A Bayesian Stochastic Frontier Model for Analysing Cost Efficiency of Commercial Banks in the US

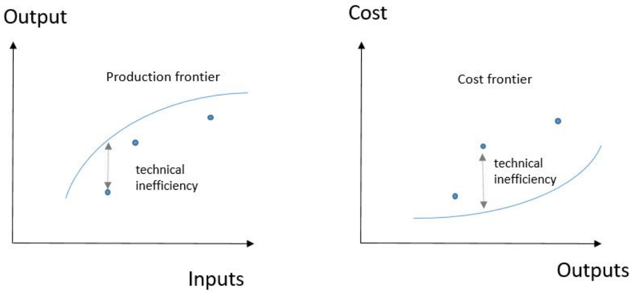
Puguang Nie　　Xibin Zhang

Monash Business School, Monash University, Australia

BNP Workshop, 4–7 December 2023

Introduction

- Stochastic frontier models introduced by Aigner et al. (1977) and Meeusen and Van Den Broeck (1977) have been often used to evaluate a firm's efficiency in productivity studies.
- A production frontier represents maximum outputs obtained from a given set of inputs.
- A cost frontier model minimises the cost for given levels of output and input.



Stochastic frontier models

Consider a panel-form cost frontier model:

$$y_{it} = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + u_i + v_{it}, \tag{1}$$

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

- y_{it} represents the logarithm of cost
- $\boldsymbol{x}_{it}$ is a $k+1$ dimensional vector of explanatory variables
- $\boldsymbol{\beta}$ is a vector of coefficients
- u_i is non-negative and time-invariant inefficiency of firm i
- v_{it} is the random error and is assumed to be iid $N(0, \sigma^2)$
- Let $\varepsilon_{it} = u_i + v_{it}$ which is regarded the composite error

Assumptions on the distribution of u_i

To estimate firm-level inefficiencies, it is often necessary to impose assumptions on the one-sided distribution of u_i.

- half normal distribution (Aigner et al., 1977),
- exponential distribution (Meeusen and Van Den Broeck, 1977),
- truncated normal distribution (Stevenson, 1980),
- Gamma distribution (Greene, 1990).

As any parametric assumption about the distribution of u_i can be subjective, some studies have proposed using

- unknown density function (Park and Simar, 1994),
- Dirichlet process prior (Griffin and Steel, 2004),
- kernel density estimator (Feng et al., 2019).

Motivation

Limitation: These studies have relaxed the distribution assumption on u_i, but they all aim to estimate u_i or estimate the density of u_i using information of $(u_1, u_2, \ldots, u_N)$, rather than using the composite errors $\varepsilon_{it} = u_i + v_{it}$.

Jondrow et al. (1982) proposed using the condition mean of u_i given ε_i, which is $E[u_i|\varepsilon_i]$, as a point prediction of u_i. This method is a popular approach to the estimation of firm-level inefficiencies.

Let $\vec{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$. We aim to

- derive a kernel density estimator for the conditional distribution of u_i given $\vec{\varepsilon}_i$; and
- generate u_i directly from this conditional density within a Markov chain Monte Carlo procedure.

Conditional density of u_i

Conditional density of u_i is given by

$$p(u_i|\vec{\varepsilon}_i) = \frac{p(u_i, \vec{\varepsilon}_i)}{p(\vec{\varepsilon}_i)}, \tag{2}$$

The joint density of $(u_i, \vec{\varepsilon}_i)$ is approximated by

$$\widehat{p}(u_i, \vec{\varepsilon}_i | h_u, h_1, h_2, \dots, h_T) \tag{3}$$

$$= \frac{1}{N-1} \frac{1}{h_u h_1 \cdots h_T} \sum_{\substack{j=1 \\ j \neq i}}^{N} k\left(\frac{u_i - u_j}{h_u}\right) \boldsymbol{K}\left(\frac{\vec{\varepsilon}_i - \vec{\varepsilon}_j}{\boldsymbol{h}}\right)$$

$$= \frac{1}{N-1} \frac{1}{h_u h_1 \cdots h_T} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left\{ k\left(\frac{u_i - u_j}{h_u}\right) \prod_{t=1}^{T} k\left(\frac{\varepsilon_{it} - \varepsilon_{jt}}{h_t}\right) \right\}, \tag{4}$$

for $i = 1, 2, \cdots, N$.

Conditional density of u_i

Marginal density of $p(\vec{\varepsilon}_i)$ is approximated by

$$\begin{aligned}
&\widehat{p}\left(\vec{\varepsilon}_i | h_1, h_2, \ldots, h_T\right) \\
&= \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{h_1 h_2 \cdots h_T} \boldsymbol{K}\left(\frac{\varepsilon_{i,1} - \varepsilon_{j,1}}{h_1}, \frac{\varepsilon_{i,2} - \varepsilon_{j,2}}{h_2}, \ldots, \frac{\varepsilon_{i,T} - \varepsilon_{j,T}}{h_T}\right) \\
&= \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left\{ \frac{1}{h_1 h_2 \cdots h_T} \prod_{t=1}^{T} k\left(\frac{\varepsilon_{it} - \varepsilon_{jt}}{h_t}\right) \right\}, \qquad (5)
\end{aligned}$$

for $i = 1, 2, \cdots, N$.

Bayesian estimation

Let $\boldsymbol{x_i} = (\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, \cdots, \boldsymbol{x_{iT}})'$, $\boldsymbol{y_i} = (y_{i1}, y_{i2}, \cdots, y_{iT})'$,
$\boldsymbol{X} = (\boldsymbol{x_1'}, \boldsymbol{x_2'}, \cdots, \boldsymbol{x_N'})'$, $\mathbf{y} = (\boldsymbol{y_1'}, \boldsymbol{y_2'}, \cdots, \boldsymbol{y_N'})'$,
$\boldsymbol{u} = (u_1, u_2, \ldots, u_N)'$.

Assuming that v_{it}, for $t = 1, , 2 \cdots, T$, are iid $N(0, \sigma^2)$, we obtain the likelihood as

$$L(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{u}, \sigma^2) = \prod_{i=1}^{N} \prod_{t=1}^{T} f\left(y_{it} \middle| \boldsymbol{\beta}, u_i, \sigma^2\right), \tag{6}$$

where

$$f\left(y_{it} \middle| \boldsymbol{\beta}, u_i, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y_{it} - \boldsymbol{x_{it}'}\boldsymbol{\beta} - u_i)^2\right]. \tag{7}$$

Note that the likelihood is conditional of u_i, for $i = 1, 2, \cdots, N$.

Bayesian estimation

- In the situation where the density of $\boldsymbol{u}$ is assumed to be a parametric density denoted as $p(\boldsymbol{u})$, the posterior would be

$$\pi\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{u}|\mathbf{y}\right) \propto L(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{u})\pi(\boldsymbol{\beta},\sigma^2)p(\boldsymbol{u}), \tag{8}$$

- We assume that $p(\boldsymbol{u}) = p(\boldsymbol{u},\varepsilon)/p(\varepsilon)$, and is approximated by

$$\widehat{p}(\boldsymbol{u}|\boldsymbol{h}) = \widehat{p}(\boldsymbol{u},\varepsilon|\boldsymbol{h})/\widehat{p}(\varepsilon|\boldsymbol{h}),$$

where $\boldsymbol{h} = (h_u, h_1, h_2, \ldots, h_T)'$ is the vector of bandwidths used by the above kernel density estimates.

- The posterior of $(\boldsymbol{\beta},\sigma^2,\boldsymbol{u},\boldsymbol{h})'$ is approximately

$$\pi\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{u},\boldsymbol{h}|\mathbf{y}\right) \propto L(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{u})\pi(\boldsymbol{\beta},\sigma^2)\widehat{p}(\boldsymbol{u}|\boldsymbol{h})\pi(\boldsymbol{h}), \tag{9}$$

- Prior choices

The prior of $(\boldsymbol{\beta},\sigma^2)$ is assumed to be

$$\pi(\boldsymbol{\beta},\sigma^2) \propto 1/\sigma^2. \tag{10}$$

Bayesian estimation

Choose an inverse Gamma prior for each squared bandwidth:

$$\pi\left(h_u^2\right) = \frac{\theta_2^{\theta_1}}{\Gamma\left(\theta_1\right)}\left(\frac{1}{h_u^2}\right)^{\theta_1+1}\exp\left(-\frac{\theta_2}{h_u^2}\right),$$

$$\pi\left(h_t^2\right) = \frac{\theta_2^{\theta_1}}{\Gamma\left(\theta_1\right)}\left(\frac{1}{h_t^2}\right)^{\theta_1+1}\exp\left(-\frac{\theta_2}{h_t^2}\right), \quad \text{for } t = 1, \ldots, T,$$

We can sequentially simulate parameters, $\boldsymbol{u}$ and $\boldsymbol{h}^2$ from the following conditional posteriors:

$$\pi\left(\boldsymbol{\beta}|\mathbf{y}, \sigma^{-2}, \boldsymbol{u}, \boldsymbol{h}^2\right) \propto f_{\text{Normal}}\left(\boldsymbol{\beta} \mid \bar{\boldsymbol{\beta}}, \overline{\mathbf{V}}\right) \tag{11}$$

$$\pi\left(\sigma^{-2}|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{h}^2\right) \propto f_{\text{Gamma}}\left(\sigma^{-2} \mid \bar{\gamma}, \bar{s}^2\right), \tag{12}$$

$$\pi\left(\boldsymbol{u}|\mathbf{y}, \boldsymbol{\beta}, \sigma^{-2}, \boldsymbol{h}^2\right) \propto \prod_{i=1}^{N} \widehat{f}(u_i|\vec{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}), \tag{13}$$

$$\pi\left(\boldsymbol{h}^2|\mathbf{y}, \boldsymbol{\beta}, \sigma^{-2}, \boldsymbol{u}\right) \propto \prod_{i=1}^{N} \widehat{f}(u_i|\vec{\boldsymbol{\varepsilon}}_{\boldsymbol{i}})p(\boldsymbol{h}^2). \tag{14}$$

Benchmark models

- Exponential distribution of u_i:

$$p\left(u_i|\lambda^{-1}\right) = \lambda \exp\left(-\lambda u_i\right), \text{for } \; i = 1, \cdots, N, \tag{15}$$

$$p\left(\lambda^{-1}\right) = f_{\text{Gamma}}\left(\lambda^{-1}|1, -\ln \kappa^*\right). \tag{16}$$

- Kernel-based density of u_i proposed by Feng et al. (2019):

$$p\left(u_i \mid \tau^2\right) = \frac{1}{(N-1)u_i} \sum_{j=1;j\neq i}^{N} \frac{1}{\tau}\phi\left(\frac{\ln u_i - \ln u_j}{\tau}\right) I\left(u_i > 0\right), \tag{17}$$

for $i = 1, 2, \cdots, N$, where τ represents the bandwidth, and $I(\cdot)$ is an indicator function, which equals 1 for a true argument and 0 otherwise.

Simulation studies

Data generating process is given by

$$y_{it} = 1 + 0.75x_{1,it} + 0.25x_{2,it} + u_i + v_{it}, \tag{18}$$

- $x_{1,it}$ and $x_{2,it}$ are randomly generated from the uniform distribution $U(0,1)$
- v_{it} is generated from $N(0, 0.2^2)$
- u_i is generated from an exponential distribution with the rate of $-\log(0.85)$
- Euclidean distance: $d = \frac{1}{N}\left(\sum_{i=1}^{N}\left(u_{\text{est},i} - u_{\text{true},i}\right)^2\right)^{1/2}$
- Spearman rank correlation: $\rho = 1 - \frac{6\sum_{i=1}^{N}(\text{Rank}_{i1} - \text{Rank}_{i2})^2}{N(N^2-1)}$

Simulation Studies

Table 1: Monte Carlo Simulation Results

	Our model	Exponential	Kernel-based
Sample size: $N = 100$ and $T = 5$			
Average Euclidean distance	0.0111	0.0071	0.0098
Spearman rank correlation	0.8201	0.7743	0.8225
Sample size: $N = 200$ and $T = 5$			
Average Euclidean distance	0.0037	0.0050	0.0033
Spearman rank correlation	0.9653	0.7761	0.9644
Sample size: $N = 100$ and $T = 20$			
Average Euclidean distance	0.0090	0.0040	0.0032
Spearman rank correlation	0.9029	0.9127	0.9080
Sample size: $N = 200$ and $T = 20$			
Average Euclidean distance	0.0029	0.0028	0.0019
Spearman rank correlation	0.9799	0.9092	0.9854

Simulation Studies

True specification

- Exponential assumption

Average Euclidean distance

- Our model and the kernel-based model cannot beat the true model for small panels (N=100)
- The kernel-based model outperforms the true model for large panels (N=200)
- Our model outperforms the true model when T=5 and is as good as the true model T=20 for large panels (N=200)

Rank Correlation

- Our model and the kernel-based model outperform the true model for short panels (T=5)
- Kernel-based model outperforms true model when N=200 and T=20; our model is best for N=200 and T=5

Simulation Studies: Distributions of inefficiency

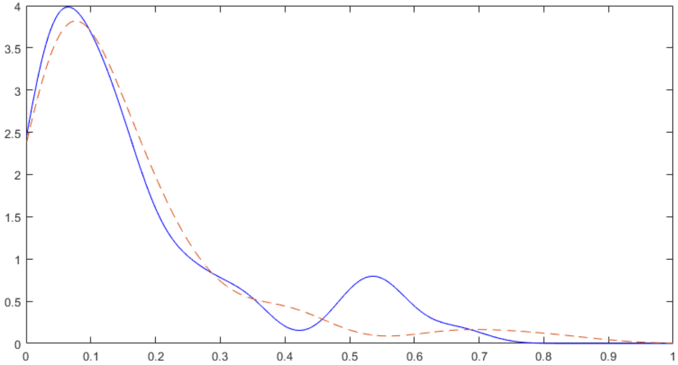


Figure 2: Posterior predictive inefficiency density for an unobserved individual. Solid blue line represents our estimate, and dashed red line represents the kernel-based estimate.

Simulation studies: Distributions of inefficiency

- Our approach and the kernel-based approach can approximate the shape of the exponential density well.
- Our model produces a certain amount of probability mass in $(0.4, 0.7)$, while the kernel-based approach does not show this feature.

Cost efficiency of commercial banks in the US

We apply our model to analyse the cost efficiency of large banks in the US. A translog cost frontier model with inputs X, outputs Y, and cost C is defined as

$$\begin{aligned}\log(C/X_L) &= a_0 + \sum_{j=1}^{L-1} a_j \log\frac{X_j}{X_L} + \frac{1}{2}\sum_{j=1}^{L-1}\sum_{k=1}^{L-1} a_{jk}\log\frac{X_j}{X_L}\log\frac{X_k}{X_L} \\ &+ \sum_{m=1}^{M} b_m \log Y_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} b_{mn}\log Y_m \log Y_n \\ &+ \frac{1}{2}\sum_{j=1}^{L-1}\sum_{m=1}^{M} g_{jm}\log\frac{X_j}{X_L}\log Y_m + \sum_{j=1}^{L-1} h_j t\log\frac{X_j}{X_L} \\ &+ \sum_{m=1}^{M}\kappa_m t\log Y_m + \tau_1 t + \frac{1}{2}\tau_2 t^2 + u_i + v_{it}, \qquad (19)\end{aligned}$$

where t is the time period, $a_{jk} = a_{kj}$ and $b_{mn} = b_{nm}$ due to symmetry.

Data

Description of data is as follows.

Table 2: Description of data

$N = 466$	Number of large banks (assets of at least \$1 billion in 1986)
$T = 80$	Quarterly data during 1986 – 2005
C	Total cost (sum of the cost of 3 inputs)
Inputs ($L = 3$)	
X_1	Quantity of labor
X_2	Quantity of purchased funds and deposits
X_3	Quantity of physical capital
Outputs ($M = 3$)	
Y_1	Consumer loans
Y_2	Securities (all non-loan financial assets)
Y_3	Non-consumer loans (industrial, commercial, and real estate loans)

Key measures

- Posterior distributions of the inefficiency for five selected banks (minimum, maximum, and 3 quartiles of the estimated inefficiencies through our model).
- Posterior predictive distribution of technical efficiency for an unobserved bank.
- Technical efficiency is defined as $\exp(-u_i)$ in a cost frontier model.
- Returns to scale:

$$RTS = \left(\sum_{m=1}^{M} \varepsilon_m \right)^{-1}, \text{ where } \varepsilon_m = \frac{\partial \log(C/X_L)}{\partial \log Y_m}.$$

- Technical change:

$$TC = -\frac{\partial \log(C/X_L)}{\partial t}.$$

Main findings: Posterior distribution of u_i

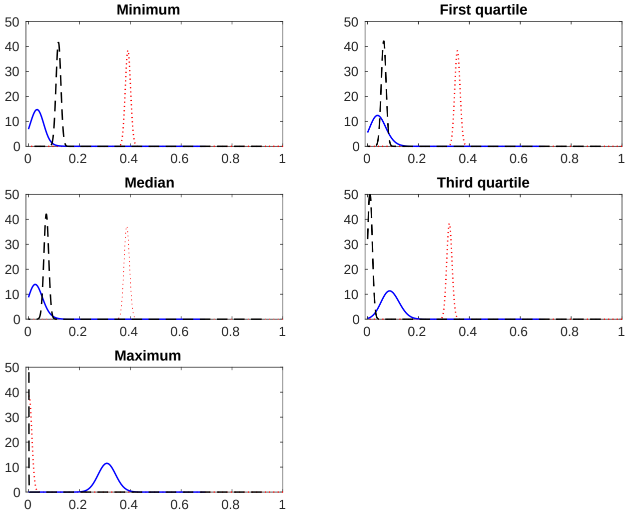


Figure 3: Posterior distributions of the inefficiency for five selected banks. Solid blue line represents our estimates, dotted red line represents the estimates via the exponential model, and dashed black line represents the estimates via the kernel-based model.

Main findings: Posterior distribution of u_i

- The exponential model tends to overestimate inefficiencies, except for the maximum inefficiency.
- As a result, it will underestimate the technical efficiency due to the restriction of exponential assumption on u_i. This phenomenon is consistent with the findings of Griffin and Steel (2004) and Feng et al. (2019).
- Inefficiencies estimated from our model and the kernel-based model are close to each other, except for the estimated maximum inefficiency.
- Densities estimated through the kernel-based model and exponential model have high peaks, suggesting these two tend to have narrow high density regions for the estimated inefficiencies.

Distribution of technical efficiency e^{-u_i}

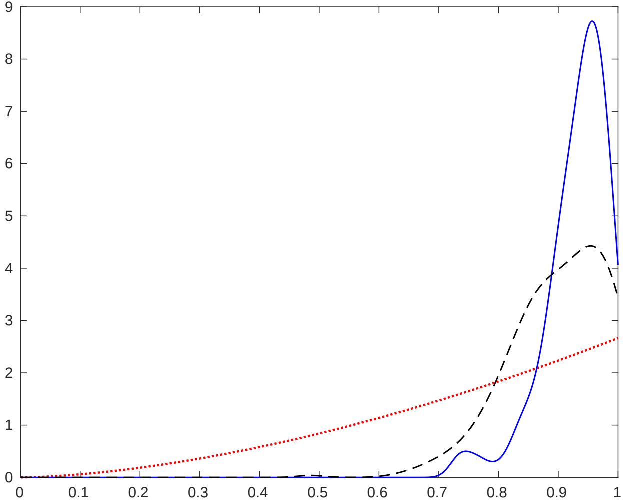


Figure 4: Posterior predictive distributions of efficiency for an unobserved bank. Solid blue line represents our estimates, dotted red line represents the exponential model, and dashed black line represents the kernel-based model.

Distribution of technical efficiency e^{-u_i}

- Our model has a very high peak around 0.95
- Our model has large probability mass on the interval $(0.8, 1)$, while the kernel-based model has a similar probability mass over a wider interval $(0.7, 1)$.
- The exponential model cannot produce similar probability mass.

Main findings: Return to scale

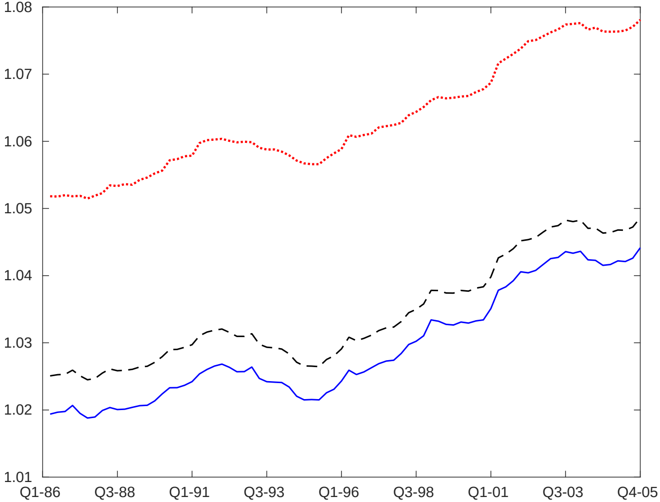


Figure 5: Point estimates of average returns to scale. Solid blue line represents our estimates, dotted red line represents the exponential stochastic frontier model, and dashed black line represents the kernel-based model.

Main Findings: Return to scale

- All the values are greater than one, indicating increasing returns to scale among large banks in the US during the sample period.
- Another evidence is that the average cost of production decreases over the sample period. This finding is consistent among the three models, but the exponential model generally produces higher values of estimates than the other two models. We tend to attribute it to the very restrictive exponential assumption.

Main Findings: Technical change

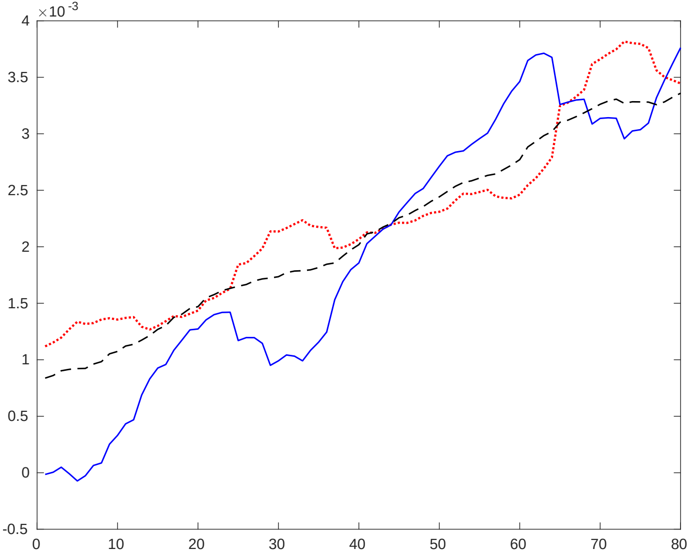


Figure 6: Point estimates of average technical change. Solid blue line represents our estimates, dotted red line represents the exponential model, and dashed black line represents the kernel-based model.

Main Findings: Technical Change

- All three models indicate an increasing trend of the technical change from 1986 Q1 to 2005 Q4.
- There are opposite moves over some short periods of time. For example, the exponential model shows an increased technical change in 1993, while our model indicates a decreased technical change.

Conclusion

- We present a nonparametric estimation of the distribution of inefficiency terms in panel stochastic frontier models
- Such estimation uses information conveyed by the composite errors rather than the inefficiencies
- It has produced obviously different distribution of the inefficinecies from the cost frontier of the US large commercial banks, in comparison to the kernel-based density estimation
- We have also obtained different measures of technical efficiency, return to scale and technical change

References

Aigner, D., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Feng, G., Wang, C., and Zhang, X. (2019). Estimation of inefficiency in stochastic frontier models: a bayesian kernel approach. *Journal of Productivity Analysis*, 51(1):1.

Greene, W. H. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46(1):141–163.

Griffin, J. E. and Steel, M. F. J. (2004). Semiparametric bayesian inference for stochastic frontier models. *Journal of Econometrics*, 123(1):121–152.

Jondrow, J., Knox Lovell, C. A., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2):233–238.

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

Park, B. U. and Simar, L. (1994). Efficient semiparametric estimation in a stochastic frontier model. *Journal of the American Statistical Association*, 89(427):929–936.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13(1):57–66.