$

$$\frac{\partial \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln y_m} = \beta_m + \sum_{m'} \beta_{mm'} \ln \breve{y}_{m'} + \sum_{j} \delta_{jm} \ln \breve{x}_j + \sum_{k(\neq p)} \varphi_{mk} \ln \breve{b}_k + \theta_{y,m} t \geq 0 \quad \forall \quad m$$

[6] Recall that the almost homogeneity property has already been imposed via the normalization.

$$
\frac{\partial \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln b_k} = \begin{cases} \gamma_k + \sum_{k'(\neq p)} \gamma_{kk'} \ln \breve{b}_{k'} + \sum_j \phi_{jk} \ln \breve{x}_j + \sum_m \varphi_{mp} \ln \breve{y}_m + \theta_{b,k} t \leq 0 \quad \forall \quad k(\neq p) \\ \\ -1 - \sum_j \alpha_j g_{x_j} + \sum_m \beta_m g_{y_m} - \sum_{k(\neq p)} \gamma_k g_{b_k} - \\ \sum_j \sum_{j'} \alpha_{jj'} \left(\ln \breve{x}_j g_{x_{j'}} + \ln \breve{x}_{j'} g_{x_j} \right) + \sum_m \sum_{m'} \beta_{mm'} \left(\ln \breve{y}_m g_{y_{m'}} + \ln \breve{y}_{m'} g_{y_m} \right) - \\ \sum_{k(\neq p)} \sum_{k'(\neq p)} \gamma_{kk'} \left(\ln \breve{b}_k g_{b_{k'}} + \ln \breve{b}_{k'} g_{b_k} \right) + \sum_j \sum_m \delta_{jm} \left(\ln \breve{x}_j g_{y_m} - \ln \breve{y}_m g_{x_j} \right) - \\ \sum_j \sum_{k(\neq p)} \phi_{jk} \left(\ln \breve{x}_j g_{b_k} + \ln \breve{b}_k g_{x_j} \right) + \sum_m \sum_{k(\neq p)} \varphi_{mk} \left(\ln \breve{b}_k g_{y_m} - \ln \breve{y}_m g_{b_k} \right) - \\ \sum_j \theta_{x,j} t g_{x_j} + \sum_m \theta_{y,m} t g_{y_m} - \sum_{k(\neq p)} \theta_{b,k} t g_{b_k} \leq 0 \quad \text{if} \quad k = p. \end{cases}
$$

The convexity of the MDTDF is imposed by ensuring that all leading principal minors of the Hessian matrix for (3.3) are non-negative. The Hessian matrix is given by

$$
\begin{bmatrix}
\alpha_{11} & \cdots & \alpha_{1J} & \delta_{11} & \cdots & \delta_{1M} & \phi_{11} & \cdots & \phi_{1(P-1)} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln x_1 \partial \ln b_P} \\
\vdots & & & & & & & & & \vdots \\
\alpha_{J1} & \cdots & \alpha_{JJ} & \delta_{J1} & \cdots & \delta_{JM} & \phi_{J1} & \cdots & \phi_{J(P-1)} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln x_J \partial \ln b_P} \\
\delta_{11} & \cdots & \delta_{J1} & \beta_{11} & \cdots & \beta_{1M} & \varphi_{11} & \cdots & \varphi_{1(P-1)} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln y_1 \partial \ln b_P} \\
\vdots & & & & & & & & & \vdots \\
\delta_{1M} & \cdots & \delta_{JM} & \beta_{M1} & \cdots & \beta_{MM} & \varphi_{M1} & \cdots & \varphi_{M(P-1)} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln y_M \partial \ln b_P} \\
\phi_{11} & \cdots & \phi_{J1} & \varphi_{11} & \cdots & \varphi_{M1} & \gamma_{11} & \cdots & \gamma_{1(P-1)} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln b_1 \partial \ln b_P} \\
\vdots & & & & & & & & & \vdots \\
\phi_{1(P-1)} & \cdots & \phi_{J(P-1)} & \varphi_{1(P-1)} & \cdots & \varphi_{M(P-1)} & \gamma_{(P-1)1} & \cdots & \gamma_{(P-1)(P-1)} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln b_{(P-1)} \partial \ln b_P} \\
\frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln x_1 \partial \ln b_P} & \cdots & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln x_J \partial \ln b_P} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln y_1 \partial \ln b_P} & \cdots & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln y_M \partial \ln b_P} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln b_1 \partial \ln b_P} & \cdots & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln b_{(P-1)} \partial \ln b_P} & \frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln b_P^2}
\end{bmatrix}
$$

where, for the ease of exposition, we have set the order number of a normalizing bad output to equal P, i.e., $b_p = b_P$, and

$$
\begin{aligned}
\frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln x_j \partial \ln b_P} &= -\sum_{j'} \alpha_{jj'} g_{x_{j'}} + \sum_m \delta_{jm} g_{y_m} - \sum_{k(\neq p)} \phi_{jk} g_{b_k} \quad \forall \quad j \\
\frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln y_m \partial \ln b_P} &= \sum_{m'} \beta_{mm'} g_{y_{m'}} - \sum_j \delta_{jm} g_{x_j} - \sum_{k(\neq p)} \varphi_{mk} g_{b_k} \quad \forall \quad m \\
\frac{\partial^2 \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln b_k \partial \ln b_P} &= \begin{cases} -\sum_{k'(\neq p)} \gamma_{kk'} g_{b_{k'}} - \sum_j \phi_{jk} g_{x_j} + \sum_m \varphi_{mp} g_{y_m} \quad \forall \quad k(\neq p) \\ \\ 2\sum_j \sum_{j'} \alpha_{jj'} g_{x_j} g_{x_{j'}} + 2\sum_m \sum_{m'} \beta_{mm'} g_{y_m} g_{y_{m'}} + \\ 2\sum_{k(\neq p)} \sum_{k'(\neq p)} \gamma_{kk'} g_{b_k} g_{b_{k'}} - 2\sum_j \sum_m \delta_{jm} g_{x_j} g_{y_m} + \\ 2\sum_j \sum_{k(\neq p)} \phi_{jk} g_{x_j} g_{b_k} - 2\sum_m \sum_{k(\neq p)} \varphi_{mk} g_{b_k} g_{y_m} \quad \text{if} \quad k = p. \end{cases}
\end{aligned}
$$

3.3 Jacobian of Transformation

We estimate (3.3)–(3.4) jointly after appending a two-sided mean-zero error term to the MDTDF in (3.3). We assume this random disturbance to be *i.i.d.* normally distributed independently of the one-sided error term, which is modeled as follows:

$$u_{\mathrm{M}} \ \sim \ \mathbb{N}_{+}\left(a_0 + a_1 t + \frac{1}{2}a_2 t^2, \sigma^2_{\mathrm{M}}\right).$$

Further, note that (3.3) is subject to an "endogeneity" problem due to the appearance of the left-hand-side "dependent variable" on the right-hand side of the equation through the "breve" variables. Addressing this problem requires the appropriate Jacobian of the transformation of the left-hand-side variable be explicitly incorporated in the log-likelihood function (Kumbhakar & Wang, 2006; Malikov et al., 2015b). The observation-specific Jacobian is given by

$$\begin{aligned}
J_{\mathrm{M}} = \frac{\partial(v_{\mathrm{M}} + u_{\mathrm{M}})}{\partial \ln b_p} = 1 &+ \sum_j \alpha_j g_{x_j} - \sum_m \beta_m g_{y_m} + \sum_{k(\neq p)} \gamma_k g_{b_k} + \\
&\sum_j \sum_{j'} \alpha_{jj'} \left(\ln \breve{x}_j g_{x_{j'}} - \ln \breve{x}_{j'} g_{x_j}\right) - \sum_m \sum_{m'} \beta_{mm'} \left(\ln \breve{y}_m g_{y_{m'}} - \ln \breve{y}_{m'} g_{y_m}\right) + \\
&\sum_{k(\neq p)} \sum_{k'(\neq p)} \gamma_{kk'} \left(\ln \breve{b}_k g_{b_{k'}} - \ln \breve{b}_{k'} g_{b_k}\right) - \sum_j \sum_m \delta_{jm} \left(\ln \breve{x}_j g_{y_m} + \ln \breve{y}_m g_{x_j}\right) + \\
&\sum_j \sum_{k(\neq p)} \phi_{jk} \left(\ln \breve{x}_j g_{b_k} - \ln \breve{b}_k g_{x_j}\right) - \sum_m \sum_{k(\neq p)} \varphi_{mk} \left(\ln \breve{b}_k g_{y_m} + \ln \breve{y}_m g_{b_k}\right) + \\
&\sum_j \theta_{x,j} t g_{x_j} - \sum_m \theta_{y,m} t g_{y_m} + \sum_{k(\neq p)} \theta_{b,k} t g_{b_k},
\end{aligned} \tag{3.5}$$

where v_M denotes the two-sided error term appended to the distance function.

4 Modeling Banking Technology

Following the wide practice in the literature [e.g., see Hughes & Mester (2014) for an excellent review], we model the bank's production technology using the so-called "intermediation approach" pioneered by Sealey & Lindley (1977), according to which the bank's balance sheet is said to capture the essential structure of its core business. More specifically, the bank's liabilities, along with its physical capital and labor, are modeled as inputs whereas its non-physical financial assets are conceptualized as outputs to the bank's production process. Liabilities include core deposits and purchased funds, and assets include commercial loans and trading securities.

In this paper, following Malikov et al. (2015b), we employ the framework for modeling banking technology that is more general than the standard one, whereby we explicitly recognize that the bank's production of economically desirable "good" outputs, such as revenue-generating earning loans, is usually accompanied by the simultaneous by-production of an economically undesirable "bad" output which takes the form of loss-generating non-performing loans. By explicitly incorporating the unintended production of non-performing loans into the model of banking production technology, not only do we accommodate an inherent feature of commercial banking whereby the issuance of interest-earning loans is always subject to credit risk associated with the borrower's failure to repay the loan but we are also capable of properly crediting bank's efficiency/productivity for their risk management efforts in a pursuit of the reduction in undesirable non-performing loans. While researchers have long ago acknowledged the importance of accounting for non-performing loans when measuring banks' performance, the popular go-to approach of conditioning the bank's technology on such loans thereby treating the latter as a "technology shifter" meant to control for risk and/or quality of loans (e.g., Berger & Mester, 1997, 2003; Altunbas, Liu & Seth, 2000; Koutsomanoli-Filippaki, Margaritis & Staikouras, 2009) is generally incapable of meaningfully crediting banks for their actively pursued reduction of non-performing loans because this approach (i) does not recognize that non-performing loans are a by-product of desirable loans and (ii) cannot accommodate the undesirability of these loans. In contrast, we can achieve the two by acknowledging that non-performing loans are, in fact, *bad outputs*.

Thus, we define the following two good outputs of the bank's production process: total loans (y_1) and total securities (y_2). Following Hughes & Mester (1998, 2013), we also include (off-balance-sheet) non-interest income (y_3) as an additional output.[7] The bank's bad output is measured by the total non-performing loans (b). The inputs are purchased funds (x_1), core deposits (x_2), labor, i.e., the number of full-time equivalent employees (x_3), and physical capital (x_4). We also include financial (equity) capital (x_5) as an additional input to the production technology. The treatment of equity capital as an input to banking production technology is consistent with Hughes & Mester's (1998) argument that banks may use the latter as a source of loanable funds and thus as a cushion against losses.

We also consider a number of variables capturing various bank's characteristics, both internal and external. These variables are intended to contextualize the economic environment in which banks operate and thus to help us identify their "target" point on the technological frontier relative to which their performance is to be measured. It is only natural to presuppose that the bank's desired input-output mix on the technological frontier under complete technical efficiency is *un*likely to be the same for all institutions. In contrast, one may expect the "target" input-output point on the technological frontier to greatly vary with the bank's characteristics reflective of its business practices in a given market environment. To accommodate this important heterogeneity across banks, we therefore let the bank's directional vector towards its "target" point on the technological frontier be a function of contextual variables which are meant to capture the heterogeneity in banks' business strategies along various dimensions including the scope and competitiveness of the market, banks' size and product mix as well as risk taking. Specifically, the chosen contextual variables are all likely to greatly influence bank's business strategies in pursuit of the maximum franchise value via, e.g., greater efficiency, access to protected markets or building valuable lending relationships (Demsetz, Saidenberg & Strahan, 1996; DeYoung & Rice, 2004). Given the differences in these contextual variables reflective of banks' idiosyncratic features, one would normally expect banks' business strategies to also significantly vary across heterogeneous banks which, in turn, would imply pronounced differences in the "target" points on the technological frontier towards which banks ultimately choose to move in order to fulfill their heterogeneous business strategies. We can model the said differences via the heterogeneous bank-year-specific varying directional vector.

The first group of contextual variables proxies the bank's (external) market environment. The included variables, along with the rationale behind them, are as follows:

- The average distance between the bank's branches and its main office (z_1): this variable captures the density of the bank's branch network in the covered geographical area and allows us to model the possibility that bank branches that are farther away from headquarters may be relatively more difficult and costly to monitor.
- The number of counties in which the bank operates (z_2): this variable quantifies the geographical scope of the bank's market and is meant to account for the possibility that banks, which

[7] In this paper, we measure off-balance-sheet income by the net non-interest income (less service charges on deposits). We acknowledge that this measure of off-balance-sheet income may be biased downward by losses. Ideally, one would want to use the gross non-interest income, which however is infeasible to construct based on the information reported in the data.

cover larger geographical areas, may exert weaker control over the operation of their branches. Also, banks operating in larger geographical areas are less likely to focus on building long-term relationships with their customers.

– The number of bank branches (z_3): similar to the above two variables, the number of branches measures the overall scope of the bank's market.

– The average Herfindahl–Hirschman Index (HHI) in counties of the bank's active presence (z_4): the variable is the average of county-level HHI indices computed using banks' total loans and measures the market concentration across all counties in which the bank operates. Banks operating in relatively more concentrated markets are expected to be exposed to weaker competitive pressures from rivals.

The second group of contextual variables measures the bank's internal characteristics and includes:

– The indicator variable for a community bank (z_5): by controlling for community banks, we can proxy for extra financial and regulatory constraints that small locally-oriented community banks face over regular commercial banks. Unlike most empirical studies that use a unidimensional criterion based on an *ad hoc* threshold for total assets, we follow the FDIC's classification in order to identify community banks.[8]

– The age of the bank (z_6): this variable captures an instance whereby banks with longer history of operations are more likely to have built long-term relationships with their customers.

– The bank's total assets (z_7): the variable is a conventional measure of the bank's size which is arguable one of the most pivotal determinants of the bank's business strategies.

– The risk-weighted asset ratio (z_8): this *ex post* proxy for the bank's exposure to credit risk is constructed as the ratio of the bank's risk-weighted assets to its total assets. By including this variable, we are able to control for both the risk taking and risk management aspects of the bank's business practices.

– The bank-wide HHI of loans (z_9): this measure of the bank's product mix computed as the HHI index of different types of loans that the bank issues is meant to quantify the revenue (and hence credit risk) diversification efforts by the banks. We consider the following types of loans included in y_1: commercial real estate loans, residential real estate loans, agriculture loans, business loans, consumer loans and other loans.

Table 1 presents summary statistics of the data. The data come from multiple sources. Financial data on all FDIC-insured commercial banks are from Call Reports available from the Federal Reserve Bank of Chicago. We match these data with the geographical data from the Summary of Deposits, an annual survey of branch office deposits for all FDIC-insured banks. Our data span the 2000–2010 period and are as of June 30th for each year. We exclude internet banks, commercial banks conducting primarily credit card activities and banks chartered outside the continental U.S. We also omit observations for which negative values for assets and equity are reported since these are likely to be the result of erroneous data reporting. The remaining sample is an unbalanced panel with 63,040 bank-year observations for 8,261 banks. We deflate all nominal stock variables to 2005 U.S. dollars using the Consumer Price Index for all urban consumers.

[8]Specifically, a bank is classified as a community bank if it has less than \$1 billion (2010 dollars) in assets or has assets above this threshold but its loan-to-asset ratio is greater than 33%, its core-deposits-to-asset ratio is greater than 50%, has between one and 75 branch offices in 2010, operates in less than four states or in less than three MSAs and has no single office with deposits exceeding \$5 billion.

Table 1. Data Summary Statistics

Variable	Mean	5th Perc.	Median	95th Perc.
Inputs				
Purchased Funds	514,337.70	2,644.68	24,381.20	310,710.46
Core Deposits	881,997.87	19,219.02	90,858.48	899,122.22
Labor	290.9	9.0	41.6	363.0
Physical Capital	17,359.89	178.09	2,679.70	29,221.86
Equity Capital	140,036.75	2,794.47	13,460.56	130,959.52
Outputs				
Loans	864,373.77	16,275.01	92,265.94	985,816.73
Securities	611,142.72	8,221.97	41,391.57	432,354.31
Non-Interest Income	26,258.79	101.41	832.03	15,183.67
Non-Performing Loans	18,651.71	21.22	723.37	16,174.63
Contextual Variables				
Ave Distance to Main Office	22.63	0.00	9.06	76.77
# of Counties Operated	3.2	1.0	1.0	7.0
# of Branches	11.7	1.0	3.0	21.0
Ave HHI in Counties Operated	0.0495	0.0001	0.0178	0.1939
Community Bank	0.9194			
Age	69.91	5.08	79.83	126.99
Total Assets	1,454,857.97	26,692.41	134,065.73	1,395,631.51
Risk-weighted Asset Ratio	0.6778	0.4679	0.6869	0.8552
Bank-wide HHI of loans	0.5241	0.3050	0.5066	0.8181

Notes: All inputs except for labor, outputs and total assets are in thousands of real 2005 U.S. dollars. Labor is the number of full-time equivalent employees; average distance to the main office is in miles; age is in years. The two "#" variables are in terms of the countables in question. The remaining variables are unit-free.

It is noteworthy that, in contrast to the dual cost-based approach to modeling banking technology which many studies in the banking literature tend to favor (e.g., Hughes & Mester, 1998, 2013; Wheelock & Wilson, 2001, 2012; Malikov, Restrepo-Tobón & Kumbhakar, 2015, to name a few), the formulation of the bank's production process based on the primal directional technology distance function, which we employ in this paper, does not require cost information including input prices which are oftentimes unavailable or measured imprecisely. The latter is particularly helpful because the cost-based estimates of the bank's technology may potentially be misleading because lower funding costs which large banks have been documented to enjoy may be mis-attributed to their underlying technology (say, in the form of scale or scope economies) as opposed to plain too-big-to-fail funding advantages embodied in large bank's costs and their input prices (e.g., see Restrepo-Tobón & Kumbhakar, 2015). The primal distance-function-based formulation of the banking technology, which we resort to in this paper, is however immune to the above problem. In addition, our approach lets us meaningfully account for equity capital which oftentimes is either excluded from the analysis of banking technology or treated as a quasi-fixed input primarily because its price is unavailable.

5 Empirical Results

We estimate the translog MDTDF in (3.3) under three alternative treatments of the directional vector:

(i) In line with the bulk of the literature (e.g., Färe et al., 2005), we pre-specify $\boldsymbol{g}$ as a "unit" direction whereby all elements of the directional vector are arbitrarily set equal to unity prior

Table 2. Bayes Factors against the Unit-Direction ADTDF Model

Functional Form	Direction	Bayes Factor
Multiplicative Distance		
Translog	Unit	19.604
Translog	Data-driven fixed	13.023
Translog	Data-driven varying	27.194
Additive Distance		
Quadratic	Unit	1.000
Quadratic	Data-driven fixed	6.106
Quadratic	Data-driven varying	7.013

to the estimation.

(ii) Following Malikov et al. (2015b), we let the data help us select the appropriate directional vector by treating $\boldsymbol{g}$ as a vector of unknown *fixed* parameters estimated simultaneously with the remaining parameters of the distance function.

(iii) To accommodate the heterogeneity among banks, as described in Section 3.1 we allow the directional vector $\boldsymbol{g}$ to meaningfully *vary* across banks by having it be an unknown function[9] of the bank's characteristics $\mathbf{z}_{it}$ estimated simultaneously with the remaining parameters of the distance function.

By estimating the above three alternative specifications of the MDTDF, we are able to empirically assess the sensitivity of the results (and the conclusions drawn based thereupon) to the potential model misspecification due to an *ad hoc* choice of the directional vector [(i) vs. (iii)] and the failure to accommodate the heterogeneity among individual banks by restricting, although no longer arbitrarily pre-specified, directional vector to be common to all banks [(ii) vs. (iii)].

In addition to the MDTDF in (3.3) which provides a proportional scale-invariant measurement of the bank's distance to its technological frontier, we also estimate its more popular counterpart in the form of the ADTDF that constitutes an additive measurement of banks' technological inefficiency. Following the wide practice in the literature, we adopt the quadratic specification for the ADTDF [for concreteness, see eq. (B.3)] which we also estimate subject to regularity conditions, including the translation property. For details on the econometric specification of the quadratic ADTDF, see Supplementary Appendix B. Analogous to the MDTDF, the quadratic ADTDF is also estimated under the three different specifications of the directional vector outlined above. By contrasting the results from our preferred MDTDF-based representation of banking technology with those based on the ADTDF, we seek to assess the potential benefits of making scale-invariant measurements of technological inefficiency.

We estimate all models via Bayesian MCMC methods the details of which are relegated to Supplementary Appendix C. While, in what follows, we discuss the results from all six models, the data lend the most support to our preferred translog MDTDF-based model with heterogeneous context-dependent directions as evidenced by the largest Bayes factor against the popular unit-direction quadratic ADTDF model (see Table 2).

Technology Estimates. Since parameters of the distance function are not that informative themselves, we instead report the estimates of more meaningful technological metrics, namely input and

[9] We select the order the polynomial sieve on the basis of the Bayes factors. For both the MDTDF- and ADTDF-based models, the data consistently favor the second-order approximation.

Table 3. Posterior Mean Estimates of the Production Technology

	Multiplicative Distance			*Additive Distance*		
	Unit Direction	Fixed Direction	Varying Direction	Unit Direction	Fixed Direction	Varying Direction
Input Elasticities						
Purchased Funds	–0.1556	–0.1356	–0.2523	0.1697	0.1734	0.3157
	(–0.2065, –0.1050)	(–0.1989, –0.0724)	(–0.3684, –0.1361)	(0.0874, 0.2522)	(0.0922, 0.2546)	(0.1646, 0.4666)
Core Deposits	–0.2125	–0.1923	–0.1224	0.3301	0.2543	0.2346
	(–0.3160, –0.1135)	(–0.2774, –0.1077)	(–0.1835, –0.0618)	(0.2889, 0.3715)	(0.1350, 0.3732)	(0.1268, 0.3424)
Labor	–0.1154	–0.0943	–0.1444	0.3100	0.1954	0.1013
	(–0.1525, –0.0756)	(–0.1382, –0.0504)	(–0.2127, –0.0762)	(0.2417, 0.3789)	(0.1074, 0.2821)	(0.0482, 0.1539)
Physical Capital	–0.2545	–0.2645	–0.2143	0.2699	0.2243	0.1033
	(–0.3506, –0.1616)	(–0.3901, –0.1393)	(–0.3147, –0.1136)	(0.2096, 0.3304)	(0.1256, 0.3230)	(0.0564, 0.1503)
Equity Capital	–0.2654	–0.2923	–0.2543	0.2897	0.1853	0.2405
	(–0.3123, –0.2254)	(–0.4282, –0.1572)	(–0.3653, –0.1424)	(0.2075, 0.3720)	(0.1029, 0.2686)	(0.1272, 0.3544)
Good Output Elasticities						
Loans	0.4755	0.3334	0.4182	–0.4143	–0.3985	–0.3133
	(0.2624, 0.6774)	(0.1708, 0.4960)	(0.2376, 0.5982)	(–0.4593, –0.3695)	(–0.5928, –0.2042)	(–0.4509, –0.1753)
Securities	0.3132	0.5200	0.3953	–0.1783	–0.2287	–0.4073
	(0.2104, 0.3966)	(0.2749, 0.7648)	(0.2299, 0.5611)	(–0.2019, –0.1547)	(–0.3364, –0.1209)	(–0.5979, –0.2161)
Non-Interest Income	0.2535	0.2203	0.2454	–0.4822	–0.4147	–0.3263
	(0.2195, 0.2899)	(0.1111, 0.3295)	(0.1214, 0.3697)	(–0.5154, –0.4489)	(–0.6202, –0.2091)	(–0.4813, –0.1717)
Bad Output Elasticities						
Non-Performing Loans	–0.3661	–0.2906	–0.4262	0.7285	0.2990	0.3339
	(–0.4092, –0.3235)	(–0.3346, –0.2465)	(–0.4676, –0.3851)	(0.4577, 0.9963)	(0.2558, 0.3421)	(0.2947, 0.3732)
Desirable Scale Elasticity						
DSE	0.9827	0.9335	0.9499	1.2752	1.0145	0.9698
	(0.6941, 1.3897)	(0.6361, 1.3556)	(0.6694, 1.3299)	(1.1184, 1.4397)	(0.7010, 1.4599)	(0.6700, 1.3789)

Note: The 95% credible intervals in parentheses.

output elasticities of the distance function. Table 3 provides the posterior mean estimates of the elasticities along with their corresponding 95% credible intervals for all six estimated models. (The signs of elasticities are expectedly reversed across the MDTDF and ADTDF given their respective monotonicity properties.) Regardless of the type of the distance function being estimated (multiplicative or additive), the results point to the stark sensitivity of the elasticity estimates to the treatment of the directional vector. For instance, consider the estimates of the elasticity of the MDTDF with respect to the undesirable non-performing loans which, owing to duality, yields an indirect measure of (the negative of) the banks' cost elasticity with respect to the said bad output. The results suggest that merely estimating the data-driven directions by treating them as unknown parameters [under the specification (ii)] may not be flexible enough to capture the "costliness" of the bad output: the absolute magnitude of the elasticity in fact goes down in this case. However, when accommodating the well-documented heterogeneity among individual banks by letting unknown directions vary across banks [specification (iii)], the model produces the elasticity estimates of substantially larger magnitudes, on average, at around 0.43 against the posterior average of 0.37 if one were to specify an *ad hoc* unit direction. Interestingly, these findings differ significantly from those obtained based on the quadratic ADTDF in which case the models estimated in the data-driven directions consistently produces estimates of the bad output elasticity that are smaller than those computed in the popular pre-specified unit direction. Also, across the two types of the distance function, we find that the ADTDF generally tends to underestimate the "costliness" of non-performing loans for banks' operations.

We next analyze the estimates of scale elasticity which provide insights about potential scale economies in the industry. Following Malikov et al. (2015), we focus on the bank's *desirable* scale elasticity measuring the equiproportional expansion in desirable outputs in response to an equiproportional increase in inputs at a given level of bad outputs in a given direction:

$$\mathrm{DSE}\left(\mathbf{x},\mathbf{y},\mathbf{b};\mathbf{g}\right) \stackrel{\mathrm{def}}{=} \left.\frac{\partial \ln \lambda_y}{\partial \ln \lambda_x}\right|_{\lambda_y=\lambda_x=1} \;:\; \vec{D}_{\mathrm{M}}(\lambda_x\mathbf{x},\lambda_y\mathbf{y},\mathbf{b};\boldsymbol{g}) \in (0,1], \tag{5.1}$$

which can be easily shown to equal $\mathrm{DSE} = -\sum_j \frac{\partial \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln x_j} / \sum_m \frac{\partial \ln \vec{D}_{\mathrm{M}}(\cdot)}{\partial \ln y_m}$.[10] Thus, by measuring the expansion in good outputs while holding the quantity of bad outputs unchanged, we effectively focus on instances whereby some of the expanded inputs are being diverted into ensuring that no additional bad outputs are by-produced with the increased production of desirable outputs. For example, banks might redirect some of their labor and capital into more carefully examining new loan applications or more actively monitoring loans upon their issuance in order to ensure that they are "performing" and thus revenue-generating. The above measure implies that the bank exhibits increasing/constant/decreasing returns to (desirable) scale if DSE is greater than/equal to/less than one. Regardless of the directional vector specification, all MDTDF-based translog models produce the posterior mean estimates of DSE between 0.93 and 0.98 with the corresponding 95% credible intervals consistently including unity thus pointing to the prevalence of constant returns to scale in the industry (see the bottom left of Table 3). The latter suggests that banks' costs may generally be invariant to the (desirable) scale of operations, which is fairly consistent with other Bayesian estimates of scale economies reported in the literature (e.g., Feng & Zhang, 2012). When estimating the quadratic ADTDF, the resulting DSE estimates are significantly less stable across different treatments of the directional vector (see the bottom right of Table 3). Specifically, while permitting the direction to be an unknown varying function of contextual variables produces the posterior mean DSE estimate that is quite similar — both quantitatively and qualitatively — to

[10] A similar DSE measure is defined for the ADTDF (see Malikov et al., 2015).

Table 4. Posterior Estimates of Technological Inefficiency

	Unit Direction	Fixed Direction	Varying Direction
		Multiplicative Distance	
Good Outputs: $\exp\{u_{\mathrm{M}}\} \in [1, \infty)$			
Mean	1.421	1.099	1.046
Median	1.431	1.080	1.032
S.d.	0.078	0.081	0.049
95% Bayes Interval	(1.230, 1.548)	(1.003, 1.303)	(1.001, 1.192)
Bad Outputs & Inputs: $1 - \exp\{-u_{\mathrm{M}}\} \in [0, 1)$			
Mean	0.294	0.085	0.042
Median	0.301	0.074	0.031
S.d.	0.041	0.062	0.040
95% Bayes Interval	(0.187, 0.354)	(0.004, 0.232)	(0.001, 0.160)
		Additive Distance	
All Outputs & Inputs: $u_{\mathrm{A}} \in [0, \infty)$			
Mean	0.136	0.128	0.067
Median	0.115	0.127	0.062
S.d.	0.102	0.061	0.034
95% Bayes Interval	(0.005, 0.383)	(0.013, 0.251)	(0.012, 0.145)

that from the alternative MDTDF model, the results are however drastically different across the two types of the distance function when the ADTDF is fitted in the *ad hoc* unit direction. In the latter case, contrary to findings based on our preferred MDTDF model with varying directions, one could conclude that, on average, banks enjoy increasing returns to scale of quite sizable magnitudes. This highlights the extreme sensitivity of the empirical results to the choice of the directional vector in which the technology distance is being measured, especially when the measurement is additive.

Inefficiency. Among all technological metrics, inefficiency is perhaps the most susceptible to the way directional vector is modeled since the latter directly effects the measurement of inefficiency by determining the location of the "target" point on the bank's technological frontier the non-zero distance to which is construed as the evidence of the bank's inefficiency. Table 4 reports the summary of posterior estimates of inefficiency across all models and specifications of the directional vector. As discussed in Section 3, in the case of the MDTDF-based formulation of banking technology, we are able to identify inefficiency in two dimensions: the good-output and input/bad-output (proportional) inefficiency. The ADTDF however produces a single (additive) measure of technological inefficiency with respect to all outputs and inputs.

Regardless whether the distance to the frontier is measured proportionally or additively, the results suggest that the failure to use the data to select the appropriate directional vector as well as to let the latter be heterogeneous across individual banks generally leads to the gross overestimation of the banks' technological inefficiency. For instance, the MDTDF estimated in an arbitrarily pre-specified unit direction produces the astounding (and quite unrealistic) results suggesting that, on average, inefficient banks have the potential to increase their good outputs by 42% while reducing the input usage and the by-production of bad outputs by 29%. Consistent with Malikov et al.'s (2015) findings, the extent of inefficiency among banks is however found to be significantly smaller if we measure the latter in the data-driven direction. In this case, the good-output and input/bad-output MDTDF-based inefficiencies are, on average, estimated to be around 10% and 8%, respectively. Employing a more flexible model which allows the directional vector to vary with

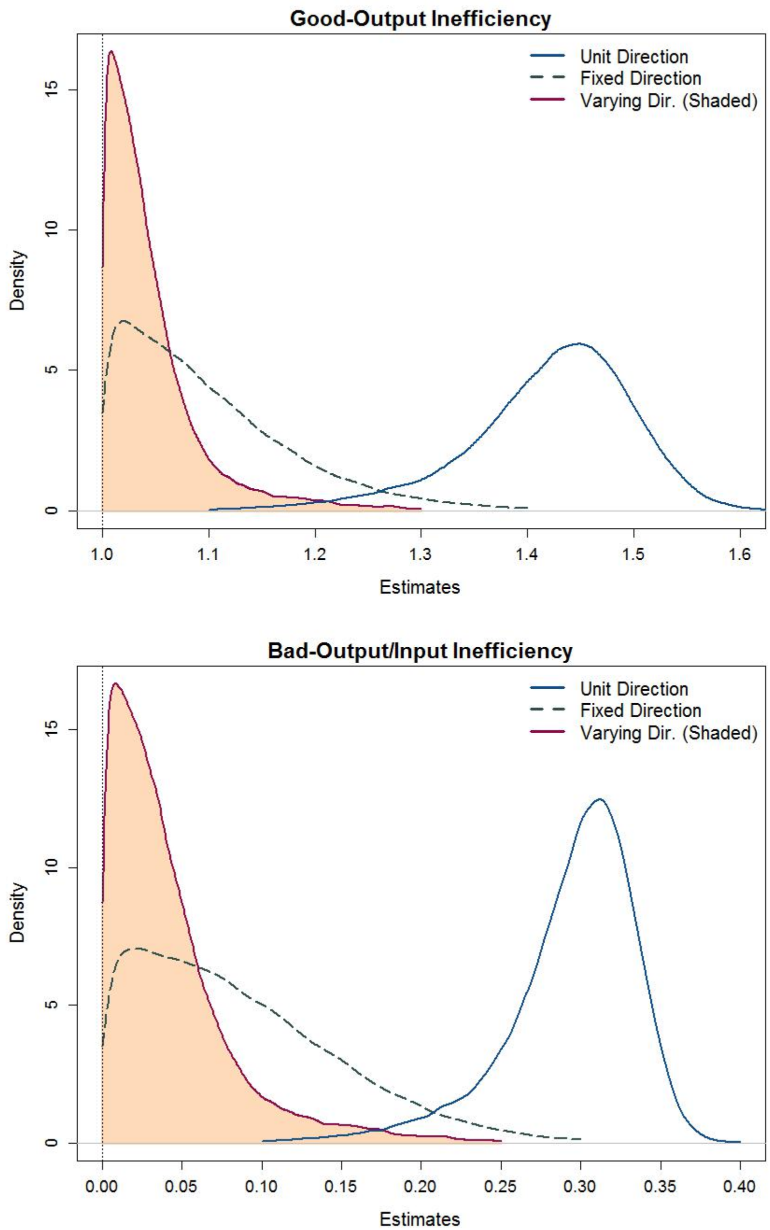

Figure 1. Kernel Densities of the MDTDF-based Inefficiency Estimates

Table 5. Posterior Mean Estimates of the Productivity Growth Decomposition, in % p.a.

	Multiplicative Distance			*Additive Distance*		
	Unit Direction	Fixed Direction	Varying Direction	Unit Direction	Fixed Direction	Varying Direction
PG	2.074	2.409	3.214	0.286	2.052	2.766
	(–1.384, 6.140)	(0.000, 4.652)	(0.001; 5.485)	(–1.604, 2.262)	(0.002, 3.113)	(–1.862; 8.041)
TC	1.975	1.201	2.098	0.320	1.200	3.094
	(–1.301, 5.474)	(0.221, 2.171)	(0.344, 3.880)	(–0.000, 0.064)	(0.618, 1.787)	(1.125, 5.052)
EC	0.005	1.337	1.339	0.000	0.980	0.001
	(–1.881, 1.915)	(–0.211, 3.219)	(0.010, 2.268)	(–1.896, 1.903)	(0.000, 1.673)	(–4.606, 4.600)

Note: The 95% credible intervals in parentheses.

banks' heterogeneous characteristics, we obtain an even more modest evidence of technological inefficiency among commercial banks: the respective posterior means of 5% and 4% in the good-output and input/bad-output dimensions. Such relatively low levels of technological inefficiency may at first appear somewhat stark given other Bayesian estimates of primal technological inefficiency among banks recently reported in the literature (e.g., Feng & Zhang, 2012). However, the latter is not surprising but rather expected since, unlike in most previous studies, our model of banking technology explicitly recognizes the undesirable nature of non-performing loans and hence does not penalize banks for diverting some of their inputs in an attempt to lower the by-production of bad outputs.

Figure 1, which plots kernel densities of the posterior estimates of the MDTDF-based inefficiency, vividly illustrates the drastic differences across alternative treatments of the directional vector. The same pattern holds for the ADTDF-based measure of inefficiency which also declines dramatically when the directional vector is no longer fixed; its magnitude however cannot be meaningfully interpreted in terms of proportional changes due its additive nature.

Productivity Growth. Table 5 presents the results for the productivity growth decomposition based on (2.8) and (A.4) for the MDTDF- and ADTDF-based representation of technology, respectively. The reported are the posterior mean estimates along with their 95% Bayes intervals for all estimated models. Regardless of the type of the distance function used, the data suggest that models allowing for a varying directional vector tend to produce larger estimates of the productivity growth in the industry likely due to their superior ability to accommodate technological heterogeneity across banks and thus properly credit banks for the improvement in their productivity. Based on our preferred MDTDF model estimated in the varying direction, over the course of the 2000–2010 period the industry has exhibited a significant productivity growth of around 3.2% per annum with about two thirds of the growth being attributed to technological change and the remaining third representing the improvement in efficiency levels. Constraining the directional vector to be common to all banks (either pre-specifying or estimating it) leads to a rather dramatic under-measurement of PG with the posterior mean estimates (in the range of 2.1–2.4% p.a.) suggesting no statistically significant temporal change in technology and/or efficiency. Interestingly, while our finding of the technical change being the dominant driver of the productivity growth is fairly consistent with the results reported in the previous Bayesian studies of banks' performance (Feng & Serletis, 2010; Feng & Zhang, 2012), the evidence in support of a rather material improvement in efficiency is however new since most these studies have documented statistically insignificant EC for banks. Again, the latter can be primarily attributed to the ability of our model, due to its flexible specification of the directional vector, to more fully capture efficiency gains across heterogeneous banks. Notably, the alternative ADTDF-based model however fails to detect any efficiency improvements even when the direction is models as a context-dependent function. In fact, this model suggests — rather questionably — that the productivity improvement has also been insignificant during the sample period, with the 95% credible interval for the posterior mean of PG including zero.

Direction. Until now, we have focused our discussion on the consequences of the failure to accommodate cross-bank technological heterogeneity on the measurement of their technology and performance. In what follows, we take a closer looks at the estimates of the said heterogeneity. More specifically, we examine the posterior estimates of the heterogeneous context-dependent directional vector which is our tool of choice for modeling bank-specific "target" points on the technological frontier the distance towards which the technology distance function is meant to measure.

Table 6 reports the posterior estimates of the estimated directional vectors for both the multiplicative and additive distance functions. In the case of the direction being treated as fixed and thereby common to all banks, the reported are posterior mean estimates of the unknown directional parameters; in the case of heterogeneous directional vectors, we naturally obtain a series of the estimates of context-dependent directional functions for which we also report the posterior interquartile ranges. Not only do the results presented in the table exemplify the extent to which the forcefully imposed assumption of a common directional vector leads to vastly mismeasured directions, but they also highlight the benefits of allowing the data select the directional vector since all directional elements are estimated at values far smaller than the go-to unit vector. Note that here, for each type of the distance function, we can compare the directions across alternative treatments of the directional vector because of the maintained "unit" normalization of the directional element corresponding to non-performing loans. However, one cannot meaningfully contrast the values of directions across the two types of the distance function (multiplicative vs. additive) due to intrinsically different forms in which they enter the models. To circumvent this problem and thus to facilitate a more telling examination of the directional vector estimates, we essentially transform elements of the directional vector into the directional *angles* they imply. More specifically, we compute the angles between the $(M+P+J)$-dimensional directional vector and each of its $(M+P+J)$ projections onto axes corresponding to all outputs and inputs. For instance, in the case of the MDTDF, the directional angle corresponding to some good output y_m is defined as follows:

$$\phi_{y_m} = \arccos\left(\frac{g_{y_m}}{||\boldsymbol{g}||}\right), \tag{5.2}$$

where $||\cdot||$ is the Euclidean norm.[11] By construction and the non-negative formulation of directional elements, the angle is limited to the first quadrant. To illustrate the usefulness of this angle-based interpretation of the directional vector, consider an example with one good output and one input only. Further, assume a unit directional vector. Hence, the directional angle corresponding to the good output is given by $\arccos\left(1/\sqrt{2}\right) = 45°$ and so is the one corresponding to the input, implying equal weights given to both variables in the firm's movement towards its "target" point on the frontier. Although in a multi-dimensional case,[12] the defined directional angles do not sum up to the usual 90°, the latter does not in any way affect the usefulness of the information contained in these angles: the smaller the angle is, the more tilted the directional vector is towards the out/in-put in question implying a relatively higher weight/priority given to the latter.

The posterior mean directional angles (measured in radians) are reported in Table 7. From the table, we see that the estimated direction varies significantly depending on whether the measurement

[11]The directional angle can be similarly defined for the ADTDF.

[12]When $(M+P+J) \geq 3$, to be precise.

Table 6. Posterior Mean Estimates of the Directional Vector

	Multiplicative Distance		*Additive Distance*	
Direction	Fixed Direction	Varying Direction	Fixed Direction	Varying Direction
		Inputs		
Purchased Funds				
Mean	0.232	0.173	0.187	0.130
IQR		(0.130, 0.216)		(0.108, 0.152)
Core Deposits				
Mean	0.314	0.127	0.227	0.082
IQR		(0.102, 0.147)		(0.065, 0.099)
Labor				
Mean	0.004	0.092	0.013	0.126
IQR		(0.078, 0.106)		(0.114, 0.137)
Physical Capital				
Mean	0.013	0.149	0.018	0.130
IQR		(0.133, 0.167)		(0.115, 0.144)
Equity Capital				
Mean	0.017	0.210	0.022	0.111
IQR		(0.172, 0.248)		(0.092, 0.127)
		Good Outputs		
Loans				
Mean	0.151	0.101	0.203	0.016
IQR		(0.081, 0.107)		(0.008, 0.024)
Securities				
Mean	0.251	0.117	0.189	0.085
IQR		(0.100, 0.134)		(0.077, 0.094)
Non-Interest Income				
Mean	0.212	0.154	0.081	0.125
IQR		(0.141, 0.165)		(0.116, 0.133)

Notes: The corresponding elements of the directional vector in the "unit direction" models are all equal to 1. Across all models, the element of the directional vector that corresponds to non-performing loans b is normalized to unity. IQR stands for the interquartile range.

of the distance to the technological frontier is proportional or additive. For instance, when allowing for the heterogeneous context-dependent directional vector, the MDTDF-based results suggest that, when seeking to reduce their inefficiency in the use of inputs, banks give the highest priority[13] to equity capital whereas the ADTDF-based model indicates the highest weight being placed on purchased funds and physical capital. However, both the MDTDF and ADTDF consistently suggest that among all inputs and outputs, banks pay the most attention to their levels of inefficiency in the by-production of the bad output: the posterior directional angle estimates are the smallest in this dimension. Notably, this finding is in stark contrast with the uniform (equal-weight) treatment non-performing loans get, along with all other outputs and inputs, if the distance function is fitted in the arbitrarily pre-specified unit direction as commonly practiced in the literature. This exemplifies the non-negligible potential for model misspecification due to an *ad hoc* choice of directional vector.

We next focus on the results for our preferred specification whereby the directional vector is modeled as a varying function of the banks' characteristics capturing their heterogeneous business

[13] As indicated by the smallest directional angle.

Table 7. Posterior Mean Estimates of Directional Angles, in radians

	Multiplicative Distance		*Additive Distance*	
Direction	Fixed Direction	Varying Direction	Fixed Direction	Varying Direction
		Inputs		
Purchased Funds				
Mean	1.364	1.410	1.396	1.446
Range		(1.181, 1.571)		(1.318, 1.568)
Core Deposits				
Mean	1.289	1.453	1.359	1.492
Range		(1.248, 1.570)		(1.378, 1.571)
Labor				
Mean	1.566	1.485	1.558	1.450
Range		(1.407, 1.570)		(1.375, 1.523)
Physical Capital				
Mean	1.558	1.432	1.554	1.446
Range		(1.326, 1.528)		(1.357, 1.535)
Equity Capital				
Mean	1.555	1.376	1.550	1.465
Range		(1.171, 1.569)		(1.364, 1.565)
		Good Outputs		
Loans				
Mean	1.436	1.477	1.381	1.554
Range		(1.343, 1.567)		(1.506, 1.570)
Securities				
Mean	1.347	1.462	1.394	1.488
Range		(1.374, 1.562)		(1.435, 1.539)
Non-Interest Income				
Mean	1.382	1.428	1.495	1.450
Range		(1.328, 1.459)		(1.376, 1.467)
		Bad Outputs		
Non-Performing Loans				
Mean	0.490	0.399	0.392	0.298
Range		(0.258, 0.590)		(0.211, 0.398)

Notes: The corresponding directional angles in the "unit direction" models are all (including that for the bad output b) equal to 1.230 radian. Also, for the reference: 1 radian $\approx 57°$.

practices. We report the posterior mean elasticities of the MDTDF's directional vector with respect to contextual variables[14] in Table 8. The included variables are found to significantly influence directions thereby buttressing the premise of heterogeneity among individual banks which ought not to be taken for granted. However, it is rather difficult to comprehend the extent to which the change in a given contextual variable ultimately affects the position of the bank's direction towards its "target" point on the frontier since all elements of the directional vector are being affected simultaneously. We tackle this issue by instead looking at the "total" effect of the change in z on directional angles corresponding to each out/in-put which accounts for both the direct and indirect effects working through other out/in-puts. For example, in order to assess how a change in, say, z_1 effects the relative priority the bank gives to some good output y_m in an attempt to minimize

[14]In case of the community bank dummy variable z_5, the reported is a semi-elasticity.

its technological inefficiency, we totally differentiate the directional angle ϕ_{y_m} defined in (5.2) with respect to $\ln(z_1)$, where we use logs to facilitate the percentage change interpretation for contextual variables, i.e.,

$$\frac{d\phi_{y_m}}{d\ln z_1} = -\left(1-\left(\frac{g_{y_m}}{||\boldsymbol{g}||}\right)^2\right)^{-1/2}\left(\frac{1}{||\boldsymbol{g}||}\frac{\partial g_{y_m}}{\partial z_1} - \frac{g_{y_m}}{||\boldsymbol{g}||^3}\boldsymbol{g}'\frac{\partial \boldsymbol{g}}{\partial z_1}\right)z_1. \quad (5.3)$$

The results on these semi-elasticities of directional angles for our preferred MDTDF-based model are presented in Table 9.[15] Note that while most estimates are seemingly of small magnitudes and thus might at first appear as being economically negligible, the latter is not necessarily the case given that an estimate as small as 0.01 implies a rather sizable 0.57° change in the directional angle. Also, it is worth reminding the reader that negative values here imply a decrease in the directional angle and thus the repositioning of the bank's directional vector more *towards* the out/in-put in question. Among the contextual variables, the county-level HHI index (z_4) measuring the competitiveness of the market is by far the most influential (magnitude-wise) factor changes in which are associated with the drastic repositioning of the bank's directional vector. Empirical evidence suggests that an increase in the HHI index signifying a rise in the market concentration (i.e., a decrease in competitiveness) is associated with banks assigning a higher priority to the reduction in inefficiency in inputs such as labor, physical and equity capital along with a smaller weight being given to the production of securities and notably non-performing loans. With regards to the latter, this finding implies that banks operating in highly concentrated markets tends to be not that much concerned with the reduction in bad outputs. Interestingly, the results suggest that the only two factors that may compel banks to prioritize the "conservation" of bad outputs both imply the expansion in the geographical scope of the bank's market: the average distance between bank branches and headquarters (z_1) and the number of counties in which the bank operates (z_2). On the other hand, we find that banks consistently prioritize good outputs and inputs over non-performing loans as they grow larger and older.

To conclude, we document non-negligible distortions in the measurement of banks' technological efficiency, productivity growth and scale elasticity when the estimated banking technology is constrained to be common to all banks under a rather unrealistic assumption of the (known or unknown) directional vector towards the bank's production frontier being fixed and not varying across individual banks. Thus, the model that accommodates the well-documented heterogeneity among banks by permitting the directional vector to be bank-specific and thus varying with banks' idiosyncratic characteristics capturing their heterogeneous business strategies is likely to provide a more robust estimation strategy. The data also seem to support the proportional measurement of technological inefficiency over the widely used additive alternative.

6 Conclusion

This paper contributes to the literature by considering the multiplicative distance-function-based formulation of the polluting technology which can meaningfully accommodate bad outputs in the spirit of the additive directional distance function while still preserving the highly desired scale-invariance property of the proportional radial distance function along with its ability to assume popular flexible functional forms such as translog and Fourier. Furthermore, the proposed MDTDF model can meaningfully accommodate technological heterogeneity across individual firms. We accomplish the latter by allowing the directional vector along which the distance toward technological

[15] For analogous results for the ADTDF-based model, see Supplementary Appendix.

Table 8. Posterior Mean Estimates of MDTDF-based Elasticities of the Varying Directional Vector

Dir.\Context.Var.	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9
Inputs									
Purchased Funds	–0.013	0.019	0.344	–0.025	–0.036	0.255	0.071	–0.043	0.255
	(0.0024)	(0.0015)	(0.0055)	(0.0052)	(0.0101)	(0.0045)	(0.0034)	(0.0131)	(0.0113)
Core Deposits	–0.032	0.015	0.154	–0.015	–0.032	0.066	0.034	–0.026	0.322
	(0.0015)	(0.0011)	(0.0340)	(0.0060)	(0.0071)	(0.0032)	(0.0028)	(0.0031)	(0.0144)
Labor	0.044	–0.017	0.007	0.053	0.026	0.044	0.076	0.055	–0.287
	(0.0320)	(0.0085)	(0.0014)	(0.0180)	(0.0030)	(0.0070)	(0.0132)	(0.0130)	(0.0320)
Physical Capital	–0.017	–0.043	0.006	0.014	0.064	0.062	0.041	0.055	–0.201
	(0.0150)	(0.0027)	(0.0017)	(0.0061)	(0.0230)	(0.0150)	(0.0071)	(0.0150)	(0.0128)
Equity Capital	0.022	–0.035	0.023	0.071	0.022	0.050	0.028	0.044	–0.177
	(0.0130)	(0.0043)	(0.0054)	(0.0130)	(0.0040)	(0.0030)	(0.0011)	(0.0054)	(0.0156)
Good Outputs									
Loans	–0.052	0.032	0.015	–0.023	–0.007	0.041	0.032	–0.021	0.128
	(0.0015)	(0.0011)	(0.0024)	(0.0017)	(0.0014)	(0.0050)	(0.0050)	(0.0030)	(0.0022)
Securities	–0.032	0.044	0.035	–0.061	–0.022	0.014	0.012	–0.025	0.144
	(0.0015)	(0.0021)	(0.0025)	(0.0022)	(0.0020)	(0.0050)	(0.0020)	(0.0047)	(0.0013)
Non-Interest Income	0.032	–0.017	0.017	0.013	0.013	0.032	0.041	–0.028	0.212
	(0.0120)	(0.0220)	(0.0050)	(0.0054)	(0.0155)	(0.0071)	(0.0081)	(0.0015)	(0.0190)

Notes: In case of the community bank dummy variable z_5, the reported is a semi-elasticity. Posterior standard errors in parentheses.

Table 9. Posterior Mean Estimates of the MDTDF-based Semi-Elasticities of Varying Directional Angles, in radian/%

Dir.\Context.Var.	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9
Inputs									
Purchased Funds	0.00209	–0.00327	–0.05375	0.08779	0.00598	–0.03944	–0.01056	0.00713	–0.07727
Core Deposits	0.00380	–0.00197	–0.01672	0.04042	0.00389	–0.00639	–0.00335	0.00309	–0.07153
Labor	–0.00377	0.00137	0.00049	–0.08826	–0.00215	–0.00273	–0.00599	–0.00465	0.04771
Physical Capital	0.00236	0.00587	0.00103	–0.03441	–0.00885	–0.00700	–0.00491	–0.00762	0.05518
Equity Capital	–0.00438	0.00671	–0.00193	–0.27635	–0.00419	–0.00752	–0.00437	–0.00859	0.06896
Good Outputs									
Loans	0.00488	–0.00312	–0.00024	0.04752	0.00075	–0.00273	–0.00245	0.00203	–0.02201
Securities	0.00347	–0.00491	–0.00241	0.13796	0.00249	–0.00021	–0.00066	0.00279	–0.02878
Non-Interest Income	–0.00463	0.00229	–0.00060	–0.03101	–0.00171	–0.00287	–0.00505	0.00412	–0.05671
Bad Outputs									
Non-Performing Loans	–0.00034	–0.00272	0.03129	0.08921	0.00200	0.02927	0.01447	0.00148	0.02546

Note: The reported are the mean changes in (the level of) directional angles in response to the unit % change in each contextual variable except for z_4, z_5 and z_9 with respect to which the reported are the level derivatives since these variables are either discrete or already measured in proportions. Also, for the reference: 0.01 radian $\approx 0.57°$.

frontier is being quantified to be an unknown function of firms' idiosyncratic characteristics. These variable are intended to contextualize both the external and internal economic environment in which firms operate and thereby help us identify firm-specific "target" points on the technological frontier relative to which firms' performance is to be measured. Thus, we rely on the data to determine the varying heterogeneous directional vector for each firm. In doing so, we are able to circumvent a well-known shortcoming of directional distance functions which produce technological measurements that are subject to uncertainty associated with the choice of the directional vector. Unlike in previous studies, our approach to modeling data-driven directions does not restrict the unknown directional vector to be fixed and thus common to all firms but rather allows it to vary across firms. We therefore allow for a greater degree of flexibility.

To highlight practical usefulness of the heterogeneous MDTDF along with its potentially significant advantages over its alternatives, we apply our methodology to study the production technology of U.S. commercial banks in 2001–2010. We document non-negligible distortions in the measurement of banks' technological efficiency, productivity growth and scale elasticity when the estimated banking technology is constrained to be common to all banks under a rather unrealistic assumption of the (known or unknown) directional vector towards the bank's production frontier being fixed. Thus, the model that accommodates the well-documented heterogeneity among banks by permitting the directional vector to be bank-specific by varying with banks' idiosyncratic characteristics thus capturing their heterogeneous business strategies is likely to provide a more robust estimation strategy. The data also seem to support the proportional measurement of technological inefficiency over the widely used additive alternative.

References

Altunbas, Y., Liu, M-H, M. P., & Seth, R. (2000). Efficiency and risk in Japanese banking. *Journal of Banking & Finance*, *24*, 1605–1628.

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

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

Atkinson, S. E., Primont, D., & Tsionas, E. G. (2014). Estimation of efficient production using optimal directions with multiple inputs and outputs. Working Paper, Southern Illinois University.

Atkinson, S. E. & Tsionas, E. G. (2015). Directional distance functions: Optimal endogenous directions. *Journal of Econometrics*. forthcoming.

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

Barnett, W. A., Geweke, J., & Wolfe, M. (1991). Seminonparametric Bayesian estimation of the Asymptotically Ideal Production Model. *Journal of Econometrics*, *49*, 5–50.

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

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

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

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

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

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

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

Dakpo, K. H., Jeanneaux, P., & Latruffe, L. (2016). Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework. *European Journal of Operational Research*, *250*, 347–359.

Daraio, C. & Simar, L. (2014). Efficiency and benchmarking with directional distances: A data driven approach. Technical Report n. 7, ISSN 2281-4299, University of Rome "La Sapienza".

Demsetz, R., Saidenberg, M., & Strahan, P. (1996). Banks with something to lose: The disciplinary role of franchise value. *Economic Policy Review*, *2*.

DeYoung, R. & Rice, T. (2004). How do banks make money? A variety of business strategies. *Economic Perspectives - Federal Reserve Bank of Chicago*, *28(4)*.

Färe, R. & Grosskopf, S. (2003). Nonparametric productivity analysis with undesirable outputs: Comment. *American Journal of Agricultural Economics*, *85*, 1070–1074.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985). *The Measurement of Efficiency of Production*. Boston, MA: Kluwer-Nijhoff Publishing.

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

Färe, R., Grosskopf, S., & Margaritis, D. (2015). Directional distance functions revisited: Selective overview and update. *Data Envelopment Analysis Journal*, *1*, 57–79.

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

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

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

Feng, G. & Serletis, A. (2010). A primal Divisia technical change index based on the output distance function. *Journal of Econometrics*, *159*, 320–330.

Feng, G. & Serletis, A. (2014). Undesirable outputs and a primal Divisia productivity index based on the directional output distance function. *Journal of Econometrics*.

Feng, G. & Zhang, X. (2012). Productivity and efficiency at large and community banks in the US: A Bayesian true random effects stochastic distance frontier analysis. *Journal of Banking & Finance*, *36*(7), 1883–1895.

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

Fernández, C., Koop, G., & Steel, M. F. J. (2005). Alternative efficiency measures for multiple-output production. *Journal of Econometrics*, *126*, 411–444.

Girolami, M. & Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, *73*, 123–214.

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

Hailu, A. & Veeman, T. S. (2000). Environmentally sensitive productivity analysis of the Canadian pulp and paper industry, 1959–1994: An input distance function approach. *Journal of Environmental Economics and Management*, *40*, 251–274.

Hailu, A. & Veeman, T. S. (2001). Non-parametric productivity analysis with undesirable outputs: An application to the Canadian pulp and paper industry. *American Journal of Agricultural Economics*, *83*, 605–616.

Hampf, B. & Krüger, J. J. (2014). Optimal directions for directional distance functions: An exploration of potential reductions of greenhouse gases. *American Journal of Agricultural Economics*. forthcoming.

Hudgins, L. B. & Primont, D. (2007). Derivative properties of directional technology distance functions. In R. Färe, S. Grosskopf, & D. Primont (Eds.), *Aggregation, Efficiency, and Measurement*. Springer.

Hughes, J. P. & Mester, L. J. (1998). Bank capitalization and cost: Evidence of scale economies in risk management and signaling. *Review of Economics and Statistics*, *80*(2), 314–325.

Hughes, J. P. & Mester, L. J. (2013). Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function. *Journal of Financial Intermediation*, *22*(4), 559–585.

Hughes, J. P. & Mester, L. J. (2014). Measuring the performance of banks: Theory, practice, evidence, and some policy implications. In A. Berger, P. Molyneux, & J. Wilson (Eds.), *Oxford Handbook of Banking* (2 ed.). Oxford: Oxford University Press. forthcoming.

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

Kumbhakar, S. C. & Tsionas, E. G. (2015). The good, the bad and the technology: Endogeneity in environmental production models. *Journal of Econometrics*. forthcoming.

Kumbhakar, S. C. & Wang, H.-J. (2006). Estimation of technical and allocative inefficiency: A primal system approach. *Journal of Econometrics*, *134*, 419–440.

Malikov, E., Kumbhakar, S. C., & Tsionas, E. G. (2015a). Bayesian approach to disentangling technical and environmental productivity. *Econometrics*, *3*, 443–465.

Malikov, E., Kumbhakar, S. C., & Tsionas, E. G. (2015b). 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*. forthcoming.

Malikov, E., Restrepo-Tobón, D., & Kumbhakar, S. C. (2015). Estimation of banking technology under credit uncertainty. *Empirical Economics*, *49*, 185–211.

Mehdiloozad, M., Sahoo, B. K., & Roshdi, I. (2014). A generalized miltiplicative directional distance function for efficiency measurement in DEA. *European Journal of Operational Research*, *232*, 679–688.

Murty, S., Russell, R. R., & Levkoff, S. B. (2012). On modeling pollution-generating technologies. *Journal of Environmental Economics and Management*, *64*, 117–135.

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

Peyrache, A. & Coelli, T. (2009). A multiplicative directional distance function. Working Paper WP02/2009, Centre for Efficiency and Productivity Analysis, University of Queensland.

Reinhard, S., Lovell, C. A. K., & Thijssen, G. (1999). Econometric estimation of technical and environmental efficiency: An application to Dutch dairy farms. *American Journal of Agricultural Economics*, *81*, 44–60.

Reinhard, S., Lovell, C. A. K., & Thijssen, G. (2000). Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA. *European Journal of Operational Research*, *121*, 287–303.

Restrepo-Tobón, D. A. & Kumbhakar, S. C. (2015). Nonparametric estimation of returns to scale using input distance functions: An application to large U.S. banks. *Empirical Economics*, *48*, 143–168.

Salnykov, M. & Zelenyuk, V. (2005). On the commensurability of directional distance functions. Technical Report 0529, IAP Statistics Network, Université Catholique de Louvain.

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

Shephard, R. W. (1953). *Cost and Production Functions*. Princeton, NJ: Princeton University Press.

Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton, NJ: Princeton University Press.

Tsionas, E. G., Malikov, E., & Kumbhakar, S. C. (2015). Do direction, normalization and the Jacobian matter in the productivity measurements based on the directional output distance functions? Working

Paper, Binghamton University.

Valadkhani, A., Roshdi, I., & Smyth, R. (2016). A multiplicative environmental DEA approach to measure efficiency changes in the world's major polluters. *Energy Economics*, *54*, 363–375.

Wheelock, D. C. & Wilson, P. W. (2001). New evidence on returns to scale and product mix among US commercial banks. *Journal of Monetary Economics*, *47*(3), 653–674.

Wheelock, D. C. & Wilson, P. W. (2012). Do large banks have lower costs? New estimates of returns to scale for US banks. *Journal of Money, Credit and Banking*, *44*(1), 171–199.

Supplementary Appendix

A Theoretical Properties of the ADTDF

The additive directional technology distance function $\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d})$ in (2.2) satisfies the following theoretical properties:[16]

(1) $\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) \geq 0 \iff (\mathbf{x},\mathbf{y},\mathbf{b}) \in \mathbb{T}$, where the zero value of $\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d})$ implies full technological efficiency; and $\vec{D}_{\mathrm{A}}(\mathbf{x},\kappa\mathbf{y},\kappa\mathbf{b};\boldsymbol{d}) \geq 0$ for $(\mathbf{x},\mathbf{y},\mathbf{b}) \in \mathbb{T}$ and $0 \leq \kappa \leq 1$;

(2) the translation property: for $\alpha \in \mathbb{R}$, we have that

$$\vec{D}_{\mathrm{A}}(\mathbf{x}-\alpha\boldsymbol{d}_x,\mathbf{y}+\alpha\boldsymbol{d}_y,\mathbf{b}-\alpha\boldsymbol{d}_b;\boldsymbol{d}) = \vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) - \alpha; \tag{A.1}$$

(3) homogeneity of degree minus one in $\boldsymbol{d}$: $\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\gamma\boldsymbol{d}) = \gamma^{-1}\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d})$ for any $\gamma > 0$;

(4) concavity of $\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d})$ in $(\mathbf{x},\mathbf{y},\mathbf{b}) \in \mathbb{T}$;

(5) positive monotonicity in inputs: if $(\mathbf{x}',\mathbf{y},\mathbf{b}) \geq (\mathbf{x},\mathbf{y},\mathbf{b}) \in \mathbb{T}$, then $\vec{D}_{\mathrm{A}}(\mathbf{x}',\mathbf{y},\mathbf{b};\boldsymbol{d}) \geq \vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d})$;

(6) negative monotonicity in good outputs: if $(\mathbf{x},\mathbf{y}',\mathbf{b}) \leq (\mathbf{x},\mathbf{y},\mathbf{b}) \in \mathbb{T}$, then $\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y}',\mathbf{b};\boldsymbol{d}) \geq \vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d})$;

(7) positive monotonicity in bad outputs: if $(\mathbf{x},\mathbf{y},\mathbf{b}') \geq (\mathbf{x},\mathbf{y},\mathbf{b}) \in \mathbb{T}$, then $\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b}';\boldsymbol{d}) \geq \vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d})$.

We can establish the formal link between the ADTDF and its multiplicative counterpart given in (2.3). Using simple algebra, one can show the connection between the two directional distance functions taking the following form:

$$\begin{aligned}
\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) \cdot \boldsymbol{d}_x &= \left(\mathbf{I}_J - \left[\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{g})\right]^{\mathrm{diag}\{\boldsymbol{g}_x\}}\right)\mathbf{x} \\
\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) \cdot \boldsymbol{d}_y &= \left(\left[\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{g})\right]^{-\mathrm{diag}\{\boldsymbol{g}_y\}} - \mathbf{I}_M\right)\mathbf{y} \\
\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) \cdot \boldsymbol{d}_b &= \left(\mathbf{I}_P - \left[\vec{D}_{\mathrm{M}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{g})\right]^{\mathrm{diag}\{\boldsymbol{g}_b\}}\right)\mathbf{b},
\end{aligned} \tag{A.2}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix.

Analogous to the derivation of the MDTDF-based productivity growth index, letting the time enter the ADTDF directly and recognizing that $(\mathbf{x},\mathbf{y},\mathbf{b})$ are all changing over time, the total differentiation of $\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b},t;\boldsymbol{d})$ with respect to t yields

$$\frac{\mathrm{d}\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b},t;\boldsymbol{d})}{\mathrm{d}t} = \sum_j \frac{\partial\vec{D}_{\mathrm{A}}(\cdot)}{\partial\log x_j}\dot{x_j} + \sum_m \frac{\partial\vec{D}_{\mathrm{A}}(\cdot)}{\partial\log y_m}\dot{y_m} + \sum_p \frac{\partial\vec{D}_{\mathrm{A}}(\cdot)}{\partial\log b_p}\dot{b_p} + \frac{\partial\vec{D}_{\mathrm{A}}(\cdot)}{\partial t}, \tag{A.3}$$

where the "dot" above a variable denotes the time derivative of its log. Dividing both sides of (A.3) by (non-zero) $\left(1+\vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b},t;\boldsymbol{d})\right)$ and some rearranging yields an equivalent expression:

$$\mathrm{PG}_{\mathrm{A}} \stackrel{\mathrm{def}}{=} -\sum_j \frac{\partial\log\left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)}{\partial\log x_j}\dot{x_j} - \sum_m \frac{\partial\log\left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)}{\partial\log y_m}\dot{y_m} - \sum_p \frac{\partial\log\left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)}{\partial\log b_p}\dot{b_p}$$

[16]The proofs are very similar to those in Chambers et al. (1996).

$$= \frac{\partial \log \left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)}{\partial t} - \frac{\mathrm{d} \log \left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)}{\mathrm{d}t}\ , \tag{A.4}$$

where the left-hand side of the equality is a Solow-type Divisia index of productivity growth which we label $\mathrm{PG_A}$, and the right-hand side of the equality yields a natural decomposition of PG_a into technical change $\mathrm{TC_A} \equiv \partial \log \left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)/\partial t$ and efficiency change $\mathrm{EC_A} \equiv -\mathrm{d} \log \left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)/\mathrm{d}t$. The ADTDF-based productivity index constitutes a weighted aggregate of percentage growth rates in inputs and desirable and undesirable outputs: $\dot{x_j}\ \forall\ j$, $\dot{y_m}\ \forall\ m$ and $\dot{b_p}\ \forall\ p$, respectively. The corresponding weights are $\partial \log \left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)/\partial \log x_j \le 0\ \forall\ j=1,\ldots,J$, $\partial \log \left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)/\partial \log y_m \ge 0$ $\forall\ m=1,\ldots,M$ and $\partial \log \left(1+\vec{D}_{\mathrm{A}}(\cdot)\right)/\partial \log b_p \le 0\ \forall\ p=1,\ldots,P$, where the signs are dictated by the monotonicity properties of the ADTDF in (2.2).

B Estimation of the ADTDF

Following Guarda et al. (2013) and Malikov et al. (2015b), we normalize the stochastic ADTDF in (2.2) in order to impose the translation property onto it for a given direction $\boldsymbol{d}$.[17] Specifically, setting α equal to one of the bad outputs, say $\alpha = b_p$, and normalizing the corresponding $d_{b_p}=1$, the translation property in (A.1) can be rewritten as

$$\vec{D}_{\mathrm{A}}\left(\mathbf{x}-b_p\boldsymbol{d}_x, \mathbf{y}+b_p\boldsymbol{d}_y, \widehat{\mathbf{b}}-b_p\widehat{\boldsymbol{d}}_b; \boldsymbol{d}\right) = \vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) - b_p, \tag{B.1}$$

where $\widehat{\boldsymbol{d}}_b \stackrel{\text{def}}{=} (d_{b_1},...,d_{b_{p-1}},d_{b_{p+1}},\ldots,d_{b_P})$ and $\widehat{\mathbf{b}}$ is as defined earlier. Note that, due to the normalization $d_{b_p}=1$, the bad output b_p now enters the left-hand side of (B.1) in the capacity of "α" from (A.1) only. The "output" b_p has however disappeared from $\vec{D}_{\mathrm{A}}(\mathbf{x}-b_p\boldsymbol{d}_x, \mathbf{y}+b_p\boldsymbol{d}_y, \widehat{\mathbf{b}}-b_p\widehat{\boldsymbol{d}}_b; \boldsymbol{d})$ because $b_p+\alpha d_{b_p} = b_p - b_p\cdot 1 = 0$. After rearranging, from (B.1) we get

$$\begin{aligned} b_p &= -\vec{D}_{\mathrm{A}}\left(\mathbf{x}-b_p\boldsymbol{d}_x, \mathbf{y}+b_p\boldsymbol{d}_y, \widehat{\mathbf{b}}-b_p\widehat{\boldsymbol{d}}_b; \boldsymbol{d}\right) + \vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) \\ &\stackrel{\text{def}}{=} -\vec{D}_{\mathrm{A}}\left(\mathbf{x}-b_p\boldsymbol{d}_x, \mathbf{y}+b_p\boldsymbol{d}_y, \widehat{\mathbf{b}}-b_p\widehat{\boldsymbol{d}}_b; \boldsymbol{d}\right) + u_{\mathrm{A}}, \end{aligned} \tag{B.2}$$

where $u_{\mathrm{A}} \stackrel{\text{def}}{=} \vec{D}_{\mathrm{A}}(\mathbf{x},\mathbf{y},\mathbf{b};\boldsymbol{d}) \ge 0$ represents the (unobserved) technological *in*efficiency. After adding the white noise term, the normalized ADTDF in (B.2) constitutes a standard stochastic frontier model. In order to estimate this stochastic ADTDF, one needs to specify its functional form. The quadratic form is the most common choice for the additive directional distance functions (e.g., Färe et al., 2005; Hudgins & Primont, 2007; Feng & Serletis, 2014; Malikov et al., 2015b; Atkinson & Tsionas, 2015). The quadratic specification owes its wide popularity in the literature to the fact that it can be easily restricted to satisfy the translation property via the set of parameter restrictions. In our instance, the translation property is imposed via the normalization. Hence, nothing seemingly precludes us from choosing other functional forms such as a quasi-translog specification[18] for the normalized ADTDF in (B.2) except in the instances when the directional vector $\boldsymbol{d}$ is such that some (or all) of the normalized inputs and outputs $(\mathbf{x}-b_p\boldsymbol{d}_x, \mathbf{y}+b_p\boldsymbol{d}_y, \widehat{\mathbf{b}}-b_p\widehat{\boldsymbol{d}}_b)$ take non-positive

[17] Note that the translation property may alternatively be imposed using the set of parameter restrictions applied to the ADTDF directly during the estimation (e.g., see Atkinson, Primont & Tsionas, 2014).

[18] The model would be *quasi*-translog because the left-hand-side variable is in the level, not the log, form.

values, which the log cannot work with. Following the tradition in the literature, here we employ the quadratic parameterization of the ADTDF:

$$\begin{aligned}
b_p = -\theta_0 - \sum_j \alpha_j \widetilde{x}_j - \sum_m \beta_m \widetilde{y}_m - \sum_{k(\neq p)} \gamma_k \widetilde{b}_k - \theta_t t - \\
\frac{1}{2}\sum_j \sum_{j'} \alpha_{jj'} \widetilde{x}_j \widetilde{x}_{j'} - \frac{1}{2}\sum_m \sum_{m'} \beta_{mm'} \widetilde{y}_m \widetilde{y}_{m'} - \frac{1}{2}\sum_{k(\neq p)} \sum_{k'(\neq p)} \gamma_{kk'} \widetilde{b}_k \widetilde{b}_{k'} - \frac{1}{2}\theta_{tt} t^2 - \\
\sum_j \sum_m \delta_{jm} \widetilde{x}_j \widetilde{y}_m - \sum_j \sum_{k(\neq p)} \phi_{jk} \widetilde{x}_j \widetilde{b}_k - \sum_m \sum_{k(\neq p)} \varphi_{mk} \widetilde{y}_m \widetilde{b}_k - \\
\sum_j \theta_{x,j} t \widetilde{x}_j - \sum_m \theta_{y,m} t \widetilde{y}_m - \sum_{k(\neq p)} \theta_{b,k} t \widetilde{b}_k + u_{\mathrm{A}},
\end{aligned} \tag{B.3}$$

where, for convenience, we have defined $\widetilde{x}_j \stackrel{\text{def}}{=} x_j - b_p d_{x_j}\ \forall\ j = 1, \ldots, J$, $\widetilde{y}_m \stackrel{\text{def}}{=} y_m + b_p d_{y_m}\ \forall\ m = 1, \ldots, M$ and $\widetilde{b}_k \equiv b_k - b_p d_{b_k}\ \forall\ k(\neq p) = 1, \ldots, P$.

We let the normalized directional vector $\widehat{\boldsymbol{d}}$ be an unknown (non-negative) nonparametric function of some R exogenous contextual variables $\mathbf{z}_{it}$ capturing bank's characteristics. We approximate this directional vector via (complete) S_n-order polynomial sieves:

$$\widehat{d}_\kappa(\mathbf{z}_{it}) \approx \exp\left(\sum_{\sum_r s_r = S_n;\ s_r \geq 0} \mu_{\kappa, s_1, \ldots, s_R} z_{1,it}^{s_1} \cdots z_{R,it}^{s_R} \right) \quad \forall\ \kappa = 1, \ldots, (J + M + P - 1). \tag{B.4}$$

We estimate the ADTDF in (B.3) [joint with (B.4)] subject to the Slutsky symmetry as well as the theoretical monotonicity and curvature restrictions.[19] Monotonicity of the (nonlinear) quadratic ADTDF in (B.3) in inputs and desirable and undesirable outputs, respectively, requires that

$$\frac{\partial \vec{D}_{\mathrm{A}}(\cdot)}{\partial x_j} = \alpha_j + \sum_{j'} \alpha_{jj'} \widetilde{x}_{j'} + \sum_m \delta_{jm} \widetilde{y}_m + \sum_{k(\neq p)} \phi_{jk} \widetilde{b}_k + \theta_{x,j} t \geq 0 \quad \forall \quad j$$

$$\frac{\partial \vec{D}_{\mathrm{A}}(\cdot)}{\partial y_m} = \beta_m + \sum_{m'} \beta_{mm'} \widetilde{y}_{m'} + \sum_j \delta_{jm} \widetilde{x}_j + \sum_{k(\neq p)} \varphi_{mk} \widetilde{b}_k + \theta_{y,m} t \leq 0 \quad \forall \quad m$$

$$\frac{\partial \vec{D}_{\mathrm{A}}(\cdot)}{\partial b_k} = \begin{cases} \gamma_k + \sum_{k'(\neq p)} \gamma_{kk'} \widetilde{b}_{k'} + \sum_j \phi_{jk} \widetilde{x}_j + \sum_m \varphi_{mp} \widetilde{y}_m + \theta_{b,k} t \geq 0 \quad \forall \quad k(\neq p) \\ \\ 1 - \sum_j \alpha_j d_{x_j} + \sum_m \beta_m d_{y_m} - \sum_{k(\neq p)} \gamma_k d_{b_k} - \\ \sum_j \sum_{j'} \alpha_{jj'} \left(\widetilde{x}_j d_{x_{j'}} + \widetilde{x}_{j'} d_{x_j} \right) + \sum_m \sum_{m'} \beta_{mm'} \left(\widetilde{y}_m d_{y_{m'}} + \widetilde{y}_{m'} d_{y_m} \right) - \\ \sum_{k(\neq p)} \sum_{k'(\neq p)} \gamma_{kk'} \left(\widetilde{b}_k d_{b_{k'}} + \widetilde{b}_{k'} d_{b_k} \right) + \sum_j \sum_m \delta_{jm} \left(\widetilde{x}_j d_{y_m} - \widetilde{y}_m d_{x_j} \right) - \\ \sum_j \sum_{k(\neq p)} \phi_{jk} \left(\widetilde{x}_j d_{b_k} + \widetilde{b}_k d_{x_j} \right) + \sum_m \sum_{k(\neq p)} \varphi_{mk} \left(\widetilde{b}_k d_{y_m} - \widetilde{y}_m d_{b_k} \right) - \\ \sum_j \theta_{x,j} t d_{x_j} + \sum_m \theta_{y,m} t d_{y_m} - \sum_{k(\neq p)} \theta_{b,k} t d_{b_k} \geq 0 \quad \text{if} \quad k = p. \end{cases}$$

The concavity of the ADTDF is imposed by ensuring that all odd-numbered (even-numbered) leading principal minors of the Hessian matrix for (B.3) are non-positive (non-negative). The

[19] Recall that the almost homogeneity property has already been imposed via the normalization.

Hessian matrix is given by

$$\begin{bmatrix}
\alpha_{11} & \dots & \alpha_{1J} & \delta_{11} & \dots & \delta_{1M} & \phi_{11} & \dots & \phi_{1(P-1)} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial x_1 \partial b_P} \\
\vdots & & & & & & & & & \vdots \\
\alpha_{J1} & \dots & \alpha_{JJ} & \delta_{J1} & \dots & \delta_{JM} & \phi_{J1} & \dots & \phi_{J(P-1)} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial x_J \partial b_P} \\
\delta_{11} & \dots & \delta_{J1} & \beta_{11} & \dots & \beta_{1M} & \varphi_{11} & \dots & \varphi_{1(P-1)} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial y_1 \partial b_P} \\
\vdots & & & & & & & & & \vdots \\
\delta_{1M} & \dots & \delta_{JM} & \beta_{M1} & \dots & \beta_{MM} & \varphi_{M1} & \dots & \varphi_{M(P-1)} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial y_M \partial b_P} \\
\phi_{11} & \dots & \phi_{J1} & \varphi_{11} & \dots & \varphi_{M1} & \gamma_{11} & \dots & \gamma_{1(P-1)} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial b_1 \partial b_P} \\
\vdots & & & & & & & & & \vdots \\
\phi_{1(P-1)} & \dots & \phi_{J(P-1)} & \varphi_{1(P-1)} & \dots & \varphi_{M(P-1)} & \gamma_{(P-1)1} & \dots & \gamma_{(P-1)(P-1)} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial b_{(P-1)} \partial b_P} \\
\frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial x_1 \partial b_P} & \dots & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial x_J \partial b_P} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial y_1 \partial b_P} & \dots & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial y_M \partial b_P} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial b_1 \partial b_P} & \dots & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial b_{(P-1)} \partial b_P} & \frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial b_P^2}
\end{bmatrix}$$

where, for the ease of exposition, we have set the order number of a normalizing bad output to equal P, i.e., $b_p = b_P$, and

$$\begin{aligned}
\frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial x_j \partial b_P} &= -\sum_{j'} \alpha_{jj'} d_{x_{j'}} + \sum_m \delta_{jm} d_{y_m} - \sum_{k(\neq p)} \phi_{jk} d_{b_k} \quad \forall \quad j \\
\frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial y_m \partial b_P} &= \sum_{m'} \beta_{mm'} d_{y_{m'}} - \sum_j \delta_{jm} d_{x_j} - \sum_{k(\neq p)} \varphi_{mk} d_{b_k} \quad \forall \quad m \\
\frac{\partial^2 \vec{D}_{\mathrm{A}}(\cdot)}{\partial b_k \partial b_P} &= \begin{cases} -\sum_{k'(\neq p)} \gamma_{kk'} d_{b_{k'}} - \sum_j \phi_{jk} d_{x_j} + \sum_m \varphi_{mp} d_{y_m} \quad \forall \quad k(\neq p) \\ \\ 2\sum_j \sum_{j'} \alpha_{jj'} d_{x_j} d_{x_{j'}} + 2\sum_m \sum_{m'} \beta_{mm'} d_{y_m} d_{y_{m'}} + \\ 2\sum_{k(\neq p)} \sum_{k'(\neq p)} \gamma_{kk'} d_{b_k} d_{b_{k'}} - 2\sum_j \sum_m \delta_{jm} d_{x_j} d_{y_m} + \\ 2\sum_j \sum_{k(\neq p)} \phi_{jk} d_{x_j} d_{b_k} - 2\sum_m \sum_{k(\neq p)} \varphi_{mk} d_{b_k} d_{y_m} \quad \text{if} \quad k = p. \end{cases}
\end{aligned}$$

Lastly, the following (observation-specific) Jacobian of the transformation is explicitly incorporated in the log-likelihood function during the estimation:

$$\begin{aligned}
J_{\mathrm{A}} = \frac{\partial(v_{\mathrm{A}} + u_{\mathrm{A}})}{\partial b_p} = 1 &- \sum_j \alpha_j d_{x_j} + \sum_m \beta_m d_{y_m} - \sum_{k(\neq p)} \gamma_k d_{b_k} - \\
& \sum_j \sum_{j'} \alpha_{jj'} \left(\widetilde{x}_j d_{x_{j'}} + \widetilde{x}_{j'} d_{x_j}\right) + \sum_m \sum_{m'} \beta_{mm'} \left(\widetilde{y}_m d_{y_{m'}} + \widetilde{y}_{m'} d_{y_m}\right) - \\
& \sum_{k(\neq p)} \sum_{k'(\neq p)} \gamma_{kk'} \left(\widetilde{b}_k d_{b_{k'}} + \widetilde{b}_{k'} d_{b_k}\right) + \sum_j \sum_m \delta_{jm} \left(\widetilde{x}_j d_{y_m} - \widetilde{y}_m d_{x_j}\right) - \\
& \sum_j \sum_{k(\neq p)} \phi_{jk} \left(\widetilde{x}_j d_{b_k} + \widetilde{b}_k d_{x_j}\right) + \sum_m \sum_{k(\neq p)} \varphi_{mk} \left(\widetilde{b}_k d_{y_m} - \widetilde{y}_m d_{b_k}\right) - \\
& \sum_j \theta_{x,j} t d_{x_j} + \sum_m \theta_{y,m} t d_{y_m} - \sum_{k(\neq p)} \theta_{b,k} t d_{b_k}. \qquad \text{(B.5)}
\end{aligned}$$

Because the ADTDF is an additive measure of the distance to the frontier and is *not* invariant to the units of measurement of its arguments $(\mathbf{x}, \mathbf{y}, \mathbf{b})$, following Färe et al. (2005) and Malikov et al. (2015b) we standardize all variables prior to the estimation by subtracting their respective sample means and dividing by their sample standard deviations. We however note that, due its proportional nature, the estimation of the MDTDF does not require such a standardization.

C Bayesian Estimation

In the case of the MDTDF, our objective is to provide statistical inferences for (3.3), possibly along with (3.4), subject to the theoretical monotonicity and curvature regularity conditions provided in Section 3.2 and the Jacobian in (3.5). In the instance of the additive distance function, the model is given by (B.3) and (B.4) subject to the corresponding theoretical regularity conditions and the Jacobian in (B.5). In either case, the model in (3.3) or (B.3) can be written as

$$y_{it} = f(\mathbf{x}_{it}, \mathbf{z}_{it}; \boldsymbol{\theta}) + v_{it} + u_{it}, \tag{C.1}$$

where y_{it} and $\mathbf{x}_{it}$ are the left-hand-side and d_x right-hand-side variables in the normalized directional distance function, respectively; $\mathbf{z}_{it}$ is the vector of d_z contextual variables; $\boldsymbol{\theta}$ is the $d_\theta \times 1$ vector of all unknown parameters of the distance function (including fixed directional parameters or the parameters in sieve approximations of varying directional functions); v_{it} is the *i.i.d.* normally distributed mean-zero random error; and u_{it} is the one-sided inefficiency term such that $u_{it} \sim \mathbb{N}_+(\mu_{it}, \sigma_u^2)$, where $\mu_{it} = a_0 + a_1 t + \frac{1}{2} a_2 t^2$.

We augment (C.1) with a reduced form to take account of the potential endogeneity of $\mathbf{x}_{it}$:

$$\mathbf{x}_{it} = \boldsymbol{\Pi}\boldsymbol{\zeta}_{it} + \mathbf{V}_{it}, \tag{C.2}$$

where $\boldsymbol{\zeta}_{it}$ is a $d_\zeta \times 1$ vector of polynomial basis functions of the contextual variables $\mathbf{z}_{it}$, and $\boldsymbol{\Pi}$ is the corresponding $d_x \times d_\zeta$ matrix of reduced-form parameters.[20] Both (C.1) and (C.2) are modeled jointly under the following assumption:

$$[v_{it}, \mathbf{V}'_{it}]' \;\sim\; \mathbb{N}_M(\mathbf{0}, \boldsymbol{\Sigma}). \tag{C.3}$$

We use the semiparametric sieve approach where the order of the sieve is determined by the marginal likelihood (for details see below). The posterior distribution of $\boldsymbol{\theta}$ may then be written as

$$p(\boldsymbol{\theta}|\boldsymbol{\Xi}) \propto \left\{ \prod_{i=1}^{n} \prod_{t=1}^{T} l_{it}(\boldsymbol{\theta}; \boldsymbol{\Xi}_{it}) J_{it}(\boldsymbol{\theta}; \boldsymbol{\Xi}_{it}) \right\} p(\boldsymbol{\theta}), \tag{C.4}$$

where $\boldsymbol{\Xi}$ generically denotes the data; $l_{it}(\boldsymbol{\theta}; \boldsymbol{\Xi}_{it})$ is the likelihood contribution for observation (i, t); $J_{it}(\boldsymbol{\theta}; \boldsymbol{\Xi}_{it})$ denotes the Jacobian of transformation; and $p(\boldsymbol{\theta})$ denotes the prior. The generic form of the likelihood contribution is

$$l_{it}(\boldsymbol{\theta}; \boldsymbol{\Xi}_{it}) = (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} \left[y_{it} - f(\mathbf{x}_{it}, \mathbf{z}_{it}; \boldsymbol{\theta}) \right]^2 \right\} \Phi\left(\lambda \frac{y_{it} - f(\mathbf{x}_{it}, \mathbf{z}_{it}; \boldsymbol{\theta})}{\sigma} \right), \tag{C.5}$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \frac{\sigma_u}{\sigma_v}$ when we do not take into account the reduced form in (C.2). If we do incorporate the reduced form, we can integrate analytically with respect to u_{it}, and the final expression is given by

$$l_{it}(\boldsymbol{\theta}; \boldsymbol{\Xi}_{it}) = (2\pi)^{-d_x/2} |\boldsymbol{\Sigma}|^{-1/2} (2\pi\sigma_u^2)^{-1/2} \Phi\left(\frac{\mu_{it}}{\sigma_u} \right)^{-1} \times \tag{C.6}$$

[20] Thus, our strategy of tackling endogeneity of $\mathbf{x}_{it}$ replies on the assumed exogeneity of $\mathbf{z}_{it}$.

$$\exp\left\{-\frac{1}{2}\left(\sigma^{11}+\frac{1}{2\sigma_u^2}\right)\left[(u_{it}-A_{it})^2+B_{it}-A_{it}^2\right]\right\}, \tag{C.7}$$

where $\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \sigma^{11} & \boldsymbol{\sigma}_1' \\ \boldsymbol{\sigma}_1 & \boldsymbol{\Sigma}^{11} \end{bmatrix}$, $A_{it} = \frac{\sigma^{11}r_{it}+\mathbf{V}_{it}'\boldsymbol{\sigma}_1+\frac{\mu_{it}}{2\sigma_u^2}}{\sigma^{11}+\frac{1}{2\sigma_u^2}}$, $B_{it} = \frac{\sigma^{11}r_{it}^2+2\mathbf{V}_{it}'\boldsymbol{\sigma}_1 r_{it}+\mathbf{V}_{it}'\boldsymbol{\Sigma}^{11}\mathbf{V}_{it}+\frac{1}{2\sigma_u^2}\mu_{it}^2}{\sigma^{11}+\frac{1}{2\sigma_u^2}}$, and $r_{it} = y_{it} - f(\mathbf{x}_{it}, \mathbf{z}_{it}; \boldsymbol{\theta})$.

In our application, the prior is flat over the domain $\mathscr{D}$ of the distance function's theoretical regularity restrictions, i.e.,

$$p(\boldsymbol{\theta}, \sigma) \propto \sigma^{-1}\mathbb{I}(\boldsymbol{\theta} \in \mathscr{D}), \tag{C.8}$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Due to the presence of regularity restrictions and the non-unit Jacobian of transformation, the likelihood is highly nonlinear thereby making the use of a Gibbs sampler be inefficient. To impose the restrictions, we therefore adopt the following strategy. Let $\overline{\mathbf{d}}$ and $\mathbf{s}$ denote sample means and standard deviations of the data $\boldsymbol{\Xi}$. We impose all restrictions at points corresponding to $\overline{\mathbf{d}} \pm h\mathbf{s}$ for h taking values from 0 to $\overline{h}$ at the 0.1 increments. Then we check whether the restrictions are satisfied at the remaining sample points. Setting $\overline{h} = 3.3$ yields acceptance of restrictions at all data points for the MDTDF; for the ADTDF, $\overline{h} = 3$.

We use the Girolami & Calderhead (2011) [hereafter GC] algorithm to update draws for $\boldsymbol{\theta}$. The algorithm uses local information about both the gradient and the Hessian of the log-posterior conditional of $\boldsymbol{\theta}$ at the existing draw. The Metropolis test is used for accepting the candidate draws. The GC algorithm is started from random initial conditions, and MCMC is executed until the convergence. Depending on the model and the subsample this takes from 2,000 to 4,000 iterations. For safety, we run 10,000 iterations. Then, we run another 50,000 MCMC iterations to obtain final results for posterior moments and densities of the parameters and functions of interest. We find that the GC algorithm performs vastly superior relative to the standard Metropolis-Hastings (MH) algorithm and that autocorrelations are much smaller.

Let $\mathscr{L}(\boldsymbol{\theta}) = \log p(\boldsymbol{\theta}|\boldsymbol{\Xi})$ denote the log-posterior of $\boldsymbol{\theta}$. Also, define $\boldsymbol{G}(\boldsymbol{\theta}) = \text{est.cov}\frac{\partial}{\partial\boldsymbol{\theta}}\log p(\boldsymbol{\Xi}|\boldsymbol{\theta})$ to be the empirical counterpart of $\boldsymbol{G}_o(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log p(\boldsymbol{\Xi}|\boldsymbol{\theta})$. Then, the Langevin diffusion is given by the following stochastic differential equation:

$$d\boldsymbol{\theta}(t) = \frac{1}{2}\widetilde{\nabla}_{\boldsymbol{\theta}}\mathscr{L}(\boldsymbol{\theta}(t))\,dt + d\mathbf{B}(t), \tag{C.9}$$

where

$$\widetilde{\nabla}_{\boldsymbol{\theta}}\mathscr{L}(\boldsymbol{\theta}(t)) = -\boldsymbol{G}^{-1}(\boldsymbol{\theta}(t)) \times \nabla_{\boldsymbol{\theta}}\mathscr{L}(\boldsymbol{\theta}(t)), \tag{C.10}$$

is the so called "natural gradient" of the Riemann manifold generated by the log-posterior with $\nabla_{\boldsymbol{\theta}}\mathscr{L}(\boldsymbol{\theta}(t))$ being the usual gradient of $\mathscr{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. The elements of the Brownian motion are

$$\boldsymbol{G}^{-1}(\boldsymbol{\theta}(t))\,d\mathbf{B}_i(t) = |\boldsymbol{G}(\boldsymbol{\theta}(t))|^{-1/2}\sum_{j=1}^{d_\theta}\frac{\partial}{\partial\boldsymbol{\theta}}\left[\boldsymbol{G}^{-1}(\boldsymbol{\theta}(t))_{ij}\,|\boldsymbol{G}(\boldsymbol{\theta}(t))|^{1/2}\right]dt + \left[\sqrt{\boldsymbol{G}(\boldsymbol{\theta}(t))}d\boldsymbol{B}(t)\right]_i. \tag{C.11}$$

The discrete form of the stochastic differential equation provides a proposal as follows

$$\widetilde{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i^o + \frac{\varepsilon^2}{2}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\nabla_{\boldsymbol{\theta}}\mathscr{L}(\boldsymbol{\theta}^o)\right\}_i - \varepsilon^2\sum_{j=1}^{d_\theta}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\frac{\partial\boldsymbol{G}(\boldsymbol{\theta}^o)}{\partial\boldsymbol{\theta}_j}\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\right\}_{ij} + \tag{C.12}$$

$$\frac{\varepsilon^2}{2}\sum_{j=1}^{d_\theta}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\right\}_{ij}\operatorname{tr}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\frac{\partial\boldsymbol{G}(\boldsymbol{\theta}^o)}{\partial\boldsymbol{\theta}_j}\right\}+\left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)}\boldsymbol{\xi}^o\right\}_i$$
$$\equiv\boldsymbol{\mu}(\boldsymbol{\theta}^o,\varepsilon)_i+\left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)}\boldsymbol{\xi}^o\right\}_i,$$

where $\boldsymbol{\theta}^o$ is the current draw, and ϵ is some constant. The proposal density is

$$q\left(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^o\right)=\mathbb{N}_{d_\theta}\left(\widetilde{\boldsymbol{\theta}},\varepsilon^2\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\right),\tag{C.13}$$

and the convergence to the invariant distribution is ensured by using the standard form MH probability, i.e.,

$$\min\left\{1,\ \frac{p\left(\widetilde{\boldsymbol{\theta}}|\cdot,\boldsymbol{\Xi}\right)q\left(\boldsymbol{\theta}^o|\widetilde{\boldsymbol{\theta}}\right)}{p\left(\boldsymbol{\theta}^o|\cdot,\boldsymbol{\Xi}\right)q\left(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^o\right)}\right\}.\tag{C.14}$$

Lastly, we select the order of sieve approximation for unknown varying directional functions as well as choose across alternative model specifications using marginal likelihoods and Bayes factors. Specifically, given a model for which the likelihood function and the prior are $L(\boldsymbol{\theta};\boldsymbol{\Xi})$ and $p(\boldsymbol{\theta})$, respectively, the posterior distribution for $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ is

$$p(\boldsymbol{\theta}|\boldsymbol{\Xi})=\frac{L(\boldsymbol{\theta};\boldsymbol{\Xi})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}}L(\boldsymbol{\theta};\boldsymbol{\Xi})p(\boldsymbol{\theta})d\boldsymbol{\theta}},$$

from where we have the identity known as the “candidate’s formula”:

$$\mathcal{M}(\boldsymbol{\Xi})=\frac{L(\boldsymbol{\theta};\boldsymbol{\Xi})p(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{\Xi})}\quad\forall\boldsymbol{\theta}\in\boldsymbol{\Theta},$$

where the marginal likelihood is $\mathcal{M}(\mathbf{X})=\int_{\boldsymbol{\Theta}}L(\boldsymbol{\theta};\boldsymbol{\Xi})p(\boldsymbol{\theta})d\boldsymbol{\theta}$. Since the identity holds for every $\boldsymbol{\theta}$, we can choose the posterior mean $\overline{\boldsymbol{\theta}}$ and assume that the posterior is nearly normal with mean $\overline{\boldsymbol{\theta}}$ and covariance $\mathbf{V}$ which can be computed from the MCMC draws. Therefore, the log marginal likelihood can be computed approximately as

$$\log\mathcal{M}(\boldsymbol{\Xi})\simeq\log L\left(\overline{\boldsymbol{\theta}};\boldsymbol{\Xi}\right)+\log p\left(\overline{\boldsymbol{\theta}}\right)+\frac{d_\theta}{2}\log(2\pi)+\frac{1}{2}\log|\mathbf{V}|.\tag{C.15}$$

To discriminate between any two models, we rely on the Bayes factor which we compute using their respective marginal likelihoods. More specifically, for two models whose marginal likelihoods are $\mathcal{M}_1(\boldsymbol{\Xi})$ and $\mathcal{M}_2(\boldsymbol{\Xi})$, the Bayes factor in favor of the first model and against the second is $BF_{1:2}=\frac{\mathcal{M}_1(\boldsymbol{\Xi})}{\mathcal{M}_2(\boldsymbol{\Xi})}$.