Economics Bulletin

Volume 39, Issue 3

Endogenous Inputs and Environmental Variables in Battese and Coelli's (1995) Stochastic Frontier Model

Takahiro Tsukamoto
Graduate School of Economics, Nagoya University

Abstract

This study provides a guideline for dealing with endogenous inputs and environmental variables in Battese and Coelli's (1995) stochastic frontier model, which has been used in a vast number of empirical studies. The presence of endogenous variables in Battese and Coelli's model leads to inconsistent parameter estimates; also, estimates in the studies that use their model may be biased. Our model uses limited information maximum likelihood methods to correct these biases. The Monte Carlo simulations provided in this article show that our estimators perform well.

This work was supported by JSPS KAKENHI Grant Number 19J12648.
Citation: Takahiro Tsukamoto, (2019) "Endogenous Inputs and Environmental Variables in Battese and Coelli's (1995) Stochastic Frontier Model", *Economics Bulletin*, Volume 39, Issue 3, pages 2228-2236
Contact: Takahiro Tsukamoto - tsukamoto.takahiro@j.mbox.nagoya-u.ac.jp.
Submitted: November 16, 2018.   **Published:** September 30, 2019.

1. Introduction

Battese and Coelli's (1995) stochastic frontier model for panel data (BC95) is widely used in empirical analysis. This is because the BC95 can simultaneously estimate the frontier and the model for technical inefficiency, thereby avoiding the bias of the two-stage approach. However, the presence of endogenous variables leads to inconsistent parameter estimates. Estimates in empirical studies that use the BC95 may therefore be biased, which indicates the necessity of new approaches to verification and improvement. To fill this gap in the literature, we propose a technique to deal with both endogenous variables in the frontier model (endogenous inputs) and endogenous variables in the model for technical inefficiency (endogenous environmental variables) in the BC95. To this effect, we use limited-information maximum-likelihood methods.

The problem of endogenous variables is an important concern in the literature on stochastic frontier analysis. The pioneering work of Kutlu (2010) deals with the endogenous inputs in Battese and Coelli's (1992) stochastic frontier model. Further, Griffiths and Hajargasht (2016) consider a Bayesian estimation of stochastic frontier models with endogenous inputs and environmental variables. However, their specification is fundamentally different from our model, because their inefficiency term correlates with only firm averages of endogenous inputs. Amsler *et al.* (2017) and Karakaplan and Kutlu (2017) address endogenous inputs and environmental variables, but they use extensions of the stochastic frontier model by Reifschneider and Stevenson (1991), Caudill and Ford (1993), and Caudill *et al.* (1995), referred to as the RSCFG model. According to the RSCFG model, the distribution of the technical inefficiency term u_i satisfies the scaling property based on the half-normal specification, namely, $u_i = u_i^O \exp(\boldsymbol{h}_i'\boldsymbol{\kappa})$, where $u_i^O \sim N^+\left(0, \sigma_{u^O}^2\right)$. Conversely, the BC95 is based on a truncated normal specification in which the inefficiency term is $u_{it} \sim N^+(\boldsymbol{z}_{it}'\boldsymbol{\delta}, \sigma_u^2)$. Kutlu (2018) addresses endogeneity problems in a distribution-free context by extending Cornwell *et al.*'s (1990) regression-based estimator.

We address both endogenous inputs and endogenous environmental variables in the BC95 by adapting the approach proposed by Amsler *et al.* (2017). Our model is neither a special nor a general case of these existing stochastic frontier models that address endogenous problems. In addition, many empirical studies use the BC95, which is one of the major stochastic frontier models. Therefore, our model makes a novel contribution to the literature on stochastic frontier analysis.

2. Estimation procedure

Consider the following stochastic production frontier model with endogenous inputs and environmental variables for (unbalanced or balanced) panel data:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},$$
$$\varepsilon_{it} := v_{it} - u_{it},$$
$$i = 1,2,\ldots,N, t = 1,2,\ldots,T_i,$$
$$\boldsymbol{x}_{it} := \begin{bmatrix} 1 \\ \boldsymbol{x}_{1it} \\ \boldsymbol{x}_{2it} \end{bmatrix}, \boldsymbol{z}_{it} := \begin{bmatrix} 1 \\ \boldsymbol{z}_{1it} \\ \boldsymbol{z}_{2it} \end{bmatrix}, \boldsymbol{q}_{it} := \begin{bmatrix} 1 \\ \boldsymbol{x}_{1it} \\ \boldsymbol{z}_{1it} \\ \boldsymbol{w}_{it} \end{bmatrix},$$

where $\boldsymbol{x}_{1it}$, $\boldsymbol{z}_{1it}$, and $\boldsymbol{w}_{it}$ are exogenous, and $\boldsymbol{x}_{2it}$ and $\boldsymbol{z}_{2it}$ are endogenous. We can easily represent the stochastic cost frontier model by replacing $-u_{it}$ with $+u_{it}$. To address the endogeneity problem, we assume reduced-form equations for the endogenous variables:

$$\boldsymbol{p}_{it} = \boldsymbol{\Pi}'\boldsymbol{q}_{it} + \boldsymbol{\xi}_{it},$$

where $\boldsymbol{p}_{it} \coloneqq \begin{bmatrix} \boldsymbol{x}_{2it} \\ \boldsymbol{z}_{2it} \end{bmatrix}, \boldsymbol{\Pi}' \coloneqq \begin{bmatrix} \boldsymbol{\Pi}'_x \\ \boldsymbol{\Pi}'_q \end{bmatrix}, \boldsymbol{\xi}_{it} \coloneqq \begin{bmatrix} \boldsymbol{\eta}_{it} \\ \boldsymbol{\tau}_{it} \end{bmatrix}$.

We define $\boldsymbol{\psi}_{it} \coloneqq \begin{bmatrix} v_{it} \\ \boldsymbol{\xi}_{it} \end{bmatrix}, \boldsymbol{\Omega} \coloneqq \begin{bmatrix} \sigma_v^2 & \boldsymbol{\Sigma}_{v\xi} \\ \boldsymbol{\Sigma}_{\xi v} & \boldsymbol{\Sigma}_{\xi\xi} \end{bmatrix}$, and then make the following assumption:

$$\boldsymbol{\psi}_{it}|\boldsymbol{q}_i \sim N(\boldsymbol{0}, \boldsymbol{\Omega}),$$
$$u_{it}|v_{it}, \boldsymbol{x}_{it}, \boldsymbol{z}_{it}, \boldsymbol{w}_{it} = u_{it}|\boldsymbol{x}_{it}, \boldsymbol{z}_{it}, \boldsymbol{w}_{it} \sim \mathrm{N}^+(\mu_{it}, \sigma_u^2),$$
$$\mu_{it} = \boldsymbol{z}'_{it}\boldsymbol{\delta},$$

where $\mathrm{N}^+(\mu_{it}, \sigma_u^2)$ refers to a non-negative normal distribution truncated at zero for which the (pre-truncated) mean and variance are μ_{it} and σ_u^2, respectively. This model allows $\boldsymbol{x}_{2it}$ and $\boldsymbol{z}_{2it}$ to be correlated with v_{it}.

The joint density of $(u_{it}, v_{it}, \boldsymbol{\xi}_{it}|\boldsymbol{q}_{it})$ is as follows:

$$f_{u,v,\xi}(u_{it}, v_{it}, \boldsymbol{\xi}_{it}|\boldsymbol{q}_{it}) = f_u(u_{it}|\boldsymbol{\xi}_{it}, \boldsymbol{q}_{it}) \cdot f_v(v_{it}|\boldsymbol{\xi}_{it}, \boldsymbol{q}_{it}) \cdot f_\xi(\boldsymbol{\xi}_{it}|\boldsymbol{q}_{it}),$$
$$f_u(u_{it}|\boldsymbol{\xi}_{it}, \boldsymbol{q}_{it}) \cdot f_v(v_{it}|\boldsymbol{\xi}_{it}, \boldsymbol{q}_{it}) = \frac{1}{2\pi\sigma_u\sigma_c\Phi\left(\frac{\mu_{it}}{\sigma_u}\right)} \exp\left[-\frac{(v_{it} - m_{it})^2}{2\sigma_c^2} - \frac{(u_{it} - \mu_{it})^2}{2\sigma_u^2}\right],$$
$$f_\xi(\boldsymbol{\xi}_{it}|\boldsymbol{q}_{it}) = (2\pi)^{-\frac{k_p}{2}}\left|\boldsymbol{\Sigma}_{\xi\xi}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\xi}'_{it}\boldsymbol{\Sigma}_{\xi\xi}^{-1}\boldsymbol{\xi}_{it}\right),$$

where k_p is a dimension of $\boldsymbol{p}_{it}$, $m_{it} \coloneqq \boldsymbol{\Sigma}_{v\xi}\boldsymbol{\Sigma}_{\xi\xi}^{-1}\boldsymbol{\xi}_{it}$, $\sigma_c^2 \coloneqq \sigma_v^2 - \boldsymbol{\Sigma}_{v\xi}\boldsymbol{\Sigma}_{\xi\xi}^{-1}\boldsymbol{\Sigma}_{\xi v}$.

Integrating out u_{it}, we obtain the joint density of $(\varepsilon_{it}|\boldsymbol{\xi}_{it}, \boldsymbol{q}_{it})$:

$$\begin{aligned}
f_\varepsilon(\varepsilon_{it}|\boldsymbol{\xi}_{it}, \boldsymbol{q}_{it}) &= \int_0^\infty f_u(u_{it}|\boldsymbol{\xi}_{it}, \boldsymbol{q}_{it}) \cdot f_v(\varepsilon_{it} + u_{it}|\boldsymbol{\xi}_{it}, \boldsymbol{q}_{it}) du_{it} \\
&= \frac{1}{\sigma}\phi\left(\frac{\mu_{it} + \varepsilon_{it} - m_{it}}{\sigma}\right) \\
&\quad \cdot \left[\int_0^\infty \frac{1}{\sqrt{2\pi}(\sigma_c\sigma_u\sigma^{-1})} \exp\left\{-\frac{1}{2(\sigma_c\sigma_u\sigma^{-1})^2}\left(u_{it} - \left(\frac{\mu_{it}\sigma_c^2 - (\varepsilon_{it} - m_{it})\sigma_c^2}{\sigma^2}\right)\right)^2\right\} du_{it}\right] \\
&\quad \cdot \left(\Phi\left(\frac{\mu_{it}}{\sigma_u}\right)\right)^{-1} \\
&= \frac{1}{\sigma}\phi\left(\frac{\mu_{it} + \varepsilon_{it} - m_{it}}{\sigma}\right) \cdot \left[\int_{-\frac{\mu_{it}\sigma_c^2 - (\varepsilon_{it} - m_{it})\sigma_u^2}{\sigma_c\sigma_u\sigma}}^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{{r_{it}}^2}{2}\right\} dr_{it}\right] \cdot \left(\Phi\left(\frac{\mu_{it}}{\sigma_u}\right)\right)^{-1} \\
&= \frac{1}{\sigma}\phi\left(\frac{\mu_{it} + \varepsilon_{it} - m_{it}}{\sigma}\right) \cdot \Phi\left(\frac{\mu_{it}\sigma_c^2 - (\varepsilon_{it} - m_{it})\sigma_u^2}{\sigma_c\sigma_u\sigma}\right) \cdot \left(\Phi\left(\frac{\mu_{it}}{\sigma_u}\right)\right)^{-1},
\end{aligned}$$

where $\sigma^2 \coloneqq \sigma_c^2 + \sigma_u^2 = \sigma_v^2 + \sigma_u^2 - \boldsymbol{\Sigma}_{v\xi}\boldsymbol{\Sigma}_{\xi\xi}^{-1}\boldsymbol{\Sigma}_{\xi v}$, $r_{it} \coloneqq \left\{u_{it} - \frac{\mu_{it}\sigma_v^2 - (\varepsilon_{it} - m_{it})\sigma_u^2}{\sigma^2}\right\}/(\sigma_c\sigma_u\sigma^{-1})$, while ϕ and Φ denote the standard normal probability density function and cumulative distribution function, respectively. Finally, the log-likelihood function for the sample observations, $(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$, is

$$
\begin{aligned}
&LL\big(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_v, \sigma_u, \boldsymbol{\Sigma}_{v\xi}, \boldsymbol{\Sigma}_{\xi\xi}, \boldsymbol{\Pi}; \boldsymbol{y}\big) \\
&\quad = -\frac{1}{2}\left(\sum_{i=1}^{N} T_i\right)\left[\ln \sigma^2 + \left(k_p + 1\right)\ln 2\pi + \ln\left|\boldsymbol{\Sigma}_{\xi\xi}\right|\right] \\
&\quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(\frac{\boldsymbol{z}_{it}'\boldsymbol{\delta} + y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{\Sigma}_{v\xi}\boldsymbol{\Sigma}_{\xi\xi}^{-1}(\boldsymbol{p}_{it} - \boldsymbol{\Pi}'\boldsymbol{z}_{it})}{\sigma}\right)^2 \\
&\quad -\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left[\ln \Phi\left(\frac{\boldsymbol{z}_{it}'\boldsymbol{\delta}}{\sigma_u}\right) - \ln \Phi\left(\frac{\boldsymbol{z}_{it}'\boldsymbol{\delta}\sigma_c^2 - \left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{\Sigma}_{v\xi}\boldsymbol{\Sigma}_{\xi\xi}^{-1}(\boldsymbol{p}_{it} - \boldsymbol{\Pi}'\boldsymbol{q}_{it})\right)\sigma_u^2}{\sigma_c\sigma_u\sigma}\right)\right] \\
&\quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left((\boldsymbol{p}_{it} - \boldsymbol{\Pi}'\boldsymbol{q}_{it})'\boldsymbol{\Sigma}_{\xi\xi}^{-1}(\boldsymbol{p}_{it} - \boldsymbol{\Pi}'\boldsymbol{q}_{it})\right).
\end{aligned}
$$

We obtain limited-information maximum-likelihood estimators by maximizing the log-likelihood function with respect to the parameters $\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_v, \sigma_u, \boldsymbol{\Sigma}_{v\xi}, \boldsymbol{\Sigma}_{\xi\xi},$ and $\boldsymbol{\Pi}$. Limited-information maximum-likelihood estimators that are not in frontier models are consistent and asymptotically normal even if the error deviates from the assumption of normal distribution. However, the consistency of our estimators requires the correctness of the distribution assumptions. This property applies to not only our model but also all stochastic frontier models estimated by maximum-likelihood methods.

As Amsler *et al.* (2016) proposed, we present a point estimator for technical efficiency u_{it} using $\boldsymbol{\xi}_{it}$ and ε_{it}. The Jondrow *et al.* (1982) type point estimator in our model is

$$
TE_i^{JLMS} = \exp\big(E(u_i|\varepsilon_{it}, \boldsymbol{\xi}_{it})\big) = \mu_{it}^* + \sigma^* \frac{\phi\left(\frac{\mu_{it}^*}{\sigma^*}\right)}{\Phi\left(\frac{\mu_{it}^*}{\sigma^*}\right)},
$$

whereas the Battese and Coelli (1988) type point estimator is

$$
TE_{it}^{BC} = \mathrm{E}(\exp(-u_{it})\,|\varepsilon_i, \boldsymbol{\xi}_i) = \exp\left[-\mu_{it}^* + \frac{1}{2}{\sigma^*}^2\right] \cdot \left\{\frac{\Phi\left(\frac{\mu_{it}^*}{\sigma^*} - \sigma^*\right)}{\Phi\left(\frac{\mu_{it}^*}{\sigma^*}\right)}\right\},
$$

where

$$
\mu_{it}^* := \frac{\boldsymbol{z}_{it}'\boldsymbol{\delta}\sigma_c^2 - \left(\varepsilon_{it} - \boldsymbol{\Sigma}_{v\xi}\boldsymbol{\Sigma}_{\xi\xi}^{-1}\boldsymbol{\xi}_{it}\right)\sigma_u^2}{\sigma^2}, \qquad \sigma^* := \frac{\sigma_u\sigma_c}{\sigma}.
$$

The estimator TE_{it}^{BC} is preferred, particularly when u_{it} is not close to zero (Kumbhakar and Lovell 2003, 78).

3. Monte Carlo simulations

In this section, we report the results of Monte Carlo simulations to examine the performance of our estimator. We generate data for the simulation using the following process:

$$
\begin{gathered}
y_{it} = \beta_0 + x_{1it}\beta_1 + x_{2it}\beta_2 + v_{it} - u_{it}, \\
\mu_{it} = \delta_0 + z_{1it}\delta_1 + z_{2it}\delta_2, \\
x_{2it} = \Pi_{x0} + x_{1it}\Pi_{x1} + z_{1it}\Pi_{x2} + w_{1it}\Pi_{x3} + w_{2it}\Pi_{x4} + \eta_{it}, \\
z_{2it} = \Pi_{z0} + x_{1it}\Pi_{z1} + z_{1it}\Pi_{z2} + w_{1it}\Pi_{z3} + w_{2it}\Pi_{z4} + \tau_{it}, \\
[x_{1it} \quad z_{1it} \quad w_{1it} \quad w_{2it}]' \sim N(\mathbf{0}, \boldsymbol{I}_4), \\
[v_{it} \quad \eta_{it} \quad \tau_{it}]' \sim N(\mathbf{0}, \boldsymbol{\Omega}), \\
u_{it} \sim \mathrm{N}^{+}(\mu_{it}, \sigma_u^2),
\end{gathered}
$$

where $\beta_0 = \beta_1 = \beta_2 = \delta_0 = \delta_1 = \delta_2 = \sigma_u^2 = 1$, $\Pi_{x0} = \Pi_{x1} = \Pi_{x2} = \Pi_{x3} = \Pi_{x4} = \Pi_{z0} = \Pi_{z1} = \Pi_{z2} = \Pi_{z3} = \Pi_{z4} = 0.1$, $\boldsymbol{\Omega} := \begin{bmatrix} \sigma_v^2 & \Sigma_{v\eta} & \Sigma_{\tau\tau} \\ \Sigma_{\eta v} & \Sigma_{\eta\eta} & \Sigma_{\eta\tau} \\ \Sigma_{\tau v} & \Sigma_{\tau\eta} & \Sigma_{\tau\tau} \end{bmatrix} = \begin{bmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}$.
x_{1it}, z_{1it}, w_{1it}, and w_{2it} are exogenous variables, whereas x_{2it} and z_{2it} are endogenous variables.

Table 1 presents the results of Monte Carlo simulations for the sample size $n = \sum_{i=1}^{N} T_i = 1{,}000$. The Monte Carlo experiments are conducted with 1,000 replications. The results show that estimators of the BC95 exhibit severe bias, whereas our model obtains good estimates in not only the frontier model and the model for technical efficiency, but also the reduced-form equations and variance–covariance matrix. Our findings suggest that the BC95 estimates may exhibit bias because of endogeneity problems, but our model can deal with both endogenous inputs and endogenous environmental variables.

Figure 1 shows the results of Monte Carlo simulations by sample size. The vertical axis represents the estimated value, and the horizontal axis means the logarithmic value of the sample size. The blue lines are estimates of our model, the black lines are estimates of the BC95, and the red lines are the true values. Twelve data sets have been created.

The estimates of our proposed model converge to correct values as the sample size grows. On the other hand, estimates of the BC95 converge to biased values. In particular, the coefficients of constant terms and endogenous variables have large biases. These results indicate that the estimates of the proposed model are consistent.

Table 2 shows the bias and RMSE in the proposed model by sample size. The estimates of our proposed model do not significantly deviate from the true values even when the sample size is about 100. When the sample size is halved, the RMSE is 1.447 times on the average, which is almost the same as the appropriate rate of $\sqrt{2}$. We can see that our model is useful even with small sample sizes.

Table 1. Results of the Monte Carlo simulations in the case of $n = 1{,}000$.

	True	BC95 Estimates (mean)	BC95 Bias	BC95 RMSE	Our model Estimates (mean)	Our model Bias	Our model RMSE
β_0	1	0.778	-0.222	0.2789	1.023	0.023	0.1804
β_1	1	0.960	-0.040	0.0510	1.000	0.000	0.0410
β_2	1	1.174	0.174	0.1796	1.007	0.007	0.0650
δ_0	1	0.878	-0.122	0.2829	1.020	0.020	0.2327
δ_1	1	1.032	0.032	0.0797	1.001	0.001	0.0538
δ_2	1	0.900	-0.100	0.1092	1.006	0.006	0.0602
σ_u^2	1	0.915	-0.085	0.1876	1.002	0.002	0.0508
σ_v^2	1	0.806	-0.193	0.2381	0.989	-0.011	0.1054
$\Sigma_{\eta v}$	0.8				0.794	-0.006	0.0604
$\Sigma_{\tau v}$	0.8				0.796	-0.004	0.0564
$\Sigma_{\eta\eta}$	1				0.994	-0.006	0.0462
$\Sigma_{\tau\eta}$	0.8				0.795	-0.005	0.0410
$\Sigma_{\tau\tau}$	1				0.993	-0.007	0.0442
Π_{x0}	0.1				0.101	0.001	0.0558
Π_{x1}	0.1				0.100	0.000	0.0318
Π_{x2}	0.1				0.099	-0.001	0.0304
Π_{x3}	0.1				0.100	0.000	0.0029
Π_{x4}	0.1				0.100	0.000	0.0029
Π_{z0}	0.1				0.101	0.001	0.0549
Π_{z1}	0.1				0.100	0.000	0.0318
Π_{z2}	0.1				0.098	-0.002	0.0315
Π_{z3}	0.1				0.100	0.000	0.0029
Π_{z4}	0.1				0.100	0.000	0.0029

Note: BC95: Battese and Coelli's (1995) stochastic frontier model for panel data; RMSE: Root mean squared error.

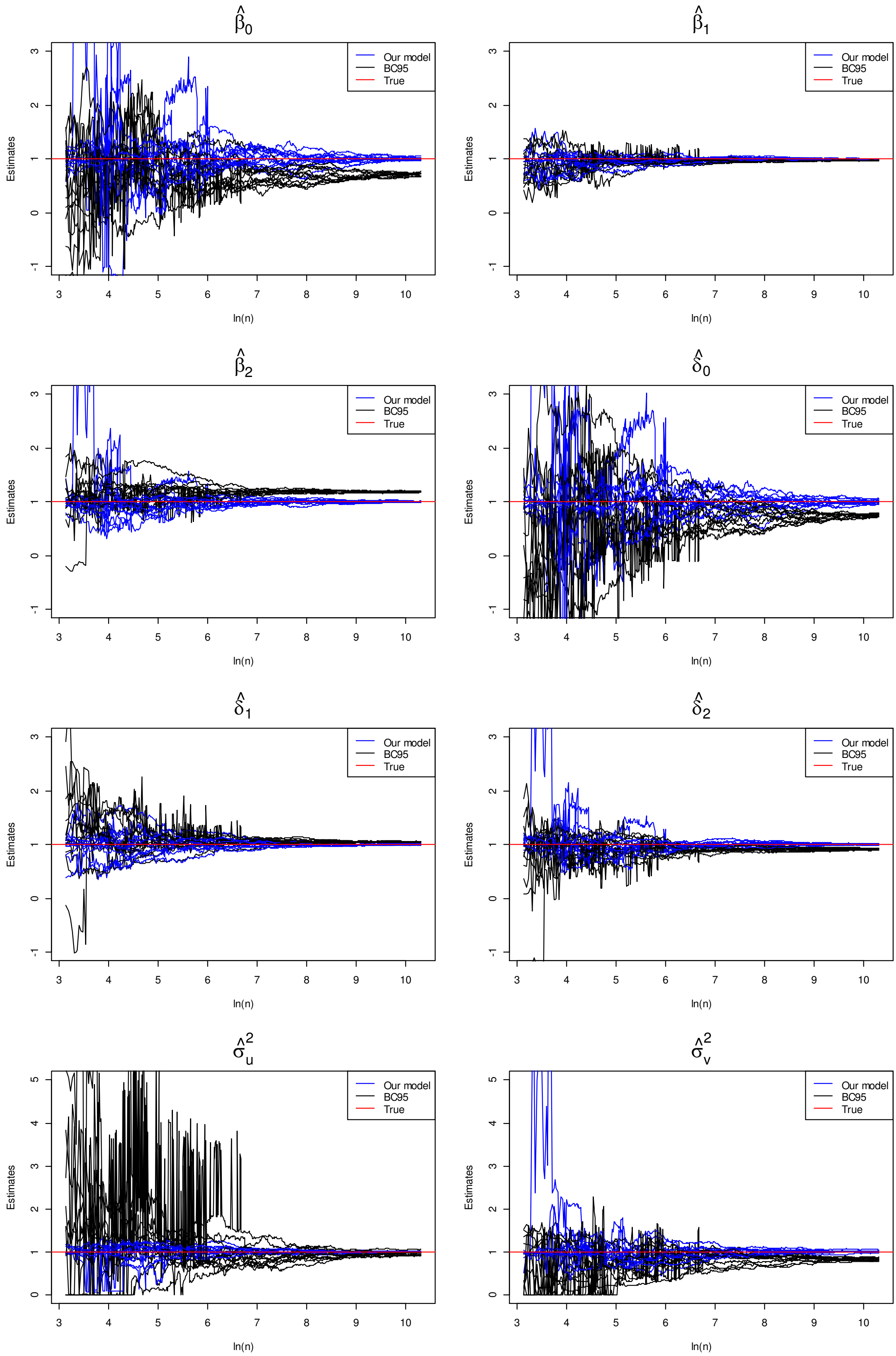

Figure 1. Results of Monte Carlo simulations by sample size.

Table 2. Bias and RMSE in the proposed model by sample size

n	100	200	400	800	1600
Bias					
β_0	0.1315	0.0934	0.0454	0.0091	0.0105
β_1	0.0043	-0.0041	-0.0026	-0.0004	-0.0003
β_2	0.0404	0.0158	0.0102	0.0007	0.0016
δ_0	0.0822	0.0733	0.0303	0.0019	0.0074
δ_1	0.0180	0.0077	0.0045	0.0022	0.0012
δ_2	0.0611	0.0212	0.0135	0.0037	0.0027
σ_u^2	-0.0004	0.0107	0.0002	-0.0033	0.0004
σ_v^2	-0.0409	-0.0445	-0.0158	-0.0056	-0.0063
$\Sigma_{\eta v}$	-0.0124	-0.0114	-0.0073	-0.0024	-0.0010
$\Sigma_{\tau v}$	0.0012	-0.0057	-0.0025	-0.0028	-0.0001
$\Sigma_{\eta\eta}$	-0.0361	-0.0186	-0.0138	-0.0051	-0.0017
$\Sigma_{\tau\eta}$	-0.0286	-0.0162	-0.0100	-0.0056	-0.0006
$\Sigma_{\tau\tau}$	-0.0396	-0.0225	-0.0103	-0.0070	-0.0008
Π_{x0}	0.0043	0.0065	0.0039	-0.0024	0.0026
Π_{x1}	-0.0048	-0.0036	-0.0020	0.0015	-0.0006
Π_{x2}	-0.0058	-0.0034	-0.0001	0.0012	-0.0012
Π_{x3}	0.0003	-0.0001	-0.0001	-0.0003	-0.0001
Π_{x4}	-0.0002	-0.0001	0.0000	0.0001	0.0000
Π_{z0}	0.0070	0.0035	0.0077	-0.0022	0.0027
Π_{z1}	-0.0074	-0.0009	-0.0018	0.0016	-0.0007
Π_{z2}	-0.0038	-0.0045	-0.0023	0.0004	-0.0014
Π_{z3}	0.0001	-0.0002	-0.0002	-0.0003	-0.0001
Π_{z4}	-0.0004	-0.0002	0.0001	0.0000	-0.0001
RMSE (Root mean squared error)					
β_0	0.6956	0.4743	0.3198	0.2003	0.1314
β_1	0.1447	0.0954	0.0662	0.0452	0.0336
β_2	0.2696	0.1674	0.1075	0.0706	0.0473
δ_0	0.8450	0.5866	0.4051	0.2549	0.1683
δ_1	0.1802	0.1289	0.0899	0.0612	0.0431
δ_2	0.2754	0.1653	0.1064	0.0693	0.0467
σ_u^2	0.2085	0.1394	0.0913	0.0588	0.0418
σ_v^2	0.3428	0.2514	0.1746	0.1155	0.0805
$\Sigma_{\eta v}$	0.1903	0.1338	0.0926	0.0616	0.0455
$\Sigma_{\tau v}$	0.1776	0.1250	0.0908	0.0590	0.0421
$\Sigma_{\eta\eta}$	0.1395	0.1023	0.0706	0.0482	0.0360
$\Sigma_{\tau\eta}$	0.1254	0.0905	0.0635	0.0436	0.0321
$\Sigma_{\tau\tau}$	0.1400	0.1003	0.0714	0.0495	0.0348
Π_{x0}	0.1717	0.1224	0.0877	0.0626	0.0436
Π_{x1}	0.1066	0.0719	0.0500	0.0363	0.0248
Π_{x2}	0.1027	0.0699	0.0504	0.0336	0.0242
Π_{x3}	0.0093	0.0066	0.0046	0.0031	0.0023
Π_{x4}	0.0096	0.0066	0.0046	0.0031	0.0022
Π_{z0}	0.1704	0.1221	0.0860	0.0615	0.0440
Π_{z1}	0.1072	0.0710	0.0494	0.0355	0.0248
Π_{z2}	0.0991	0.0718	0.0501	0.0341	0.0248
Π_{z3}	0.0094	0.0064	0.0046	0.0032	0.0023
Π_{z4}	0.0096	0.0064	0.0045	0.0032	0.0023

4. Concluding remarks

In this study, we attempted to treat endogenous inputs and environmental variables in the BC95, which has been used in a vast number of empirical studies. We developed good limited-information maximum-likelihood estimators, thereby fixing bias in the BC95. Our model is useful for verification and improvement of empirical studies that use the BC95.

As is true of all stochastic frontier models with maximum-likelihood methods, the desired properties of the estimators depend on the correctness of the distribution assumptions of the error terms. This is a drawback of the maximum-likelihood method and also indicates the necessity of the proposed model. If the inefficiency can reasonably be expected to follow the half-normal distribution, $u_i = u_i^O \exp(\boldsymbol{h}_i'\boldsymbol{\kappa})$, where $u_i^O \sim N^+\left(0, \sigma_{u^O}^2\right)$, one can use the model proposed by Amsler *et al.* (2017). However, if one can reasonably expect the inefficiency to follow the truncated normal distribution, $u_{it} \sim N^+(\boldsymbol{z}_{it}'\boldsymbol{\delta}, \sigma_u^2)$, the proposed model should preferably be used. In some cases, it is not known before the analysis which distribution assumption is appropriate. Thus, it may be necessary to conduct a comparison after the analysis using multiple models. Stochastic frontier models assuming a truncated normal distribution, including the BC95, are used for many empirical analyses. Our model should be one of the models to be estimated.

Finally, our study is limited in that we applied our model to simulated data and not real-world data. Application of our model to real-world data is thus a topic for future research.

References

Amsler, C., A. Prokhorov, and P. Schmidt (2016) "Endogeneity in Stochastic Frontier Models" *Journal of Econometrics* **190**(2), 280-288. doi:10.1016/j.jeconom.2015.06.013

Amsler, C., A. Prokhorov, and P. Schmidt (2017) "Endogenous Environmental Variables in Stochastic Frontier Models" *Journal of Econometrics* **199**(2), 131-140. doi:10.1016/j.jeconom.2017.05.005

Battese, G. E. and T. J. Coelli (1988) "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data" *Journal of Econometrics* **38**(3), 387-399. doi:10.1016/0304-4076(88)90053-X

Battese, G. E. and T. J. Coelli (1992) "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India" *Journal of Productivity Analysis* **3**(1–2), 153-169. doi:10.1007/BF00158774

Battese, G. E. and T. J. Coelli (1995) "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data" *Empirical Economics* **20**(2), 325-332. doi:10.1007/BF01205442

Caudill, S. B. and J. M. Ford (1993) "Biases in Frontier Estimation Due to Heteroscedasticity" *Economics Letters* **41**(1), 17-20. doi:10.1016/0165-1765(93)90104-K

Caudill, S. B., J. M. Ford, and D. M. Gropper (1995) "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity" *Journal of Business & Economic Statistics* **13**(1), 105-111. doi:10.2307/1392525

Cornwell, C., P. Schmidt, and R. C. Sickles (1990) "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels" *Journal of Econometrics* **46** (1–2), 185-200. doi:10.1016/0304-4076(90)90054-W

Griffiths, W. E. and G. Hajargasht (2016) "Some Models for Stochastic Frontiers with Endogeneity" *Journal of Econometrics* **190**(2), 341-348. doi:10.1016/j.jeconom.2015.06.012

Jondrow, J., C. K. Lovell, I. S. Materov, and P. Schmidt (1982) "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model" *Journal of Econometrics* **19**(2–3), 233-238. doi:10.1016/0304-4076(82)90004-5

Karakaplan, M. U. and L. Kutlu (2017) "Handling Endogeneity in Stochastic Frontier Analysis" *Economics Bulletin* **37**(2), 889-901. http://www.accessecon.com/Pubs/EB/2017/Volume37/EB-17-V37-I2-P79.pdf

Kumbhakar, S. C. and C. K. Lovell (2003) *Stochastic Frontier Analysis*, Cambridge University Press: Cambridge.

Kutlu, L. (2010) "Battese-Coelli Estimator with Endogenous Regressors" *Economics Letters* **109**(2), 79-81. doi:10.1016/j.econlet.2010.08.008

Kutlu, L. (2018) "A Distribution-Free Stochastic Frontier Model with Endogenous Regressors" *Economics Letters* **163**, 152-154. doi:10.1016/j.econlet.2017.12.026

Reifschneider, D. and R. Stevenson (1991) "Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency" *International Economic Review* **32**(3), 715-723. doi:10.2307/2527115