Bayesian Estimation of a Semiparametric Stochastic Frontier Model with Persistent and Transient Inefficiencies

Puguang Nie[1,2], Christopher F. Parmeter[3], Valentin Zelenyuk[4], Xibin Zhang[2]

[1]Faculty of Science and Technology, Beijing Normal – Hong Kong Baptist University, Zhuhai, China

[2]Department of Econometrics and Business Statistics, Monash University, Australia

[3]Miami Herbert Business School, University of Miami, Miami FL, USA

[4]School of Economics, The University of Queensland, Brisbane, Australia

July 26, 2024
Revised: July 31, 2025

Abstract: We propose kernel estimators for the densities of persistent and transient inefficiencies in a generalized panel data stochastic frontier model. This approach relaxes the conventional distributional assumptions imposed on the inefficiencies, thereby aiming to increase robustness. A Bayesian sampling algorithm is developed to estimate all model parameters and unobserved entities, as well as the bandwidths needed to construct the densities of persistent and transient inefficiencies. The advantages of our estimator and sampling algorithm are demonstrated through numerical studies based on pseudo-samples. We find that our estimation approach outperforms conventional Bayesian estimation, providing more accurate estimates of firm level inefficiencies. We also apply the approach to analyze the cost frontier and related characteristics for a panel of large US commercial banks, and discuss differences relative to the traditional Bayesian framework. Our estimator reveals that the high-density region of transient inefficiency varies over time, whereas the conventional approach does not possess such a feature. Moreover, the contribution of persistent inefficiency to the total inefficiency revealed by our approach is slightly more than half of that suggested by the conventional framework.

Key words: Bayesian sampling, cost efficiency, high-density region, kernel conditional density estimation, random effects.

JEL classification: C11, C14, C23, G21

1 Introduction

Stochastic frontier models, introduced by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), have been extensively used in evaluating efficiency and productivity of firms across many industries (see, for example, Lee and Tyler, 1978; Greene, 2005, and recent reviews by Kumbhakar et al., 2022a, 2022b). A production frontier represents the maximal output that can be obtained from a given set of inputs while a cost frontier captures minimum costs given fixed output levels and input prices. To account for the maximal (minimal) nature of the frontier, a stochastic frontier model features a composite error term, which is a convolution of a symmetric component capturing statistical noise and a non-negative component, capturing inefficiency. To estimate firm-level inefficiencies and the frontier itself, certain distributional assumptions on the two components of the composite error are needed. The risk of distributional misspecification makes it essential to examine how inefficiencies are modeled in stochastic frontier frameworks.

The distribution of inefficiency is often specified as a member of a parametric family. Aigner et al. (1977) considered a Half-Normal distribution, while Meeusen and van den Broeck (1977) proposed specifying an Exponential distribution for inefficiency. Later, Stevenson (1980) proposed using a Truncated Normal distribution and Greene (1990) chose a general Gamma distribution. To estimate inefficiencies, Jondrow et al. (1982) suggested using a firm-specific estimator based on their conditional mean given the composite error (which itself depends on the distributional assumptions for both pieces of the composite error).

As a parametric assumption of the distribution of inefficiency is subjective, semi- and nonparametric estimation approaches of the distribution of inefficiency have been investigated. Schmidt and Sickles (1984) were the first to eschew distributional assumptions on inefficiency, but needed to require inefficiency to be time invariant, an onerous assumption with even moderately sized panels. Their approach is also limited by its inability to provide an absolute measure of inefficiency, instead comparing firms to a best practice firm. Park and Simar (1994) suggested a nonparametric specification of inefficiency, but left the frontier function parametrically specified. Their asymptotic results on the estimation of inefficiency require the panel to be sufficiently long. Park, Sickles and Simar (1998, 2003) continued studies in this direction with extensions allowing dependence between input variables and inefficiency. Tran and Tsionas (2009) derived a two-step procedure to estimate the parameters of the frontier using the least squares method and the conditional mean of inefficiency using a local linear estimator. Further, Kneip et al. (2012) assumed that inefficiency is a random function of time and estimated them via a linear combination of a small number of common functions calculated from the data. Under these nonparametric approaches, one could only estimate the mean of the individual level of inefficiency. This is one limitation that we will address in our work here.

A Bayesian approach has the advantage that non-observable inefficiencies as unknown identities, can be estimated through a sampling procedure, during which parameters are estimated. Bayesian methods have found a range of interesting implementations in the frontier estimation literature. Among the first attempts was the work of van den Broeck et al. (1994), who conducted Bayesian inference for a stochastic frontier model, implementing it through Bayes factors among different distribution assumptions (including Half-Normal, Truncated Normal, and Gamma distributions) on inefficiencies. They used Monte Carlo simulation together with importance sampling to carry out the required numerical integration. Koop et al. (1994) developed Bayesian techniques for analyzing cost efficiencies with a flexible form of the cost function and further described the Gibbs sampling method applied in stochastic frontier models (Koop et al., 1995). Shortly after, Koop et al. (1997) discussed a stochastic frontier model with fixed or random effects within a Bayesian framework. They postulated a flat prior for the coefficients in the frontier, the error variance and unobserved firm heterogeneity in the fixed effects panel stochastic frontier model while they used an informative prior for inefficiency in the random effects panel data model.

Fernández et al. (1997) examined whether posterior inference can be conducted with an improper prior, as well as provided theoretical foundations for Bayesian inference in stochastic frontier models. Tsionas (2000, 2001) specified a Normal-Gamma and a Truncated Normal distribution for inefficiencies and conducted Bayesian estimation of stochastic frontier models using the Gibbs sampler. However, both of these Bayesian studies assume a parametric distribution of inefficiencies, which may limit their flexibility and robustness in capturing the true inefficiency distribution in empirical applications.

To avoid parametric prior density assumptions on the inefficiency, Griffin and Steel (2004) proposed using a Dirichlet process based technique to estimate the distribution of inefficiency within a sampling framework. An important contribution of their approach was to combine the advantages of the nonparametric and Bayesian approaches in the estimation of inefficiencies and their distribution. In a recent study, Feng, Wang and Zhang (2019) proposed approximating the distribution of inefficiencies through a transformed Parzen-Rosenblatt kernel density estimator and using it as the prior density function. Although both approaches have relaxed parametric assumptions on the distribution of inefficiency, both methods estimate the distribution of inefficiency based on their own information, rather than information contained in the composite errors.[1]

Bayesian estimation of the four component model proposed by Colombi et al. (2014), initially developed by Tsionas and Kumbhakar (2014), is in its nascency. Further, semi- or nonparametric estimation of this model is even sparser. Here, building upon and expanding the ideas of Feng et al. (2019), we seek to use recently developed insights on the Bayesian construction of the inefficiency distributions, both persistent and transient, to estimate the four component model in a straightforward approach. In our study, we present a new nonparametric estimator for the two inefficiency distributions using information contained in the realized composite errors and the unobserved heterogeneity. These are the key differences between our approach to estimation and the method proposed by Feng et al. (2019).

Our starting point is that time-varying inefficiency is a part of the composite error, and the estimation of inefficiency and its density rely on the composite error (Amsler et al., 2024). Similarly, persistent inefficiency is part of unobserved heterogeneity and the estimation of it relies on that heterogeneity. Even though both forms of inefficiency are unknown, the composite errors and unobserved heterogeneity can be calculated once the production frontier has been estimated. Consequently, the conditional density

[1] A recent contribution by Tsionas et al. (2023) is a twist on this theme, using efficiency estimates constructed from DEA to form a prior.

of inefficiency is approximated by the joint density of the inefficiency and the realized composite errors divided by the marginal density of the realized composite errors, where both the joint density and the marginal density are approximated by the corresponding kernel estimates.

Our main contribution is the development of a semiparametric Bayesian approach for the general four-component stochastic frontier panel data model. In particular, we provide a new, robust, approximation to the distributions of both types of inefficiency, building up the likelihood function and posterior. We also develop a sampling algorithm to estimate the parameters of the frontier function, the error variance and various bandwidths involved in the approximation of the conditional density of the inefficiencies. We sample the coefficients that appear in the parametric frontier, and then calculate the composite errors and draw the inefficiencies from their conditional density. From the perspective of panel data models, our approach to estimation can be classified within the random effects framework because our likelihood is built based on the assumption that the inefficiencies are not correlated with the explanatory variables. Nonetheless, while sampling the smoothing parameters involved in the conditional density of transient inefficiencies, we have to rely on realized composite errors and consequently the model is effectively treated as a fixed effects panel data model (see, for example, Rendon, 2013).

We use four pseudo-samples to examine the performance of our estimator in comparison with Tsionas and Kumbhakar's (2014) estimator. Even though the pseudo-sample is generated Tsionas and Kumbhakar (2014), our estimator shows superior performance against theirs. The advantages of our estimator are further illustrated by considering the translog stochastic frontier model that analyzes cost efficiency of a panel of large US commercial banks. Although both our estimator and that of Tsionas and Kumbhakar's (2014) produce similar parameter estimates, the corresponding 95% Bayesian credible intervals obtained via our approach are narrower than those from Tsionas and Kumbhakar (2014). Moreover, our approach finds that for a bank with median level inefficiency, the 95% high-density region of the transient inefficiency varies over time. However, such time-varying features are not discovered using Tsionas and Kumbhakar's (2014) approach. Moreover, the contribution of persistent inefficiency to the total inefficiency of a hypothetical bank estimated by our approach is roughly half of that from Tsionas and Kumbhakar's (2014) estimator.

The rest of this paper is organized as follows. We present a conditional kernel density estimator of inefficiencies and develop a sampling algorithm for estimation purposes in the next section. In Section 3, we use two pseudo-samples to examine the performance of our approach against two competing estimators. In Section 4, we compare our approach to the fully parametric approach of Tsionas and Kumbhakar (2014) using a panel of US large commercial banks. Section 5 concludes the paper.

2 Semiparametric stochastic frontier model

2.1 Unknown distribution of inefficiencies

Tsionas and Kumbhakar (2014) proposed a panel data stochastic cost frontier model that disentangles unobserved firm heterogeneity from persistent and transient technical inefficiency. The model is expressed as

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \eta_i + u_{it} + v_{it}, \tag{1}$$

where $\alpha_i \sim \text{iid}\,\mathcal{N}(0, \sigma_\alpha^2)$ and $v_{it} \sim \text{iid}\,\mathcal{N}(0, \sigma_v^2)$ with σ_α and σ_v being unknown (and strictly positive and finite) parameters, for firms $i = 1, 2, \ldots, N$ and time $t = 1, 2, \ldots, T$. In this model, y_{it} represents the logarithmic cost, $\boldsymbol{x}_{it}$ is a vector of p explanatory variables for the ith firm at time t, $\boldsymbol{\beta}$ is the corresponding vector of coefficients, α_i represents the firm-specific random effects capturing the unobserved heterogeneity (time-invariant and independent from inefficiency), η_i represents the persistent (long-run) inefficiency component (non-negative), v_{it} represents the idiosyncratic random error, and u_{it} is a non-negative random variable that represents the short-run or transient inefficiency, which can be interpreted as the percentage increase in cost of firm i due to temporal inefficiency.

The short-term inefficiency u_{it} is assumed to be time-variant and uncorrelated with explanatory variables and v_{it}. The time-invariant restriction is unrealistic sometimes, especially in a panel with a long time horizon. To relax this restriction, Cornwell, Schmidt and Sickles (1990) introduced time-varying inefficiencies as the sum of the time-invariant inefficiency and a quadratic function of time t. Kumbhakar (1990) and Battese and Coelli (1992) specified time-varying inefficiencies as the product of time-invariant inefficiency and (different) non-linear functions of time t. In such studies, the time-varying inefficiencies are treated as random effects, and estimation is carried out through maximization of parametric likelihood functions.

Let $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{u}$ denote respectively, the collections of y_{it}, $\boldsymbol{x}_{it}$ and u_{it}, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_N)'$. Under the assumption of iid Gaussian errors of v_{it}, the likelihood is

$$L(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{u}, \sigma_\alpha^2, \sigma_v^2, \boldsymbol{X}) = \left(2\pi\sigma_v^2\right)^{-\frac{NT}{2}} \prod_{i=1}^{N} \prod_{t=1}^{T} \exp\left\{-\frac{1}{2\sigma_v^2}\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i - \eta_i - u_{it}\right)^2\right\},$$

which Tsionas and Kumbhakar (2014) used to develop their sampling algorithms together with distributional assumptions of η_i, u_{it} and α_i.

A key issue to proceed with the above likelihood function is to obtain the densities of u_{it} and η_i. In Tsionas and Kumbhakar (2014), they assumed that $\eta_i \sim \mathrm{iid}\, \mathcal{N}^+\left(0, \sigma_\eta^2\right)$ and $u_{it} \sim \mathrm{iid}\, \mathcal{N}^+\left(0, \sigma_u^2\right)$, where σ_η and σ_u are unknown (and positive and finite) parameters. We note that the composite errors, $\eta_i + u_{it} + v_{it}$, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, contain information about the distributional properties of η_i and u_{it}. Therefore, instead of imposing distributional assumptions on η_i and u_{it}, we propose to approximate the densities of η_i and u_{it} by kernel density estimates of their respective conditional densities, where the bandwidths are treated as parameters to be estimated along with the remaining model parameters through a Bayesian sampling procedure.[2]

2.2 Conditional densities

2.2.1 Conditional density of u_{it}

The key insight to our approach is that the composite error $u_{it} + v_{it}$ contains information about the distribution of u_{it} (Amsler et al., 2024). Let $\varepsilon_{it} = u_{it} + v_{it}$, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. We approximate the density of u_{it} by the conditional density of u_{it} given

[2]Although α_i might be estimated in a similar way as η_i, we have not done so because the random effects are usually not of primary interest in the context of panel stochastic frontier models.

$(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$, denoted as $f(u_{it}|\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$, which is defined as

$$f_{u|\varepsilon}\left(u_{it}|\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}\right) = \frac{f_{u,\varepsilon}\left(u_{it}, \varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}\right)}{f_{\varepsilon}\left(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}\right)},$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. This conditional density is well defined conditional on realized ε_{it}.[3]

We approximate the joint density of $(u_{it}, \varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$ by

$$\widehat{f}_{u,\varepsilon}\left(u_{it}, , \varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT} \mid h_u, \boldsymbol{h}_{\varepsilon}\right) = \frac{1}{TN} \sum_{j=1}^{N} \sum_{t=1}^{T} \boldsymbol{K}_{\boldsymbol{h}_{\varepsilon}} \left(\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_j\right) k_{h_u} \left(u_{it} - u_{jt}\right), \tag{2}$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$, $\boldsymbol{K}_{\boldsymbol{h}_{\boldsymbol{\varepsilon}}}(\cdot)$ is a multivariate kernel with $\boldsymbol{h}_{\varepsilon} = (h_1, h_2, \ldots, h_T)'$ a vector of bandwidths, and $k_{h_u}(\cdot)$ is a univariate kernel with bandwidth h_u. For simplicity, we use $\boldsymbol{h}$ to represent the vector of the above-mentioned bandwidths, $h_u, h_1, h_2, \ldots, h_T$, and $\boldsymbol{h}^2$ to denote the vector of squared bandwidths.[4]

In the literature on multivariate density estimation, the multivariate kernel function is often specified as a product of univariate kernel functions, each of which is allowed to have its own bandwidth (see, for example, Li and Racine, 2006). As discussed by Zhang et al. (2014), when a kernel density estimator such as (2) is plugged into a likelihood function for the purpose of estimating bandwidths, we also need to exclude the jth observation that makes $\varepsilon_{it} = \varepsilon_{jt}$ or $u_{it} = u_{jt}$ (for all t) from (2) in order to exclude cases with $k(0/h_t)$ or $k(0/h_u)$, which will make the corresponding bandwidth arbitrary. Let

$$T_{\varepsilon i} = \{t : \varepsilon_{jt} \neq \varepsilon_{it}, \text{ for } t = 1, 2, \ldots, T\} \text{ and } J_{ui} = \left\{j : u_{jt} \neq u_{it}, \exists t \in \{1, 2, \ldots, T\}\right\},$$

for $i = 1, 2, \ldots, N$. Let N_u denote the number of terms excluded from the summation in (2). Therefore, the leave-one-group-out estimator given by (2) is then expressed as[5]

$$\widehat{f}(u_{it}, \boldsymbol{\varepsilon}_i|\boldsymbol{h}) = \frac{1}{NT - N_u} \frac{1}{h_u h_1 h_2 \cdots h_T} \sum_{j \in J_{ui}} \sum_{t=1}^{T} \left\{ k\left(\frac{u_{it} - u_{jt}}{h_u}\right) \prod_{t \in T_{\varepsilon i}} k\left(\frac{\varepsilon_{it} - \varepsilon_{jt}}{h_t}\right) \right\}. \tag{3}$$

[3]If the density of u_{it} were approximated by the kernel density estimate based on $(u_{i1}, u_{i2}, \ldots, u_{iT})$ it would be similar in fashion to Feng et al. (2019). However, as Feng et al. (2019) assume that inefficiency is time invariant, the resulting estimate would not use information contained in $(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$.

[4]The kernel-form conditional density given by (2) is defined in the frequentist domain. Although $(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$ is a vector of random entities, it is known conditional on $\boldsymbol{\beta}$, α_i, η_i and initial values of u_{it}, and the u_{it}'s are known conditional on $\boldsymbol{\beta}$, α_i and η_i. We will develop a Bayesian sampling algorithm, through which we will sample all these unknown entities conditionally.

[5]It is not enough to simply leave one observation out, because a zero argument of any kernel function will make the corresponding bandwidth arbitrary.

A popular choice of a kernel function is the probability density function (PDF) of the standard Gaussian distribution, which we use throughout this paper.

The joint density of $\boldsymbol{\varepsilon}_i$ is approximated by

$$\widehat{f}_\varepsilon\left(\boldsymbol{\varepsilon}_i|\boldsymbol{h}_\varepsilon\right) = \frac{1}{N-N_\varepsilon}\frac{1}{h_1h_2\cdots h_T}\sum_{\substack{j=1\\j\neq i}}^{N}\left\{\prod_{t\in T_{\varepsilon i}}k\left(\frac{\varepsilon_{it}-\varepsilon_{jt}}{h_t}\right)\right\}, \tag{4}$$

where N_ε denotes the number of terms excluded from the function and the bandwidths $h_1, h_2, \ldots, h_T$ are defined as the same as those in (3).

The conditional density of u_{it} given the realized composite error terms is

$$\widehat{f}_{\boldsymbol{h}}(u_{it}|\boldsymbol{\varepsilon}_i,\boldsymbol{h}) = \frac{\widehat{f}(u_{it},\boldsymbol{\varepsilon}_i|\boldsymbol{h})}{\widehat{f}_\varepsilon(\boldsymbol{\varepsilon}_i|\boldsymbol{h}_\varepsilon)}, \quad \text{for } i=1,2,\ldots,N, \text{ and } t=1,2,\ldots,T. \tag{5}$$

Each of these kernel estimates integrates to one and satisfies positivity, and is thus a density function.

Our idea is to approximate the density of u_{it} by the conditional density of u_{it} conditional on $(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$. The resulting kernel density estimator involves the univariate kernel for u_{it} and multivariate kernel for $(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$ and thus is a high-dimensional density estimation. An alternative approach is to approximate the density of u_{it} by the conditional density of u_{it} conditional on ε_{it}, which results in a low dimensional kernel density estimator using univariate kernels. However, the use of a multivariate kernel density of $(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$ allows accommodation of possible serial dependence of u_{it} over time (Amsler et al., 2014). We will discuss this alternative, low-dimensional approach later in a subsection of comparison studies using pseudo-samples.

2.2.2 Conditional density of η_i

Let $\theta_{it} = \eta_i + v_{it}$ and $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{iT})'$, and let

$$J_{\eta i} = \{j : \eta_j \neq \eta_i, \text{for } j=1,2,\ldots,N\} \text{ and } T_{\theta i} = \{t : \theta_{jt} \neq \theta_{it}, \text{for } t=1,2,\ldots,T\},$$

for $i = 1, 2, \cdots, N$. The joint density of $(\eta_i, \boldsymbol{\theta}_i)$ is approximated by a kernel estimator:

$$\widehat{f}(\eta_i,\boldsymbol{\theta}_i|\boldsymbol{b}) = \frac{1}{N-N_\eta}\frac{1}{b_\eta b_1 b_2\cdots b_T}\sum_{j\in J_{\eta i}}\left\{k\left(\frac{\eta_i-\eta_j}{b_\eta}\right)\prod_{t\in T_{\theta i}}k\left(\frac{\theta_{it}-\theta_{jt}}{b_t}\right)\right\}, \tag{6}$$

for $i = 1, 2, \ldots, N$, where N_η is the number of terms excluded from the density function, and $\boldsymbol{b}$ represents the vector of the above-mentioned bandwidths, $b_\eta, b_1, b_2, \ldots, b_T$, and $\boldsymbol{b}^2$ to denote the vector of squared bandwidths.

The joint density of $\boldsymbol{\theta}_i$ is approximated by

$$\widehat{f}_\theta\left(\boldsymbol{\theta}_i|\boldsymbol{b}_\theta\right) = \frac{1}{N - N_\theta} \frac{1}{b_1 b_2 \cdots b_T} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left\{ \prod_{t \in T_{\theta i}} k\left(\frac{\theta_{it} - \theta_{jt}}{b_t}\right) \right\}, \tag{7}$$

where N_θ is the number of terms excluded from the density function, and $\boldsymbol{b}_\theta$ represents the vector of bandwidths, $b_1, b_2, \ldots, b_T$.

The conditional density of η_i given θ_{it} is

$$\widehat{f}_{\boldsymbol{b}}(\eta_i|\boldsymbol{\theta}_i, \boldsymbol{b}) = \frac{\widehat{f}(\eta_i, \boldsymbol{\theta}_i|\boldsymbol{b})}{\widehat{f}_\theta(\boldsymbol{\theta}_i|\boldsymbol{b}_\theta)}, \quad \text{for } i = 1, 2, \ldots, N. \tag{8}$$

It might be possible to select optimal bandwidths using a likelihood cross-validation method if u_{it} and $\boldsymbol{\varepsilon}_i$, or η_i and $\boldsymbol{\theta}_i$ are observable. However, due to their non-observable nature, we follow Zhang et al. (2006) and treat bandwidths as parameters, which are estimated through Markov chain Monte Carlo (MCMC) sampling techniques.[6]

2.3 Bayesian estimation

A Bayesian approach has an advantage of being able to estimate firm-specific inefficiencies via classical frequentist approaches. Flexible inference without parametric assumptions about the distribution of inefficiencies could be achieved by placing a prior distribution on infinite dimensional spaces. Ferguson (1973) introduced a popular nonparametric Dirichlet prior process, which was employed by Griffin and Steel (2004) to model the distribution of inefficiencies in stochastic frontier models. Walker, Damien,

[6] As a referee rightly noted, our approach could be elaborated further by also modeling the determinants of persistent and time-varying efficiency (as in Lai and Kumbhakar, 2018a,b). If an environmental variable is to be included in the model, for example, via $\sigma^2_{\eta,i} = \exp(\gamma z_i)$, our sampling framework can be modified to sample the additional parameter as well. One approach to incorporating information from the environmental variable is to derive the kernel density estimator of $\eta_i \exp(-0.5\gamma z_i)$ instead of the current conditional density estimator. Alternatively, we can allow the bandwidth for the kernel function of η_i to be made dependent on $\exp(\gamma z_i)$, because the variance of η_i derived from the kernel estimation of its marginal density depends on the squared bandwidth. Both approaches allow the variance of u_{it} to be influenced by an environmental variable (and can be generalized to accommodate multiple environmental variables). As our empirical example does not focus on environmental variables, we leave a full accounting of this approach for future research.

Laud and Smith (1999) provided a survey of Bayesian nonparametric methods, including the mixture of a Dirichlet process and Pólya tree priors. Another nonparametric method is based on kernel density estimation, whose performance is determined by bandwidths. Bandwidth selection is critical in kernel density estimation and nonparametric regression (see, for example Zhang, King and Hyndman, 2006; Zhang, Brooks and King, 2009; Zhang, King and Shang, 2014, for discussion on bandwidth selection using Bayesian approaches).

The stochastic frontier model considered in this paper is semiparametric in the sense that we do not impose parametric assumptions on the distributions of inefficiency. Our method is based on kernel density estimation, which is used to derive the conditional density estimator of inefficiencies. We then derive a posterior for the semiparametric model and describe sampling steps in the MCMC procedure.

2.3.1 Priors and posteriors

In the four-component stochastic frontier model, Tsionas and Kumbhakar (2014) assumed that $u_{it} \sim \mathcal{N}^+(0, \sigma_u^2)$ and $\eta_i \sim \mathcal{N}^+(0, \sigma_\eta^2)$. Instead of these parametric assumptions, we suggest using kernel estimators presented by (5) and (8). We follow the general model structure of Tsionas and Kumbhakar (2014), but implement a semiparametric approach by replacing their parametric assumptions on the inefficiency distributions with nonparametric kernel estimators. From the perspective of distributional specification, the parametric assumptions in Tsionas and Kumbhakar (2014) reflect subjective choices that may not align well with the underlying data. In contrast, our approach employs data-driven, nonparametric (kernel-based) estimates of the inefficiency distributions, which are expected to offer greater flexibility and robustness.

Following the discussion of O'Donnell and Coelli (2005), we choose a joint improper prior for $\boldsymbol{\beta}$:

$$p(\boldsymbol{\beta}) \propto 1.$$

According to the prior choice discussed by Tsionas and Kumbhakar (2014), we assume

$$\frac{\bar{q}_\kappa}{\sigma_\kappa^2} \sim \chi^2\left(\bar{\lambda}_\kappa\right), \text{ for } \kappa = \alpha \text{ and } v,$$

where $\bar{\lambda}_\kappa$ denotes the degree of freedom of the χ^2 distribution, and hyperparameters are chosen as $\bar{q}_\kappa = 10^{-4}$ and $\bar{\lambda}_\kappa = 1$, for $\kappa = \alpha$ and v.

Following the discussion of Zhang, Brooks and King (2009), we choose the prior of each squared bandwidth to be:

$$p\left(b_\eta^2\right) \propto \frac{1-\exp\left(-b_\eta^2/2\right)}{\exp\left(b_\eta^2/2\right)},\ p\left(b_t^2\right) \propto \frac{1-\exp\left(-b_t^2/2\right)}{\exp\left(b_t^2/2\right)}, \quad \text{for } t=1,2,\ldots,T,$$
$$p\left(h_u^2\right) \propto \frac{1-\exp\left(-h_u^2/2\right)}{\exp\left(h_u^2/2\right)},\ p\left(h_t^2\right) \propto \frac{1-\exp\left(-h_t^2/2\right)}{\exp\left(h_t^2/2\right)}, \quad \text{for } t=1,2,\ldots,T.$$

Such priors are able to prevent bandwidths from getting too large or too small during sampling iterations, and consequently, let the data choose bandwidths. Thus, the joint prior of all squared bandwidths is

$$p(\boldsymbol{b}^2) = p(b_\eta^2)\prod_{t=1}^{T} p(b_t^2),\ \ p(\boldsymbol{h}^2) = p(h_u^2)\prod_{t=1}^{T} p(h_t^2).$$

According to Bayes' theorem, the posterior of $(\boldsymbol{\beta}, \sigma_\alpha^2, \sigma_v^2, \boldsymbol{u}, \boldsymbol{h}^2, \boldsymbol{b}^2)'$ is

$$\begin{aligned}&\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \sigma_\alpha^2, \sigma_v^2, \boldsymbol{u}, \boldsymbol{h}^2, \boldsymbol{b}^2 | \boldsymbol{y}, \boldsymbol{X}) \propto \\ &L(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{u}, \sigma_\alpha^2, \sigma_v^2, \boldsymbol{X}) p(\boldsymbol{\beta}) p(\boldsymbol{\alpha}|\sigma_\alpha^2) p(\sigma_\alpha^2) p(\sigma_v^2) p(\boldsymbol{u}|\boldsymbol{h}^2) p(\boldsymbol{h}^2) p(\boldsymbol{\eta}|\boldsymbol{b}^2) p(\boldsymbol{b}^2),\end{aligned} \tag{9}$$

where $p(\boldsymbol{u}|\boldsymbol{h}^2)$ is approximated by (5) and $p(\boldsymbol{\eta}|\boldsymbol{b}^2)$ is approximated by (8). If such approximated kernel estimators were treated as frequentist-domain estimators and bandwidths were selected using likelihood cross-validation, the resulting Bayesian sampling procedure would be a hybrid Bayes-frequentist approach in the sense of Yuan (2009). Nonetheless, in our Bayesian framework, bandwidths are treated as parameters and are sampled conditional on updates to the other unknown entities. As a consequence, our Bayesian framework might also be regarded as a Bayes-frequentist approach.[7]

The details on conditional posteriors are presented in Appdendix A. We choose the estimates of u_{it} and η_i from Tsionas and Kumbhakar (2014) as their respective initial values in the MCMC procedure. To impose non-negativity for the inefficiencies and bandwidths, we use the random-walk Metropolis algorithm to sample each component of $\boldsymbol{\eta}$, $\boldsymbol{b}^2$, $\boldsymbol{u}$ and $\boldsymbol{h}^2$ in log transformations, where tuning parameters are selected in an adaptive way, as discussed by Garthwaite et al. (2016). Therefore, the components of $(\boldsymbol{\beta}, \sigma_\alpha^2, \sigma_v^2, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{u}, \boldsymbol{b}^2, \boldsymbol{h}^2)$ are sequentially sampled through conditional posteriors described by (A.1)–(A.8).

[7] We thank an anonymous reviewer for pointing this out.

3 Performance assessment using pseudo samples

In our simulation study, all pseudo-samples are generated from the same four-component stochastic frontier model of Tsionas and Kumbhakar (2014), but under varying scenarios designed to test different structural features of inefficiency. The purpose of this design is not to compare models *per se*, but rather to assess and compare the performance of various estimators when confronted with different forms of misspecification. To this end, we focus on a set of carefully constructed data-generating processes that differ in the behavior and distribution of persistent and transient inefficiencies.

We evaluate five estimators: our proposed semiparametric Bayesian estimator, the fully parametric Bayesian estimator of Tsionas and Kumbhakar (2014), and three variants of our estimator designed to isolate specific model components. These variants are: (i) the imposition of time-invariant inefficiency only in the model; (ii) the imposition of time-varying inefficiency only in the model; and (iii) use of a simplified conditional density based solely on contemporaneous composite errors. While the comparisons of some of these estimators are relegated to the appendix for brevity, we emphasize three key ones in the main analysis: our full estimator, the benchmark estimator of Tsionas and Kumbhakar (2014), and our simplified approach based on univariate (contemporaneous) conditioning. This structure allows us to systematically assess the value of our kernel-based density approximation and its flexibility in recovering inefficiency distributions under various forms of misspecification.

3.1 Competing estimators

We examine the performance of our estimator in comparison with several alternative estimation approaches as follows.

Estimator A is Tsionas and Kumbhakar's (2014) Bayesian estimator for the four-component panel stochastic frontier model given in (1).

Estimator B is our estimator, but applied to the panel stochastic frontier model with time-invariant inefficiencies only:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \eta_i + v_{it}, \tag{10}$$

where $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$ represents random effects. We propose to approximate the joint density function $f(\eta_i, \boldsymbol{\theta}_i)$ by

$$\widehat{f}(\eta_i, \boldsymbol{\theta}_i | b_\eta, \boldsymbol{b}_\theta) = \frac{1}{N-1} \frac{1}{b_\eta b_1 b_2 \dots b_T} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left\{ k\left(\frac{\eta_i - \eta_j}{b_\eta}\right) \prod_{t=1}^{T} k\left(\frac{\theta_{it} - \theta_{jt}}{b_t}\right) \right\}, \tag{11}$$

with $\boldsymbol{\theta}_i = \eta_i + v_{it}$. Moreover, we propose to approximate the density of η_i by $f(\eta_i|\boldsymbol{\theta}_i)$, which is the conditional density of η_i given the composite error $\boldsymbol{\theta}_i$, and is approximated by the corresponding kernel density estimator. The resulting sampling algorithm is a simplified version of the algorithm developed for our proposed estimation strategy.

Estimator C is our estimator, but applied to the panel stochastic frontier model without persistent inefficiencies:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + u_{it} + v_{it}, \tag{12}$$

where $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$ represents random effects, u_{it}, for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$, are iid with their density being approximated by (5).

Estimator D is is our estimator for the four component panel stochastic frontier model given in (1) where the density of u_{it} is approximated by the conditional density of u_{it} on ε_{it} and is expressed as

$$\widehat{f}(u_{it}|\varepsilon_{it}, h_u, h_\varepsilon) = \frac{\widehat{f}(u_{it}, \varepsilon_{it}|h_u, h_\varepsilon)}{\widehat{f}(\varepsilon_{it}|h_\epsilon)}. \tag{13}$$

The numerator and denominator are approximated respectively, by

$$\widehat{f}(u_{it}, \varepsilon_{it}|h_u, h_\varepsilon) = \frac{1}{NT-1} \frac{1}{h_u h_\varepsilon} \sum_{j=1}^{N} \sum_{\substack{s=1 \\ s \neq t \text{ if } j=i}}^{T} k\left(\frac{u_{it} - u_{js}}{h_u}\right) k\left(\frac{\varepsilon_{it} - \varepsilon_{js}}{h_\epsilon}\right), \tag{14}$$

and

$$\widehat{f}(\varepsilon_{it}|h_\varepsilon) = \frac{1}{NT-1} \frac{1}{h_u h_\varepsilon} \sum_{j=1}^{N} \sum_{\substack{s=1 \\ s \neq t \text{ if } j=i}}^{T} k\left(\frac{\varepsilon_{it} - \varepsilon_{js}}{h_\epsilon}\right). \tag{15}$$

The estimator of the four component model here uses a kernel conditional density estimate of u_{it} that is conditional on ε_{it} rather than the collection of ε_{it} across time, making it a simplified version of our proposed approach. In comparison to estimator

D, we anticipate that our approach will effectively address possible unmodeled serial dependence of u_{it} over time.

The errors v_{it}, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, are iid $\mathcal{N}(0, \sigma_v^2)$, where σ_v is unknown (and positive and finite) parameters. In summary, the difference between our estimator and the four competing estimators are described in Table 1.

Table 1: A summary of competing estimators in the simulation study.

Models	Differences to our approach
Estimator A	Tsionas and Kumbhakar's (2014) estimator, where the estimation of persistent and transient inefficiencies depends on their distributional assumptions.
Estimator B	Only persistent inefficiencies are are estimated based on conditional kernel density estimation.
Estimator C	Only transient inefficiencies are estimated based on conditional kernel density estimation.
Estimator D	The distribution of u_{it} is approximated by the conditional kernel density estimator only conditioning on ε_{it}.

3.2 Data generating process

We now conduct a simulation study to examine the performance of our stochastic frontier panel data estimator in comparison with several competing approaches. The data generating process is

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + \eta_i + u_{it} + v_{it}, \tag{16}$$

where $\alpha_i \sim \mathcal{N}(0, 0.1^2)$ and $v_{it} \sim \mathcal{N}(0, 0.1^2)$. The explanatory variables x_{it} is randomly generated from the standard normal distribution. We set $\beta_0 = \beta_1 = 1$ and the sample size to be $N = 100$ and $T = 10$. When generating inefficiency terms, we consider the following four scenarios:

Scenario 1: We generate $\eta_i \sim \mathcal{N}^+(0, 0.5^2)$ independently across i, and $u_{it} \sim \mathcal{N}^+(0, 0.5^2)$ independently across i and t. We evaluate the performance of our approach in comparison to the estimator proposed by Tsionas and Kumbhakar (2014) where their model is correctly specified.

Scenario 2: We generate $\eta_i \sim \mathcal{N}^+(0, 0.5^2)$ independently across i, and define $u_{it} = u_{i0} \times g(t)$ following Kumbhakar (1990), where $u_{i0} \sim \mathcal{N}^+(0, 0.5^2)$ is generated independently across i, and $g(t) = (1 + \exp(0.05t + 0.04t^2))^{-1}$. This setup allows u_{it} to exhibit time-varying variance. The competing estimator is the one proposed by Tsionas and Kumbhakar (2014), and their assumption regarding the distribution of u_{it} is misspecified in this context. This scenario enables us to evaluate the robustness of our estimator under distributional misspecification of transient inefficiencies.[8]

Scenario 3: We generate $\eta_i \sim \chi^2(1)$ independently across i, and define $u_{it} = u_{i0} \times g(t)$, where $u_{i0} \sim \mathcal{N}^+(0, 0.5^2)$ is generated independently across i and $g(t) = (1 + \exp(0.05t + 0.04t^2))^{-1}$. The competing estimator is the one proposed by Tsionas and Kumbhakar (2014), and their assumptions of the distributions of η_i and u_{it} are both misspecified here. This scenario allows us to assess the robustness of our proposed approach under joint distributional misspecification.

Scenario 4: We generate $\eta_i \sim \mathcal{N}^+(0, 0.5^2)$ independently across i, and define $u_{it} = 0.1 + 0.8u_{i,t-1}$, where $u_{i,0} \sim \mathcal{N}^+(0, 0.5^2)$ is generated independently across i. This setup allows u_{it} to possess serial correlation over time. This scenario enables us to assess the performance of our estimator in comparison to Estimator D in the presence of unmodeled temporal dependence in u_{it}.

3.3 Assessment criteria

Following Feng, Wang and Zhang (2019), we employ two measures to evaluate the correlation between the estimated inefficiencies and the corresponding true values.

[8]This simple data generating process allows u_{it} to vary over time, while requiring only a distributional assumption for u_{i0}. If more complex time-varying dynamics are of interest, one may consider the data generating processes discussed by Lai and Kumbhakar (2020; 2023).

The first is Spearman's rank correlation between two vectors of ranks:

$$\rho_t = 1 - \frac{6}{N\left(N^2-1\right)} \sum_{i=1}^{N} \left(\text{Rank}_{it}^{(\text{est})} - \text{Rank}_{it}^{(\text{true})}\right)^2, \text{ for } t = 1, 2, \ldots, T, \tag{17}$$

where $\text{Rank}_{it}^{(\text{est})}$ and $\text{Rank}_{it}^{(\text{true})}$ are the ranks of the ith firm at time t according to the estimated inefficiencies and their corresponding true inefficiencies, respectively. The resulting correlation coefficient ρ_t is between -1 and 1. The closer it is to 1, the better the corresponding estimator performs. We calculate the average of the time-varying rank correlations for u_{it} which is denoted as $\bar{\rho}$. We also calculate the rank correlations for η_i which is denoted as ρ_η.

The second is the average Euclidean distance between the estimated transient inefficiencies denoted as $u_{it}^{(\text{est})}$, and the true transient inefficiencies denoted as $u_{it}^{(\text{true})}$, and is defined as

$$d_t = \frac{1}{N}\Big(\sum_{i=1}^{N} \big(u_{it}^{(\text{est})} - u_{it}^{(\text{true})}\big)^2\Big)^{1/2}, \text{ for } t = 1, 2, \ldots, T. \tag{18}$$

The smaller the Euclidean distance, the better the corresponding estimator performs. We calculate the average (over time) of the time-varying Euclidean distance which is denoted as $\bar{d}$. For the persistent inefficiency, the average Euclidean distance is calculated by

$$d_\eta = \frac{1}{N}\Big(\sum_{i=1}^{N} \big(\eta_i^{(\text{est})} - \eta_i^{(\text{true})}\big)^2\Big)^{1/2}, \tag{19}$$

where $\eta_i^{(\text{est})}$ is the estimated persistent inefficiency and $\eta_i^{(\text{true})}$ is the true persistent inefficiency.

3.4 Discussion of the estimators

As the number of transient inefficiencies to be estimated is equal to the number of generated inefficiencies, randomness has a significant impact on Tsionas and Kumbhakar's (2014) estimation of u_{it} and its distribution. This is evident from the posterior conditional distribution of u_{it}, which is the half-normal distribution with mean $\widehat{u}_{it}$. In contrast, our proposed kernel conditional density estimator borrows information across both i and

t. By construction, it is able to utilize the information contained in all the generated inefficiencies, $\widehat{u}_{it}$, for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. This explains why our kernel conditional density estimate outperforms Tsionas and Kumbhakar's (2014) estimate even if their distributional assumption of u_{it} is correctly specified. Our approach offers an advantage of robustness in empirical studies where the true distribution of u_{it} is unknown.

Further, even though Tsionas and Kumbhakar's (2014) estimator is correctly specified, we might still expect that our estimator possesses improved behavior following the argument in Amsler et al. (2024). The decomposition of variance of a given inefficiency is $V(u) = V\left(E[u|\xi]\right) + E\left[Var(u|\xi)\right]$. As we increase the size of the conditioning set from ξ to ξ^* (which is what our estimator proposes), then necessarily $E\left[Var(u|\xi^*)\right] \leq E\left[Var(u|\xi)\right]$ while $V\left(E[u|\xi^*]\right) \geq V\left(E[u|\xi]\right)$. Thus, even under correct specification, the fact that our estimator leverages more information in the conditioning set is what could lead to improved performance in the simulations.

Put differently, one advantage of our kernel-based estimator is that it estimates the joint conditional density $f(u_{it}|\varepsilon_{i1}, \ldots, \varepsilon_{iT})$, using information pooled across both firms and time periods. In contrast, the approach in Tsionas and Kumbhakar (2014) relies on parametric distributional assumptions and conditional likelihoods that are constructed separately for each firm and time period. As a result, their method does not borrow information across the panel in the same way and may be more vulnerable to misspecification of the inefficiency distribution.

The flexibility of our method to learn from the entire panel structure is particularly valuable in empirical settings where the shape of the inefficiency distribution is unknown or complex. This advantage will be clearly demonstrated in Scenario 2. Nonetheless, kernel-induced densities are not independent across firms or time, as the method averages discrepancies across both dimensions. As such an average allocates large weights to the neighbors of each data point, the kernel estimation induces less dispersion in the inefficiency distributions, resulting in a more concentrated probability mass than that of the Half-Normal density.

3.5 Results obtained through simulated samples

Each of our sampling procedures consists of 15,000 iterations with the first 5,000 iterations being discarded to allow for burn-in. The mixing performance, or loosely speaking the convergence status, of each simulated chain is monitored by the simulation inefficiency factor (SIF), which is approximated by the ratio of the variance of the sample mean over the variance of the sample mean from a hypothetical sampler taking draws independently from the posterior (see, for example, Kim, Shephard and Chib, 1998; Zhang, Brooks and King, 2009). The closer the SIF value is to 1, the better the mixing performance. In our experience, a sampling algorithm usually achieves reasonable mixing performance when its SIF values are below 100. The SIF values of the simulated chains derived under the four scenarios respectively, are summarized and tabulated in the Appendix Table B1. The low SIF values indicate that our samplers have achieved very good mixing performance.

3.5.1 Scenario 1

When the pseudo-sample is generated under the assumption that $u_{it} \sim \mathcal{N}^+(0, \sigma_u^2)$, Tsionas and Kumbhakar's (2014) estimator of the four component model is correctly specified. However, our estimator does not need an assumption about the distribution of u_{it} but possesses flexibility to estimate the unspecified distribution of u_{it}. The posterior mean estimates are summarized and tabulated in Table 2. In terms of estimation of β_0 and β_1, which are the parameters of the linear frontier, our estimator and Estimator D outperform Tsionas and Kumbhakar's (2014) estimator and the other two estimators based on incorrect specification (Estimators B and C). When estimating the two variance parameters, σ_α and σ_v, our estimator produced more accurate estimates than Tsionas and Kumbhakar's (2014) estimator. Moreover, Estimator C produced a more accurate estimate of σ_v than Tsionas and Kumbhakar's (2014) estimator.

As shown in Figure 1, when assessing the accuracy of the estimated u_{it}, we found that the rank correlation coefficients derived via our approach are higher than those obtained via Tsionas and Kumbhakar's (2014) estimator for all t values. Nonetheless, their estimator produced much higher rank correlation coefficients than the other two competing estimators for all t values. Moreover, using the measure of the average rank correlation coefficients across t, we found that our preferred estimator produces $\bar{\rho} = 0.9915$ and outperforms Tsionas and Kumbhakar's (2014) estimator, and that both estimators perform much better than Estimator C. As for the persistent inefficiency, our estimator produces $\rho_\eta = 0.9880$ and performs better than Tsionas and Kumbhakar's (2014) estimator, which produces $\rho_\eta = 0.9157$.

In order to measure the overall accuracy of the estimated inefficiencies, we calculated the average of estimated persistent and transient inefficiencies for each model, and compared them with the average of the true inefficiencies, respectively. The row starting with $\bar{u}$ of Table 2 shows that Tsionas and Kumbhakar's (2014) estimator performs slightly better than ours, and that both estimators perform much better than Estimator C. The row starting with $\bar{\eta}$ shows that our estimator performs slightly better than Tsionas and Kumbhakar's (2014) estimator when estimating the persistent inefficiencies.

Under the Euclidean distance measure d_t, the lower panel of Figure 1 shows that our estimator outperforms Tsionas and Kumbhakar's (2014) estimator because the resulting distance values derived via our approach are always clearly smaller than those obtained via Tsionas and Kumbhakar's (2014) estimator for all t values. Moreover, according to the values of $\bar{d}$ and d_η shown in Table 2, our estimator clearly outperforms Tsionas and Kumbhakar's (2014) estimator. As such, under the measures of d_t, $\bar{d}$ and d_η, our proposed approach outperforms Tsionas and Kumbhakar's (2014) estimator, which in turn outperforms Estimator C.

In this scenario, Estimator D demonstrates marginally superior performance compared to our approach in terms of $\bar{\rho}$ and $\bar{d}$. However, it underperforms relative to our estimator when measuring ρ_η and d_η. Notably, our approach has the potential advantage of modeling the unobserved dependence of u_{it}. We will examine this aspect in the following scenarios.

Thus, even though the pseudo-sample is generated from the four component model, our Bayesian semiparametric model has produced clearly better results than the correctly specified model. Ignoring the incorporation of persistent inefficiencies, our estimator leads to inaccurate estimates of the linear frontier, but the ranking of inefficiencies remains almost unchanged. However, ignoring the incorporation of transient inefficiencies clearly worsens the performance of our estimator in terms of both ranking and frontier accuracy.

3.5.2 Scenario 2

When the pseudo-sample is generated under $u_{it} = u_{i0} \times (1 + \exp(0.05t + 0.04t^2))^{-1}$ with u_{i0} being generated from $\mathcal{N}^+(0, \sigma_u^2)$, the competing Estimator A is incorrectly specified in terms of the distribution of u_{it}. Meanwhile, our estimator does not require a distributional assumption of u_{it} and as a consequence, is robust.

The results for this scenario are summarized in Table 3 and Figure 2. In particular, one can see that the estimate of β_1 for each estimator is very close to the true value, although Tsionas and Kumbhakar's (2014) estimator performs slightly worse than the other three estimators. When estimating the intercept of the linear frontier, our estimator outperforms Tsionas and Kumbhakar's (2014) estimator, which in turn outperforms the other two competing estimators due to their reliance on an incorrect parametric specification. Our estimator also outperforms Tsionas and Kumbhakar's (2014) estimator in the estimation of σ_α and σ_v, while the estimate of σ_v obtained under the incorrectly specified Estimator C is as accurate as that obtained under our proposed approach.

In order to measure the overall accuracy of the random effects, persistent inefficiency, and transient inefficiency, we calculated the average of their estimated values obtained through each estimator, and then compared them with the averages of the true values, respectively. Overall, the values of $\bar{\alpha}$, $\bar{\eta}$ and $\bar{u}$ also reported in Table 3 show that our estimator and Estimator D perform better than the other three estimators under model misspecification.

Figure 2 shows that the rank correlation coefficients for u_{it} derived from our estimator are higher than those obtained using the estimator of Tsionas and Kumbhakar (2014) for all values of t. Moreover, the rank correlation of persistent inefficiencies calculated from our estimator is $\rho_\eta = 0.9819$, which is higher than that obtained using Tsionas and Kumbhakar's (2014) estimator, suggesting that our approach yields improved performance.

Under the Euclidean distance measures d_t and $\bar{d}$, the lower panel of Figure 2 shows that our estimator has produced smaller distance values than Tsionas and Kumbhakar's (2014) estimator for all t values. In addition, our estimator produced a distance value of $\bar{d} = 0.0007$ and $d_\eta = 0.0050$, whereas Tsionas and Kumbhakar's (2014) estimator produced an average distance value of $\bar{d} = 0.0036$ and $d_\eta = 0.0126$. Thus, our proposed approach outperforms Tsionas and Kumbhakar's (2014) estimator, which performs slightly better than the other two estimators.

On the one hand, ignoring the incorporation of persistent inefficiencies in our estimator does not lead to obvious changes under the rank correlation and Euclidean distance measures. However, it does result in inaccurate estimates of the intercept of the linear frontier. On the other hand, ignoring the incorporation of transient inefficiencies clearly worsens the performance of our estimator.

3.5.3 Scenario 3

When the pseudo-sample is generated with $\eta_i \sim \chi^2(1)$ and $u_{it} = u_{i0} \times (1 + \exp(0.05t + 0.04t^2))^{-1}$, the model proposed by Tsionas and Kumbhakar (2014) is misspecified. It is therefore worthwhile to assess the performance of our estimator in comparison to theirs, as our approach does not rely on such distributional assumptions. The results obtained from this scenario are summarized in Table 4 and Figure 3, where we primarily focus on comparing our estimator with Tsionas and Kumbhakar (2014).

From Table 4 we can see that for the estimation of the intercept and β_1, our estimator performs slightly better than the estimator of Tsionas and Kumbhakar (2014). The average values of α, η and u obtained from our estimator are all closer to the true values than those derived from Tsionas and Kumbhakar's (2014) estimator. In addition, the average rank correlation coefficients $\bar{\rho}$ and ρ_η calculated from our estimator are higher than those from Tsionas and Kumbhakar's (2014) estimator. Similarly, the average Euclidean distance measures $\bar{d}$ and d_η are smaller when using our estimator, indicating better performance. Figure 3 further supports these findings by presenting the rank correlation and Euclidean distance over time.

Overall, the results from this scenario suggest that when the four component model is misspecified for a pseudo sample, our estimator outperforms Tsionas and Kumbhakar's

(2014). This indicates that our estimator and the associated sampling algorithm are robust to the distributional assumptions of η_i and u_{it}.

3.5.4 Scenario 4

In our proposed Bayesian framework, we approximate the density of u_{it} using the conditional density of u_{it}, which is conditional on either the multivariate density of $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ or the univariate density of ε_{it}. Let the resulting two estimators be denoted as $\widehat{f}(u_{it}|\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ and $\widehat{f}(u_{it}|\varepsilon_{it})$, respectively. When the pseudo-sample is generated with u_{it} being generated from $\mathcal{N}^+(0, \sigma_u^2)$, the univariate density leads to more accurate estimates of the transient inefficiencies than the multivariate counterpart. However, the opposite is true when the pseudo-sample is generated with u_{it} being correlated.

To further evaluate the accuracy of the transient inefficiency estimates, we generated an additional pseudo-sample where $u_{i,t} = 0.1 + 0.8u_{i,t-1}$, with $u_{i,1}$ being drawn from $\mathcal{N}^+(0, \sigma_u^2)$. This specification induces strong serial correlation in u_{it}. The resulting distance measures, rank correlations, and parameter estimates are summarized in Table 5. The results show that the multivariate density estimate $\widehat{f}(u_{it}|\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ leads to more accurate estimates of the transient inefficiencies than its univariate counterpart. This finding confirms that using $\widehat{f}(u_{it}|\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ as an approximation to the density of u_{it} effectively accommodates the unobserved serial dependence of the transient inefficiencies.

4 Cost efficiency of large US commercial banks

We utilize our semiparametric model to analyze the cost efficiency of a panel of large US commercial banks. Over the past three decades, there has been considerable interest in studying the efficiency of the banking industry, as it is one of the largest and most important sectors in the US economy. Both data envelopment analysis (Sherman and Gold, 1985) and stochastic frontier models with various specifications (Adams et al., 1999; Kumbhakar and Tsionas, 2005; Kumbhakar et al., 2007; Malikov et al., 2016; Tsionas et al., 2023) have been employed in previous studies to assess various types of efficiency of banks.

Technological and financial innovations have contributed to the rise of larger and more complex organizations in the US commercial banking industry. This structural transformation has raised important questions about how efficiently these institutions operate. Studies such as Feng and Zhang (2012, 2014); Feng et al. (2019) have examined the productivity of large US banks using stochastic distance frontier methods. In this context, cost efficiency reflects both the internal organization of banks and, more broadly, the effectiveness of public service provision by the government. Improving cost efficiency can reduce resource waste in banking operations and support overall economic development. Thus, analyzing cost efficiency not only provides a measure of bank performance but may also offer guidance for future improvements of their business practices in the sector.

4.1 Data and models

We used the annual data of commercial banks from the Reports of Income and Condition published by the Federal Reserve Bank of Chicago (see Feng and Serletis, 2009, for a description of the data). The data contains 141 continuously operating large banks, which have at least 3 billion dollars in assets in year 2000 dollars. The sample period is from 1998 to 2005, and hence we have observations for $N = 141$ banks over $T = 8$ years. We consider three input prices: the wage rate for labor, the interest rate for borrowed funds and the price of physical capital; three outputs: consumer loans, non-consumer loans, and securities; as well as the total cost. All variables are constructed following Berger and Mester (2003).

We employ a translog cost frontier model (more details about the translog function can be found in Sickles and Zelenyuk, 2019, and references therein) with L input prices

$X_1, \dots, X_L$, M outputs $Y_1, \dots, Y_M$, and cost C for each bank i at time $t = 1, 2, \dots, T$:

$$
\begin{aligned}
\log(C_{it}/X_{L,it}) =& a_0 + \sum_{j=1}^{L-1} a_j \log \frac{X_{j,it}}{X_{L,it}} + \frac{1}{2} \sum_{j=1}^{L-1} \sum_{k=1}^{L-1} a_{jk} \log \frac{X_{j,it}}{X_{L,it}} \log \frac{X_{k,it}}{X_{L,it}} \\
&+ \sum_{m=1}^{M} c_m \log Y_{m,it} + \frac{1}{2} \sum_{m=1}^{M} \sum_{n=1}^{M} c_{mn} \log Y_{m,it} \log Y_{n,it} \\
&+ \frac{1}{2} \sum_{j=1}^{L-1} \sum_{m=1}^{M} g_{jm} \log \frac{X_{j,it}}{X_{L,it}} \log Y_{m,it} + \sum_{j=1}^{L-1} w_j t \log \frac{X_{j,it}}{X_{L,it}} \\
&+ \sum_{m=1}^{M} \delta_m t \log Y_m + \gamma_1 t + \frac{1}{2} \gamma_2 t^2 + \alpha_i + \eta_i + u_{it} + v_{it},
\end{aligned} \tag{20}
$$

where $a_{jk} = a_{kj}$ and $c_{mn} = c_{nm}$, for all j, k, m, n, according to symmetry. In this situation, homogeneity of degree one in input prices is imposed (as required by economic theory).

In matrix notation we rewrite equation (20) as

$$
y_{it} = \boldsymbol{x}'_{it} \boldsymbol{\beta} + \alpha_i + \eta_i + u_{it} + v_{it}, \tag{21}
$$

for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$, where y_{it} is the normalized cost, $\boldsymbol{x}_{it}$ is a vector of all variables on the right hand side of (20) including the constant term, $\boldsymbol{\beta}$ is the vector of corresponding coefficients of the translog function, α_i is the firm-specific random effect, and η_i and u_{it} represent persistent and transient inefficiency, respectively, and whose distributions are unknown in our semiparametric model.

4.2 Parameter estimates and mixing performance

We estimated (20) using our proposed sampling algorithm, and for comparison purposes, we also estimated the model under the assumptions discussed by Tsionas and Kumbhakar (2014). Our sampling procedure consists of 50,000 iterations, where the first 20,000 iterations are discarded to allow for burn-in. The estimated coefficients of our semiparametric model appear in Table 6, while a summary of parameter estimates using Tsionas and Kumbhakar's (2014) estimator is presented in Table 7. We found that the signs of the estimated coefficients are mostly consistent across the two models, yet the magnitudes of the coefficients are in some cases substantially different. Moreover, according to the 95% Bayesian credible intervals for the parameters derived under both models, we found that our estimator tends to produce narrower intervals than Tsionas and Kumbhakar's (2014) estimator. Taken together, some differences in magnitudes and in the credible intervals, also lead to different conclusions about the significance of some coefficient from zero (see for example, c_{23} and δ_3).

The estimates of the bandwidths used in our density estimation are reported in Appendix Table B2 along with the corresponding SIF values. As all the SIF values of the parameters and bandwidths are clearly below 100, the sampling algorithm achieved a reasonable mixing performance.

4.3 Transient and persistent inefficiencies

4.3.1 Inefficiencies of a median-inefficiency bank

We consider a bank with median estimated transient inefficiency, that is the median among

$$\bar{u}_{i\cdot}^{\text{(est)}} = \frac{1}{T}\sum_{t=1}^{T} u_{it}^{\text{(est)}}, \quad \text{for} \quad i = 1, 2, \ldots, N.$$

Using our estimator, we obtained time-varying posterior densities of transient inefficiency for the bank with the median inefficiency, identified based on the average inefficiency over time. These density plots are presented in Figure 4. Under Tsionas and Kumbhakar's (2014) estimator, the graphs of the time-varying posterior density of transient inefficiency for the same bank are presented in Figure 5. During the sample period, the transient inefficiency derived from Tsionas and Kumbhakar's (2014) estimator places the majority of probability mass inside the interval $(0, 0.6)$, whereas our proposed approach places a similar amount of probability mass inside a narrower interval, $(0.12, 0.24)$. This result is key, suggesting much less variation in transient efficiency of a bank at the median level of inefficiency over the 8 year period than what would be found assuming Half-Normal distribution for both persistent and transient inefficiency.

With our proposed approach, we obtained the box and whisker plot of the transient inefficiency of the median bank in each year, and such plots are presented in the upper panel of Figure B.1. The high posterior density region of the transient inefficiency increased during 1998–2000, after which it started to decrease until the year 2002. However, during 2002–2005, the 95% high-density region (HDR) steadily moved upward. In contrast to this pattern, with Tsionas and Kumbhakar's (2014) estimator, the HDR of the posterior density of the median bank exhibits little variation over the years according to the box and whisker plot presented in the lower panel of Figure B.1.

Figure 5 presents the graphs of the posterior density of persistent inefficiency of the median bank obtained through our estimator, and that obtained through Tsionas and Kumbhakar's (2014) estimator; The 95% HDRs are (0.01,0.05) and (0, 0.1), respectively. This finding means that more often than not, our estimator tends to predict a lower level of persistent inefficiency than Tsionas and Kumbhakar's (2014) estimator for a median-inefficiency bank.

4.3.2 Inefficiencies of a hypothetical bank

Averaging over the MCMC draws for our proposed approach, we obtained the posterior estimates of η_i and u_{it} denoted respectively, as $\hat{\eta}_i$ and $\hat{u}_{it}$, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. Based on the collection of $\hat{\eta}_i$, we obtained the kernel estimate of the density of persistent inefficiency of an unobserved bank, which we call a hypothetical bank. Similarly, based on the collection of $\hat{u}_{it}$, we derived the kernel estimate of the density of transient inefficiency of the hypothetical bank. The two density estimates are plotted in the upper panel of Figure 6.

The kernel density estimates of the transient inefficiency derived through both estimators are quite close to each other. The 95% HDR of the density estimated with our approach is (0.075, 0.443) with its mode being located at 0.1602, whereas the HDR of the density estimated following Tsionas and Kumbhakar (2014) is (0.074, 0.430) with its median being located at 0.1560.

The kernel estimate of the density of the persistent inefficiency derived under our approach differs from that derived using Tsionas and Kumbhakar's (2014) estimator. The 95% HDR derived through our estimator is (0.012, 0.031) with its mode being 0.0205. However, the 95% HDR derived with Tsionas and Kumbhakar's (2014) estimator is (0.014, 0.069) with a mode of 0.0342 (50% larger).

The kernel density estimates of total inefficiency, which is defined as the sum of the posterior means of persistent and transient inefficiencies, for a hypothetical bank under both models are plotted in Figure 7. Our estimator yields a 95% highest density region (HDR) over the interval (0.088, 0.458), with a mode of 0.1811, while the model of Tsionas and Kumbhakar (2014) produces a 95% HDR over the interval (0.087, 0.491), with a mode of 0.1853. Our estimated density of total inefficiency results in a narrower HDR than that derived from Tsionas and Kumbhakar's (2014) posterior estimates of persistent and transient inefficiencies. The shape and central location of our kernel density estimate for total inefficiency are similar to those based on Tsionas and Kumbhakar's (2014) posterior estimates. This similarity likely arises from the fact that total inefficiency reflects the overall discrepancy between observed cost and the cost frontier, and both models results in similar frontier estimates.

According to our approach, persistent inefficiency contributes 9.44% to the mode of the total inefficiency. Meanwhile, under Tsionas and Kumbhakar's (2014) estimator, persistent inefficiency contributes 16.19% to the mode of total inefficiency.

Further, we compare our kernel density approximations with the estimated Half-Normal densities of inefficiencies derived by plugging in the posterior averages of the parameters from the model of Tsionas and Kumbhakar (2014). In Figure 8, the dashed red curves represent the prior assumptions about transient and persistent inefficiencies. These two curves are clearly different from our density estimates, which are based on the realized inefficiencies.

For transient inefficiencies, the kernel density approximation concentrates its mass within the interval (0.1, 0.3), whereas the Half-Normal density spreads it over a broader range, (0, 0.5). For persistent inefficiencies, the kernel density places its high-density region in the interval (0.01, 0.03), while the Half-Normal density again spans a wider interval, (0, 0.1). In both cases, the kernel density approximation provides a more informative representation of the underlying inefficiency distributions than the Half-Normal assumption.

4.3.3 Prediction accuracy

We compare the prediction accuracy of our estimator with Tsionas and Kumbhakar's (2014) estimator. We leave 10% of banks (14 banks) out for the purpose of assessing prediction performance and use the other 90% of banks (127 banks) for estimation. Denote N_{Train} and N_{Test} as the number of banks in the training set and testing set, respectively. Let y^*_{kt} represent the dependent variable in the testing set for $k = 1, 2, \ldots, N_{\text{Test}}$ and $t = 1, 2, \ldots, T$.

To evaluate the prediction accuracy, we calculate the mean squared prediction error (MSPE) at year t:

$$\text{MSPE}_t = \frac{1}{N_{\text{Test}}} \sum_{k=1}^{N_{\text{Test}}} (\widehat{y}^*_{kt} - y^*_{kt})^2,$$

where $\widehat{y}^*_{kt}$ is the predicted response. Figure 9 presents the calculated MSPE_t over time. Overall, the MSPE_t derived from our estimator is smaller than that obtained from Tsionas and Kumbhakar's (2014) estimator, except for the three years of 2000, 2001 and 2005. Nonetheless, the average MSPE derived from our approach is 0.1123, while the one derived from Tsionas and Kumbhakar's (2014) estimator is 0.1204, indicating that our approach leads to more accurate prediction.

4.4 Returns to scale and technical change

Over the past two decades, a considerable number of studies have examined returns to scale (RTS) at large U.S. banks, driven by the growing dominance of these institutions in the banking industry. RTS can be measured in terms of elasticities of cost:

$$RTS_{it} = 1/\sum_{m=1}^{M} \varepsilon_{mit}, \text{ where } \varepsilon_{mit} = \frac{\partial \log(C_{it}/X_{Lit})}{\partial \log Y_{mit}}.$$

Technical change (TC) has been of interest to economists as well. Usually, a time trend is involved in the translog cost frontier model by including linear and quadratic components of time, and the interactions of the time trend with input prices and outputs of the cost frontier. For such specifications, a measure of TC is given by

$$TC_{it} = -\frac{\partial \log(C_{it}/X_{Lit})}{\partial t}.$$

Figure 10 presents the point estimates of average annual RTS and TC, calculated by averaging over individuals for each year, as well as the corresponding 95% Bayesian credible intervals, respectively. In general, the estimates of RTS and TC derived from the two models are quite close to each other and follow similar patterns. However, the 95% Bayesian credible intervals derived from our estimator are substantially narrower than the corresponding ones from Tsionas and Kumbhakar's (2014) estimator, suggesting more accurate estimation with our proposed approach.

The estimated values of average RTS during 1998–2005 are slightly larger than one under each model, indicating that, on average, there is some evidence for increasing returns to scale among large commercial banks in the US during the sample period. The conclusion is consistent between the two models, but Tsionas and Kumbhakar's (2014) estimator generally shows slightly higher estimates from the year 2000. Interestingly, both models suggest that the average RTS has increased over time, from around 1.05 in 1998 for both models to around 1.08 in 2005 for our estimator (1.10 for Tsionas and Kumbhakar's (2014) estimator). Meanwhile, the average TC estimates fluctuate (roughly) between 2% and 5% per annum, with a peak in 2000 (that is, right before the dot-com bubble crush and the 2001 recession) and a trough in 2004, followed by a quick recovery to 5% in 2005.

5 Conclusion

We propose a new Bayesian approach to estimate persistent and transient inefficiencies, along with their distributions, in the four-component panel stochastic frontier model. This model was originally studied by Tsionas and Kumbhakar (2014) using Bayesian methods under the assumption that both inefficiency components follow Half-Normal distributions. Rather than impose such distributional assumptions, we approximate the joint density of transient inefficiency and the composite error term—defined as the sum of transient inefficiency and Gaussian noise varying over time and space. The density of transient inefficiency is then obtained by dividing this joint density by the marginal density of the composite error. A similar approach is used to approximate the density of persistent inefficiency, with all bandwidths treated as parameters. We construct a proportional posterior and develop a sampling algorithm accordingly. Using pseudo-samples, we demonstrate that our method exhibits a strong mixing performance and outperforms the model of Tsionas and Kumbhakar (2014) as well as four alternative specifications.

We applied our estimator, as well as Tsionas and Kumbhakar's (2014), to analyze the cost efficiency of a panel of large US commercial banks. Both models produced similar estimates of most parameters, but not all. Moreover, the corresponding 95% Bayesian credible intervals derived through our estimator were mostly narrower than those derived from Tsionas and Kumbhakar's (2014) estimator, suggesting that our estimates are more precise; in some cases leading to different conclusions about the parameters.

For a large bank with a median level inefficiency, the 95% high-density region of the transient inefficiency varies over time discovered by our approach. However, this time-varying feature is not obvious for the transient inefficiency estimated through Tsionas and Kumbhakar's (2014) estimator. Moreover, our Bayesian semiparametric econometric approach tends to predict lower levels of persistent inefficiency than Tsionas and Kumbhakar (2014).

When studying inefficiencies of a hypothetical bank, we have found that the densities of transient inefficiency derived through the two models are quite close, but the estimated density of persistent inefficiency derived through our approach differs substantially from that derived through Tsionas and Kumbhakar's (2014) estimator. Upon deriving the kernel density of total inefficiency, we have found that persistent inefficiency contributes 9.44% to the mode of total inefficiency according to our estimator. On the other hand, persistent inefficiency contributes 16.19% to the mode of total inefficiency. The estimates of the average annual RTS and TC derived from the two models follow similar patterns, however, the 95% Bayesian credible intervals derived from our estimator are narrower than corresponding ones from Tsionas and Kumbhakar's (2014) estimator, suggesting more accurate estimation with our approach.

Acknowledgements

The authors would like to thank three anonymous reviewers and the Editor for their constructive comments that improved the quality of this paper.

References

Adams, R.M., Berger, A.N., Sickles, R.C., 1999. Semiparametric approaches to stochastic panel frontiers with applications in the banking industry. Journal of Business & Economic Statistics 17, 349–358. doi:10.2307/1392292.

Aigner, D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21–37. doi:10.1016/0304-4076(77)90052-5.

Amsler, C., James, R., Prokhorov, A., Schmidt, P., 2024. Improving predictions of technical inefficiency, in: Parmeter, C.F., Tsionas, M.G., Wang, H.J. (Eds.), Essays in Honor of Subal Kumbhakar. Emerald Publishing Limited. volume 46, pp. 309–328. doi:10.1108/S0731-905320240000046011.

Amsler, C., Prokhorov, A., Schmidt, P., 2014. Using copulas to model time dependence in stochastic frontier models. Econometric Reviews 33, 497–522. doi:10.1080/07474938.2013.825126.

Battese, G., Coelli, T., 1992. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. Journal of Productivity Analysis 3, 153–169. doi:10.1007/BF00158774.

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, 57–95. doi:10.1016/S1042-9573(02)00006-2.

van den Broeck, J., Koop, G., Osiewalski, J., Steel, M.F.J., 1994. Stochastic frontier models: A Bayesian perspective. Journal of Econometrics 61, 273–303. doi:10.1016/0304-4076(94)90087-6.

Colombi, R., Kumbhakar, S., Martini, G., Vittadini, G., 2014. Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. Journal of Productivity Analysis 42, 123–136. doi:10.1007/s11123-014-0386-y.

Cornwell, C., Schmidt, P., Sickles, R.C., 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46, 185–200. doi:10.1016/0304-4076(90)90054-W.

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, 105–138. doi:10.1002/jae.1021.

Feng, G., Wang, C., Zhang, X., 2019. Estimation of inefficiency in stochastic frontier models: A Bayesian kernel approach. Journal of Productivity Analysis 51, 1–19. doi:10.1007/s11123-018-0542-x.

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 and Finance 36, 1883–1895. doi:10.1016/j.jbankfin.2012.02.008.

Feng, G., Zhang, X., 2014. Returns to scale at large banks in the US: A random coefficient stochastic frontier approach. Journal of Banking and Finance 39, 135–145. doi:10.1016/j.jbankfin.2013.10.012.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. The Annals of Statistics 1, 209–230. doi:10.1214/aos/1176342360.

Fernández, C., Osiewalski, J., Steel, M.F.J., 1997. On the use of panel data in stochastic frontier models with improper priors. Journal of Econometrics 79, 169–193. doi:10.1016/S0304-4076(97)88050-5.

Garthwaite, P.H., Fan, Y., Sisson, S.A., 2016. Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro Process. Communications in Statistics – Theory and Methods 45, 5098–5111. doi:10.1080/03610926.2014.936562.

Greene, W.H., 1990. A Gamma-distributed stochastic frontier model. Journal of Econometrics 46, 141–163. doi:10.1016/0304-4076(90)90052-U.

Greene, W.H., 2005. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126, 269–303. doi:10.1016/j.jeconom.2004.05.003.

Griffin, J.E., Steel, M.F.J., 2004. Semiparametric Bayesian inference for stochastic frontier models. Journal of Econometrics 123, 121–152. doi:10.1016/j.jeconom.2003.11.001.

Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19, 233–238. doi:10.1016/0304-4076(82)90004-5.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. The Review of Economic Studies 65, 361–393. doi:10.1111/1467-937X.00050.

Kneip, A., Sickles, R.C., Song, W., 2012. A new panel data treatment for heterogeneity in time trends. Econometric Theory 28, 590–628. doi:10.1017/S026646661100034X.

Koop, G., Osiewalski, J., Steel, M., 1994. Bayesian efficiency analysis with a flexible form: The AIM cost function. Journal of Business & Economic Statistics 12, 339–346. doi:10.1080/07350015.1994.10524549.

Koop, G., Osiewalski, J., Steel, M.F.J., 1997. Bayesian efficiency analysis through individual effects: Hospital cost frontiers. Journal of Econometrics 76, 77–105. doi:10.1016/0304-4076(95)01783-6.

Koop, G., Steel, M.F.J., Osiewalski, J., 1995. Posterior analysis of stochastic frontier models using Gibbs sampling. Computational Statistics 10, 353–373.

Kumbhakar, S.C., 1990. Production frontiers, panel data, and time-varying technical inefficiency. Journal of Econometrics 46, 201–211. doi:10.1016/0304-4076(90)90055-X.

Kumbhakar, S.C., Park, B.U., Simar, L., Tsionas, E.G., 2007. Nonparametric stochastic frontiers: A local maximum likelihood approach. Journal of Econometrics 137, 1–27. doi:10.1016/j.jeconom.2006.03.006.

Kumbhakar, S.C., Parmeter, C.F., Zelenyuk, V., 2022a. Stochastic Frontier Analysis: Foundations and Advances I. Springer Nature Singapore, Singapore. pp. 331–370. doi:10.1007/978-981-10-3455-8_9.

Kumbhakar, S.C., Parmeter, C.F., Zelenyuk, V., 2022b. Stochastic Frontier Analysis: Foundations and Advances II. Springer Nature Singapore, Singapore. pp. 371–408. doi:10.1007/978-981-10-3455-8_11.

Kumbhakar, S.C., Tsionas, E.G., 2005. Measuring technical and allocative inefficiency in the translog cost system: A Bayesian approach. Journal of Econometrics 126, 355–384. doi:10.1016/j.jeconom.2004.05.006.

Lai, H.P., Kumbhakar, S.C., 2018a. Endogeneity in panel data stochastic frontier model with determinants of persistent and transient inefficiency. Economics Letters 162, 5–9. doi:10.1016/j.econlet.2017.10.003.

Lai, H.P., Kumbhakar, S.C., 2018b. Panel data stochastic frontier model with determinants of persistent and transient inefficiency. European Journal of Operational Research 271, 746–755. doi:10.1016/j.ejor.2018.04.043.

Lai, H.P., Kumbhakar, S.C., 2020. Estimation of a dynamic stochastic frontier model using likelihood-based approaches. Journal of Applied Econometrics 35, 217–247. doi:10.1002/jae.2746.

Lai, H.P., Kumbhakar, S.C., 2023. Panel stochastic frontier model with endogenous inputs and correlated random components. Journal of Business & Economic Statistics 41, 80–96. doi:10.1080/07350015.2021.2001341.

Lee, L.F., Tyler, W.G., 1978. The stochastic frontier production function and average efficiency: An empirical analysis. Journal of Econometrics 7, 385–389. doi:10.1016/0304-4076(78)90061-1.

Li, Q., Racine, J.S., 2006. Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton, New Jersey.

Malikov, E., Kumbhakar, S.C., Tsionas, M.G., 2016. A cost system approach to the stochastic directional technology distance function with undesirable outputs: The case of US banks in 2001–2010. Journal of Applied Econometrics 31, 1407–1429. doi:10.1002/jae.2491.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18, 435–444. doi:10.2307/2525757.

O'Donnell, C.J., Coelli, T.J., 2005. A Bayesian approach to imposing curvature on distance functions. Journal of Econometrics 126, 493–523. doi:10.1016/j.jeconom.2004.05.011.

Park, B.U., Sickles, R.C., Simar, L., 1998. Stochastic panel frontiers: A semiparametric approach. Journal of Econometrics 84, 273–301. doi:10.1016/S0304-4076(97)00087-0.

Park, B.U., Sickles, R.C., Simar, L., 2003. Semiparametric-efficient estimation of AR(1) panel data models. Journal of Econometrics 117, 279–309. doi:10.1016/S0304-4076(03)00149-0.

Park, B.U., Simar, L., 1994. Efficient semiparametric estimation in a stochastic frontier model. Journal of the American Statistical Association 89, 929–936. doi:10.1080/01621459.1994.10476826.

Rendon, S.R., 2013. Fixed and random effects in classical and Bayesian regression. Oxford Bulletin of Economics and Statistics 75, 460–476. doi:10.1111/j.1468-0084.2012.00700.x.

Schmidt, P., Sickles, R.C., 1984. Production frontiers and panel data. Journal of Business & Economic Statistics 2, 367–374. doi:10.2307/1391278.

Sherman, H.D., Gold, F., 1985. Bank branch operating efficiency: Evaluation with data envelopment analysis. Journal of Banking and Finance 9, 297–315. doi:10.1016/0378-4266(85)90025-1.

Sickles, R.C., Zelenyuk, V., 2019. Measurement of Productivity and Efficiency: Theory and Practice. Cambridge University Press. doi:10.1017/9781139565981.

Stevenson, R.E., 1980. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13, 57–66. doi:10.1016/0304-4076(80)90042-1.

Tran, K.C., Tsionas, E.G., 2009. Estimation of nonparametric inefficiency effects stochastic frontier models with an application to British manufacturing. Economic Modelling 26, 904–909. doi:10.1016/j.econmod.2009.02.011.

Tsionas, E., 2000. Full likelihood inference in Normal-Gamma stochastic frontier models. Journal of Productivity Analysis 13, 183–205. doi:10.1023/A:1007845424552.

Tsionas, E.G., 2001. Posterior analysis of stochastic frontier models with truncated normal errors. Computational Statistics 16, 559–575. doi:10.1007/s180-001-8330-0.

Tsionas, E.G., Kumbhakar, S.C., 2014. Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model. Journal of Applied Econometrics 29, 110–132. doi:10.1002/jae.2300.

Tsionas, M., Parmeter, C.F., Zelenyuk, V., 2023. Bayesian artificial neural networks for frontier efficiency analysis. Journal of Econometrics 236, 105491. doi:10.1016/j.jeconom.2023.105491.

Walker, S.G., Damien, P., Laud, P.W., Smith, A.F.M., 1999. Bayesian nonparametric inference for random distributions and related functions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61, 485–527. doi:10.1111/1467-9868.00190.

Yuan, A., 2009. Bayesian frequentist hybrid inference. The Annals of Statistics 37, 2458–2501. doi:10.1214/08-AOS649.

Zhang, X., Brooks, R.D., King, M.L., 2009. A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation. Journal of Econometrics 153, 21–32. doi:10.1016/j.jeconom.2009.04.004.

Zhang, X., King, M.L., Hyndman, R.J., 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. Computational Statistics and Data Analysis 50, 3009–3031. doi:10.1016/j.csda.2005.06.019.

Zhang, X., King, M.L., Shang, H.L., 2014. A sampling algorithm for bandwidth estimation in a nonparametric regression model with a flexible error density. Computational Statistics and Data Analysis 78, 218–234. doi:10.1016/j.csda.2014.04.016.

Table 2: Estimates derived via the proposed approach and competing estimators under Scenario 1 in the simulation study.

		Estimate (standard deviation)				
	True value	Our estimator	Estimator A	Estimator B	Estimator C	Estimator D
β_0	1.0	0.9994 (0.0120)	1.0141 (0.0166)	1.4181 (0.0337)	1.4103 (0.0315)	1.0061 (0.0139)
β_1	1.0	1.0056 (0.0038)	1.0118 (0.0089)	1.0102 (0.0107)	1.0062 (0.0039)	1.0071 (0.0034)
σ_α	0.1	0.1149 (0.0091)	0.1562 (0.0350)	0.3212 (0.0258)	0.3215 (0.0234)	0.1361 (0.0108)
σ_η	0.5	—	0.4859 (0.0408)	—	—	—
σ_u	0.5	—	0.5024 (0.0182)	—	—	—
σ_v	0.1	0.1143 (0.0027)	0.1236 (0.0222)	0.3208 (0.0076)	0.1133 (0.0030)	0.0992 (0.0024)
$\bar{\alpha}$	0.0011 (0.1000)	0.0003 (0.1093)	0.0084 (0.0728)	-0.0011 (0.3030)	0.3191 (0.0011)	0.0007 (0.1313)
$\bar{\eta}$	0.4081 (0.3110)	0.4123 (0.3155)	0.3916 (0.2511)	0.3973 (0.0595)	—	0.4065 (0.3077)
$\bar{u}$	0.4009 (0.3038)	0.4022 (0.3107)	0.4001 (0.2816)	—	0.4042 (0.3104)	0.4008 (0.3036)
$\bar{\rho}$	1.0	0.9915	0.9032	—	0.9920	0.9988
$\bar{d}$	0.0	0.0017	0.0034	—	0.0018	0.0006
ρ_η	1.0	0.9880	0.9157	0.1666	—	0.9682
d_η	0.0	0.0060	0.0110	0.0308	—	0.0095

Note: True values of $\bar{\alpha}$, $\bar{\eta}$, and $\bar{u}$ are the averages of the corresponding simulated true values with their respective standard deviations given in parentheses.

Table 3: Estimates derived via the proposed approach and competing estimators under Scenario 2 in the simulation study.

		Estimate (standard deviation)				
	True value	Our estimator	Estimator A	Estimator B	Estimator C	Estimator D
β_0	1.0	0.9708 (0.0115)	0.9337 (0.0111)	1.4118 (0.0346)	1.4195 (0.0297)	0.9734 (0.0106)
β_1	1.0	0.9961 (0.0034)	0.9914 (0.0048)	0.9930 (0.0046)	0.9955 (0.0034)	0.9950 (0.0034)
σ_α	0.1	0.1082 (0.0085)	0.1900 (0.0335)	0.3559 (0.0259)	0.3491 (0.0251)	0.0986 (0.0083)
σ_η	0.5	—	0.4958 (0.0425)	—	—	—
σ_u	0.5	—	0.2123 (0.0140)	—	—	—
σ_v	0.1	0.1008 (0.0024)	0.1014 (0.0126)	0.1362 (0.0032)	0.1025 (0.0024)	0.1013 (0.0024)
$\bar{\alpha}$	-0.0259 (0.0941)	-0.0003 (0.1034)	0.0090 (0.1066)	-0.0022 (0.3517)	0.0019 (0.3472)	0.0002 (0.0925)
$\bar{\eta}$	0.4438 (0.3305)	0.4496 (0.3230)	0.3992 (0.2465)	0.1014 (0.0557)	—	0.4476 (0.3440)
$\bar{u}$	0.0890 (0.1083)	0.0910 (0.1126)	0.1691 (0.1020)	—	0.0895 (0.1106)	0.0898 (0.1118)
$\bar{\rho}$	1.0	0.9944	0.5417	—	0.9929	0.9952
$\bar{d}$	0.0	0.0007	0.0036	—	0.0008	0.0006
ρ_η	1.0	0.9819	0.9355	-0.2396	—	0.9863
d_η	0.0	0.0050	0.0126	0.0488	—	0.0042

Note: True values of $\bar{\alpha}$, $\bar{\eta}$, and $\bar{u}$ are the averages of the simulated corresponding true values with their respective standard deviations given in parentheses.

Table 4: Estimates derived via the proposed approach and competing estimators under Scenario 3 in the simulation study.

		Estimate (standard deviation)				
	True value	Our estimator	Estimator A	Estimator B	Estimator C	Estimator D
β_0	1.0	0.9705 (0.0319)	0.9207 (0.0543)	2.1402 (0.0901)	2.1457 (0.0716)	0.9593 (0.0183)
β_1	1.0	0.9961 (0.0033)	0.9934 (0.0070)	0.9960 (0.0047)	0.9957 (0.0033)	0.9963 (0.0033)
σ_α	0.1	0.3087 (0.0271)	0.4992 (0.2094)	1.5953 (0.1135)	1.5949 (0.1145)	0.1902 (0.0143)
σ_η	$\sqrt{2}$	—	1.8832 (0.1572)	—	—	—
σ_u	0.5	—	0.2384 (0.0703)	—	—	—
σ_v	0.1	0.0994 (0.0024)	0.1568 (0.0493)	0.1398 (0.0033)	0.0994 (0.0023)	0.0991 (0.0023)
$\bar{\alpha}$	-0.0259 (0.0941)	0.0092 (0.3055)	-0.1037 (0.2902)	-0.0184 (1.5975)	-0.0143 (1.6000)	-0.0019 (0.1870)
$\bar{\eta}$	1.1534 (1.5983)	1.1518 (1.7057)	1.2196 (1.3706)	0.1046 (0.0647)	—	1.1742 (1.6442)
$\bar{u}$	0.0948 (0.1156)	0.0948 (0.1155)	0.1897 (0.0867)	—	0.0949 (0.116)	0.0949 (0.1157)
$\bar{\rho}$	1.0	0.9999	0.5244	—	0.9999	0.9999
$\bar{d}$	0.0	0.0001	0.0041	—	0.0001	0.0001
ρ_η	1.0	0.9957	0.9522	-0.0491	—	0.9981
d_η	0.0	0.0298	0.0269	0.1908	—	0.0175

Note: True values of $\bar{\alpha}$, $\bar{\eta}$, and $\bar{u}$ are the averages of the simulated corresponding true values with their respective standard deviations given in parentheses.

Figure 1: Rank correlation coefficient and Euclidean distance for transient inefficiencies under Scenario 1, where Tsionas and Kumbhakar's (2014) estimator for the model is correctly specified.

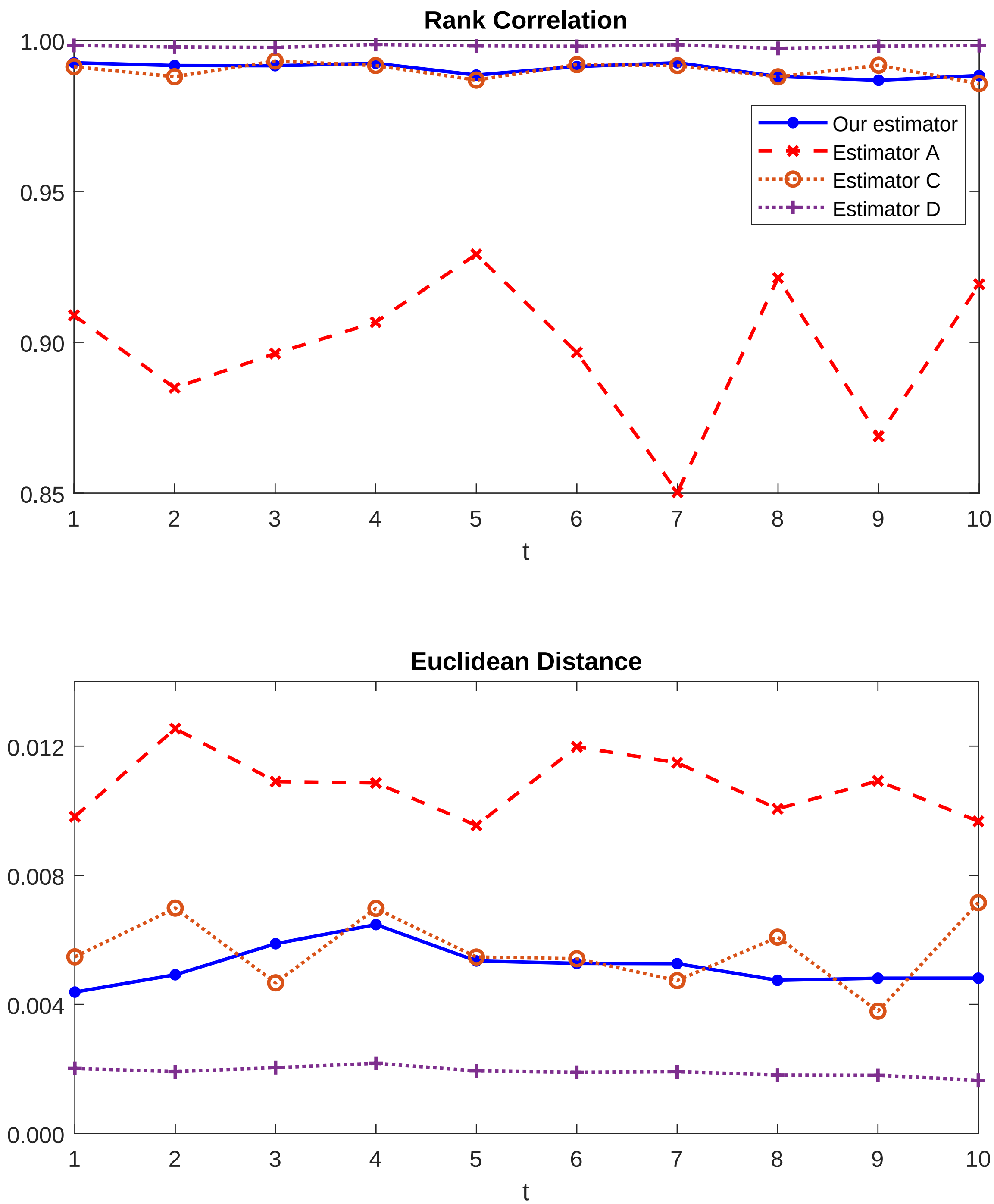


Figure 2: Rank correlation coefficient and Euclidean distance for transient inefficiencies under Scenario 2, where Tsionas and Kumbhakar's (2014) estimator for the model is misspecified for u_{it}.

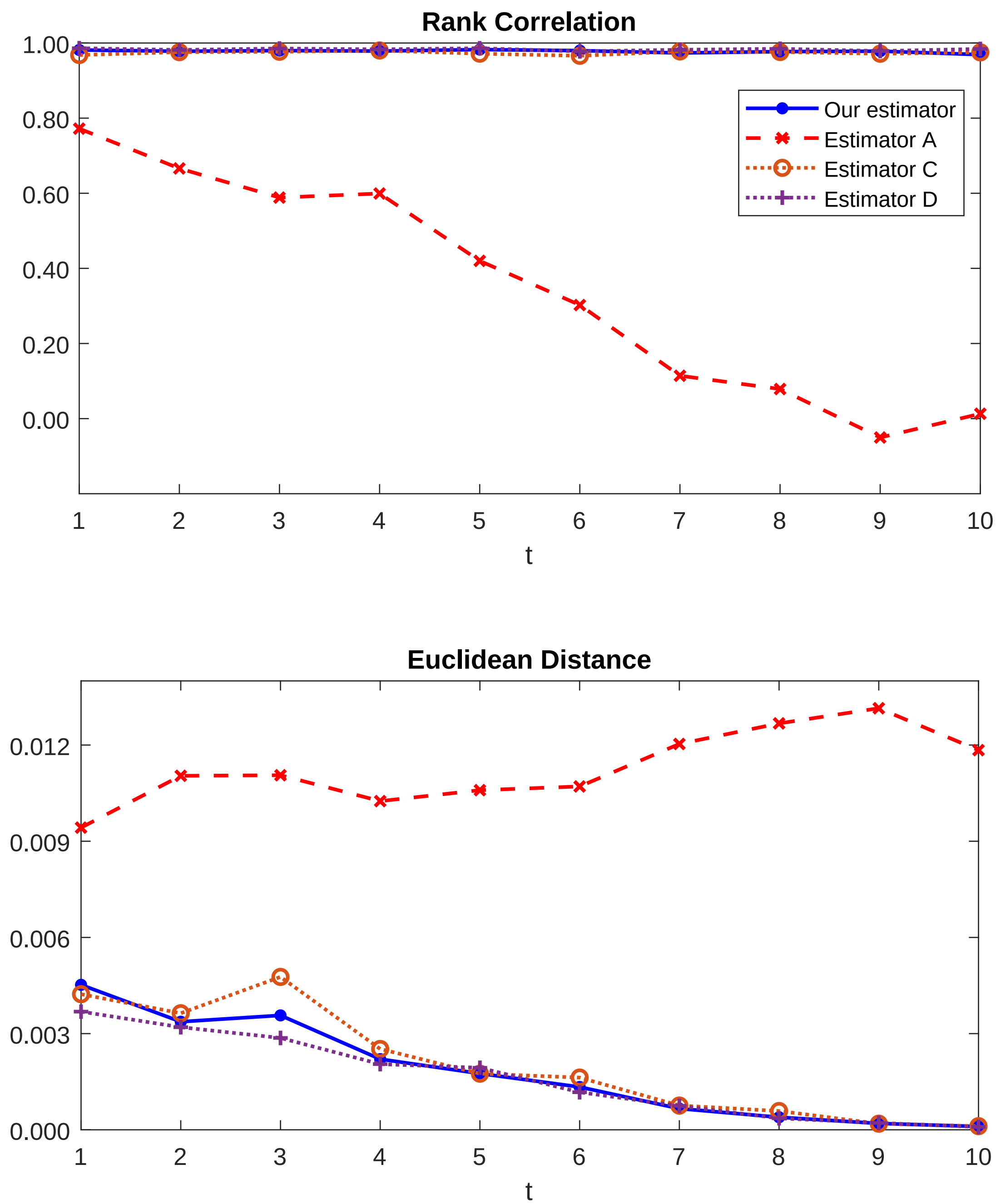


Figure 3: Rank correlation coefficient and Euclidean distance for transient inefficiencies under Scenario 3, where Tsionas and Kumbhakar's (2014) estimator for the model is misspecified for both η_i and u_{it}.

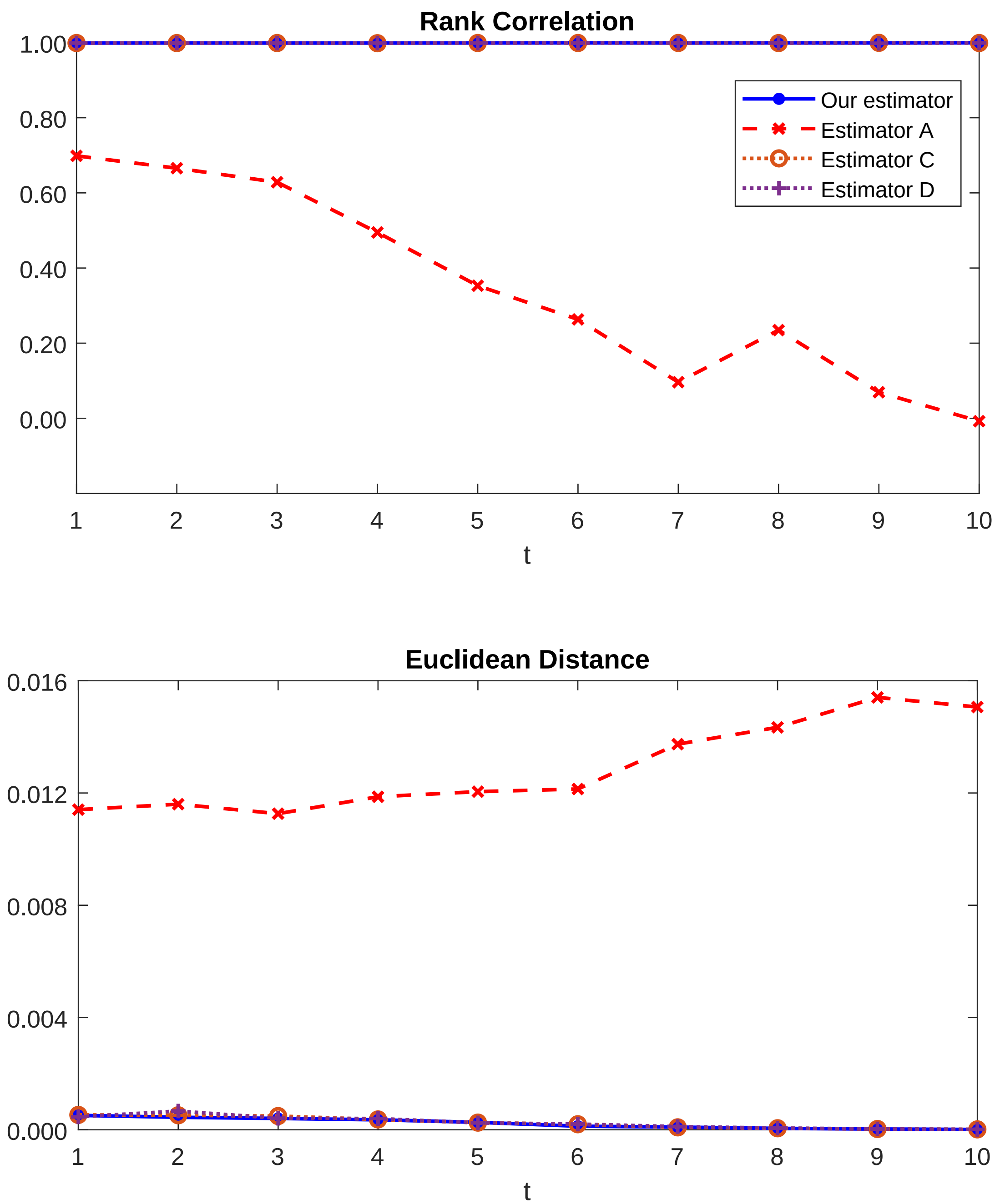
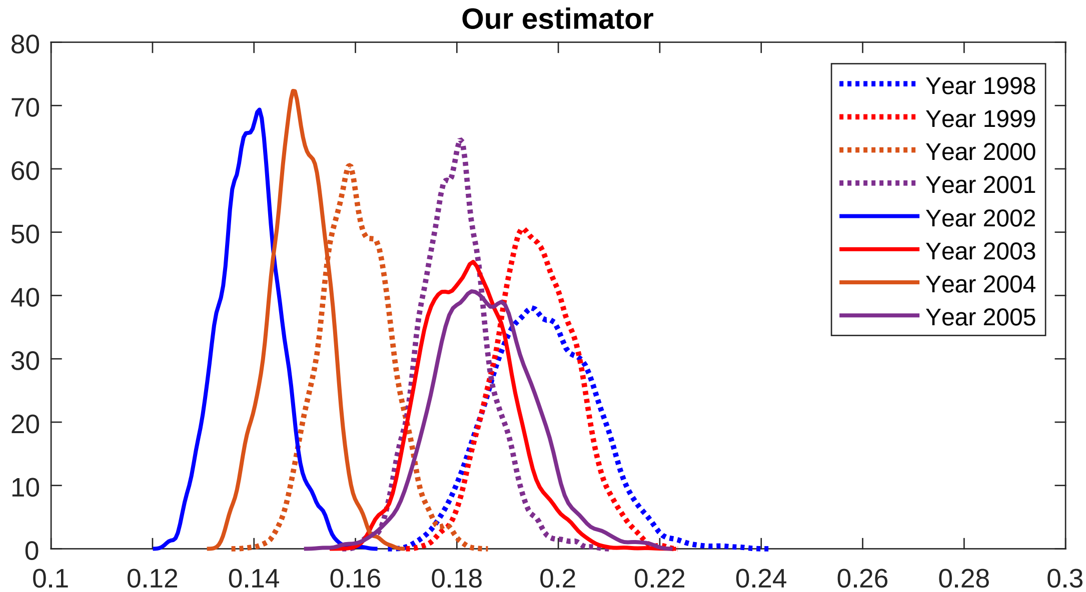


Figure 4: Time-varying posterior density of transient inefficiency for a large US bank (1998–2005) with a median inefficiency.



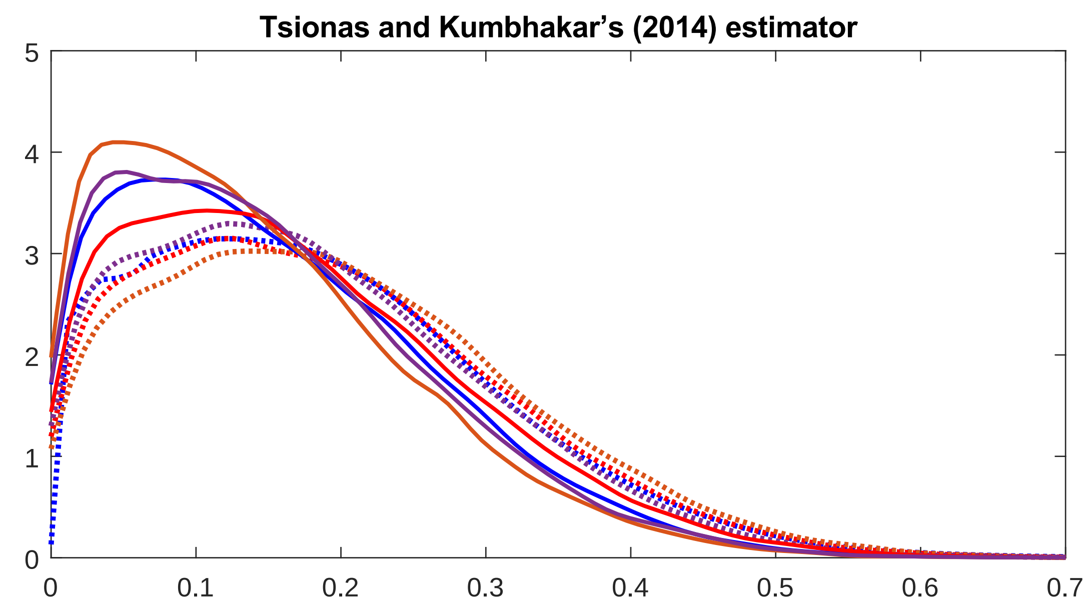


Table 5: Estimates derived via the proposed approach and competing estimators under Scenario 4 in the simulation study.

	True value	Estimate (standard deviation)	
		Our estimator	Estimator D
β_0	1.0	0.9829 (0.0113)	0.9812 (0.0112)
β_1	1.0	0.9956 (0.0034)	0.9964 (0.0033)
σ_α	0.1	0.1051 (0.0086)	0.1057 (0.0082)
σ_η	0.5	—	—
σ_u	0.5	—	—
σ_v	0.1	0.0999 (0.0023)	0.0968 (0.0023)
$\bar{\alpha}$	-0.0259 (0.0941)	-0.0003 (0.0992)	0.0002 (0.1007)
$\bar{\eta}$	0.4438 (0.3305)	0.4397 (0.3259)	0.4399 (0.3204)
$\bar{u}$	0.4627 (0.1629)	0.4625 (0.1623)	0.4635 (0.1623)
$\bar{\rho}$	1.0	0.9950	0.9663
$\bar{d}$	0.0	0.0003	0.0009
ρ_η	1.0	0.9851	0.9819
d_η	0.0	0.0053	0.0052

Note: True values of $\bar{\alpha}$, $\bar{\eta}$, and $\bar{u}$ are the averages of the simulated corresponding true values with their respective standard deviations given in parentheses.

Figure 5: Posterior density of persistent inefficiency for a large US bank (1998-2005) with median inefficiency.

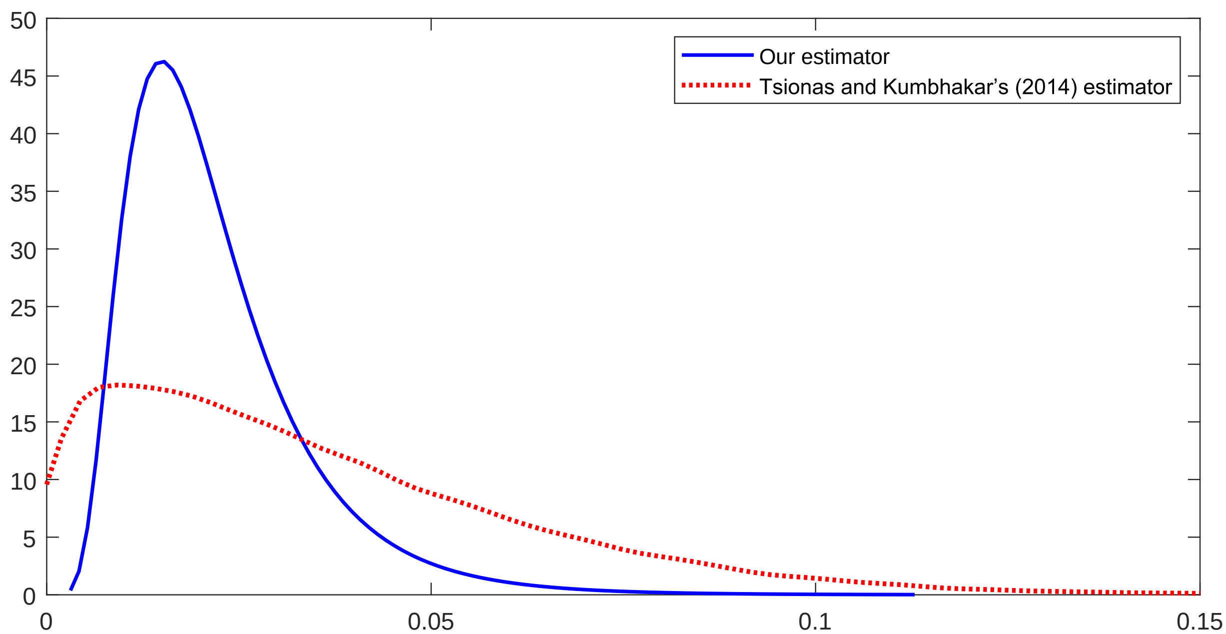


Table 6: Parameter estimates from our proposed approach for the cost frontier of large US banks (1998-2005).

Parameters	Posterior mean	95% Credible interval	SIF
a_0	1.2658	(1.1906, 1.3408)	7.3
a_1	0.3167	(0.2527, 0.3813)	4.4
a_2	0.6688	(0.6233, 0.7143)	3.4
c_1	0.0536	(0.0292, 0.0780)	5.7
c_2	0.4754	(0.4397, 0.5122)	5.5
c_3	0.4546	(0.4117, 0.4982)	6.3
a_{11}	-0.0283	(-0.0713, 0.0152)	4.0
a_{12}	0.0727	(0.0426, 0.1026)	4.2
a_{22}	0.0706	(0.0600, 0.0811)	2.9
c_{11}	-0.0797	(-0.0944, -0.0648)	2.9
c_{12}	-0.0068	(-0.0096, -0.004)	2.0
c_{13}	0.0281	(0.0208, 0.0354)	9.0
c_{22}	-0.0677	(-0.0860, -0.0489)	8.1
c_{23}	0.0105	(-0.0097, 0.0305)	6.9
c_{33}	0.1604	(0.1399, 0.1806)	5.6
g_{11}	-0.2248	(-0.2671, -0.1810)	5.2
g_{12}	0.1036	(0.0761, 0.1307)	5.9
g_{21}	0.0632	(0.0326, 0.0939)	5.9
g_{22}	0.0085	(-0.0424, 0.0602)	5.0
g_{13}	-0.0761	(-0.0923, -0.0599)	3.9
g_{23}	0.0469	(0.0233, 0.0702)	4.2
w_1	-0.1477	(-0.1988, -0.0974)	4.5
w_2	0.0534	(0.0254, 0.0818)	3.9
δ_1	-0.0006	(-0.0030, 0.0017)	3.9
δ_2	0.0095	(0.0051, 0.0139)	2.4
δ_3	-0.0100	(-0.0147, -0.0052)	3.1
γ_1	0.0484	(0.0413, 0.0553)	2.4
γ_2	-0.0303	(-0.0354, -0.0253)	3.0
σ_α	0.1522	(0.1329, 0.1746)	5.2
σ_v	0.0946	(0.0903, 0.0991)	4.6

Table 7: Parameter estimates from Tsionas and Kumbhakar's (2014) estimator for the cost frontier of large US banks (1998–2005).

Parameters	Posterior mean	95% Credible interval	SIF
a_0	1.3137	(1.1679, 1.4589)	2.6
a_1	0.1937	(0.0694, 0.3181)	2.2
a_2	0.6939	(0.6001, 0.7881)	2.4
c_1	0.0989	(0.0572, 0.1403)	2.3
c_2	0.4537	(0.3859, 0.5219)	2.1
c_3	0.4230	(0.3483, 0.4976)	2.0
a_{11}	-0.0001	(-0.0797, 0.0787)	2.7
a_{12}	0.0360	(-0.0243, 0.0978)	2.6
a_{22}	0.0725	(0.0496, 0.0953)	2.5
c_{11}	-0.0806	(-0.1161, -0.0448)	2.1
c_{12}	-0.0088	(-0.0158, -0.0018)	2.3
c_{13}	0.0376	(0.0285, 0.0467)	2.5
c_{22}	-0.1041	(-0.1299, -0.0778)	2.8
c_{23}	0.0316	(0.0020, 0.0613)	3.1
c_{33}	0.1802	(0.1482, 0.2122)	2.7
g_{11}	-0.2042	(-0.2742, -0.1350)	2.4
g_{12}	0.0694	(0.0276, 0.1111)	2.6
g_{21}	0.1032	(0.0543, 0.1525)	3.0
g_{22}	-0.0583	(-0.1475, 0.0302)	2.6
g_{13}	-0.0750	(-0.1055, -0.0447)	2.9
g_{23}	0.0298	(-0.0152, 0.0741)	2.7
w_1	-0.1575	(-0.2466, -0.0685)	2.5
w_2	0.0552	(0.0007, 0.1102)	2.7
δ_1	-0.0020	(-0.0072, 0.0031)	2.3
δ_2	0.0032	(-0.0075, 0.0141)	2.5
δ_3	-0.0051	(-0.016, 0.0059)	2.6
γ_1	0.0636	(0.0473, 0.0801)	2.4
γ_2	-0.0344	(-0.0459, -0.0229)	2.2
σ_α	0.0432	(0.0245, 0.0657)	25.3
σ_η	0.0489	(0.0272, 0.0735)	28.0
σ_u	0.2429	(0.1953, 0.2832)	34.0
σ_v	0.2000	(0.1824, 0.2195)	23.7

Figure 6: Kernel density estimates of inefficiencies for a hypothetical bank.

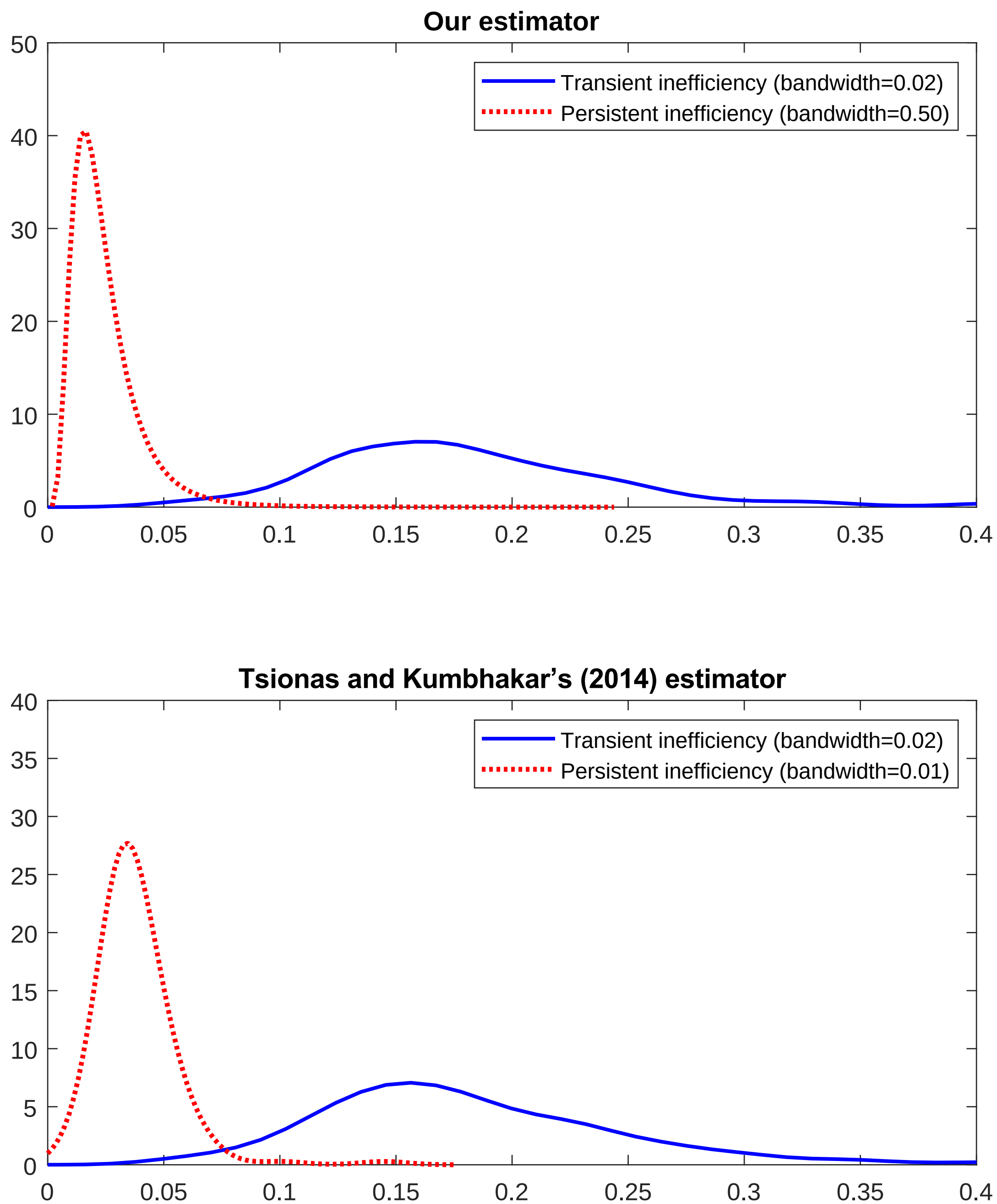


Figure 7: Kernel density estimates of total inefficiency that is the sum of the persistent and transient inefficiencies for a hypothetical bank.

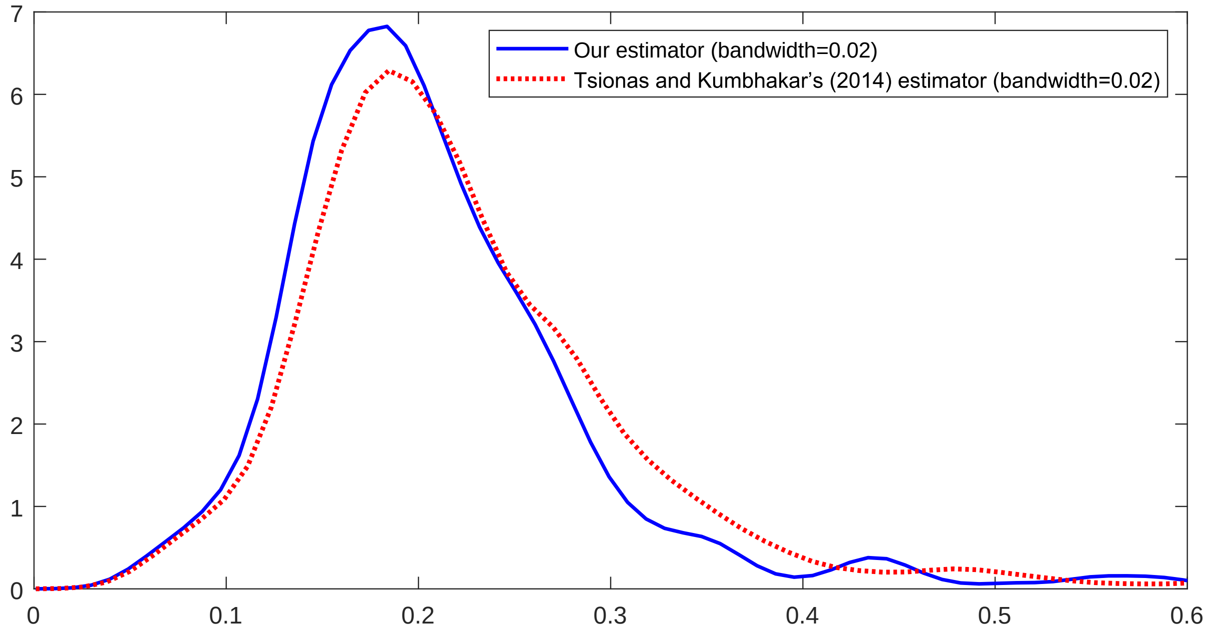


Figure 8: Our estimated densities of transient and persistent inefficiencies in comparison with the parametric densities using parameters estimated following Tsionas and Kumbhakar (2014).

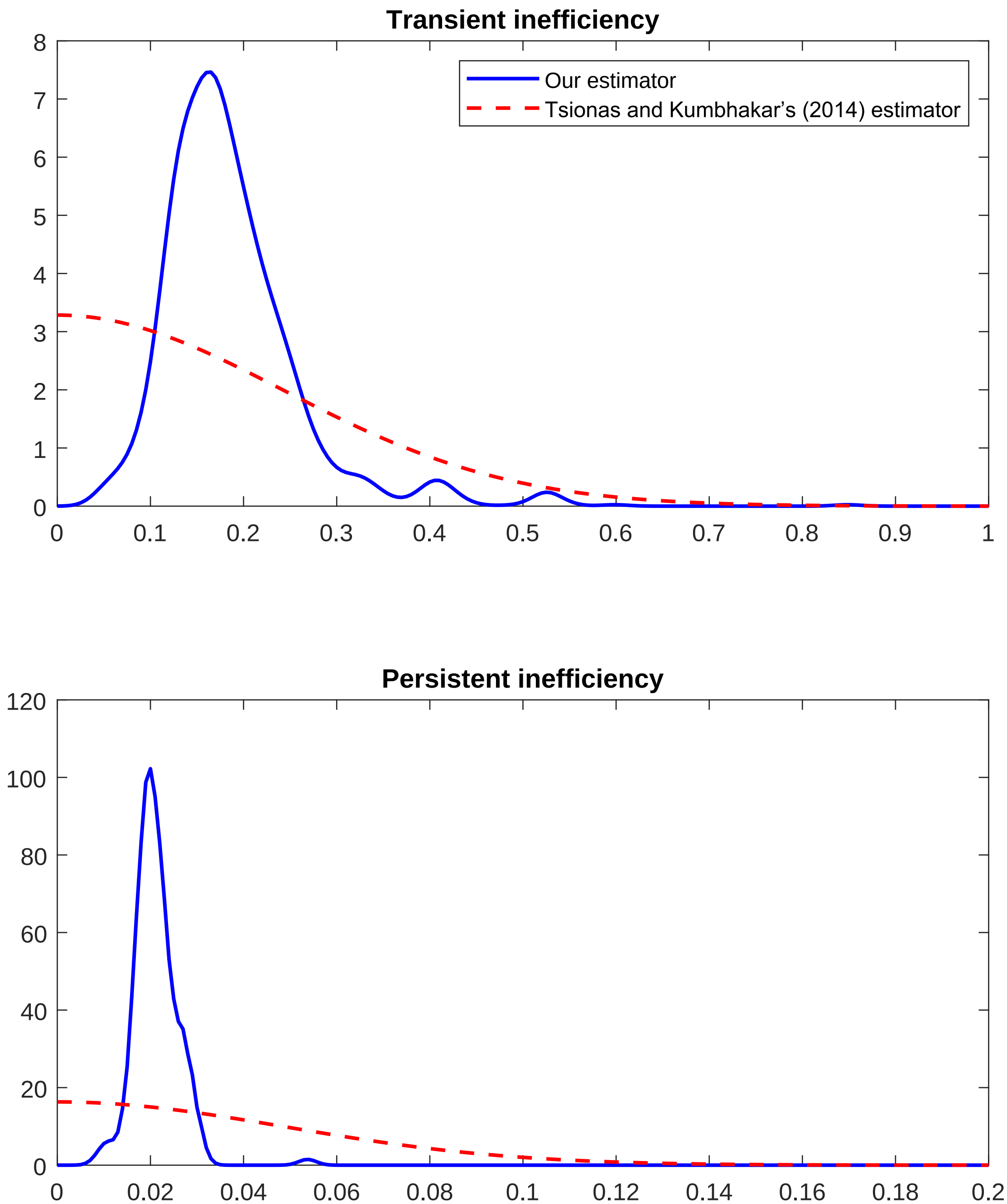


Figure 9: Out-of-sample mean squared prediction errors for the predicted responses of 14 banks through the estimated translog model given by (20).

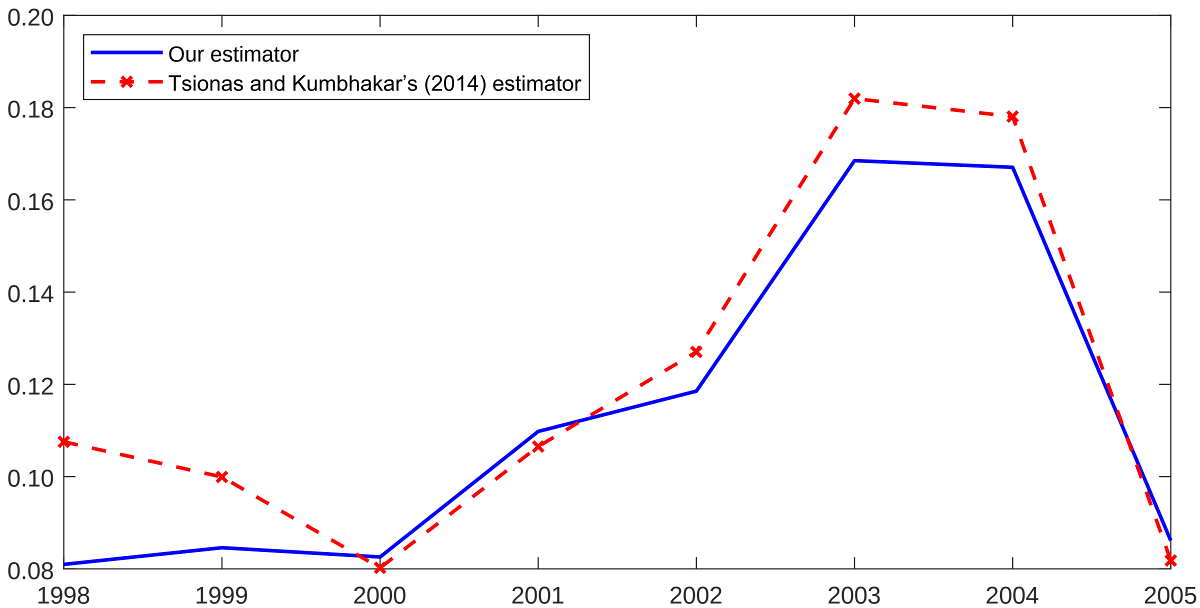
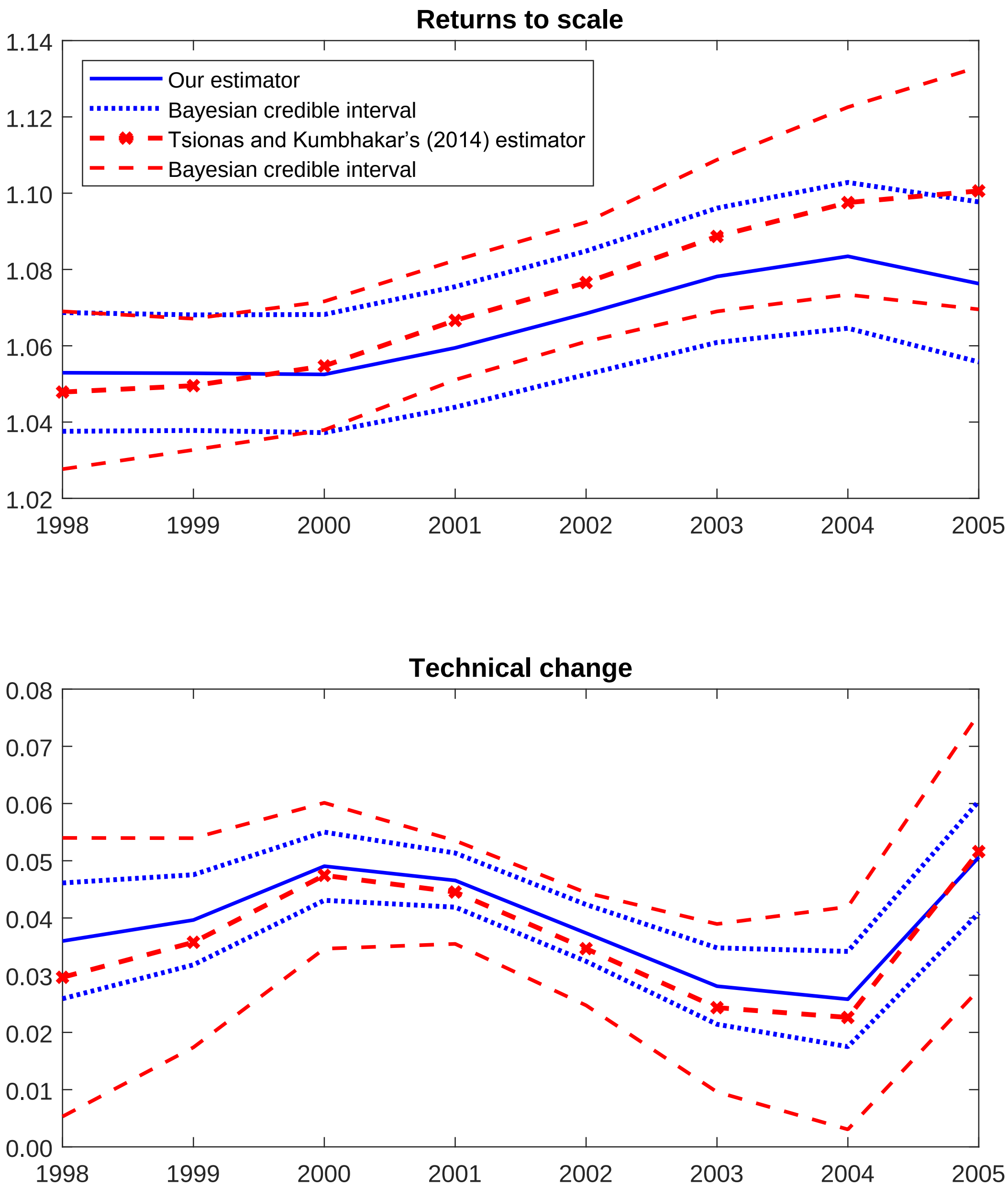


Figure 10: Estimates and the corresponding 95% Bayesian credible intervals of average returns to scale and technical change, derived respectively, from our proposed approach and Tsionas and Kumbhakar's (2014) estimator, of the cost frontier of large US banks (1998-2005).



Appendices

A Details about conditional posteriors

In order to estimate the firm-specific random effects, we develop an algorithm that is based on the sampling algorithm of Tsionas and Kumbhakar (2014). The conditional posterior of β is

$$\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{\eta},\sigma_\alpha^2,\sigma_v^2,\boldsymbol{u},\boldsymbol{h}^2,\boldsymbol{b}^2,\boldsymbol{y},\boldsymbol{X}\sim\mathcal{N}(\widetilde{\beta},\widetilde{V}), \tag{A.1}$$

where

$$\widetilde{V}=\left[\frac{1}{\sigma_v^2}\sum_{i=1}^N\boldsymbol{x}_i'\boldsymbol{x}_i\right]^{-1},\quad \text{and}\quad \widetilde{\beta}=\widetilde{V}\left[\frac{1}{\sigma_v^2}\sum_{i=1}^N\boldsymbol{x}_i'\left(\boldsymbol{y}_i-\alpha_i\boldsymbol{l}_T-\eta_i\boldsymbol{l}_T-\boldsymbol{u}_i\right)\right],$$

with $\boldsymbol{x}_i$ being a $T\times(p+1)$ matrix expressed as $(\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},\cdots,\boldsymbol{x}_{iT})'$, $\boldsymbol{y}_i=(y_{i1},y_{i2},\cdots,y_{iT})'$ and $\boldsymbol{l}_T$ being a vector of dimension T with all components being one.

Further, the conditional posterior of σ_v^2 is

$$\sigma_v^{-2}\left\{\bar{q}_v+\sum_{i=1}^N\sum_{t=1}^T\left(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{\beta}-\alpha_i-\eta_i-u_{it}\right)^2\right\}\Bigg|\ \boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\eta},\boldsymbol{u},\boldsymbol{y},\boldsymbol{X}\sim\chi^2\left(NT+\bar{\lambda}_v\right). \tag{A.2}$$

Next, the conditional posterior of α_i is

$$\alpha_i|\sigma_\alpha^2,\sigma_v^2,\eta_i,\boldsymbol{u}_i,\boldsymbol{h}^2,\boldsymbol{b}^2,\boldsymbol{y}_i,\boldsymbol{x}_i\sim\mathcal{N}\left(\bar{\alpha}_i,\bar{s}^2\right), \tag{A.3}$$

where $\bar{s}^2=\sigma_\alpha^2\left(1+\sigma_\alpha^2\boldsymbol{l}_T{}'\Sigma^{-1}\boldsymbol{l}_T\right)^{-1}$, $\bar{\alpha}_i=\bar{s}^2\boldsymbol{l}_T{}'\Sigma^{-1}(\boldsymbol{y}_i-\boldsymbol{x}_i'\boldsymbol{\beta}-\eta_i-\boldsymbol{u}_i)$ and $\Sigma=\sigma_v^2\boldsymbol{I}_T$.

Further, σ_α^2 can be drawn from the following conditional posteriors:

$$\sigma_\alpha^{-2}(\bar{q}_\alpha+\boldsymbol{\alpha}'\boldsymbol{\alpha})\big|\ \boldsymbol{\beta},\boldsymbol{\eta},\sigma_\eta^2,\sigma_v^2,\boldsymbol{u},\boldsymbol{h}^2,\boldsymbol{y},\boldsymbol{X}\sim\chi^2\left(N+\bar{\lambda}_\alpha\right), \tag{A.4}$$

Then, we draw η_i from the conditional density

$$\eta_i|\boldsymbol{\beta},\boldsymbol{\alpha},\sigma_\alpha^2,\sigma_v^2,\boldsymbol{h}^2,\boldsymbol{b}^2,\boldsymbol{y},\boldsymbol{X}\sim\widehat{f}_{\boldsymbol{b}}(\eta_i|\boldsymbol{\theta}_i,b_\eta,\boldsymbol{b}_\theta) \tag{A.5}$$

for $i=1,2,\dots,N$. Then, the conditional posterior of each component of $\boldsymbol{b}^2$ is

$$\begin{aligned}
&b_\eta^2|\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\eta},\sigma_\alpha^2,\sigma_v^2,\boldsymbol{u},\boldsymbol{y},\boldsymbol{X}\sim\prod_{i=1}^N\widehat{f}_{\boldsymbol{b}}(\eta_i|\boldsymbol{\theta}_i,b_\eta,\boldsymbol{b}_\theta)p(\boldsymbol{b}^2),\ \text{and}\\
&b_t^2|\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\eta},\sigma_\alpha^2,\sigma_v^2,\boldsymbol{u},\boldsymbol{y},\boldsymbol{X}\sim\prod_{i=1}^N\widehat{f}_{\boldsymbol{b}}(\eta_i|\boldsymbol{\theta}_i,b_\eta,\boldsymbol{b}_\theta)p(\boldsymbol{b}^2),\quad t=1,2,\dots,T.
\end{aligned} \tag{A.6}$$

Finally, we approximate the conditional posterior of u_{it} by

$$u_{it}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \sigma_\alpha^2, \sigma_v^2, \boldsymbol{h}^2, \boldsymbol{b}^2, \boldsymbol{y}, \boldsymbol{X} \sim \widehat{f}_{\boldsymbol{h}}(u_{it}|\boldsymbol{\varepsilon}_i, h_u, \boldsymbol{h}_\varepsilon), \tag{A.7}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$.

The conditional posterior of each component of $\boldsymbol{h}^2$ is

$$h_u^2|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \sigma_\alpha^2, \sigma_\eta^2, \sigma_v^2, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{X} \sim \prod_{i=1}^{N}\prod_{t=1}^{T} \widehat{f}_{\boldsymbol{h}}(u_{it}|\boldsymbol{\varepsilon}_i, h_u, \boldsymbol{h}_\varepsilon) p(\boldsymbol{h}^2), \text{ and}$$

$$h_t^2|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \sigma_\alpha^2, \sigma_\eta^2, \sigma_v^2, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{X} \sim \prod_{i=1}^{N}\prod_{t=1}^{T} \widehat{f}_{\boldsymbol{h}}(u_{it}|\boldsymbol{\varepsilon}_i, h_u, \boldsymbol{h}_\varepsilon) p(\boldsymbol{h}^2), \quad t = 1, 2, \ldots, T. \tag{A.8}$$

B Additional tables and figures

Figure B.1: Box and whisker plots of transient inefficiencies for a large US bank (1998-2005) with the median inefficiency at different years.

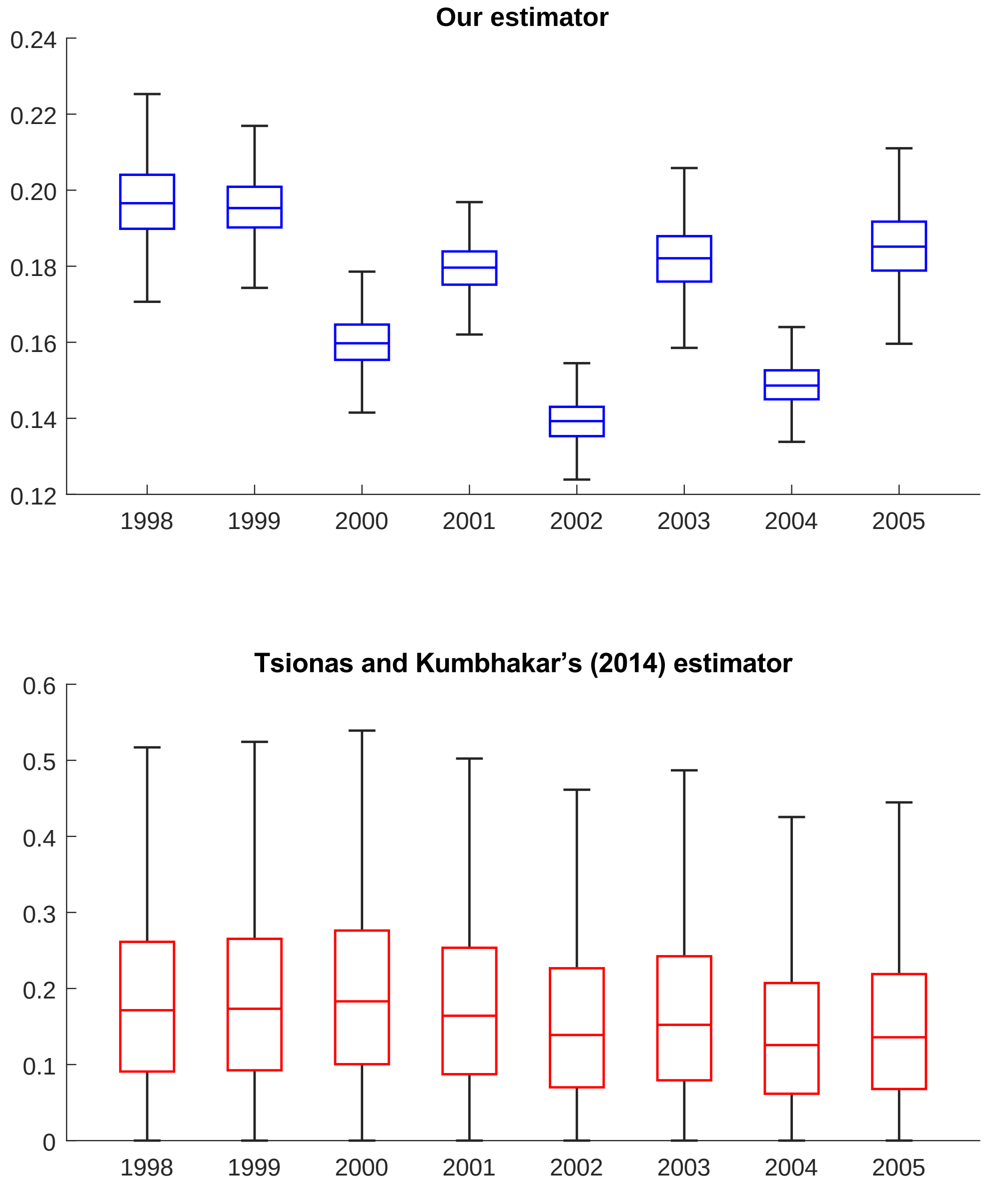


Table B1: Simulation inefficiency factor for each parameter in our simulation studies.

	Scenario 1					Scenario 2				
	Our Estimator	Estimator A	Estimator B	Estimator C	Estimator D	Our Estimator	Estimator A	Estimator B	Estimator C	Estimator D
β_0	11.5	15.7	11.7	18.4	15.0	12.4	14.8	18.1	18.6	12.0
β_1	1.2	7.3	1.1	1.1	1.3	1.3	3.5	1.2	1.3	1.3
σ_α	1.8	13.0	2.2	1.2	2.7	1.7	9.2	1.3	1.1	3.8
σ_η	—	3.4	—	—	—	—	2.9	—	—	—
σ_u	—	10.0	—	—	—	—	14.9	—	—	—
σ_v	1.4	18.5	1.2	4.4	1.6	1.1	17.7	1.1	1.1	1.3
h_u	19.2	—	—	14.6	12.2	16.8	—	—	15.7	8.0
h_1	16.8	—	—	16.2	9.3	16.8	—	—	16.8	8.1
h_2	17.6	—	—	17.6	—	16.7	—	—	17.6	—
h_3	17.2	—	—	14.4	—	18.7	—	—	16.7	—
h_4	18.1	—	—	16.7	—	16.2	—	—	16.4	—
h_5	18.2	—	—	16.2	—	15.6	—	—	15.9	—
h_6	15.5	—	—	14.1	—	16.2	—	—	16.6	—
h_7	17.5	—	—	16.8	—	17.0	—	—	15.4	—
h_8	18.0	—	—	17.3	—	15.8	—	—	15.7	—
h_9	17.6	—	—	17.1	—	16.6	—	—	15.6	—
h_{10}	17.7	—	—	17.6	—	17.0	—	—	15.6	—
b_η	17.5	—	12.7	—	17.6	19.3	—	10.2	—	19.0
b_1	16.8	—	16.6	—	17.3	16.9	—	17.3	—	16.8
b_2	17.6	—	18.4	—	15.8	16.9	—	15.6	—	17.5
b_3	17.4	—	17.4	—	16.7	14.2	—	17.8	—	18.2
b_4	17.5	—	18.3	—	16.6	16.8	—	16.7	—	18.6
b_5	16.1	—	18.5	—	16.1	16.1	—	17.7	—	17.9
b_6	17.9	—	18.8	—	17.5	18.2	—	16.0	—	17.6
b_7	16.7	—	17.6	—	16.2	18.3	—	16.5	—	18.3
b_8	18.4	—	18.6	—	18.6	19.5	—	17.4	—	18.5
b_9	19.3	—	18.0	—	12.5	16.4	—	17.7	—	13.2
b_{10}	17.3	—	17.4	—	16.6	18.6	—	17.0	—	16.8

	Scenario 3					Scenario 4				
	Our Estimator	Estimator A	Estimator B	Estimator C	Estimator D	Our Estimator	Estimator A	Estimator B	Estimator C	Estimator D
β_0	18.8	19.6	19.7	19.8	16.7	12.4	—	—	—	13.1
β_1	1.2	3.8	1.2	1.3	1.3	1.3	—	—	—	1.6
σ_α	7.1	17.3	1.0	1.0	2.2	2.9	—	—	—	1.8
σ_η	—	3.7	—	—	—	—	—	—	—	—
σ_u	—	19.7	—	—	—	—	—	—	—	—
σ_v	1.2	19.7	1.1	1.4	1.1	1.2	—	—	—	1.9
h_u	16.2	—	—	16.5	9.6	18.2	—	—	—	11.6
h_1	12.8	—	—	14.2	8.7	16.9	—	—	—	8.5
h_2	19.7	—	—	16.9	—	17.2	—	—	—	—
h_3	18.2	—	—	15.4	—	18.0	—	—	—	—
h_4	18.7	—	—	17.5	—	17.5	—	—	—	—
h_5	18.9	—	—	16.0	—	16.7	—	—	—	—
h_6	17.9	—	—	17.5	—	17.5	—	—	—	—
h_7	18.0	—	—	16.8	—	17.2	—	—	—	—
h_8	16.9	—	—	17.5	—	16.7	—	—	—	—
h_9	17.6	—	—	18.3	—	15.4	—	—	—	—
h_{10}	18.6	—	—	17.4	—	16.8	—	—	—	—
b_η	17.7	—	10.6	—	17.3	17.7	—	—	—	17.4
b_1	15.8	—	16.5	—	18.7	16.6	—	—	—	17.4
b_2	17.2	—	17.0	—	18.5	14.8	—	—	—	18.8
b_3	18.2	—	18.6	—	18.3	14.7	—	—	—	18.2
b_4	19.1	—	16.9	—	19.6	15.7	—	—	—	16.9
b_5	17.7	—	17.2	—	17.9	14.3	—	—	—	17.5
b_6	19.2	—	18.0	—	18.2	16.0	—	—	—	16.6
b_7	18.6	—	17.7	—	18.5	15.2	—	—	—	16.5
b_8	17.2	—	17.1	—	18.6	17.0	—	—	—	17.4
b_9	18.5	—	17.6	—	19.0	16.7	—	—	—	17.3
b_{10}	18.9	—	17.5	—	18.6	15.4	—	—	—	16.3

Table B2: Bandwidth estimates from our estimator for the cost frontier of large US banks (1998–2005).

Parameters	Posterior mean	95% Credible interval	SIF
h_u	0.0149	(0.0098, 0.0208)	42.8
h_1	1.6017	(0.2195, 4.0787)	37.0
h_2	0.4968	(0.0699, 2.8135)	42.3
h_3	1.3053	(0.0931, 3.8154)	35.7
h_4	0.2748	(0.0547, 2.4713)	51.3
h_5	1.0831	(0.1407, 3.5606)	37.4
h_6	0.0831	(0.0615, 0.1149)	41.1
h_7	0.8506	(0.1241, 3.3292)	43.4
h_8	0.1587	(0.0964, 0.3085)	47.0
b_η	0.0019	(0.0012, 0.0028)	32.2
b_1	1.5807	(0.4071, 3.1868)	30.5
b_2	1.6170	(0.4580, 3.1802)	31.6
b_3	1.5574	(0.3970, 3.0848)	32.9
b_4	1.5896	(0.4817, 3.1578)	30.6
b_5	1.6145	(0.4588, 3.2455)	33.1
b_6	0.0586	(0.0333, 0.0929)	18.6
b_7	1.5576	(0.3976, 3.1516)	32.1
b_8	1.6084	(0.4456, 3.1661)	31.0