

# Estimation of stochastic frontier models based on multimodel inference

Cliff J. Huang · Hung-pin Lai

Published online: 13 December 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** In most empirical studies, once the best model has been selected according to a certain criterion, subsequent analysis is conducted conditionally on the chosen model. In other words, the uncertainty of model selection is ignored once the best model has been chosen. However, the true data-generating process is in general unknown and may not be consistent with the chosen model. In the analysis of productivity and technical efficiencies in the stochastic frontier settings, if the estimated parameters or the predicted efficiencies differ across competing models, then it is risky to base the prediction on the selected model. Buckland et al. (Biometrics 53:603–618, 1997) have shown that if model selection uncertainty is ignored, the precision of the estimate is likely to be overestimated, the estimated confidence intervals of the parameters are often below the nominal level, and consequently, the prediction may be less accurate than expected. In this paper, we suggest using the model-averaged estimator based on the multimodel inference to estimate stochastic frontier models. The potential advantages of the proposed approach are twofold: incorporating the model selection uncertainty into statistical inference; reducing the model selection bias and variance of the frontier and technical efficiency estimators. The approach is demonstrated empirically via the estimation of an Indian farm data set.

**Keywords** Model selection · Model-averaged estimator · Stochastic frontier · Technical efficiency

**JEL Classification** C52 · D24

## 1 Introduction

Model selection is commonly used in empirical studies and serves as a tool to select the best model among the competing ones. Once a model has been chosen, subsequent analyses are conditional on the selected model without considering the uncertainty regarding model selection. However, model selection can be highly variable and the selected model is not necessarily true by definition.<sup>1</sup> Even if the data-generating-process (DGP) is fixed, different model selection criteria may result in different choices of models. It is often the case that no single model is clearly superior to the other models. Therefore, it may be risky to base the empirical analysis or prediction on a single model. In this paper, we propose using model averaging to analyze stochastic frontier (SF) models as an alternative to a single model inference.

Since the pioneered works of Aigner et al. (1977), SF analysis has been widely used in productivity and efficiency studies to describe and estimate models of the production frontier. However, there are only limited systematic treatments of tests or model selection criteria in the existing SF literatures. Most empirical applications of SF analysis apply some off-the-shelf model selection scheme of “goodness-of-fit” checking of residuals, or “significant test” of coefficients to arrive at some “best model” that is thought to adequately approximate the true model.

---

An earlier version of this paper was presented at the Asia-Pacific Productivity Conference 2010 in Taipei, Taiwan.

---

C. J. Huang  
Department of Economics, Vanderbilt University, Nashville,  
TN, USA

H. Lai (✉)  
Department of Economics, National Chung Cheng University,  
Chia-Yi, Taiwan  
e-mail: ecdhpl@ccu.edu.tw

---

<sup>1</sup> Box (1976) is quick to point out that “All models are wrong, but some are useful.”

In studying the household maize production in Kenya, Liu and Myers (2009) propose an  $R^2$  type measure to assess a model's overall explanatory power of SF regression setting. In model selection, Schmidt and Lin (1984) and Alvarez et al. (2006) propose the use of the likelihood ratio (LR) test, the Lagrange multiplier (LM) test, and the Wald test on zero coefficient restrictions for SF model selection. Lai and Huang (2010), on the other hand, apply the Akaike (1973, 1974) and Takeuchi (1976) information criteria to model selection. However, regardless what method of model selection is used, the model selection uncertainty is always present, as the same data are used for both model selection and the associated parameter estimation and inference. If model selection uncertainty is ignored, the subsequent productivity and efficiency analyses of the selected "best model" would be invalid. In the estimation of the labor input coefficient  $\hat{\beta}$ , for example, the sampling variance of the estimator, given the selected best model, should have two components, the conditional variance given the selection mode,  $\text{var}(\hat{\beta}|\text{selected model})$ , and the variance of the selected model itself as it is an estimator of the true. Pötscher (1991) has shown that the distribution of estimators and test statistics are dramatically affected by the act of model selection. Lee and Pötscher (2005) also pointed out that ignoring the variance of the model selection component may lead to invalid inference. On most empirical studies, the reported confidence intervals tend to be too short, and a hypothesis rejected at an announced 5% significance level might actually have been tested at a rather higher level. It is, therefore, sometimes advantageous, as proposed in this paper, for the SF analysis to average estimators across several competing models, rather than relying on only the single "best model".

A model-averaged estimator is a weighted average of estimators obtained from competing models. Model averaging has several practical and theoretical advantages: it incorporates model selection uncertainty into statistical inference; it reduces the model selection bias effects in parameter estimation; it lessens the estimated variance while controlling the omitted variable bias (see Hansen 2005). The subsequent statistical inferences on the model-averaged estimator are conducted based on the entire set of competing models.

In this paper, we propose using the model-averaged estimator to analyze the SF models. The basic framework of SF model selection and the model-averaged estimator is described in Sect. 2. Three model selection criteria, the Akaike information criterion (AIC, Akaike 1973), Takeuchi's information criterion (TIC, Takeuchi 1976), and the Bayesian information criterion (BIC, Schwarz 1978), are used in constructing the weights of the competing models. A comparison of the model-averaged estimators based on the AIC, BIC, and TIC weights are examined by

the bootstrap method. The implementation of the model-averaged estimator, the frontier regression and the technical inefficiencies, are discussed in Sect. 3. A demonstration and application to the Indian farm data set used by Battese and Coelli (1992, 1995), Wang and Schmidt (2002), Alvarez et al. (2006), and Lai and Huang (2010) are given in Sect. 4. A conclusion is given in Sect. 5.

## 2 Model-averaged estimator

Let  $\{y_i, x_i, z_i\}_{i=1}^n$  be an independent data set on  $n$  firms. A general form of the SF model can be written as

$$y_i = g(x_i; \beta) + v_i - u_i, \quad (1)$$

where  $y_i$  denotes the logarithm of output, and the production frontier  $g(x_i; \beta)$  is a function of logged input vector  $x_i \in R^P$  with the parameter vector  $\beta$ . Following standard SF modeling, the random error  $v_i$  represents statistical noise and is assumed to be independent to the non-negative random error  $u_i$  representing the production inefficiency. The other variables  $z_i$  denote the exogenous factors that may affect  $u_i$ . In the case of Cobb-Douglas production frontier, the function  $g(x_i; \beta)$  is specified as linear in  $x_i$ ; in the case of translog production frontier,  $g(x_i; \beta)$  is both linear and quadratic in  $x_i$ . Following Battese and Coelli (1988), the technical efficiency (TE) of the  $i$ th firm can be defined as the conditional expectation,

$$\text{TE}_i = E(e^{-u_i} | \varepsilon_i), \quad (2)$$

where  $\varepsilon_i = v_i - u_i$  is the composite error.

Let  $f(y_i|x_i, z_i; \theta)$  be the conditional probability density function (PDF) of  $y_i|x_i, z_i$  where  $\theta = (\beta^T, \theta_v^T, \theta_u^T)^T$  denotes the vector of parameters in model (1),  $\theta_v$  and  $\theta_u$  represent the vectors of parameters in the conditional PDFs of  $v$  and  $u$ , respectively. Therefore, the maximum likelihood (ML) estimator of  $\theta$  can be defined as

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta \in \Theta} \ln L_n(\theta) \\ &= \arg \max_{\theta \in \Theta} \sum_{i=1}^n \ln f(y_i|x_i, z_i; \theta). \end{aligned} \quad (3)$$

Once the ML estimator  $\hat{\theta} = (\hat{\beta}^T, \hat{\theta}_v^T, \hat{\theta}_u^T)^T$  is obtained, the production frontier  $g(x_i; \beta)$  of the  $i$ th firm is predicted by

$$\hat{g}(x_i; \beta) = g(x_i; \hat{\beta}), \quad (4)$$

and the TE is estimated by

$$\widehat{\text{TE}}_i = E\left(e^{-u_i} | \hat{\varepsilon}_i; \hat{\theta}\right), \quad (5)$$

where  $\hat{\varepsilon}_i = y_i - x_i^T \hat{\beta}$  is the predicted value of the composite error  $\varepsilon_i$ .

Now, suppose there are  $J$  competing SF models under consideration,

$$y_i = g^j(x_i; \beta^j) + v_i^j - u_i^j, \quad j = 1, 2, \dots, J. \tag{6}$$

For example,  $g^1(x_{1i}, x_{2i}; \beta^1)$  is a Cobb-Douglas function with two inputs, labor and capital inputs, while  $g^2(x_{1i}, x_{2i}; \beta^2)$  is a translog function, and  $g^3(x_{1i}, x_{2i}; \beta^3)$  is a Fourier flexible function. The distributions of the noise component  $v_i^j$  and the inefficiency component  $u_i^j$  may differ from model to model with the associated parameter vectors denoted as  $\theta_v^j$  and  $\theta_u^j$  respectively.<sup>2</sup> The parameter vector of the Cobb-Douglas specification  $\beta^1 = (\beta_0^1, \beta_1^1, \beta_2^1)$  consists of an intercept and the two coefficients associated with  $x_{1i}$  and  $x_{2i}$ . For the translog specification, the parameter vector  $\beta^2 = (\beta_0^2, \beta_1^2, \beta_2^2, \beta_{11}^2, \beta_{22}^2, \beta_{12}^2)$  consists of three additional coefficients associated with the squared and cross-product of  $x_{1i}$  and  $x_{2i}$ ; and for the Fourier flexible specification,  $\beta^3 = (\beta_0^3, \beta_1^3, \beta_2^3, \beta_{11}^3, \beta_{22}^3, \beta_{12}^3, \text{others})$  includes additional coefficients associated with the Fourier series in  $x_{1i}$  and  $x_{2i}$ .<sup>3</sup> In the following analysis, we use  $M_j$  to denote the  $j$ th competing model and  $\hat{\theta}^j = (\hat{\beta}^{jT}, \hat{\theta}_v^{jT}, \hat{\theta}_u^{jT})^T$  to denote the ML estimator of  $\theta^j$  defined in (3) under model  $M_j$ . Given the predicted composite error  $\hat{\varepsilon}_i^j = (y_i - x_i^T \hat{\beta}^j)$ , the estimated TE of the  $i$ th firm under  $M_j$  is  $\widehat{TE}_i^j = E(e^{-u_i} | \hat{\varepsilon}_i^j, \hat{\theta}^j)$ , and the predicted frontier is  $\hat{g}(x_i; \beta^j) = g(x_i; \hat{\beta}^j)$ .

Suppose that we are interested in a certain parameter of the SF model in (1), for example, the coefficient  $\beta_1$  of  $x_{1i}$ , the labor input. In a typical model selection procedure, once the  $j$ th model has been selected, the estimator  $\hat{\beta}_1^j$  of the coefficient associated with  $x_{1i}$  of the selected model is considered as the “best” estimator of  $\beta_1$ . The estimator  $\hat{\beta}_1^j$  is called the estimator-post-selection. An alternative to selecting one “best” model and the best estimator involves forming a weighted average of the  $J$  estimators  $\hat{\beta}_1^j$  from all competing models  $M_j$  ( $j = 1, 2, \dots, J$ ), i.e.,  $\hat{\beta}_1 = \sum_{j=1}^J \pi_j \hat{\beta}_1^j$  with the scaled weights  $\pi_j$  assigned to model  $M_j$  such that  $\sum_j \pi_j = 1$ . The estimator  $\hat{\beta}_1$  is called the model-averaged estimator. Other than in parameter estimation, the model averaging is also applicable to other predictions of the SF model, such as the firm-specific frontier  $g(x_i; \beta)$  or

<sup>2</sup> For example, among the relevant discussions on SF models, the distribution of the inefficiency component is specified either as  $N^+(\delta_0 + \delta^T z_i, e^{2(\gamma_0 + \gamma^T z_i)})$  or as  $N^+(e^{\delta_0 + \delta^T z_i}, e^{2(\gamma_0 + \gamma^T z_i)})$ . The former specification is called the generalized linear mean (GLM) model and the latter specification is called the generalized exponential mean (GEM) model (Lai and Huang 2010). As detailed in Sect. 4, various settings of the distributional parameters generate most commonly specified SF models.

<sup>3</sup> See Gallant (1982) for the specification of the Fourier flexible function.

the firm-specific technical efficiency  $TE_i$ . The estimator-post-selection of  $g(x_i; \beta)$  is  $g(x_i; \hat{\beta}^j)$ , while the model-averaged estimator is  $\hat{g}(x_i; \beta) = \sum_{j=1}^J \pi_j g(x_i; \hat{\beta}^j)$ . Similarly, the estimator-post-selection of  $TE_i$  is  $\widehat{TE}_i^j$ , while the model-averaged estimator is  $\widehat{TE}_i = \sum_{j=1}^J \pi_j \widehat{TE}_i^j$ .

In general, let  $\phi_i = \phi(x_i, z_i; \theta)$  be a parameter of interest in a SF model. Here the parameter is defined more broadly as it may be a coefficient  $\beta$  which is constant, or it may be a firm-specific frontier  $g(x_i; \beta)$  or a firm-specific TE  $E(e^{-u_i} | \varepsilon_i)$  which is a function of the exogenous variables  $z_i$  that affect the inefficiency component  $u_i$ . If we follow Buckland et al. (1997), the model-averaged estimator of the parameter  $\phi_i$  is

$$\hat{\phi}_i = \sum_{j=1}^J \pi_j \hat{\phi}_i^j, \tag{7}$$

where  $\hat{\phi}_i^j$  is the  $j$ th model estimator of  $\phi_i$  with the assigned weight  $\pi_j$  such that  $\sum_j \pi_j = 1$ . The weight  $\pi_j$  is interpreted as the weight of evidence in favor of model  $M_j$ .

We like to emphasize that the above proposed model-average estimator can only be constructed when the parameter of interest  $\phi_i$  are comparable in all  $J$  competing models. If the  $J$  competing models are nested and  $\phi_i$  is missing from the model  $j$ , say the coefficient of the variable  $x_i$  with coefficient  $\phi_i^j$  is missing, then the comparability implies that  $\phi_i^j = 0$  is specified, and hence we set  $\hat{\phi}_i^j = 0$  in (7) when computing the averaged estimator.<sup>4</sup> In general, the  $J$  competing models may be non-nested in the model’s coefficient vector  $\theta^j$ , or the parameter of interest  $\phi_i^j$  is a function of model’s coefficients, say,  $\phi_i^j = \phi(x_i, z_i; \theta^j)$ . In this case, the model-average estimator is applicable to the comparable parameters  $\phi_i^j$  and is not applicable to the non-comparability of  $\theta^j$ . One example is the model-average estimator of the firm-specific TE,<sup>5</sup>  $\widehat{TE}_i = \sum_{j=1}^J \pi_j E(e^{-u_i} | \hat{\varepsilon}_i^j, \hat{\theta}^j)$ .

Some commonly used weights are based on the Kullback–Leibler (K–L) information criterion of model selection. The Akaike information criterion of Akaike (1973) for model  $M_j$  is defined as

$$AIC_j = -2 \ln L_n(\hat{\theta}^j) + 2k_j, \tag{8}$$

where  $k_j$  is the number of parameters in  $M_j$ . This equation suggests that the AIC model-averaged weights are

<sup>4</sup> This is exactly the case we confront when we estimate the marginal effect of the inefficiency determinants on the mean of the technical efficiency in our empirical example in Sect. 4. Model  $M_2$  does not specify any inefficiency determinants and thus the all marginal effects are set as zeros. Please see Table 1.

<sup>5</sup> In our empirical, eight models are estimated but model 1 ( $M_1$ ) and model 8 ( $M_8$ ) are not nested to each other.

$$\pi_j^{\text{AIC}} = \frac{\exp(-\frac{1}{2}\text{AIC}_j)}{\sum_{k=1}^J \exp(-\frac{1}{2}\text{AIC}_k)} = \frac{\exp(-\frac{1}{2}\Delta_j^{\text{AIC}})}{\sum_{k=1}^J \exp(-\frac{1}{2}\Delta_k^{\text{AIC}})}, \tag{9}$$

where  $\Delta_j^{\text{AIC}} = \text{AIC}_j - \text{AIC}_{\min}$  measures the relative AIC differences between model  $M_j$  and the best model among the  $J$  competing models.<sup>6</sup> A smaller value of  $\Delta_j^{\text{AIC}}$  corresponds to a better model, and the best competing models has  $\Delta_j^{\text{AIC}} = 0$ . Therefore,  $\Delta_j^{\text{AIC}}$  serves as a relevant measure of ranking the competing models. Furthermore, since the likelihood value  $L_n(\hat{\theta}^j)$  is proportional to  $\exp(-\frac{1}{2}\Delta_j)$ , Akaike (1983) called the term  $\exp(-\frac{1}{2}\Delta_j)$  the relative likelihood of model, and the Akaike weight  $\pi_j^{\text{AIC}}$  measures the relative importance of model  $M_j$  among all  $J$  competing models.

Alternatively, the Bayesian information criterion (BIC) of Schwarz (1978) for model  $M_j$  is defined as

$$\text{BIC}_j = -2 \ln L_n(\hat{\theta}^j) + k_j \ln n, \tag{10}$$

where  $\Delta_j^{\text{BIC}} = \text{BIC}_j - \text{BIC}_{\min}$ . The corresponding BIC model-averaged weights are

$$\pi_j^{\text{BIC}} = \frac{\exp(-\frac{1}{2}\text{BIC}_j)}{\sum_{k=1}^J \exp(-\frac{1}{2}\text{BIC}_k)} = \frac{\exp(-\frac{1}{2}\Delta_j^{\text{BIC}})}{\sum_{k=1}^J \exp(-\frac{1}{2}\Delta_k^{\text{BIC}})}. \tag{11}$$

The third model-averaged weights are defined in terms of the Takeuchi information criterion (TIC) of Takeuchi (1976),

$$\text{TIC}_j = -2 \ln L_n(\hat{\theta}^j) + 2 \text{tr} [H(\hat{\theta}^j)I(\hat{\theta}^j)^{-1}], \tag{12}$$

where

$$I(\hat{\theta}^j) = \frac{1}{n} \sum_{i=1}^n - \frac{\partial^2 \ln f(y_i|x_i, z_i; \hat{\theta}^j)}{\partial \theta^j \partial \theta^{jT}}$$

is the Fisher information matrix and

$$H(\hat{\theta}^j) = \frac{1}{n} \sum_{i=1}^n \left[ \frac{\partial \ln f(y_i|x_i, z_i; \hat{\theta}^j)}{\partial \theta^j} \frac{\partial \ln f(y_i|x_i, z_i; \hat{\theta}^j)}{\partial \theta^{jT}} \right]$$

is the Hessian matrix. The TIC model-averaged weights are

$$\pi_j^{\text{TIC}} = \frac{\exp(-\frac{1}{2}\text{TIC}_j)}{\sum_{k=1}^J \exp(-\frac{1}{2}\text{TIC}_k)} = \frac{\exp(-\frac{1}{2}\Delta_j^{\text{TIC}})}{\sum_{k=1}^J \exp(-\frac{1}{2}\Delta_k^{\text{TIC}})}. \tag{13}$$

and  $\Delta_j^{\text{TIC}} = \text{TIC}_j - \text{TIC}_{\min}$ .

<sup>6</sup> It is merely for the computational purpose of subtracting the minimum  $\text{AIC}_{\min}$  to avoid numerical problems with very large or very small arguments inside the exponential function.

Results when comparing AIC, BIC, and TIC performance depend on the nature of the DGP, on whether the competing models contains the truth, on the sample sizes, and also on the objective of the researcher—select the true model or select the K–L best approximating model. The model selection criteria of AIC, TIC, and BIC, take the form of a penalized log-likelihood function with various penalty functions. By comparing the penalty terms in (8), (10) and (12), AIC and TIC may perform poorly in model selection if there are too many parameters in relation to the sample size in the competing models. The BIC of (10) is quite similar to the AIC of (8) with a strong penalty for a more complex competing model with more parameters. Furthermore, if the candidate models contain the truth, BIC selection is consistent in the sense that it detects “the true model” with probability tending to 1 when the sample size increases, but AIC and TIC are not consistent. On the other hand, the selection criteria of AIC and TIC are efficient in the sense that the criteria reach the minimum of mean squared prediction errors as the sample size tends to infinity. Accordingly, when the true model exists and it is one of the candidate models under consideration, BIC is suggested. However, if the true model is complicated or is not one of the candidate models under consideration, then one can only approximate the truth by a set of candidate models. In this case, one may estimate the K-L distance by AIC and TIC and determine which approximating model is best. Both AIC and BIC are more suitable for model selection among nested models, but TIC is not restricted to the nested cases. More specifically, AIC can be treated as an approximation to TIC since  $\text{tr}[H(\hat{\theta}^j)I(\hat{\theta}^j)^{-1}] = k_j$  if the “information equality” holds. See Takeuchi (1976) for further details. Consequently, the choice of criteria may depend on the empirical economist’s belief or objective. Applications of AIC and TIC on selection and test of SF models are examined in Lai and Huang (2010).

Once the model-averaged estimator defined in (7) is obtained, it remains to find the estimator for the variance of the model-averaged estimator. Suppose that the weights  $\pi_j$  are given by either  $\pi_j^{\text{AIC}}$ ,  $\pi_j^{\text{BIC}}$ , or  $\pi_j^{\text{TIC}}$ , and the parameter estimators  $\hat{\phi}_i^j$  and  $\hat{\phi}_i^k$  under models  $M_j$  and  $M_k$  are correlated with the correlation coefficient  $\rho_{jk}$ . By decomposing the mean squares of the model-averaged estimator  $\hat{\phi}_i$ , the variance of  $\hat{\phi}_i$  can be shown to be<sup>7</sup>

<sup>7</sup> See Buckland et al. (1997) for the derivation.

$$\text{var}(\hat{\phi}_i) = \sum_{j=1}^J \pi_j^2 \left[ \text{var}(\hat{\phi}_i^j | M_j) + (\phi_i^j - \phi_i)^2 \right] + \sum_{j \neq k} \sum \pi_j \pi_k \rho_{jk} \sqrt{\left[ \text{var}(\hat{\phi}_i^j | M_j) + (\phi_i^j - \phi_i)^2 \right] \left[ \text{var}(\hat{\phi}_i^k | M_k) + (\phi_i^k - \phi_i)^2 \right]}, \tag{14}$$

where  $\text{var}(\hat{\phi}_i^j | M_j)$  is the variance under  $M_j$  and  $(\phi_i^j - \phi_i)$  denotes the model misspecification bias. In practice, however, both  $\rho_{jk}$  and  $\text{var}(\hat{\phi}_i^j | M_j)$  are unknown and need to be estimated. Suppose  $\widehat{\text{var}}(\hat{\phi}_i^j | M_j)$  and  $\hat{\rho}_{jk}$  are the sample estimates of  $\text{var}(\hat{\phi}_i^j | M_j)$  and the sample correlation coefficient of  $\hat{\phi}_i^j$  and  $\hat{\phi}_i^k$ , respectively. The corresponding estimate of the variance of the average estimator is then

$$\widehat{\text{var}}(\hat{\phi}_i) = \sum_{j=1}^J \pi_j^2 \left[ \widehat{\text{var}}(\hat{\phi}_i^j | M_j) + (\hat{\phi}_i^j - \hat{\phi}_i)^2 \right] + \sum_{j \neq k} \sum \pi_j \pi_k \hat{\rho}_{jk} \sqrt{\left[ \widehat{\text{var}}(\hat{\phi}_i^j | M_j) + (\hat{\phi}_i^j - \hat{\phi}_i)^2 \right] \left[ \widehat{\text{var}}(\hat{\phi}_i^k | M_k) + (\hat{\phi}_i^k - \hat{\phi}_i)^2 \right]}. \tag{15}$$

Empirically, it is likely that the estimates  $\hat{\phi}_i^j$  and  $\hat{\phi}_i^k$  under models  $M_j$  and  $M_k$  are highly correlated as the same data are used in the estimation. By setting  $\hat{\rho}_{jk} = 1$ , we have the upper bound of  $\widehat{\text{var}}(\hat{\phi}_i)$ ,

$$\widehat{\text{var}}(\hat{\phi}_i) \leq \left[ \sum_{j=1}^J \pi_j \sqrt{\widehat{\text{var}}(\hat{\phi}_i^j | M_j) + (\hat{\phi}_i^j - \hat{\phi}_i)^2} \right]^2. \tag{16}$$

### 3 Stochastic frontier model-averaged estimator

Two set of parameters are important in the estimation of SF model in (1). One is the frontier  $g(x_i; \beta)$ , which requires the estimation of the parameters  $\beta$ , and the other is the TE  $\text{TE}_i = E(e^{-u_i} | \varepsilon_i)$ , which requires the estimation of the parameters  $\theta = (\beta^T, \theta_v^T, \theta_u^T)^T$ . In this section, we separately propose the model-averaged estimation of  $\beta$  and  $\text{TE}_i$ .

Suppose  $\beta_t \in \beta$  is the  $t$ th parameter of the frontier function in  $g(x_i; \beta)$ . By (7) the model-averaged estimator is

$$\hat{\beta}_t = \sum_{j=1}^J \pi_j \hat{\beta}_t^j \tag{17}$$

where  $\hat{\beta}_t^j$  is the estimator under model  $M_j$ . However, the computation of the variance of  $\hat{\beta}_t$  requires the estimate of the sample correlation coefficient  $\hat{\rho}_{jk}(\beta_t)$  of  $\hat{\beta}_t^j$  and  $\hat{\beta}_t^k$  between model  $M_j$  and model  $M_k$ .

In the previous works of Buckland et al. (1997) and Burnham and Anderson (1998), they suggest simplifying (15) by assuming  $\hat{\rho}_{jk}(\beta_t) = 1$  to obtain the upper bound as in (16), or setting  $\hat{\rho}_{jk}(\beta_t) = \rho$  for all  $t, j$  and  $k$ . Estimation of  $\hat{\rho}_{jk}(\beta_t)$  is clearly infeasible in the ML approach as for each data set there is only one ML estimate  $\hat{\beta}_t^j$  for a model. Burnham and Anderson (1998) have suggested using the bootstrap approach as an alternative method to obtain the distribution of the averaged estimator  $\hat{\beta}_t$  and the variance

$\widehat{\text{var}}(\hat{\beta}_t)$  from the bootstrap samples. This approach requires first estimating the  $J$  competing models and then averaging over those bootstrap estimators in order to obtain the averaged estimator and its distribution. A large number of bootstrapped samples is obviously needed to obtain a good approximation.<sup>8</sup> It is a computation-intensive approach.

To lessen the computation time, we propose an alternative bootstrap approach to estimating the standard error of the averaged estimator. To compute the variance of the averaged estimator  $\widehat{\text{var}}(\hat{\beta}_t)$  in (15), only the estimation of the correlation coefficients  $\hat{\rho}_{jk}(\beta_t)$  between the competing models is required. We therefore propose bootstrapping only the correlation coefficients between the frontier coefficients. The main difference between our bootstrap approach and that of Burnham and Anderson (1998) is that model averaging is not required in our bootstrap procedure because we do not intend to obtain the distribution of the averaged estimator. But the question of how many bootstrap samples are required to obtain a good result will be examined later on by a Monte Carlo simulation.

The proposed bootstrap procedure is stated as follows. A bootstrapped sample is generated by independently

<sup>8</sup> Burnham and Anderson (1998) recommend 10,000 bootstrap samples and at least 1,000 are required.



drawing from the original data with a replacement, and then using the bootstrapped data to estimate the  $J$  competing models. Denote  $\hat{\beta}_t^j(b)$  as the estimates of  $\beta_t^j$  from the  $b$ th bootstrapped sample. By repeating the same procedure for  $B$  times, the estimator of the sample correlation coefficient of  $\hat{\beta}_t^j$  and  $\hat{\beta}_t^k$  can be computed,

$$\widehat{\text{var}}(\widehat{\text{TE}}_i) = \sum_{j=1}^J \pi_j^2 \left[ \widehat{\text{var}}(\widehat{\text{TE}}_i^j | M_j) + (\widehat{\text{TE}}_i^j - \widehat{\text{TE}}_i)^2 \right] + \sum_{j \neq k} \sum_{j=1}^J \pi_j \pi_k \hat{\rho}_{jk}(\text{TE}_i) \times \sqrt{\left[ \widehat{\text{var}}(\widehat{\text{TE}}_i^j | M_j) + (\widehat{\text{TE}}_i^j - \widehat{\text{TE}}_i)^2 \right] \left[ \widehat{\text{var}}(\widehat{\text{TE}}_i^k | M_k) + (\widehat{\text{TE}}_i^k - \widehat{\text{TE}}_i)^2 \right]}. \tag{22}$$

$$\hat{\rho}_{jk}(\beta_t) = \frac{\sum_{b=1}^B (\hat{\beta}_t^j(b) - \bar{\beta}_t^j) (\hat{\beta}_t^k(b) - \bar{\beta}_t^k)}{\sqrt{\sum_{b=1}^B (\hat{\beta}_t^j(b) - \bar{\beta}_t^j)^2} \sqrt{\sum_{b=1}^B (\hat{\beta}_t^k(b) - \bar{\beta}_t^k)^2}}, \tag{18}$$

where  $\bar{\beta}_t^j$  is the sample mean of  $\hat{\beta}_t^j(b)$ ,  $b = 1, \dots, B$ . Equations (15) and (18) together suggest that the variance of the averaged estimator for  $\beta_t^j$  can be estimated by

$$\widehat{\text{var}}(\hat{\beta}_t) = \sum_{j=1}^J \pi_j^2 \left[ \widehat{\text{var}}(\hat{\beta}_t^j | M_j) + (\hat{\beta}_t^j - \hat{\beta}_t)^2 \right] + \sum_{j \neq k} \sum_{j=1}^J \pi_j \pi_k \hat{\rho}_{jk}(\beta_t) \sqrt{\left[ \widehat{\text{var}}(\hat{\beta}_t^j | M_j) + (\hat{\beta}_t^j - \hat{\beta}_t)^2 \right] \left[ \widehat{\text{var}}(\hat{\beta}_t^k | M_k) + (\hat{\beta}_t^k - \hat{\beta}_t)^2 \right]}. \tag{19}$$

Similarly, we may define the model-averaged TE estimator of the  $i$ th firm to be the weighted average of the  $J$  competing estimators,

$$\widehat{\text{TE}}_i = \sum_{j=1}^J \pi_j \widehat{\text{TE}}_i^j, \tag{20}$$

where  $\widehat{\text{TE}}_i^j$  defined in (5) is the estimator for firm  $i$ 's TE under Model  $M_j$ . Following standard SF modeling, the distribution of the symmetric error  $v_i^j$  is assumed to be  $N(0, \sigma_v^{j2})$  and is independent of the nonnegative error  $u_i^j$  which follows the truncated-normal distribution as  $N^+(\mu^j, \sigma_u^{j2})$ , i.e., truncated from below at zero and with the mode at  $\mu^j$ . Given the ML estimators  $\hat{\theta}^j = (\hat{\beta}^j, \hat{\sigma}_v^{j2}, \hat{\mu}^j, \hat{\sigma}_u^{j2})$ , we have

$$\widehat{\text{TE}}_i^j = \frac{1 - \Phi(\hat{\sigma}_*^j - \hat{\mu}_{*i}^j / \hat{\sigma}_*^j)}{1 - \Phi(-\hat{\mu}_{*i}^j / \hat{\sigma}_*^j)} e^{-\hat{\mu}_{*i}^j + \frac{1}{2} \hat{\sigma}_*^{j2}}, \tag{21}$$

where  $\hat{\mu}_{*i}^j = (-\hat{\sigma}_u^{j2} \hat{\varepsilon}_i^j + \hat{\mu}^j \hat{\sigma}_v^{j2}) / \hat{\sigma}^{j2}$ ,  $\hat{\sigma}_*^{j2} = \hat{\sigma}_v^{j2} \hat{\sigma}_u^{j2} / \hat{\sigma}^{j2}$ ,  $\hat{\sigma}^{j2} = (\hat{\sigma}_v^{j2} + \hat{\sigma}_u^{j2})$ , and  $\hat{\varepsilon}_i^j = y_i - x_i^T \hat{\beta}^j$  is the predicted value of the composite error  $\varepsilon_i^j$ .

Following (15), the variance of the model-averaged TE estimator is

Obviously, the estimation of the variance of  $\widehat{\text{TE}}_i$  in (22) requires the estimation of the conditional variances,  $\widehat{\text{var}}(\widehat{\text{TE}}_i^j | M_j)$ , and the cross-model correlation coefficient  $\hat{\rho}_{jk}(\text{TE}_i)$  between  $\widehat{\text{TE}}_i^j$  and  $\widehat{\text{TE}}_i^k$ . Since the TE is highly nonlinear in  $\hat{\theta}^j = (\hat{\beta}^j, \hat{\sigma}_v^{j2}, \hat{\mu}^j, \hat{\sigma}_u^{j2})$ , the computation of the conditional variance,  $\widehat{\text{var}}(\widehat{\text{TE}}_i^j | M_j)$ , may have to resort to some numerical approximations. However, we have no

basis for estimating the cross-model correlation coefficient  $\hat{\rho}_{jk}(\text{TE}_i)$  other than the bootstrap. Because of the computational intensity required in the computation of individual firm-specific variances  $\widehat{\text{var}}(\widehat{\text{TE}}_i)$ , only the sample estimate of the model-averaged TE is illustrated without the evaluation of the variances.<sup>9</sup>

<sup>9</sup> Alternatively, we could assume that the correlation between the two models' prediction on efficiency is independent of firms, i.e.,  $\rho_{jk}(\text{TE}_i) = \rho_{jk}(\text{TE}_t) = \rho_{jk}(\text{TE})$ , for all firms,  $i, t = 1, \dots, n$ . Thus, we suggest the estimation of  $\rho_{jk}(\text{TE})$  based on the estimates of the  $n$  firm-specific technical efficiencies,  $\hat{\rho}_{jk}(\text{TE}) =$

$\frac{\sum_{i=1}^n (\widehat{\text{TE}}_i^j - \widehat{\text{TE}}_i^k) (\widehat{\text{TE}}_i^k - \widehat{\text{TE}}_i^j)}{\sqrt{\sum_{i=1}^n (\widehat{\text{TE}}_i^j - \widehat{\text{TE}}_i^j)^2} \sqrt{\sum_{i=1}^n (\widehat{\text{TE}}_i^k - \widehat{\text{TE}}_i^k)^2}}$ , where  $\widehat{\text{TE}}_i^j$  is the sample mean of  $\widehat{\text{TE}}_i^j$ , for  $I = 1, \dots, n$ .

### 4 An empirical illustration

In this section, the proposed model-averaged estimation is applied to a SF production function using Indian farm data. The same data has been used by Battese and Coelli (1992, 1995), Wang (2002), Alvarez et al. (2006), and Lai and Huang (2010). The farm’s output variable  $Y$  is the total value of output. The other variables, *Land*, *PILand*, *Labor*, *Bullock*, *Cost*, *Age*, *School*, and *Year*, are either the input variable or the exogenous variable that might affect the TE. *Land* is the total area of irrigated and unirrigated land under operation; *PILand* is the proportion of operated land that is irrigated; *Labor* is the total hours of family and hired labor; *Bullock* is the hours of bullock labor; *Cost* is the value of other inputs, including fertilizer, manure, pesticides, machinery, etc.; *Age* is the age of the primary decision-maker in the farming operation; *School* is the years of formal schooling of the primary decision maker; and *Year* is the year of the observations involved. The following empirical analysis is programmed in Stata 10.0 software.

Among the relevant discussions on SF models with truncated-normal distributions, various assumptions are imposed on the distributions of the one-sided error. In the previous study of Lai and Huang (2010), they consider the following eight types of settings of  $u_i$  in the SF models, including:

$M_1$ : The generalized exponential mean (GEM) model (Alvarez et al. 2006)

$$u_i \sim N^+(e^{\delta_0 + \delta^T z_i}, e^{2(\gamma_0 + \gamma^T z_i)}).$$

$M_2$ : The Stevenson Model (Stevenson 1980)

$$u_i \sim N^+(\mu, \sigma_u^2).$$

$M_3$ : The scaled Stevenson model (Wang and Schmidt 2002)

$$u_i = e^{z_i^T \delta} \cdot \omega_i \quad \text{with} \quad \omega_i \stackrel{i.i.d.}{\sim} N^+(\mu, \sigma_u^2).$$

$M_4$ : The KGMHLBC model (Kumbhakar et al. 1991; Huang and Liu 1994; Battese and Coelli 1995)

$$u_i \sim N^+(\mu e^{\delta^T z_i}, \sigma_u^2).$$

$M_5$ : The RSCFG- $\mu$  model

$$u_i \sim N^+(\delta_0, \sigma_u^2 e^{2\gamma^T z_i}).$$

$M_6$ : The RSCFG model (Reifschneider and Stevenson 1991; Caudill and Ford 1993; Caudill et al. 1995)

$$u_i \sim N^+(0, e^{2(\gamma_0 + \gamma^T z_i)}).$$

$M_7$ : The HL model (Huang and Liu 1994)

$$u_i \sim N^+(\delta_0 + \delta^T z_i, e^{2\gamma_0}).$$

$M_8$ : The generalized linear mean (GLM) model (Lai and Huang 2010)

$$u_i \sim N^+(\delta_0 + \delta^T z_i, e^{2(\gamma_0 + \gamma^T z_i)}).$$

The specifications of the frontier function  $g^j(x_i; \beta^j)$  are identical for all eight models except that distributions of  $u_i^j$  are different as indicated in  $M_j, j = 1, \dots, 8$ .

The input variables  $x$  in our empirical model include:  $\ln(Land)$ ,  $PILAND$ ,  $\ln(Labor)$ ,  $\ln(Bullock)$ ,  $\ln(Cost)$ ,  $Year$ ,  $\ln(Labor) \times Year$ ,  $School$ ; and the exogenous variables  $z$  used in the inefficiency term include:  $Age$ ,  $School$ , and  $Year$ . This set of variables has been selected by Lai and Huang (2010) after conducting the AIC and TIC model selection criterion, and also Voung’s LR test. Therefore, we concentrate only on the analyses of SF models with different distribution assumptions on the one-sided error  $u$ , while keeping the input variables fixed. The estimated parameters of models  $M_1$ – $M_8$  are summarized in Table 1. The descriptive statistics of the technical efficiencies and the sample average of the marginal effect of the inefficiency determinants on the TE,  $\partial TE_i^j / \partial z_i$ , are also reported in the bottom of Table 1. Since  $TE_i^j$  is a complicated function of the parameters, the close form of  $\partial TE_i^j / \partial z_i$  is almost intractable. We, therefore, use the numerical method to compute the derivative and obtain these estimates.

The correlation coefficients of the estimated technical efficiencies under models  $M_1$ – $M_8$  are summarized in Table 2. All values are close to one indicates that the predictions of technical efficiencies do not vary much under the eight models.

Three different weights based on AIC, BIC, and TIC are considered in the model-averaged estimators. Table 3 shows the values of  $AIC_j$ ,  $BIC_j$ , and  $TIC_j$  under models  $M_1$ – $M_8$  and their respective weights,  $\pi_j^{AIC}$ ,  $\pi_j^{BIC}$ , and  $\pi_j^{TIC}$ . Both AIC and TIC select  $M_8$  as the best model while BIC selects  $M_2$  as the best model. However, if  $M_8$  is dropped, then both AIC and BIC choose  $M_6$  as the best model while TIC suggests  $M_1$  as the best model. It seems that different model selection criteria may suggest different models, and rank the competing models in different order. In other words, different model selection criteria may select different best empirical models for inference and thus have inconsistent conclusions. Although both AIC and TIC choose the same best model, the relative importance of

**Table 1** The estimates of models  $M_1$ – $M_8$

Variable	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$
<b>Frontier</b>								
$\ln(Land)$	0.2546*** <sup>a</sup> (0.0760) <sup>b</sup>	0.2954*** (0.0677)	0.2972*** (0.0700)	0.2988*** (0.0709)	0.2962*** (0.0704)	0.2909*** (0.0729)	0.3005*** (0.0701)	0.2547*** (0.0736)
$PI_{Land}$	0.1900 (0.1793)	0.2458 (0.1736)	0.2361 (0.1755)	0.2307 (0.1769)	0.2360 (0.1754)	0.2316 (0.1779)	0.2386 (0.1754)	0.1764 (0.1762)
$\ln(Labor)$	1.1601*** (0.0808)	1.1103*** (0.0812)	1.1193*** (0.0818)	1.1014*** (0.0785)	1.1205*** (0.0821)	1.1300*** (0.0809)	1.1102*** (0.0813)	1.1553*** (0.0826)
$\ln(Bullock)$	−0.4258*** (0.0614)	−0.4103*** (0.0586)	−0.4129*** (0.0603)	−0.3942*** (0.0604)	−0.4129*** (0.0604)	−0.4186*** (0.0616)	−0.4035*** (0.0595)	−0.4234*** (0.0598)
$\ln(Cost)$	0.0167 (0.0128)	0.0094 (0.0125)	0.0141 (0.0126)	0.0148 (0.0128)	0.0141 (0.0126)	0.0154 (0.0127)	0.0142 (0.0125)	0.0155 (0.0126)
$\ln(Labor) \times Year$	0.0029* (0.0013)	0.0049*** (0.0011)	0.0028* (0.0013)	0.0031* (0.0016)	0.0026* (0.0013)	0.0026* (0.0013)	0.0026 (0.0014)	0.0028* (0.0013)
<i>School</i>	−0.0070 (0.0096)	−0.0021 (0.0090)	−0.0020 (0.0097)	0.0027 (0.0097)	−0.0019 (0.0099)	−0.0031 (0.0096)	0.0012 (0.0101)	−0.0060 (0.0098)
Constant	1.6759*** (0.3452)	1.7353*** (0.3368)	1.7657*** (0.3397)	1.8409*** (0.3332)	1.7650*** (0.3402)	1.7969*** (0.3429)	1.7601*** (0.3366)	1.6319*** (0.3414)
$\mu$								
<i>Age</i>	−0.0755 (0.0412)	–	–	0.0288 (0.0301)	–	–	−0.0058 (0.0163)	−0.0449 (0.0278)
<i>School</i>	−0.2736 (0.2013)	–	–	0.1828 (0.1303)	–	–	0.0272 (0.0831)	−0.1974 (0.1505)
<i>Year</i>	0.2392 (0.1896)	–	–	−0.5036 (0.2633)	–	–	−0.2261 (0.2510)	0.1304 (0.1103)
Constant	0.2891 (1.2579)	−2.8040 (6.2633)	−1.7263 (3.4535)	−2.4782 (2.2780)	−0.9399 (1.6886)	–	−0.3258 (1.7290)	1.0991 (0.6654)
$\sigma_u^{2c}$								
<i>Age</i>	0.0049 (0.0107)	–	–	–	−0.0063 (0.0072)	−0.0076 (0.0084)	–	0.0266** (0.0097)
<i>School</i>	0.0074 (0.0514)	–	–	–	−0.0130 (0.0411)	−0.0165 (0.0473)	–	0.1026 (0.0586)
<i>Year</i>	−0.1990*** (0.0521)	–	–	–	−0.1126* (0.0482)	−0.1433** (0.0475)	–	−0.2230*** (0.0508)
Constant	−0.8158 (0.6797)	0.0623 (1.7281)	0.6029 (1.2908)	−1.4952*** (0.1792)	0.2760 (0.8771)	−0.2152 (0.5384)	−0.3150 (1.0928)	−1.2733 (0.8728)
$\sigma_v^{2d}$								
$c_v$	−3.5094*** (0.3194)	−3.1637*** (0.2680)	−3.2728*** (0.3039)	−3.5524*** (0.3304)	−3.2737*** (0.2981)	−3.4817*** (0.3003)	−3.2219*** (0.2873)	−3.2717*** (0.2649)
<b>Scale</b>								
<i>Age</i>	–	–	−0.0043 (0.0051)	–	–	–	–	–
<i>School</i>	–	–	−0.0084 (0.0284)	–	–	–	–	–
<i>Year</i>	–	–	−0.0778** (0.0293)	–	–	–	–	–



**Table 1** continued

Number of parameters	17	11	14	14	14	13	14	17
Estimated TEs								
Mean	0.6955	0.7581	0.7439	0.6891	0.7446	0.7037	0.7530	0.7449
SD	0.1636	0.1494	0.1581	0.1647	0.1589	0.1639	0.1571	0.1595
Min.	0.2090	0.2617	0.2299	0.2214	0.2296	0.2131	0.2345	0.2279
Max.	0.9322	0.9412	0.9374	0.9356	0.9381	0.9352	0.9400	0.9332
Marginal effects <sup>c</sup>								
Age	0.0005579	0	0.0000715	−0.0001693	0.0002566	0.0001834	0.0001192	0.0007775
School	0.0022343	0	0.0001381	−0.0010758	0.0005284	0.0003964	−0.0005589	0.0039937
Year	0.0020867	0	0.0012805	0.0029641	0.0045615	0.0034490	0.0046471	0.0036095

<sup>a</sup> \*, \*\*, and \*\*\* Represent the *p*-values that are less than 0.1, 0.05, and 0.01, respectively

<sup>b</sup> Numbers in parentheses are the standard errors

<sup>c</sup>  $\sigma_u^2 = e^{2(\gamma_0 + \gamma^T z_i)}$

<sup>d</sup>  $\sigma_v^2 = e^{c_v}$

<sup>e</sup> The sample means of the marginal effects are reported

**Table 2** The correlation coefficients of the predicted TEs of models  $M_1$ – $M_8$

	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$
$M_1$	1							
$M_2$	0.9735	1						
$M_3$	0.9892	0.9834	1					
$M_4$	0.99	0.9753	0.987	1				
$M_5$	0.9886	0.9802	0.9998	0.9856	1			
$M_6$	0.9964	0.9752	0.995	0.9948	0.9946	1		
$M_7$	0.9818	0.9782	0.9983	0.9838	0.9985	0.9905	1	
$M_8$	0.9913	0.9757	0.9926	0.9751	0.9928	0.988	0.9876	1

**Table 3** The estimated values of AIC, BIC, and TIC of models  $M_1$ – $M_8$  and the corresponding weights

Model	AIC	BIC	TIC	$\pi_j^{AICa}$	$\pi_j^{BIC}$	$\pi_j^{TIC}$
$M_1$	200.053* <sup>b</sup>	261.289	166.161***	0.0883	0.0001	0.1431***
$M_2$	202.775	242.398****	180.850	0.0226	0.8254****	0.0001
$M_3$	200.134	250.564*	172.231*	0.0848*	0.0139*	0.0069*
$M_4$	204.791	255.220	176.881	0.0083	0.0014	0.0007
$M_5$	199.937**	250.367**	172.034**	0.0936**	0.0154**	0.0076**
$M_6$	199.193***	246.021***	173.288	0.1358***	0.1349***	0.0041
$M_7$	201.097	251.527	173.194	0.0524	0.0086	0.0042
$M_8$	196.530****	257.766	162.637****	0.5141****	0.0004	0.8333****

<sup>a</sup>  $\pi_j^{AIC}$ ,  $\pi_j^{BIC}$ ,  $\pi_j^{TIC}$  Represent the weights assigned to *j*th model

<sup>b</sup> \*\*\*\*, \*\*\*, \*\*, \* The best, second best, third, and fourth among the top four best models

models are evaluated differently. For instance, the ranking of the relative importance among the top four models from the highest to the lowest are:

AIC ranking: ( $M_8, M_6, M_5, M_1$ );

BIC ranking: ( $M_2, M_6, M_5, M_3$ );

TIC ranking: ( $M_8, M_1, M_5, M_3$ ).

Table 4 gives the relative frequencies of models selected by AIC, BIC, and TIC under various replications of the bootstrapped sample. Each bootstrapped sample is obtained by sampling from the original data with replacement until

**Table 4** The relative frequencies of models selected by AIC, BIC, and TIC

	AIC				BIC				TIC			
	150	300	500	1,000	150	300	500	1,000	150	300	500	1,000
$M_1$	17.33	18.67	19.8	18.5	1.33	1	1.8	1.8	25.33	27.67	29	28
$M_2$	0.67	2.33	2.6	2.5	52	51.67	50.2	47.4	0	0	0	0
$M_3$	6	4.67	4.8	5.2	6	5.33	4.2	4.5	6	4.67	4.6	5.1
$M_4$	10	9.33	10	9.6	10	9	9.2	8.7	2.67	5	4.8	4.3
$M_5$	6	7.67	7	8.6	4	4.33	4.6	5.6	4.67	5.67	5	6.1
$M_6$	8.67	10	9.8	10	22	23.33	24.2	26	0	0.67	0.4	0.3
$M_7$	4	3.33	3.8	4.8	1.33	2	2.8	2.5	2.67	2.33	2.2	2.4
$M_8$	47.33	44	42.2	40.8	3.33	3.33	3	3.5	58.67	54	54	53.8

**Table 5** The averaged estimates of the coefficients in the frontier equation and descriptive statistics of the averaged technical inefficiencies

Variable/criterion	AIC	BIC	TIC
Estimated coefficients in the frontier part: $\hat{\beta} = \sum_{j=1}^J \pi_j \hat{\beta}^j$			
$\ln(Land)$	0.2708	(0.0545) <sup>a</sup> (0.0755) <sup>b</sup>	0.2949 (0.0669) (0.0685)
$PILand$	0.2010	(0.1324) (0.1787)	0.2435 (0.1716) (0.1744)
$\ln(Labor)$	1.1422	(0.0610) (0.0838)	1.1133 (0.0797) (0.0815)
$\ln(Bullock)$	-0.4195	(0.0447) (0.0605)	-0.4114 (0.0579) (0.0591)
$\ln(Cost)$	0.0151	(0.0095) (0.0127)	0.0104 (0.0125) (0.0127)
$\ln(Labor) \times year$	0.0028	(0.0009) (0.0013)	0.0045 (0.0012) (0.0014)
$School$	-0.0044	(0.0065) (0.0100)	-0.0022 (0.0089) (0.0092)
Constant	1.6928	(0.0470) (0.3482)	1.7448 (0.1053) (0.3384)
Estimated technical efficiencies: $\widehat{TE}_i = \sum_{j=1}^J \pi_j \widehat{TE}_i^j$			
Mean	0.7351		0.7502 0.7377
SD	0.1594		0.1512 0.1599
Min.	0.2256		0.2539 0.2252
Max.	0.9326		0.9395 0.9325
Estimated marginal effects of the inefficiency determinants on the mean of TE <sup>c</sup>			
Age	0.0005089		0.00003080 0.00073136
School	0.0023275		0.00005892 0.00365145
Year	0.0033120		0.00059873 0.00338582

<sup>a</sup> Numbers in parentheses are the estimated standard errors, which are obtained based on the matrix of correlation coefficients calculated by 1,000 bootstrapped data sets

<sup>b</sup> Numbers in parentheses are the upper bound of the estimated standard errors

<sup>c</sup> The averaged sample means of the marginal effects are reported

the bootstrapped sample size is the same as that of the original. The number of replications denoted by  $B$ , is 150, 300, 500, and 1,000, respectively. Several findings are

observed. First, the order of these models is quite stable under different sizes of  $B$  for the three criteria. Second, the ranking of the top four best models are:

**Table 6** The estimated standard errors of the averaged coefficients in the frontier equation under different bootstrapped sample

	AIC				BIC				TIC			
	150	300	500	1,000	150	300	500	1,000	150	300	500	1,000
ln(Land)	0.0483	0.0511	0.0532	0.0545	0.0664	0.0668	0.0668	0.0669	0.0616	0.0624	0.0635	0.0643
PILand	0.1239	0.1318	0.1335	0.1324	0.1705	0.1714	0.1718	0.1716	0.1516	0.1540	0.1547	0.1545
ln(Labor)	0.0552	0.0603	0.0615	0.0610	0.0790	0.0796	0.0798	0.0797	0.0700	0.0715	0.0722	0.0721
ln(Bullock)	0.0470	0.0465	0.0457	0.0447	0.0576	0.0579	0.0580	0.0579	0.0537	0.0530	0.0527	0.0524
ln(Cost)	0.0098	0.0095	0.0096	0.0095	0.0124	0.0125	0.0125	0.0125	0.0113	0.0111	0.0111	0.0111
ln(Labor) × year	0.0009	0.0009	0.0008	0.0009	0.0012	0.0012	0.0012	0.0012	0.0011	0.0011	0.0011	0.0011
School	0.0063	0.0064	0.0065	0.0065	0.0088	0.0089	0.0089	0.0089	0.0082	0.0083	0.0083	0.0083
Constant	0.2122	0.2152	0.2160	0.2169	0.3206	0.3238	0.3246	0.3244	0.2880	0.2887	0.2892	0.2896

AIC ranking: ( $M_8, M_1, M_6, M_4$ );

BIC ranking: ( $M_2, M_6, M_4, M_5$ );

TIC ranking: ( $M_8, M_1, M_5, M_3$ ).

Among these three criteria, the order of the eight models under consideration obtained by TIC in Table 4 is virtually consistent with the order of weights calculated according to TIC in Table 3. However, the orders of models based AIC and BIC in Tables 3 and 4 are not consistent. Furthermore, the top two models suggested by AIC weight are  $M_8$  and  $M_6$ , with  $M_8$  and  $M_1$  for the bootstrapped data. Therefore, it seems that TIC is a relative stable criterion compared with the other two criteria. Third, BIC tends to select a model with fewer parameters such as  $M_2$  with 11 parameters, and rank lower the model with more parameters, such as,  $M_1$  and  $M_8$  with 17 parameters each.

The model-averaged estimates of the coefficients in the frontier equation from models  $M_1$ – $M_8$ , and the descriptive statistics of the model-averaged technical efficiencies computed via (17) and (20) are given in Table 5, where the standard errors of the model-averaged estimates in the frontier function are estimated on the basis of 1,000 bootstrapped data sets. The corresponding upper bounds of these averaged estimates are also obtained according to (16) by setting  $\hat{\rho}_{jk}(\beta_t) = 1$  for all  $j, k$ , and  $t$ .

If the number of competing models is large, it may require much computation time when one conducts the bootstrap procedure. In order to examine whether the number of bootstrapped data sets significantly changes the estimated standard errors, we calculated the standard errors for the number of bootstrapped data sets  $B = 150, 300, 500$ , and 1,000. The estimated results are summarized in Table 6. It seems that when  $B = 300$ , we already have quite good results, which are very close to what we obtained when  $B = 1,000$ . Since our objective of using bootstrapped data is not to obtain the distribution, but to compute the correlation coefficient between the coefficient estimates, computation of (18) does not actually require a large bootstrapped number.

### 5 Conclusion

In this paper, we suggest using the model-averaged estimator based on the multimodel inference in estimating the SF model. The potential advantages of the proposed approach include: (1) the incorporation of the model selection uncertainty into statistical inference; (2) averaging across the frontier coefficients and the predicted technical efficiencies can reduce the model selection bias effects and thus reduce estimated variance.

We demonstrate our approach by utilizing Indian farm data. Three model selection criteria, AIC, BIC, and TIC, are considered in our empirical example. Our results show that among these criteria, TIC seems to be relatively stable in ordering the competing models. Moreover, the main advantage of TIC is that it does not require the true model to be among the competing models, but AIC and BIC do. BIC intends to select as the best model the one that contains the fewest parameters. Finally, we also suggest using the bootstrapped data to estimate the correlation coefficient between the estimated coefficients over the competing models, and show that the estimation does not require a large bootstrapped number.

**Acknowledgments** We thank Robin Sickles for his helpful comments. We also thank a referee for carefully reading the original version of this manuscript and for preparing helpful and constructive critiques which served to greatly improve the exposition and content. Huang gratefully acknowledges the Institute for Advanced Studies in Humanities and Social Sciences, National Taiwan University, for the research support. Lai gratefully acknowledges the National Science Council of Taiwan (NSC-99-2410-H-194-020) for the research support.

### References

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of Stochastic Frontier production models. *J Econom* 6:21–37  
 Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds)

- 2nd international symposium on information theory, Akademiai Kiadó, Budapest, pp 267–281
- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Akaike H (1983) Information measures and model selection. *Int Stat Inst* 44:277–291
- Alvarez A, Amsler C, Orea L, Schmidt P (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *J Prod Anal* 25:201–212
- Battese G, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *J Econom* 38:387–399
- Battese G, Coelli TJ (1992) Frontier production functions, technical efficiency and panel data: with applications to paddy farmers in India. *J Prod Anal* 3:153–169
- Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empir Econ* 20(2):325–332
- Box GEP (1976) Science and statistics. *J Am Stat Assoc* 71:791–799
- Buckland ST, Burnham KP, Augustin NH (1997) Model selection: an integral part of inference. *Biometrics* 53:603–618
- Burnham KP, Anderson DR (1998) *Model selection and multimodel inference*. Springer, New York
- Caudill SB, Ford JM (1993) Biases in Frontier estimation due to heteroskedasticity. *Econ Lett* 41:17–20
- Caudill SB, Ford JM, Gropper DM (1995) Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity. *J Bus Econ Stat* 13:105–111
- Gallant AR (1982) Unbiased determination of production technologies. *J Econom* 20:285–323
- Hansen BE (2005) Challenges for econometric model selection. *Econom Theory* 21:60–68
- Huang C, Liu JT (1994) Estimation of a non-neutral stochastic frontier production function. *J Prod Anal* 5:171–180
- Kumbhakar SC, Ghosh S, McGuckin JT (1991) A generalized production Frontier approach for estimating determinants of inefficiency in US dairy farms. *J Bus Econ Stat* 9:279–286
- Lai HP, Huang CJ (2010) Likelihood ratio test for model selection of stochastic frontier models. *J Prod Anal* 34:3–13
- Lee H, Pötscher BM (2005) Model selection and inference: facts and fiction. *Econom Theory* 21:21–59
- Liu Y, Myers R (2009) Model selection in stochastic frontier analysis with an application to maize production in Kenya. *J Prod Anal* 31:33–46
- Pötscher BM (1991) The effect of model selection on inference. *Econom Theory* 7:163–185
- Reifschneider D, Stevenson R (1991) Systematic departures from the Frontier: a framework for the analysis of firm inefficiency. *Int Econ Rev* 32:715–723
- Schmidt P, Lin TF (1984) Simple Tests of alternative specifications in Stochastic Frontier models. *J Econom* 24(3):349–361
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Stevenson RE (1980) Likelihood functions for generalized Stochastic Frontier estimation. *J Econom* 13:57–66
- Takeuchi K (1976) Distribution of information statistics and a criterion of model fitting. *Suri-Kagaku (Math Sci)* 153:12–18 (in Japanese)
- Wang HJ (2002) Heteroscedasticity and non-monotonic efficiency effects of a Stochastic Frontier model. *J Prod Anal* 18:241–253
- Wang HJ, Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *J Prod Anal* 18:129–144