Estimation of a panel stochastic frontier model with unobserved common shocks

Hsu, Chih-Chiang and Lin, Chang-Ching and Yin, Shou-Yung

Department of Economics, National Central University, Institute of Economics, Academia Sinica, Department of Economics, National Central University

March 2012

Estimation of a Panel Stochastic Frontier Model with Unobserved Common Shocks

Chih-Chiang Hsu
Department of Economics
National Central University

Chang-Ching Lin
Institute of Economics
Academia Sinica

Shou-Yung Yin
Department of Economics
National Central University

(Preliminary; do not cite or quote without permission)

Estimation of a Panel Stochastic Frontier Model with Unobserved Common Shocks

Abstract

This paper develops panel stochastic frontier models with unobserved common correlated effects. The common correlated effects provide a way of modeling cross-sectional dependence and represent heterogeneous impacts on individuals resulting from unobserved common shocks. Traditional panel stochastic frontier models do not distinguish between common correlated effects and technical inefficiency. In this paper, we propose a modified maximum likelihood estimator (MLE) that does not require estimating unobserved common correlated effects. We show that the proposed method can control the common correlated effects and obtain consistent estimates of parameters and technical efficiency for the panel stochastic frontier model. Our Monte Carlo simulations show that the modified MLE has satisfactory finite sample properties under a significant degree of cross-sectional dependence for relatively small T. The proposed method is also illustrated in applications based on a cross country comparison of the efficiency of banking industries.

JEL classification: C23

Keywords: fixed effects, common correlated effects, factor structure, cross-sectional dependence, stochastic frontier

1 Introduction

Panel data sets have been increasingly used in stochastic frontier models to analyze the inefficiency or efficiency of firms, banks and some government system we concerned. However, the conventional assumption of cross-sectional independence of error structure in stochastic frontier model might be suspect. One potential source of cross-sectional dependence is the global or economy-wise shocks, which might have various impacts on different firms/units, such as changes in interest rates and taxation, oil shocks, financial crises, or aggregate technological innovations. This type of cross-sectional dependence is usually referred to as common correlated effects in the literature and usually modeled by factor structure, a linear combination of common factors. Ignoring these unobserved common shocks can make the estimators of the parameters of slope and efficiency biased. Nevertheless, an endogeneity problem may arise because these common shocks may affect both firms' input decisions and their outputs. [1]

Conventional panel stochastic frontier analysis have relied on linear panel models with fixed or random effects without imposing distributional assumption on inefficiency. (Schmidt and Sickles (1984), Cornwell, Schmidt and Sickles (1990), Han, Orea and Schmidt (2005) and Lee (2006)). Although such methods are easy to implement and can measure the relative inefficiency by comparing the individual effects at each time period, they treat inefficiency is time-invariant and any time-invariant across unit heterogeneity might be wrongly counted for inefficiency. Recently, Ahn, Lee and Schmidt (2007) generalized the specification of inefficiency by imposing a factor structure. Since factor structure can also capture unobserved common correlated effects, treating all effects from factor structure as inefficiency can also be unreasonable. For example, we may conclude that some local and small banks suffering less from financial shocks are more efficient than international banks. The alternative approach, adopted the Aigner, Lovell and Schmidt's model (1977) to panel data, has distributional assumptions but allows the inefficiency to vary over time (Battese

[1] To solve the endogeneity problem, Olley and Pakes (1996) and Levinshon and Petrin (2003) showed that investment and intermediate goods can be used as the proxies of these unobserved state variables, however, may not be valid in the cost function analysis.

and Coelli (1988), Kumbhakar (1990), and Wang and Schmidt (2002)). Greene (2003, 2005) and Wang and Ho (2010) further modified this approach to overcome the identification problem between individual effects and inefficiency via Maximum-Likelihood estimation. Nevertheless, to our best knowledge, no previous study has taken cross-sectional dependence into account along with this approach.

In this paper, we propose to incorporate the unobserved correlated common shocks with stochastic frontier models to capture cross-sectional dependence and try to identify the inefficiency, and unobserved common correlated effects (and time-invariant across unit heterogeneity). Following the spirit of Pesaran (2006), our estimation involves a transformation to filter out the unobserved common shocks and then to estimate slope coefficients and parameters in inefficiency function by maximizing the log-likelihood function. There are four features of our method. First, our method inherits the advantage of transformation proposed by Pearsan (2006), by which we can consistently estimate the parameters in the model without explicitly estimating the common correlated effects or factor structure. Moreover, while the asymptotic requirements of sample size for measuring inefficiency in Ahn, Lee and Schmidt (2007)[2] is large N and T, it is only need large N and fixed T in our method.[3] Second, we use the scale function proposed by Wang and Schmidt (2002) to explain the inefficiency. Such specification enables us to directly investigate the underlying determinants of inefficiency and to obtain meaningful policy inferences to improve efficiency. Third, we can estimate the inefficiency spotlessly. In other words, inefficiency can be identified from unobserved common shocks which might not explain inefficiency. Fourth, this method is still valid regardless of the presence of fixed effects or common correlated effects in the model. In addition, this method can be applied to cost inefficiency analysis. Furthermore, we investigate the small sample property via Monte Carlo simulation. We compare the bias from the estimation procedure proposed by Wang and Ho (2002) and our method. Simulation results show that our proposed method outperforms Wang and Ho (2002) when model exists the unobserved common shocks. The bias

[2]Both Pesaran's (2006) and Ahn, Lee and Schmidt's (2006) methods are based on the assumption that the number of factors is less than the number of regressors.

[3]Similar restriction of sample size will arise if we adopt the method by Bai (2009).

is often quite small in our method.

We also apply our approach to analyze the inefficiency of the banking industry in OECD countries. In recent years, research on the variations in bank efficiency has expanded, (see, for example, Lensink, Meesters and Naaborg (2008) and Sun and Change (2010)). While these studies discuss bank efficiency in two different ways, they do not deal with common correlated effects. In contrast, our empirical application focuses on the bank efficiency after filtering out common correlated effects by our proposed approach.

The remainder of this paper is organized as follows. In Section 2, we describe the setup of the stochastic frontier model with common effects and discuss the assumptions, our estimation approach, and the asymptotic properties of the proposed estimator. Section 3 studies the small sample properties using Monte Carlo simulations. Our empirical study is discussed in Section 4. Section 5 concludes the paper. The mathematic proofs of the analytical properties are provided in the Appendix.

2 Model, Assumptions and Estimation

2.1 The Model

Consider the following stochastic frontier model with common correlated effects

$$y_{it} = \alpha_i + x_{it}'\beta + \lambda_i' f_t + v_{it} - u_{it}, \qquad i = 1, \dots, N,\ t = 1, \dots, T, \tag{1}$$

$$x_{it} = \mathbf{A}_i + \tau_i' f_t + e_{it} \tag{2}$$

$$u_{it} = h_{it} u_i^* = h(z_{it}'\delta) u_i^*, \tag{3}$$

where y_{it} is the natural logarithm of output of firm i in period t, x_{it} is a $(k \times 1)$ vector of the natural logarithm of inputs in this production system, α_i denotes the unobserved individual effects, the common correlated effects are modeled by the product of f_t, which includes r unobserved common factors, and corresponding factor loadings λ_i, and v_{it} and u_{it} are the idiosyncratic errors and the term which measures inefficiency, respectively. The regressors are formed by equation (2), where τ_i denotes a $(r \times k)$ vector of factor loadings and, therefore, our specification allows not only for cross-sectional dependence but also for the correlation between common factors and regressors. The random variable e_{it} is idiosyncratic error and are mutually independent of v_{it} and u_{it}. Finally, as shown in equation (3), we let u_{it} equal to a positive function $h_{it} = h(z_{it}'\delta)$ times $u_i^* \sim N^+(\mu, \sigma_u^2)$, in which both μ and σ_u^2 do not dependent on observed variables z_{it}. This specification is referred as to scaling property by Wang and Schmidt (2002) and allows us to directly estimate slope coefficients and capture inefficiency in a one-step procedure. [4] More importantly, since f_t is unobserved, it is difficult to check whether all f_t are related to inefficiency, in other words, whether all f_t can be regard as a measure of inefficiency. Instead, our specification that separating common correlated effects and $u_{it} = h(z_{it}'\delta)u_i^*$, enables us to directly investigate the effect of observed variables z_{it} on inefficiency and then obtain meaningful policy inferences to improve efficiency.[5]

The common correlated effects in the above model are mainly used to capture the heterogeneous impacts of unobserved common random shocks, such as a dramatic global economic decline. There is room for further investigation into the assumption of the correlation between x_{it} and λ_i or f_t. While Pesaran (2006) assumed that x_{it} is correlated with f_t alone, Ahn et al. (2006, 2007) assumed x_{it} is correlated with λ_i and then rectify the endogeneity caused by the correlation between regressors x_{it} and the factor structure. However, they retain the ambiguity of the identification of common correlated effects and inefficiency, i.e., treating $\lambda_i' f_t - u_{it}$, based on our specification, as inefficiency. It is also worth noticing that the conventional fixed effects stochastic frontier models proposed by Greene (2005) and Wang and Ho (2010) are special cases of our specification with $f_t = 1$. [6] Obviously, the model without common correlated effects will reduce to the canonical

[4]The detailed features of the scaling property are discussed in Wang and Schmidt (2002).

[5]Here $h(z_{it}'\delta)$ is used to capture idiosyncratic inefficiency. Notice that even though z_{it} might contain both systematic and idiosyncratic components, if z_{it} is not collinear with $\lambda_i' f_t$, we can still use that observed variable to capture the effects of z_{it} on idiosyncratic inefficiency. Particularly, inefficiency may contain both idiosyncratic and systematic components, we can still separate them into systematic part, $\lambda_i^{*'} f_t^*$, and idiosyncratic part $h(z_{it}'\delta)$.

[6]Although the fixed effects model is a special case of the common correlated effects model, without loss of generality, α_i is still treated as a parameter here, which is potential specific-heterogeneity uncorrelated with f_t, v_{it} and u_{it}.

production stochastic frontier model proposed by Aigner et al. (1977).

In the following, we will introduce how to estimate the proposed model with common correlated effects and establish the asymptotic properties under some suitable assumptions.

2.2 Estimation

Since we take account of common correlated effects in our stochastic frontier model, the correlation between the common correlated effects and regressors makes the estimation of our model nontrivial. Here, we construct a transformation to control for common correlated effects and we then apply the maximum likelihood approach to consistently estimate the parameters.

First, we construct a matrix $M_w = I_T - H_w(H_w' H_w)^{-1} H_w'$, where $H_w = (\mathbf{D}, \bar{\mathbf{Y}}_w, \bar{\mathbf{h}}_w \mu_{**})$, $\mathbf{D}$ is a $(T \times 1)$ vector of ones, $\bar{\mathbf{Y}}_w = (\bar{\mathbf{y}}_w, \bar{\mathbf{X}}_w)$ is the cross-sectional average of $(\mathbf{y}_i, \mathbf{X}_i)$ under the weight ω_i, $\bar{\mathbf{h}}_w$ denotes the cross-sectional average of h_{it}, and $\mu_{**} = (\mu + \frac{\phi(\frac{-\mu}{\sigma_u})}{1-\Phi(\frac{-\mu}{\sigma_u})} \sigma_u)$ is the mean of the truncated normal $u_i^* \sim N^+ \left(\mu, \sigma_u^2\right)$. Here, Φ and ϕ represent the cumulative density function and probability density function of a standard normal distribution, respectively. The rank of M_w, which depends on the dimension of $H_w = (\mathbf{D}, \bar{\mathbf{Z}}_w, \bar{\mathbf{h}}_w \mu_{**})$, is $T - \dim(H_w) = T - s$.

We then transform equation (1) by pre-multiplying M_w

$$M_w \mathbf{y}_i = M_w \mathbf{X}_i \beta + M_w \varepsilon_i + M_w \mathbf{F} \lambda_i, \tag{4}$$

where $M_w \varepsilon_i = M_w v_i - M_w u_i$, $M_w v_i \sim N\left(0, \Pi\right)$, $\Pi = \sigma_v^2 M_w$, $M_w u_i = M_w h\left(\mathbf{z}_i' \delta\right) u_i^*$, and $\mathbf{F} = (f_1', f_2', ..., f_T')$ is a $(T \times r)$ matrix. Since M_w is an idempotent matrix, we solve the non-invertible problem with M_w based on the method of Khatri (1968). In addition, following Wang and Ho (2010), we obtain the marginal log-likelihood function for each individual

$$\begin{aligned} \ln L_i = & -\frac{1}{2}\left(T - s\right)\left(\ln\left(2\pi\right) + \ln \sigma_v^2\right) - \frac{1}{2}\left(\varepsilon_i + \mathbf{F}\lambda_i\right)' M_w \Pi^- M_w \left(\varepsilon_i + \mathbf{F}\lambda_i\right) \\ & + \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) + \ln\left(\sigma_* \Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \ln\left(\sigma_u \Phi\left(\frac{\mu}{\sigma_u}\right)\right), \end{aligned} \tag{5}$$

where

$$\mu_* = \frac{\mu/\sigma_u^2 - (\varepsilon_i + \mathbf{F}\lambda_i)' M_w \Pi^- M_w h_i}{h_i' M_w \Pi^- M_w h_i + 1/\sigma_u^2} \tag{6}$$

$$\sigma_*^2 = \frac{1}{h_i' M_w \Pi^- M_w h_i + 1/\sigma_u^2}. \tag{7}$$

To estimate the parameters, we maximize the sum of $\ln L_i$ over individuals.

Notice that the above equations are designed for the production system. For the cost system, the main model should be modified as

$$y_{it} = \alpha_i + x_{it}'\beta + \lambda_i' f_t + v_{it} + u_{it}, \tag{8}$$

where y_{it} now denotes the total cost of firm i in period t. The individual log-likelihood function is similar to equation (5) except that

$$\mu_* = \frac{\mu/\sigma_u^2 + (\varepsilon_i + \mathbf{F}\lambda_i)' M_w \Pi^- M_w h_i}{h_i' M_w \Pi^- M_w h_i + 1/\sigma_u^2}.$$

2.3 The Properties of the Proposed Estimator

In this section, we will first present the proof to show that M_w can filter out common correlated effects as $N \to \infty$ and then show the consistency of the proposed method. To establish the asymptotic properties, the following assumptions are used throughout the paper:

Assumptions:

1. The error structure contains v_{it}, e_{it} and u_i^*, which are distributed independently of each other and of the regressors x_{it}, z_{it}, $\forall\ i, t$. We also assume that

   $$v_{it} \sim N(0, \sigma_v^2)$$
   $$u_i^* \sim N^+(\mu, \sigma_u^2),$$

   where the variances σ_v^2 and σ_u^2 are bounded.

2. The common factors d_t and f_t are covariance stationary with absolute summable autocovariances, distributed independently of v_{it}, e_{it} and u_i^*, $\forall\ i, t$.

3. The unobserved factor loadings λ_i with mean λ and τ_i with mean τ are mutually independent and of v_{it}, e_{it} , u_i^*, and the common factors d_t, f_t, $\forall\ i, t$. In particular, $\|\lambda_i\|$ and $\|\tau_i\|$ are finite with finite second moment.

4. The function of the determinants $h(z_{it}'\delta)$ should be assumed to have finite first and second moments and to be distributed independently of v_{it}, e_{it} and u_i^*.

5. There are cross-sectional weights w_i that satisfy (i)$w_i = O(1/N)$, (ii) $\sum_{i=1}^N w_i = 1$ and (iii) $\sum_{i=1}^N |w_i| < K$. Therefore, the weighted average of the cross-sectional variable can be defined as $\bar{r}_{wt} = \sum_{i=1}^N w_i r_{it}$.

Assumption 1 is a standard distributional assumption for stochastic frontier model. Assumptions 2, 3, 4 and 5 are similar to the assumptions used in Pesaran (2006) for panel models with multi-factor error structures.

In order to show the consistency of our estimator, we first rewrite the stochastic frontier model with common correlated effects in equations (1)–(3) as

$$\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} = \begin{bmatrix} 1 & \beta' \\ 0 & I_k \end{bmatrix} \begin{bmatrix} \alpha_i \\ \mathbf{A}_i \end{bmatrix} d_t + \begin{bmatrix} 1 & \beta' \\ 0 & I_k \end{bmatrix} \begin{bmatrix} \lambda_i' \\ \tau_i' \end{bmatrix} f_t - \begin{bmatrix} u_{it} \\ 0 \end{bmatrix} + \begin{bmatrix} v_{it} + \beta' e_{it} \\ e_{it} \end{bmatrix}$$

$$\Rightarrow \mathbf{Y}_{it} = \mathbf{B}_i' d_t + \mathbf{C}_i' f_t - \mathbf{U}_{it} + \xi_{it}$$

where $\mathbf{A}_i$ is a $k \times 1$ vector of $(\alpha_i, ..., \alpha_i)'$, and $d_t = 1$.

Next, we take the cross-sectional average under the weight w_i, and then we have

$$\Rightarrow \bar{\mathbf{Y}}_{wt} = \bar{\mathbf{B}}_w' d_t + \bar{\mathbf{C}}_w' f_t - \bar{\mathbf{U}}_{wt} + \bar{\xi}_{wt}, \tag{9}$$

where $\bar{\mathbf{U}}_{wt} = (\sum_{i=1}^N w_i h_{it} u_i^*, 0)'$. Following the proof of Pesaran (2006), we obtain $\bar{\xi}_{wt} \xrightarrow{p} 0$ and $\bar{\mathbf{C}}_w \xrightarrow{p} \mathbf{C}$ as $N \to \infty$, where $\mathbf{C} = \begin{bmatrix} \lambda & \tau \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta & I_k \end{bmatrix}$. Then, under the assumption $\text{rank}(\bar{\mathbf{C}}_w) = r \leq k+1$, $\forall i$, we obtain

$$f_t - (\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}(\bar{\mathbf{Y}}_{wt} - \bar{\mathbf{B}}_w' d_t + \bar{\mathbf{U}}_{wt}) \xrightarrow{p} 0 \tag{10}$$

Equation (10) gives us the set $\{\mathbf{D}, \bar{\mathbf{y}}_w, \bar{\mathbf{X}}_w, \bar{\mathbf{u}}_w\}$ which can be regarded as the proxies of the factor structure. Here, this result still hold even we relax the assumption that the term to capture inefficiency $h_{it} = h(z_{it}'\delta)$ is uncorrelated with f_t.[7] Notice that $\bar{\mathbf{U}}_{wt}$ contains the average of the inefficiency terms, u_i^*, which is unobserved in the stochastic frontier model. Hence, it cannot be directly applied to represent the factor structure. However, under Assumptions 1-4, Lemma 1 in appendix shows that $\sum_{i=1}^N w_i h_{it} u_i^*$ converges to $\sum_{i=1}^N w_i h_{it}\mu_{**} = \bar{h}_{wt}\mu_{**}$ as $N \to \infty$. Therefore, equation (10) can be replaced by

$$f_t - (\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}\left(\bar{\mathbf{Y}}_{wt} - \bar{\mathbf{B}}_w' d_t + \begin{bmatrix} \bar{h}_{wt}\mu_{**} \\ 0 \end{bmatrix}\right) \xrightarrow{p} 0 \tag{11}$$

as $N \to \infty$. That is the reason M_w can be constructed by $I_T - H_w(H_w'H_w)^{-1}H_w'$ with $H_w = (\mathbf{D}, \bar{\mathbf{Y}}_w, \bar{\mathbf{h}}_w\mu_{**})$. This result indicates that common correlated effects in the stochastic frontier model will be eliminated as $N \to \infty$ after taking the transformation.

Due to the above property, we provided the following proposition to show that the marginal log-likelihood function in equation (5) will asymptotic to the marginal log-likelihood function using the transform matrix M_w^*, where $M_w^* = I_T - H_w^*(H_w^{*'}H_w^*)^{-1}H_w^{*'}$, where $H_w^* = (\mathbf{D}, \bar{\mathbf{Y}}_w, \bar{\mathbf{h}}_w u_i^*)$.[8]

Proposition 1. *Under Assumptions 1–5,*

$$\frac{\sum_{i=1}^N w_i \ln L_i}{T} = \frac{\sum_{i=1}^N w_i \ln L_i^*}{T} + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Thus, the maximum likelihood method can therefore consistently estimate the parameters under fixed T and large N.

Compared with the GMM procedure proposed by Ahn et al. (2006), our estimation has two desirable properties. First, we can just focus on z_{it} that is concerned with measuring

[7]The positive function $h(\cdot)$ makes us have the difficulty to rewrite h_{it} as a linear factor structure and have a matrix representation with equations (1)–(3). But equation (10) still hold, the sacrifice is the limit of number of factors still can not greater than $k + 1$.

[8]Note that the identification of δ in the positive function $h(z_{it}'\delta)$ requires the full column rank condition of $\left(\mathbf{D}, \bar{\mathbf{X}}_w, \bar{\mathbf{h}}_w u_i^*\right)$, that is the condition rules out the perfect multi-collinearity situation. The first and obvious case is that the function $h(\cdot)$ contains only one time-invariant regressor to explain inefficiency in order that the transform matrix will orthogonal to $\bar{\mathbf{h}}_w$. The second case is that $\bar{\mathbf{h}}_w$ is a linear combination of $\bar{\mathbf{X}}_w$.

inefficiency and treat other unobserved inefficiencies as part of the common correlated effects which can be filtered out by our transformation. However, the GMM method lets inefficiency and common correlated effects into a mess. Second, our estimates are suitable for fixed-T-and-large-N and large-T-large-N panel data. In the latter case, however, the numbers of parameters and many instruments in the GMM method increase with T. In addition, the recovery of the parameters is involved in the GMM procedure.

2.4 The Inefficiency Index

It is important to measure the inefficiency index in applications. However, how can the inefficiency index be estimated after the CCE[9] transformation? To obtain the solution, we follow by Wang and Ho (2010), who use the conditional expectation estimator proposed by Jondrow et al. (1982), namely, $E(u_i|\varepsilon_i)$ evaluated at $\varepsilon_i = \hat{\varepsilon}_i$. In the same manner, the inefficiency index in our estimation is the conditional expectation of u_{it} on the vector of the transformed $\varepsilon_i = \mathbf{v}_i - \mathbf{u}_i$, i.e., $M_w\varepsilon_i$. Note that $M_w\varepsilon_i$ can be evaluated at $\widehat{M_w\varepsilon_i}$, and following Wang and Ho (2010), the conditional inefficiency index is

$$E\left(u_{it}|M_w\varepsilon_i\right) = h(z_{it}'\delta)\left[\mu_* + \frac{\phi\left(\frac{\mu_*}{\sigma_*}\right)\sigma_*}{\Phi\left(\frac{\mu_*}{\sigma_*}\right)}\right] \tag{12}$$

3 Monte Carlo Simulations

In this section, we use Monte Carlo simulations to investigate the finite sample properties of our proposed estimator. We first consider the following stochastic production frontier model for $i = 1, \ldots, N$ and $t = 1, \ldots, T$:

$$y_{it} = \alpha_i + x_{it}\beta + \lambda_i' f_t + v_{it} - \exp(z_{it}'\delta)u_i^* \tag{13}$$

$$x_{it} = \mathbf{A}_i + \tau_i' f_t + e_{it}, \tag{14}$$

where $\alpha_i \sim U(0,1)$, x_{it} is a regressor, $f_t \sim N(0, \sigma_f)$ is a common factor, $\sigma_f = 0.2$, factor loadings λ_i and τ_i follow $N(1, 0.2)$, z_{it} consists of $z_{it,1} \sim N(0,1)$ and $z_{it,2} = t$, which

[9]In brief, we denote CCE as the abbreviation for the common correlated effects/common correlated effects transformation.

impleis the inefficiency is time-varying, $v_{it} \sim N(0, \sigma_v^2)$, $u_i^* \sim N^+(0, \sigma_u^2)$, v_{it} and u_i^* are mutually independent, and $e_{it} \sim N(0,1)$. The parameter values are

$$(\beta, \delta_1, \delta_2, \sigma_v^2, \sigma_u^2, \mu) = (0.5, 0.5, 0.1, 0.1, 0.2, 0.5).$$

$N = \{50, 100, 200, 400\}$, $T = \{5, 10, 20\}$, and the number of replications is 1,000 in all simulations.

To demonstrate the importance of our transformation in the presence of common correlated effects, we also compared our method with the estimation proposed by Wang and Ho (2010), which only takes the fixed effects into account by means of the within transformation. Hereafter, we let Within denote the latter method and let CCE denote our estimator.

Our simulation results are reported in Table 1. As we can see, CCE tends to have a smaller bias than Within for all parameters over all combinations of (N, T) except δ_2 when $T = 5$. Moreover, CCE uniformly has a smaller RMSE than Within as $T \geq 10$. Even when $T = 5$, the RMSE ratios, ψ =RMSE(Within)/RMSE(CCE), increase with the increase in N. For example, the ψ of $\hat{\delta}$ is 0.614 when $(N, T) = (50, 5)$ and increases to 1.036, which indicates that CCE has a smaller RMSE than Within by 3.6%, when $(N, T) = (50, 5)$. It is also worth noting that the bias and the RMSE of CCE decline as T or N increases for all parameters. By contrast, due to failing to control for the common correlated effects, the Within estimators of β and δ are still biased and cannot be improved even when T or N is large.

For robustness, we further consider the finite sample performance for different degrees of cross-sectional correlation by adjusting the magnitude of σ_f. In particular, we consider two settings with $\sigma_f^2 = 0.1$ and 1, respectively. As we can see from model (1), when σ_f is smaller, our model is closer to the model with fixed effects only and the common correlated effects become less important. Furthermore, instead of letting $z_{it,2} = t$ in $h(z_{it}'\delta)$, we consider group-specific inefficiency by letting $z_{it,2}$ be a group dummy such that $z_{it,2} = 1$ for any unit in Group 2; otherwise $z_{it,2} = 0$. The members in Group 1 are randomly assigned in each repetition with the number of units $N_1 = \lfloor U(0.3, 0.7) \times N \rfloor$, regardless of whether $\lfloor A \rfloor$ is the integer closest to A. The other group has $N - N_1$ units.

The group membership is known in advance. The parameters in this set of simulations take the following values

$$(\beta, \delta_1, \delta_2, \sigma_v^2, \sigma_u^2, \mu) = (0.5, 0.5, 0.1, 0.1, 0.2, 0.5).$$

The results are summarized in Tables 2 and 3 with $T = \{10, 20\}$, respectively. Since we have similar patterns to the previous simulation, that is the bias and the RMSE of CCE decline as T or N increases, we do not report the case when $T = 5$. Details regarding the results of $T = 5$ can be obtained from the authors on request. It will be clear from these results that the bias for Within seems to be less serious as $\sigma_f^2 = 0.1$, and becomes more significant as $\sigma_f^2 = 1$. More importantly, the performance of our approach is generally better than Within approach even when $\sigma_f^2 = 0.1$, which demonstrates that our method is still robust even when the common correlated effects are small in the data. In particular, the estimates of σ_v^2 and σ_u^2 for the Within approach seem to be overestimated in the presence of the common correlated effects. On the contrary, CCE provides less unbiased estimates even when $\sigma_f^2 = 0.1$.

We next turn to the experiment which takes account of x_{it} and u_{it} are both correlated with the factors. In this simulation, it is convenient to set $u_{it} = \exp(\mathbf{z_{it}}'\delta)u_i^*$ to ensure u_{it} is positive and assume

$$\mathbf{z}_{it} = \pi_i' f_t + \mathbf{e}_{z,it}, \tag{15}$$

to make u_{it} is correlated with f_t. We still have two variables $z_{1,it}$ and $z_{2,it}$ which can affect u_{it}. Particularly, the factor loadings $\pi_{i,1}$ and $\pi_{i,2}$ follow $N(1, 0.4)$ and $N(1, 0.2)$ respectively, $f_t \sim N(0, 0.6)$ to let factor is important in this model, and each of $\mathbf{e}_{z,it}$ follows $N(0, 1)$. x_{it} is similar to the former setting. The parameters in this set of simulations take the following values

$$(\beta, \delta_1, \delta_2, \sigma_v^2, \sigma_u^2, \mu) = (0.5, 0.2, -0.1, 0.1, 0.1, 0.4).$$

Table 4 summarizes the simulation results. A general finding is that our proposed method is relatively much better than Within in all combinations. The bias is almost 0 in CCE except σ_u^2, whereas the bias of Within are serious not only in β but also δ's. Notice that the small bias of σ_u^2 in CCE will decrease as N increasing. On the contrary, the bias of σ_u^2 in Within is enormous, and it is not surprising because Within do not control the common correlated effects, and the components from the biased $\hat{h}_{it}$ will induce large variation of u_i^*.

In general, the simulation shows the clear results that the estimation without control common correlated effects will bias the estimates. We also conduct a similar simulation for the cost frontier model, which is not reported here. Its pattern again confirms the importance of taking the common correlated effects into account in a stochastic frontier model and are similar to the findings summarized in Tables 1–4. Due to space limitations, these results are available from the authors upon request.

4 Empirical Study

The existing body of research on "bank efficiency" has grown rapidly. Lensink et al. (2008), Berger, Hasan and Zhou (2009) and Sun and Chang (2010) provide different aspects to measure bank efficiency. However, it is not clear how these aspects determine efficiency when common correlated effects are taken into account. Our empirical study therefore applies an approach that uses the proposed CCE transformation to deal with common correlated effects even when the efficiency terms are directly unobserved.

4.1 Data

We evaluate the cost efficiency in OECD countries by using the proposed transformation allowing for the common correlated effects in the stochastic frontier model. The conventional intermediation approach to measuring the cost faced by a bank is used in this study. Total cost is defined as the sum of interest expense and non-interest expense. Following Berger et al. (2009) and Sun and Chang (2010), we consider the following output variables in the cost function: total loans (TL), other earning assets (OEA), total deposits (TD) and liquid assets (LA). We additionally consider the price of capital (PC) and funds (PF), defined by the ratio of interest expenses to total deposits and the ratio of non-interest expenses to total fixed assets, respectively, as our input prices. In order to guarantee linear homogeneity in input prices of the cost function, we re-scale TC and PC by PF.

The cost function used here is

$$\ln\left(\frac{\text{TC}}{PF}\right)_{it} = \beta_0 \ln\left(\frac{\text{PC}}{PF}\right)_{it} + \beta_1 \ln \text{TL}_{it} + \beta_2 \ln \text{OEA}_{it} + \beta_3 \ln \text{TD}_{it} + \beta_4 \ln \text{LA}_{it} + \lambda_i f_t + v_{it} + u_{it}. \tag{16}$$

To allow the inefficiency across banks to be measured by explanatory variables, we use the scaling function proposed by Wang (2002). The specification of the scaling function is as follows

$$h(z'_{it}\delta) = \exp(\delta_1 \ln \text{TA} + \delta_2 \text{ETA} + \delta_3 \text{ROAA} + \delta_4 \text{Year 2008} + \text{Country dummy}), \tag{17}$$

where TA denotes the total assets, ETA denotes the equity to assets, and ROAA denotes the return on average assets. These three variables are commonly used to control the efficiency. TA measures the relationship between the efficiency and the size of the bank. ETA can represent the equity position of a bank and avoid the scale bias making large banks more efficient (Berger and Mester, 1997). In addition, ETA may reflect the risk preference of a manager of a bank. ROAA can be regard as a proxy for manager ability. A year dummy variable is also included to capture the averaged effect of the global crisis in 2008 across banks. Furthermore, we also put the country dummy in the scaling function to measure the different inefficiency in these countries.

We consider a balanced panel data set covering 1996-2009 with 311 commercial banks from nine countries: Austria, Belgium, Canada, Denmark, France, Germany, Switzerland, the United Kingdom, and the United States. The data are taken from Bankscope and are inflation-adjusted. Except for ETA and ROAA, all the other variables are transformed into natural logs. Table 5 presents the descriptive statistics of these variables.

4.2 Empirical Results

The empirical results obtained by our approaches are summarized in the right panel of Table 6. We report not only the estimates of the coefficients in the cost function β's, but also the estimates of the parameters in the inefficiency equation δ's. For comparison, we additionally show the results based on the Within approach proposed by Wang and Ho (2010) in the left panel of Table 6.

Consider the coefficients in the cost function using our approach first. The coefficient of the input prices (PC/PF) is positive at the 1% significance level, which indicates that a higher capital cost results in a higher total cost and is similar to the empirical results of Lensink et al. (2008) and Sun and Chang (2010). As expected, the output variables, such as TL, TD and LA, also have positive effects on the total cost. While the estimated coefficient of OEA is negative, it is not significantly different from zero. The empirical results from the Within approach are qualitatively similar to those based on our CCE approach. However, the former tends to deliver larger estimated coefficients of PC/PF and TL than our approach.

Next, we turn our focus to the coefficients of the inefficiency equation. The coefficient for TA, equal to -0.054, is negative and significant at the 1% level, which implies that larger banks are on average more efficient than smaller banks as TA is regarded as a proxy for the bank size. The estimated sign of this coefficient is different from that in Han, Orea and Schmidt (2005) and Sun and Chang (2010). However, Delis and Papanikolaou (2009) pointed out that the relationship between bank size and efficiency is inverse U-shaped, which implies that the efficiency increases with size and then decreases thereafter. In our data, almost 90% of banks are small and medium sized and, therefore, are more likely to have a positive relationship with efficiency.[10] In addition, our results indicate that an increase in ETA will raise inefficiency, which can be explained in two ways. First, ETA can be regarded as a proxy for the risk-preference of a manager. A higher equity position reveals that the manager is risk-averse and might not be good at using financial leverage to increase the size of bank, which indicates that the manager may not seek to minimize the cost. Second, inefficiency will lead to a lower profit and put equity in a high position.

The coefficient of the bank's ROAA is negative at the 1% significance level, which implies that banks with a higher ROAA are generally more efficient than those with a

[10]Following Berger et al. (2009), the classification of bank size is defined as follows. The bank's size is small if its assets are less than or equal to $1 billion, its size is medium if the bank's assets are greater than $1 billion but less than $20 billion, and the bank is large if its assets are greater than $20 billion.

lower ROAA. Our result is consistent with Lensink et al. (2008).[11] Furthermore, the year 2008 dummy variable for capturing the financial crisis has positive but small effect on efficiency based on our approach, which might be potentially due to the fact that other variables already reflect the financial crisis, and the crisis is due to the financial institutions' highly leveraged behavior and increasing risky investment before 2008. However, in late 2008, financial institutions try to survive on this crisis under rigorous management even the emergency bailout was proposed. By comparing with other years, the rigorous management may provide a small positive effect on efficiency in 2008.

Comparing the results from different approaches further reflects the importance of controlling the common correlated effects in the frontier model. Notice that our CCE approach is consistent and has satisfactory finite sample performance even when there do not exist any or only small cross-sectional correlation effects as shown in the previous sections. Thus, the large estimated value of σ_u^2 based on the Within approach appears to reflect the fact that ignoring common correlated effects might result in higher uncertainty in the inefficiency term.

Finally, to understand bank efficiency across countries, we also take account of the country dummy to capture the country effect. The results are quite different in these two approaches, for example, the negative coefficients of Canada, Denmark and UK in CCE imply these countries are more efficient. However, these coefficients in Within model are positive. We also provide the conditional cost efficiency, defined as the average cost efficiency in each country in Figure 1 and 2. The conditional efficiency in Within model reveals that Canada and Denmark are less efficient. On the contrary, in CCE model, it seems that banks in Belgium are relatively efficient while those in France are relatively inefficient. Furthermore, the index are close to 0 without taking account of the unobserved common shock, and it is intuitive if these common shock are regard as inefficiency. Figure 3 further illustrates the efficiency pattern over time in each country.[12] By contrast, bank

[11] This result is different from that of Sun and Chang (2010). While it might be caused by endogeneity, the ROAA should exhibit a negative relationship with inefficiency as pointed out by Lensink et al. (2008).

[12] Since the time-varying variables' coefficients are close to 0 in Within model and have the flat pattern of efficiency, we do not report here.

efficiencies in most countries tend to improve over time before 2006, and dramatically decline in every country during the global crisis in 2008.

5 Concluding Remark

Many studies are conducted to reveal the fact that it is important to distinguish fixed effects from inefficiency. However, such research fails to consider the possibility that the specific-heterogeneity can be regarded as common correlated effects. This paper therefore provides a stochastic model with the incorporation of the factor structure and adopts the method proposed by Pesaran (2006) to eliminate the factor structure. The factor structure can be eliminated as long as the cross-sectional dimension is sufficiently large. With this transformation, we can use the observed variables to explain the inefficiency and directly estimate the inefficiency which is not influenced by the unobserved factor structure. Since the inefficiency is conditional upon the estimated residuals, our approach can provide a reliable result in the estimation of inefficiency.

Table 1: Simulation results with cross-section dependence

	$T=5$					$T=10$					$T=20$				
	With-in		CCE			Within		CCE			Within		CCE		
$N=50$	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.125	0.150	-0.002	0.058	2.596	0.146	0.159	0.000	0.021	7.695	0.155	0.162	0.000	0.012	13.170
$\hat{\delta_1}$	-0.010	0.127	0.008	0.208	0.614	-0.002	0.080	-0.002	0.060	1.335	0.000	0.025	0.000	0.015	1.683
$\hat{\delta_2}$	0.002	0.095	0.032	0.122	0.778	-0.002	0.021	0.001	0.013	1.565	0.000	0.005	0.000	0.002	2.729
$\hat{\sigma}_v^2$	0.166	0.202	-0.013	0.030	6.663	0.191	0.209	0.000	0.009	23.053	0.199	0.209	0.006	0.009	23.839
$\hat{\sigma}_u^2$	0.049	0.239	0.039	0.279	0.856	0.031	0.159	0.007	0.116	1.372	0.006	0.086	-0.003	0.070	1.232
$\hat{\mu}$	0.068	0.263	0.014	0.285	0.924	0.020	0.208	-0.001	0.154	1.347	-0.007	0.137	-0.002	0.113	1.221
	Within		CCE			Within		CCE			Within		CCE		
$N=100$	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.129	0.155	0.000	0.040	3.921	0.147	0.159	0.000	0.014	11.573	0.154	0.161	0.000	0.008	19.771
$\hat{\delta_1}$	-0.027	0.109	-0.005	0.147	0.739	-0.002	0.071	0.001	0.039	1.800	0.001	0.023	0.000	0.010	2.203
$\hat{\delta_2}$	-0.006	0.086	0.020	0.095	0.903	-0.002	0.019	0.000	0.010	1.906	0.000	0.005	0.000	0.001	3.499
$\hat{\sigma}_v^2$	0.177	0.214	-0.009	0.022	9.859	0.194	0.211	0.000	0.006	33.560	0.201	0.210	0.003	0.005	40.385
$\hat{\sigma}_u^2$	0.060	0.218	0.059	0.256	0.853	0.019	0.111	0.003	0.073	1.514	0.005	0.069	-0.003	0.051	1.348
$\hat{\mu}$	0.096	0.231	-0.004	0.240	0.963	0.026	0.173	0.004	0.106	1.642	-0.003	0.111	-0.001	0.079	1.412

(continued)

	Within		CCE			Within		CCE			Within		CCE		
$N = 200$	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.131	0.153	0.001	0.028	5.409	0.147	0.159	0.000	0.010	16.189	0.154	0.160	0.000	0.006	28.865
$\hat{\delta_1}$	-0.026	0.094	-0.007	0.105	0.903	0.004	0.062	0.001	0.030	2.100	0.002	0.023	0.000	0.007	3.260
$\hat{\delta_2}$	-0.006	0.078	0.010	0.078	0.998	-0.003	0.018	0.000	0.007	2.478	0.000	0.005	0.000	0.001	4.584
$\hat{\sigma}_v^2$	0.179	0.212	-0.005	0.015	13.772	0.195	0.212	0.000	0.004	48.627	0.200	0.209	0.002	0.003	63.266
$\hat{\sigma}_u^2$	0.051	0.185	0.061	0.216	0.853	0.015	0.093	0.003	0.055	1.708	0.002	0.056	-0.002	0.036	1.548
$\hat{\mu}$	0.087	0.202	-0.015	0.196	1.027	0.009	0.147	-0.003	0.076	1.944	-0.003	0.093	0.001	0.057	1.630
	Within		CCE			Within		CCE			Within		CCE		
$N = 400$	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.126	0.148	0.000	0.019	7.817	0.147	0.158	0.000	0.007	23.143	0.155	0.161	0.000	0.004	40.098
$\hat{\delta_1}$	-0.026	0.085	-0.003	0.082	1.036	0.000	0.059	0.001	0.021	2.794	0.000	0.022	0.000	0.005	4.404
$\hat{\delta_2}$	-0.005	0.076	0.010	0.073	1.032	-0.002	0.017	0.000	0.006	3.025	0.000	0.005	0.000	0.001	5.839
$\hat{\sigma}_v^2$	0.173	0.205	-0.004	0.011	18.678	0.194	0.211	0.000	0.003	67.751	0.202	0.210	0.001	0.002	105.087
$\hat{\sigma}_u^2$	0.044	0.152	0.043	0.175	0.868	0.011	0.084	0.000	0.036	2.319	0.002	0.050	-0.002	0.026	1.895
$\hat{\mu}$	0.082	0.180	-0.028	0.159	1.131	0.015	0.132	-0.005	0.052	2.521	0.006	0.080	0.002	0.043	1.849

[1] In brief, we denote Within as the abbreviation of the within-transformation and CCE as the abbreviation for the common correlated effects transformation.
[2] ψ is the ratio of RMSE(Within)/RMSE(CCE).
[3] The true values of the parameter set are $\beta = 0.5, \delta_1 = 0.5, \delta_2 = 0.1, \sigma_v^2 = 0.1, \sigma_u^2 = 0.2$, and $\mu = 0.5$.

Table 2: Simulation results with cross-sectional dependence under different σ_f (T=10)

	$\sigma_f^2 = 0.1$					$\sigma_f^2 = 1$				
$N = 50$	Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.087	0.097	-0.001	0.020	4.963	0.433	0.446	-0.000	0.019	24.016
$\hat{\delta_1}$	0.001	0.075	0.006	0.077	0.971	0.014	0.126	0.002	0.074	1.694
$\hat{\delta_2}$	0.001	0.200	-0.003	0.232	0.862	0.015	0.280	0.004	0.216	1.299
$\hat{\sigma}_v^2$	0.109	0.119	-0.001	0.009	13.655	0.592	0.604	-0.001	0.009	67.373
$\hat{\sigma}_u^2$	0.017	0.152	0.009	0.151	1.008	0.077	0.258	0.017	0.158	1.629
$\hat{\mu}$	0.007	0.182	0.010	0.181	1.006	-0.038	0.226	-0.003	0.177	1.272
$N = 100$	Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.089	0.098	-0.000	0.014	7.041	0.427	0.439	-0.001	0.014	31.009
$\hat{\delta_1}$	0.001	0.052	0.001	0.050	1.037	0.002	0.096	-0.001	0.051	1.871
$\hat{\delta_2}$	0.009	0.130	0.001	0.162	0.805	0.002	0.182	0.004	0.162	1.125
$\hat{\sigma}_v^2$	0.112	0.123	-0.000	0.006	19.624	0.596	0.607	-0.000	0.006	98.686
$\hat{\sigma}_u^2$	0.009	0.103	0.009	0.105	0.982	0.082	0.229	0.011	0.107	2.143
$\hat{\mu}$	-0.009	0.128	-0.001	0.125	1.026	-0.040	0.192	-0.004	0.136	1.412
$N = 200$	Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.089	0.096	-0.000	0.009	10.234	0.430	0.441	0.000	0.010	45.506
$\hat{\delta_1}$	0.002	0.037	0.001	0.037	0.995	0.001	0.068	-0.002	0.037	1.845
$\hat{\delta_2}$	0.003	0.094	0.008	0.114	0.824	-0.007	0.133	-0.003	0.115	1.154
$\hat{\sigma}_v^2$	0.111	0.121	0.000	0.004	28.167	0.597	0.608	-0.000	0.004	135.998
$\hat{\sigma}_u^2$	0.005	0.068	0.002	0.070	0.978	0.083	0.195	0.009	0.076	2.560
$\hat{\mu}$	-0.008	0.088	-0.003	0.089	0.987	-0.060	0.152	0.003	0.087	1.754
$N = 400$	Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.088	0.094	-0.000	0.007	13.411	0.426	0.438	0.000	0.007	65.428
$\hat{\delta_1}$	0.003	0.026	0.001	0.024	1.068	-0.003	0.049	0.002	0.025	1.980
$\hat{\delta_2}$	0.002	0.067	-0.000	0.079	0.843	-0.007	0.094	0.004	0.077	1.229
$\hat{\sigma}_v^2$	0.109	0.119	-0.000	0.003	38.680	0.594	0.606	-0.000	0.003	193.373
$\hat{\sigma}_u^2$	0.002	0.049	0.000	0.043	1.124	0.086	0.178	0.003	0.046	3.831
$\hat{\mu}$	-0.009	0.062	-0.002	0.063	0.986	-0.060	0.119	-0.008	0.060	1.963

[1] ψ is the ratio of RMSE(Within)/RMSE(CCE).
[2] The true values of the parameter set are $\beta = 0.5, \delta_1 = 0.5, \delta_2 = 0.5, \sigma_v^2 = 0.1, \sigma_u^2 = 0.2$, and $\mu = 0.5$.

Table 3: Simulation results with cross-sectional dependence under different σ_f (T=20)

	$\sigma_f^2 = 0.1$					$\sigma_f^2 = 1$				
$N = 50$	Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.089	0.094	0.000	0.011	8.272	0.447	0.453	0.000	0.012	38.062
$\hat{\delta_1}$	-0.002	0.044	-0.003	0.035	1.257	-0.000	0.089	-0.002	0.038	2.332
$\hat{\delta_2}$	0.000	0.171	-0.002	0.194	0.885	0.002	0.209	-0.006	0.193	1.084
$\hat{\sigma}_v^2$	0.110	0.116	-0.000	0.005	22.063	0.626	0.631	-0.000	0.005	122.391
$\hat{\sigma}_u^2$	0.010	0.110	-0.001	0.101	1.096	0.054	0.207	-0.000	0.102	2.030
$\hat{\mu}$	0.001	0.141	0.015	0.130	1.080	-0.013	0.171	0.011	0.131	1.310
$N = 100$	Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.089	0.093	0.000	0.008	11.114	0.443	0.448	0.000	0.008	53.979
$\hat{\delta_1}$	0.000	0.033	0.000	0.025	1.299	0.000	0.062	0.001	0.025	2.512
$\hat{\delta_2}$	0.002	0.118	-0.003	0.135	0.875	-0.008	0.140	-0.006	0.133	1.055
$\hat{\sigma}_v^2$	0.110	0.116	-0.000	0.004	31.880	0.629	0.635	-0.000	0.004	169.412
$\hat{\sigma}_u^2$	0.003	0.078	-0.002	0.069	1.127	0.041	0.148	-0.002	0.069	2.133
$\hat{\mu}$	-0.004	0.094	0.003	0.091	1.037	-0.018	0.125	0.004	0.087	1.428
$N = 200$	Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.088	0.092	-0.000	0.006	16.425	0.442	0.447	-0.000	0.006	77.077
$\hat{\delta_1}$	-0.000	0.023	-0.000	0.017	1.314	-0.001	0.044	-0.000	0.017	2.558
$\hat{\delta_2}$	0.003	0.085	0.004	0.090	0.940	-0.003	0.101	-0.001	0.094	1.081
$\hat{\sigma}_v^2$	0.110	0.115	-0.000	0.003	44.603	0.631	0.636	0.000	0.003	244.367
$\hat{\sigma}_u^2$	0.005	0.052	-0.001	0.045	1.165	0.040	0.116	0.000	0.049	2.383
$\hat{\mu}$	-0.008	0.057	-0.001	0.053	1.067	-0.028	0.087	0.002	0.053	1.644
$N = 400$	Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.088	0.091	-0.000	0.004	22.551	0.444	0.449	-0.000	0.004	107.029
$\hat{\delta_1}$	0.001	0.015	-0.000	0.012	1.291	0.001	0.030	0.000	0.013	2.374
$\hat{\delta_2}$	0.007	0.059	0.004	0.065	0.919	-0.003	0.075	-0.002	0.066	1.126
$\hat{\sigma}_v^2$	0.109	0.114	-0.000	0.002	62.508	0.635	0.639	0.000	0.002	354.291
$\hat{\sigma}_u^2$	0.002	0.035	0.001	0.032	1.097	0.031	0.082	0.001	0.033	2.524
$\hat{\mu}$	-0.010	0.040	-0.002	0.035	1.140	-0.031	0.067	-0.002	0.037	1.815

[1] ψ is the ratio of RMSE(Within)/RMSE(CCE).
[2] The true values of the parameter set are $\beta = 0.5, \delta_1 = 0.5, \delta_2 = 0.5, \sigma_v^2 = 0.1, \sigma_u^2 = 0.2$, and $\mu = 0.5$.
[3] The bias is defined by (Estimated value − True Value).

Table 4: Simulation results with cross-section dependence, case about x and z are correlated with factor (T=20)

	Within		CCE		
$N = 50$	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.292	0.534	0.000	0.012	45.393
$\hat{\delta_1}$	-0.139	0.431	0.001	0.066	6.514
$\hat{\delta_2}$	0.102	0.373	0.001	0.037	10.198
$\hat{\sigma}_v^2$	1.507	9.953	-0.001	0.005	1837.170
$\hat{\sigma}_u^2$	34721.845	68882.385	1.412	13.975	4928.831
$\hat{\mu}$	0.038	0.225	0.036	0.234	0.959
	Within		CCE		
$N = 100$	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.272	0.446	0.000	0.008	54.281
$\hat{\delta_1}$	-0.140	0.342	0.001	0.054	6.370
$\hat{\delta_2}$	0.089	0.304	0.000	0.028	10.693
$\hat{\sigma}_v^2$	0.917	3.523	-0.001	0.004	968.563
$\hat{\sigma}_u^2$	39725.677	76044.209	0.180	0.866	87785.493
$\hat{\mu}$	0.054	0.214	0.008	0.174	1.227
	Within		CCE		
$N = 200$	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.287	0.486	-0.000	0.006	85.221
$\hat{\delta_1}$	-0.124	0.372	0.003	0.039	9.653
$\hat{\delta_2}$	0.093	0.288	-0.001	0.020	14.582
$\hat{\sigma}_v^2$	1.135	5.096	-0.000	0.003	2004.083
$\hat{\sigma}_u^2$	32871.244	65416.402	0.078	0.444	147318.408
$\hat{\mu}$	0.039	0.213	-0.011	0.106	2.004
	Within		CCE		
$N = 400$	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.249	0.390	-0.000	0.004	95.194
$\hat{\delta_1}$	-0.136	0.349	-0.000	0.027	12.743
$\hat{\delta_2}$	0.098	0.230	0.000	0.014	16.608
$\hat{\sigma}_v^2$	0.856	2.493	-0.000	0.002	1348.341
$\hat{\sigma}_u^2$	40286.341	76224.203	0.050	0.319	238843.857
$\hat{\mu}$	0.048	0.213	-0.007	0.050	4.282

[1] ψ is the ratio of RMSE(Within)/RMSE(CCE).

[2] The true values of the parameter set are $\beta = 0.5, \delta_1 = 0.2, \delta_2 = -0.1, \sigma_v^2 = 0.1, \sigma_u^2 = 0.4$, and $\mu = 0.5$.

[3] The bias is defined by (Estimated value − True Value).

Table 5: Statistics of variables used in the cost function

Variables	Mean	Std. Dev.	Min	Max
Total Cost	1.22×10^3	5.24×10^3	3	7.49×10^4
Output quantities				
Total loans	1.23×10^4	5.29×10^4	8	8.61×10^5
Other earning assets	9.96×10^3	5.62×10^4	3	9.68×10^5
Total deposits	1.34×10^4	6.35×10^4	4	1.06×10^6
Liquid assets	5.93×10^4	3.56×10^4	2	7.80×10^5
Input prices				
Price of capital	5.51	15.77	0.17	370.17
Price of funds	0.27	1.97	1.89×10^{-3}	44.44
Other variables' quantity and ratios				
Total assets	2.67×10^4	1.37×10^5	44	3.13×10^6
Return on average assets	1.10	1.27	-16.9	17.2
Equity to assets	9.46	5.07	-3.5	62.95

[1] The variables in total cost and output quantities are measured in US $ millions.
[2] There are a total of 4,354 bank-year observations.

Table 6: Estimation results for the cost frontier

	Within		CCE	
	$\hat{\beta}$	Std. Dev.	$\hat{\beta}$	Std. Dev.
Effects on cost function				
ln(PC/PF)	0.321***	0.043	0.188 ***	0.003
ln(TL)	0.142***	0.220	0.058 ***	0.006
ln(OEA)	0.065***	0.057	-0.007	0.004
ln(TD)	0.795***	0.228	0.852 ***	0.005
ln(LA)	0.005	0.067	0.012 ***	0.003
t	0.015***	0.010		
Effects on inefficiency				
ln(TA)	-0.007***	0.001	-0.054***	0.007
ETA	0.001***	2.1×10^{-4}	0.071***	0.001
ROAA	-4.5×10^{-4}***	1.3×10^{-4}	-0.070***	0.003
Year08	-0.004***	0.001	-0.019***	0.001
Austria	0.251	0.466	-1.257***	0.024
Belgium	-0.134	0.282	-0.750***	0.005
Canada	0.646***	0.213	-0.209***	0.003
Denmark	0.561***	0.172	-1.201***	0.008
France	0.701***	0.053	1.017***	0.008
Germany	0.643	0.477	-0.144***	0.003
Switzerland	0.385**	0.179	0.011***	0.001
UK	0.314**	0.156	-1.132***	0.007
USA	0.113	0.123	0.415***	0.014
σ_v^2	0.034***	9.6×10^{-5}	0.005***	6.6×10^{-6}
σ_u^2	447.062***	161.515	0.015***	1.4×10^{-4}
μ	1.2×10^{-4}***	2.6×10^{-13}	0.006***	2.0×10^{-7}
log L	5020.754		9287.145	

[1] * Significant at the 10% level, ** Significant at the 5% level and *** Significant at the 1% level.

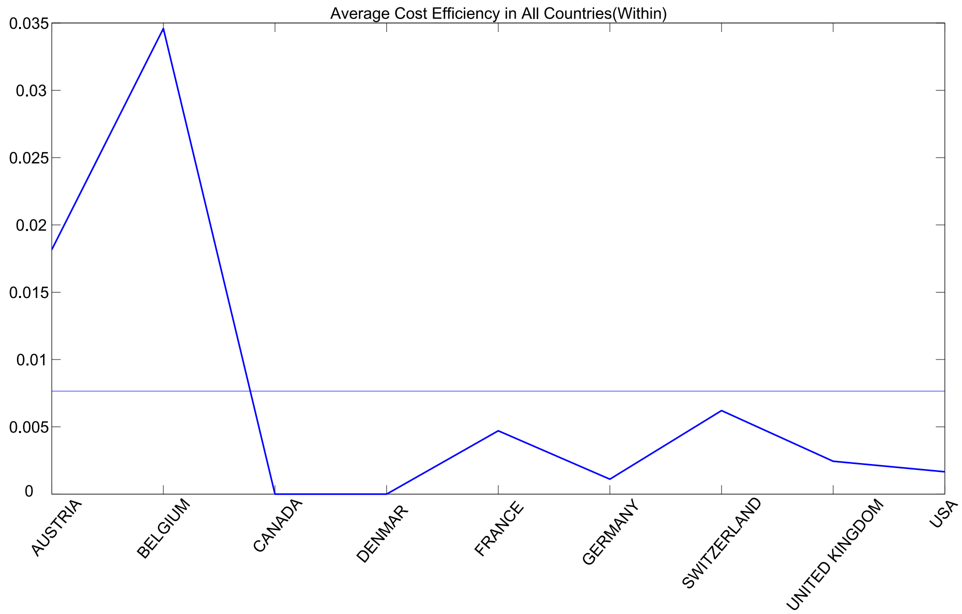

Figure 1: Average Cost Efficiency in All Countries(Within)

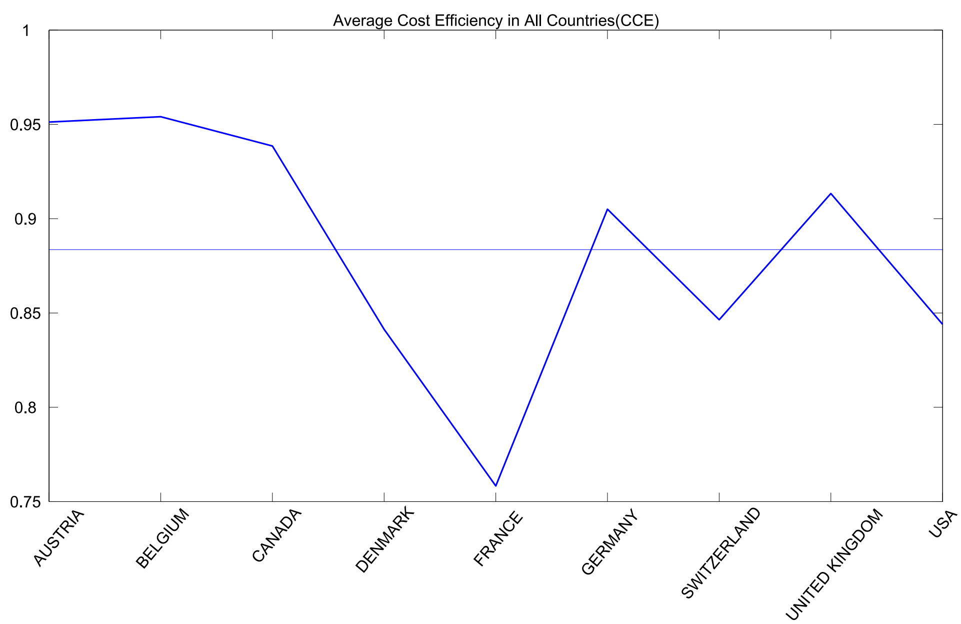

Figure 2: Average Cost Efficiency in All Countries(CCE)

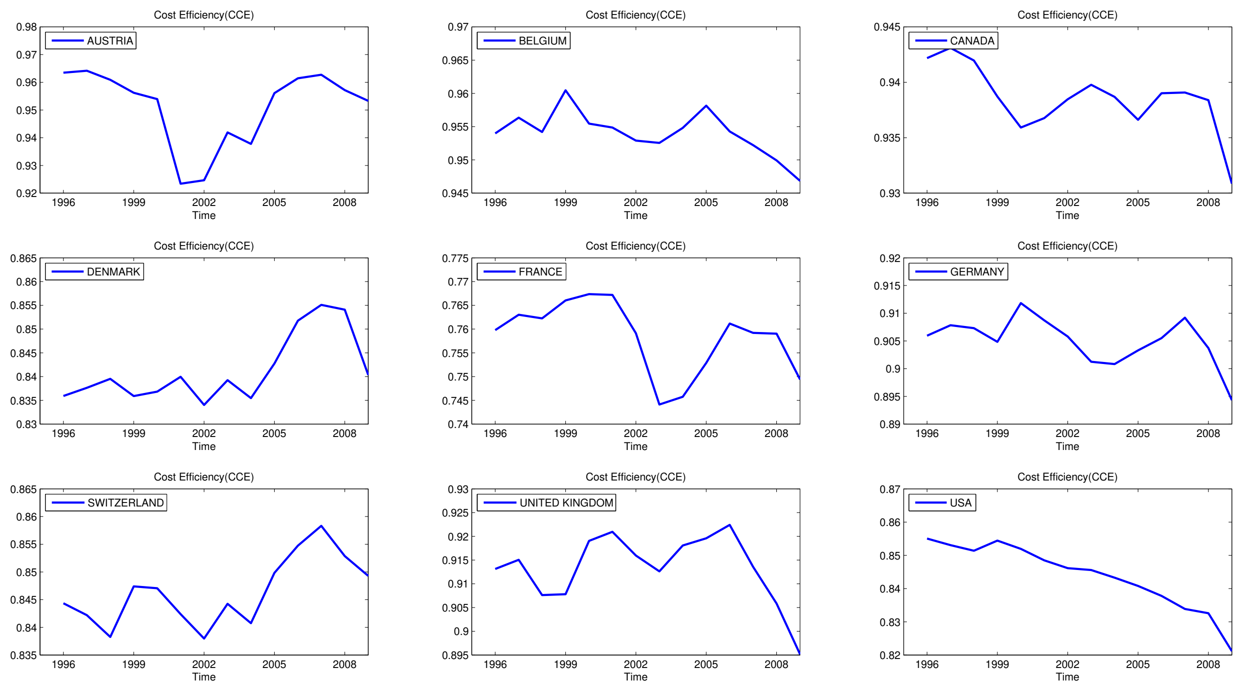

Figure 3: The Pattern of Cost Efficiency in All Countries(CCE)

Appendix

We first rewrite equations (1)-(3) as following:

$$\mathbf{y}_i = \mathbf{D}\alpha_i + \mathbf{X}_i\beta_i + \mathbf{F}\lambda_i + \varepsilon_i$$

$$\varepsilon_i = \mathbf{v}_i - \mathbf{u}_i$$

$$\mathbf{X}_i = \mathbf{D}\mathbf{A}_i' + \mathbf{F}\tau_i + \mathbf{e}_i$$

where $\mathbf{y}_i$ and ε_i are $T \times 1$ vectors, $\mathbf{X}_i$ is a $T \times k$ matrix, the common factors $\mathbf{F} = (f_1, f_2, ..., f_T)'$ is a $T \times r$ matrix with r common factors, α_i captures individual-specific time-invarious heterogeneity, $\mathbf{D} = \mathbf{1}_T$ is a $T \times 1$ vector with ones, and $\mathbf{A}_i = (\alpha_{i1}, \ldots, \alpha_{ik})'$ is an $n \times k$ vector.

Let

$$\xi_{it} = \begin{bmatrix} v_{it} + \beta' e_{it} \\ e_{it} \end{bmatrix}$$

and let $\bar{y}_{tw} = \sum_{i=1}^N w_i y_{it}$, $\bar{x}_{tw} = \sum_{i=1}^N w_i x_{it}$ and $\bar{\xi}_{tw} = \sum_{i=1}^N w_i \xi_{it}$. Additionally, let $\bar{\mathbf{y}}_w = (\bar{y}_{1w}, \ldots, \bar{y}_{Tw})'$ $\bar{\mathbf{X}}_w = (\bar{x}'_{1w}, \ldots, \bar{x}'_{Tw})'$ and $\bar{\xi}_w = (\bar{\xi}'_{1w}, \ldots, \bar{\xi}'_{Tw})'$. Based on Pesaran (2006), to proxy the common factors in our model, we could use

$$H_w^* = (\mathbf{D}, \bar{\mathbf{y}}_w, \bar{\mathbf{X}}_w, \sum_{i=1}^N w_i \mathbf{h}_i u_i^*)$$

if u_i^* were observed. However, u_i^* is unobserved practically. To resolve this problem, we propose using $\bar{\mathbf{h}}_w \mu_{**}$ as a proxy of $\sum_{i=1}^N w_i \mathbf{h}_i u_i^*$.

Lemma 1. *Under Assumptions 1–5, for each t,*

$$\sum_{i=1}^N w_i h_{it} u_i^* - \sum_{i=1}^N w_i h_{it} \mu_{**} = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Proof of Lemma 1

Since u_i^* is independent of h_{it} and u_j^* and u_i^* are mutually independent for all $i \neq j$ under Assumption 4, it can be shown that

$$E(h_{it} u_i^*) = E(h_{it}) E(u_i^*) = E(h_{it}) \mu_{**},$$

$$E[h_{it} h_{jt} (u_i^* - \mu_{**})(u_j^* - \mu_{**})] = E(h_{it} h_{jt}) E(u_i^* - \mu_{**}) E(u_j^* - \mu_{**}) = 0$$

Therefore, we obtain

$$E\left(\sum_{i=1}^{N} w_i h_{it} u_i^* - \sum_{i=1}^{N} w_i h_{it} \mu_{**}\right) = \sum_{i=1}^{N} w_i E(h_{it}) E(u_i^*) - \sum_{i=1}^{N} w_i E(h_{it}) \mu_{**} = 0$$

and

$$\begin{aligned} Var\left(\sum_{i=1}^{N} w_i h_{it} u_i^* - \sum_{i=1}^{N} w_i h_{it} \mu_{**}\right) &= E\left[\sum_{i=1}^{N} w_i^2 h_{it}^2 (u_i^* - \mu_{**})^2\right] \\ &= \sum_{i=1}^{N} w_i^2 E(h_{it}^2) E\left[(u_i^* - \mu_{**})\right]^2 = O\left(\frac{1}{N}\right) \end{aligned}$$

Thus, the desired result follows. □

Proof of Proposition 1

So far we have proved Lemma 1, we still need to prove Proposition 1. First, we define two objective functions

$$\begin{aligned} \ln L_i &= -\frac{1}{2}\left[(T-s)\left(\ln(2\pi) + \ln \sigma_v^2\right) + \frac{1}{2}(\varepsilon_i + \mathbf{F}\lambda_i)' M_w \Pi^- M_w (\varepsilon_i + \mathbf{F}\lambda_i) - \left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right)\right] \\ &\quad + \ln\left(\sigma_* \Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \ln\left(\sigma_u \Phi\left(\frac{\mu}{\sigma_u}\right)\right), \\ \ln L_i^* &= -\frac{1}{2}\left[(T-s)\left(\ln(2\pi) + \ln \sigma_v^2\right) - \frac{1}{2}\varepsilon_i' M_w^* \Pi^- M_w^* \varepsilon_i - \left(\frac{\mu_{**}^2}{\sigma_{**}^2} - \frac{\mu^2}{\sigma_u^2}\right)\right] \\ &\quad + \ln\left(\sigma_{**} \Phi\left(\frac{\mu_{**}}{\sigma_{**}}\right)\right) - \ln\left(\sigma_u \Phi\left(\frac{\mu}{\sigma_u}\right)\right). \end{aligned}$$

where

$$\begin{aligned} \mu_* &= \frac{\mu/\sigma_u^2 - (\varepsilon_i + \mathbf{F}\lambda_i)' M_w h_i}{h_i' M_w h_i + 1/\sigma_u^2}, \qquad \sigma_*^2 = \frac{1}{h_i' M_w h_i + 1/\sigma_u^2} \\ \mu_{**} &= \frac{\mu/\sigma_u^2 - \varepsilon_i' M_w^* h_i}{h_i' M_w^* h_i + 1/\sigma_u^2}, \qquad \sigma_{**}^2 = \frac{1}{h_i' M_w^* h_i + 1/\sigma_u^2} \\ M_w^* &= I - H_w^* (H_w^{*\prime} H_w^*)^{-1} H_w^{*\prime}. \end{aligned}$$

Thus, to complete the proof of Proposition 1, we must show that

$$\frac{1}{T}\sum_{i=1}^{N} w_i \ln L_i = \frac{1}{T}\sum_{i=1}^{N} w_i \ln L_i^* + O_p\left(\frac{1}{\sqrt{N}}\right)$$

by the following lemmas.

Lemma 2. *Under Assumptions 1–5,*

(B1) $\frac{\mathbf{D}'\mathbf{v}_i}{T} = O_p(\frac{1}{\sqrt{T}})$ *(B2)* $\frac{\mathbf{F}'\mathbf{v}_i}{T} = O_p(\frac{1}{\sqrt{T}})$ *(B3)* $\frac{\mathbf{D}'\mathbf{u}_i}{T} = O_p(1)$

(B4) $\frac{\mathbf{F}'\mathbf{u}_i}{T} = O_p(1)$ *(B5)* $\frac{1}{T}\sum_{j=1}^{N} w_j \mathbf{h}_j' u_j^* \mathbf{u}_i = O_p(1)$ *(B6)* $\frac{\bar{\xi}_w^{*\prime}\mathbf{v}_i}{T} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$

(B7) $\frac{\mathbf{D}'\mathbf{h}_i}{T} = O_p(1)$ *(B8)* $\frac{\mathbf{F}'\mathbf{h}_i}{T} = O_p(1)$ *(B9)* $\frac{1}{T}\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^* \mathbf{h}_i = O_p(1)$

(B10) $\frac{\bar{\xi}_w^{*\prime}\mathbf{F}}{T} = O_p(\frac{1}{\sqrt{NT}})$ *(B11)* $\frac{\bar{\xi}_w^{*\prime}\bar{\xi}_w^{*}}{T} = O_p(\frac{1}{N})$ *(B12)* $\frac{\bar{\xi}_w^{*\prime}\mathbf{D}}{T} = O_p(\frac{1}{\sqrt{NT}})$

Proof of Lemma 2

First, we let $\bar{\xi}_w^* = (\mathbf{0}, \bar{\xi}_w, \mathbf{0})$, and under Assumptions 2 and 4, that $(\mathbf{D}, \mathbf{F}, \mathbf{h}_i, \mathbf{v}_i)$ are covariance stationary, Lemma 2 can be shown based on the proofs of Pesaran (2006). □

Lemma 3. *Under Assumptions 1–5,*

(C1) $\frac{1}{T}\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^* \mathbf{v}_i = O_p\left(\frac{1}{\sqrt{NT}}\right)$.

(C2) $\frac{\bar{\xi}_w^{*\prime}\mathbf{u}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right)$.

(C3) $\frac{\bar{\xi}_w^{*\prime}\mathbf{h}_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right)$.

(C4) $\frac{1}{T}\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^* \bar{\xi}_w^* = O_p\left(\frac{1}{\sqrt{NT}}\right)$.

Proof of Lemma 3

Consider (*C1*) first. Since $\mathbf{v}_i$ is zero mean and independent of h_i and u_i^* for all i's and t's by Assumption 1, we have

$$E\left(\frac{1}{T}\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^* \mathbf{v}_i\right) = \frac{1}{T} E\left(\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^*\right) E(\mathbf{v}_i) = 0$$

Also,

$$\begin{aligned}
Var\left(\frac{1}{T}\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^* \mathbf{v}_i\right) &= \frac{1}{T^2} E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N} w_i^2 h_{it} h_{is} u_i^{*2} v_{it} v_{is}\right) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N} w_i^2 E(h_{it}h_{is}) E(u_i^{*2}) E(v_{it}v_{is}) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{N} w_i^2 E(h_{it}^2) E(u_i^{*2}) E(v_{it}^2) \\
&= O\left(\frac{1}{N}\right)\frac{1}{T^2}\sum_{t=1}^{T} E(h_{it}^2) E(u_i^{*2}) E(v_{it}^2) = O\left(\frac{1}{NT}\right)
\end{aligned}$$

where the third equality comes from the fact that $E(v_{it}v_{is}) = 0$ for $t \neq s$, and the last equality comes from the fact that $E(h_{it}^2), E(u_i^{*2})$, and $E(v_{it}^2)$ are finite by Assumption 1 and 4.

Next consider (*C2*). Under the independent assumption of e_{it} and u_{it}^*, and by letting $\bar{\xi}_{w,l}^* = (\bar{\xi}_{w,1,l}^*, \bar{\xi}_{w,2,l}^*, ..., \bar{\xi}_{w,T,l}^*)'$ denotes the l-th element of $\bar{\xi}_w^*$, we only consider the case which the element is not 0. We have

$$E\left(\frac{\bar{\xi}_{w,l}^{*\prime}\mathbf{u}_i}{T}\right) = \frac{1}{T}E\left(\bar{\xi}_{w,l}^{*\prime}\right)E\left(\mathbf{u}_i\right) = 0$$

furthermore, by Cauch-Schwarz inequality and the fact that $E(h_{it}^2)$ and $E(u_i^{*2})$ are bounded

$$\begin{aligned}
Var\left(\frac{\bar{\xi}_{w,l}^{*\prime}\mathbf{u}_i}{T}\right) &= \frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}E(\bar{\xi}_{w,t,l}^*\bar{\xi}_{w,s,l}^*)E(u_{it}u_{is})\right) \\
&= \frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}E(\bar{\xi}_{w,t,l}^*\bar{\xi}_{w,s,l}^*)E(h_{it}h_{is})E(u_i^{*2})\right) \\
&= O\left(\frac{1}{N}\right)\frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}E(h_{it}h_{is})E(u_i^{*2})\right) \\
&\leq O\left(\frac{1}{N}\right)\frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}E(h_{it}^2)^{\frac{1}{2}}E(h_{is}^2)^{\frac{1}{2}}\right) \\
&\leq O\left(\frac{1}{N}\right)\frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}K\right) = O\left(\frac{1}{N}\right)
\end{aligned}$$

where K is a finite positive constant, and the second equality comes from Lemma 2(*B11*). Similarly, we can obtain (*C3*).

Finally, consider (*C4*). Similarly, let $\bar{\xi}_{w,l}^* = (\bar{\xi}_{w,1,l}^*, \bar{\xi}_{w,2,l}^*, ..., \bar{\xi}_{w,T,l}^*)'$ denotes the l-th element of $\bar{\xi}_w^*$. Since h_{it}, u_i^*, ε_{it} and e_{it} are mutually independent and their second moments are bounded, we have

$$E\left(\sum_{i=1}^{N}w_i\mathbf{h}_i'u_i^*\bar{\xi}_{w,l}^*\right) = \sum_{i=1}^{N}w_iE(\mathbf{h}_i')E(u_i^*)E(\bar{\xi}_{w,l}^*) = 0 \tag{18}$$

and

$$\begin{aligned}
E(\bar{u}_{wt}^2) &= E\left[\left(\sum_{i=1}^{N} w_i h_{it} u_i^*\right)^2\right] \\
&= E\left(\sum_{i=1}^{N} w_i^2 h_{it}^2 u_i^{*2} + \sum_{i,j=1,i\neq j}^{N} w_i w_j h_{it} h_{jt} u_i^* u_j^*\right) \\
&= \sum_{i=1}^{N} w_i^2 E(h_{it}^2) E(u_i^{*2}) + \sum_{i,j=1,i\neq j}^{N} w_i w_j E(h_{it} h_{jt}) E(u_i^*) E(u_j^*) \\
&= \sum_{i=1}^{N} w_i^2 E(h_{it}^2) Var(u_i^*) + \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j E(h_{it} h_{jt}) E(u_i^*) E(u_j^*) \\
&\leq \sum_{i=1}^{N} w_i^2 E(h_{it}^2) Var(u_i^*) + \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j E(h_{it}^2)^{\frac{1}{2}} E(h_{jt}^2)^{\frac{1}{2}} E(u_i^*) E(u_j^*) = O(1)
\end{aligned}$$

by Cauch-Schwarz inequality. Thus,

$$\begin{aligned}
Var\left(\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^* \bar{\xi}_{w,l}^*\right) &= Var\left(\sum_{t=1}^{T} \bar{u}_{wt} \bar{\xi}_{w,t,l}^*\right) = \sum_{t=1}^{T} Var\left(\bar{\xi}_{wt} \bar{u}_{w,t,l}\right) \\
&= \sum_{t=1}^{T} E(\bar{\xi}_{w,t,l}^2) E(\bar{u}_{wt}^2) = O\left(\frac{T}{N}\right), \qquad (19)
\end{aligned}$$

where the second equality comes from the fact that for all $t \neq s$

$$\begin{aligned}
&Cov\left(\bar{\xi}_{w,t,l}^* \bar{u}_{wt}, \bar{\xi}_{w,s,l}^* \bar{u}_{ws}\right) = E\left(\bar{\xi}_{w,t,l}^* \bar{u}_{wt} \bar{\xi}_{w,s,l}^* \bar{u}_{ws}\right) - E\left(\bar{\xi}_{w,t,l}^* \bar{u}_{wt}\right) E\left(\bar{\xi}_{w,s,l}^* \bar{u}_{ws}\right) \\
&= E\left(\bar{\xi}_{w,t,l}^*\right) E\left(\bar{\xi}_{w,s,l}^*\right) E\left(\bar{u}_{wt} \bar{u}_{ws}\right) - E\left(\bar{\xi}_{w,t,l}^*\right) E\left(\bar{u}_{wt}\right) E\left(\bar{\xi}_{w,s,l}^*\right) E\left(\bar{u}_{ws}\right) = 0.
\end{aligned}$$

Together with (18) and (19), we obtain

$$Var\left(\frac{1}{T}\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^* \bar{\xi}_w^*\right) = O\left(\frac{1}{NT}\right)$$

hence, $\frac{1}{T}\sum_{i=1}^{N} w_i \mathbf{h}_i' u_i^* \bar{\xi}_w^* = O_p\left(\frac{1}{\sqrt{NT}}\right)$. □

Lemma 4. *Under Assumptions 1-5,*

(D1) $\frac{1}{T}\sum_i^N w_i\varepsilon_i' M_w \varepsilon_i = \frac{1}{T}\sum_i^N w_i\varepsilon_i' M_w^* \varepsilon_i + O_p\left(\frac{1}{\sqrt{N}}\right)$.

(D2) $\frac{1}{T}\sum_i^N w_i\lambda_i' \mathbf{F}' M_w \varepsilon_i = O_p\left(\frac{1}{\sqrt{N}}\right)$.

(D3) $\frac{1}{T}\sum_i^N w_i\lambda_i' \mathbf{F}' M_w \lambda_i \mathbf{F} = O_p\left(\frac{1}{\sqrt{N}}\right)$.

(D4) $\frac{1}{T}\sum_i^N w_i\mathbf{h}_i' M_w \mathbf{h}_i = \frac{1}{T}\sum_i^N w_i\mathbf{h}_i' M_w^* \mathbf{h}_i + O_p\left(\frac{1}{\sqrt{N}}\right)$.

(D5) $\frac{1}{T}\sum_i^N w_i\varepsilon_i' M_w \mathbf{h}_i = \frac{1}{T}\sum_i^N w_i\varepsilon_i' M_w^* \mathbf{h}_i + O_p\left(\frac{1}{\sqrt{N}}\right)$.

Proof of Lemma 4

Recall $M_w = \left(I_T - H_w\left(H_w' H_w\right)^{-1} H_w'\right)$, $\mathbf{G} = (\mathbf{D}, \mathbf{F}, \bar{\mathbf{U}}_w)$ and $\bar{\xi}_w^* = \left(\mathbf{0}, \bar{\xi}_w, \mathbf{0}\right)$. Therefore $H_w = (\mathbf{D}, \bar{\mathbf{y}}_w, \bar{\mathbf{X}}_w, \sum_{i=1}^N w_i \mathbf{h}_i \mu_{**})$ can be written by

$$H_w = \left(\mathbf{G} + O_p\left(\frac{1}{\sqrt{N}}\right)\right)\bar{\mathbf{P}}_w + \bar{\xi}_w^* = \mathbf{G}\bar{\mathbf{P}}_w + \bar{\xi}_w^* + O_p\left(\frac{1}{\sqrt{N}}\right) = H_w^* + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{20}$$

where

$$\bar{\mathbf{P}}_w = \begin{bmatrix} 1 & \bar{\mathbf{B}}_w & 0 \\ 0 & \bar{\mathbf{C}}_w & 0 \\ 0 & \mathrm{I}_{k+1} & 1 \end{bmatrix}$$

Then we consider *(D1)*

$$\frac{1}{T}\sum_{i=1}^N w_i \varepsilon_i' M_w \varepsilon_i = \frac{1}{T}\sum_{i=1}^N w_i \varepsilon_i' \varepsilon_i - \frac{1}{T}\sum_{i=1}^N w_i \varepsilon_i' H_w\left(H_w' H_w\right)^{-1} H_w' \varepsilon_i.$$

By (20) and the fact that $\varepsilon_{it} = v_{it} - h_{it}u_i^*$, we have

$$\frac{\varepsilon_i' H_w}{T} = \frac{\varepsilon_i' H_w^*}{T} + \frac{1}{T}\sum_{t=1}^T \varepsilon_{it} O_p\left(\frac{1}{\sqrt{N}}\right) = \frac{\varepsilon_i' H_w^*}{T} + \frac{1}{T}\sum_{t=1}^T \left(v_{it} - h_{it}u_i^*\right) O_p\left(\frac{1}{\sqrt{N}}\right). \tag{21}$$

Since v_{it} is a covariance stationary process with zero mean and independent over t's and across i's, we have

$$Var\left(\frac{1}{T}\sum_{t=1}^T v_{it}\right) = \left(\frac{1}{T^2}\right)\sum_{t=1}^T E(v_{it}^2) = O(1/T)$$

Thus, $\frac{1}{T}\sum_{t=1}^{T} v_{it} = O_p\left(\frac{1}{\sqrt{T}}\right)$. Similarly, by Assumption 1 and 4, that u_i^* and h_{it} have finite second moments and are mutually independent, we have

$$E\left(\frac{1}{T}\sum_{t=1}^{T} h_{it}u_i^*\right) = O(1), \quad Var\left(\frac{1}{T}\sum_{t=1}^{T} h_{it}u_i^*\right) = \frac{1}{T^2}E\left(\sum_{t=1}^{T}(h_{it}u_i^*)\right) = O\left(\frac{1}{T}\right),$$

Thus, $\frac{1}{T}\sum_{t=1}^{T} h_{it}u_i^* = O_p(1)$. Therefore, (21) can be rewritten as

$$\frac{\varepsilon_i H_w}{T} = \frac{\varepsilon_i H_w^*}{T} + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{22}$$

Notice also that the last equality holds due to $H_w^* = \mathbf{G}\bar{\mathbf{P}}_w + \bar{\xi}_w^*$, where $\mathbf{G}$ is covariance stationary, and $\bar{\mathbf{P}}_w$ is bounded. We have

$$\frac{H_w' H_w}{T} = \frac{1}{T}\left(H_w^* + O_p\left(\frac{1}{\sqrt{N}}\right)\right)'\left(H_w^* + O_p\left(\frac{1}{\sqrt{N}}\right)\right) = \frac{H_w^{*'}H_w^*}{T} + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{23}$$

Additionally, by lemma2 (*B1–5*) and lemma3 (*C1*), it can be shown that

$$\frac{H_w^{*'}\varepsilon_i}{T} = O_p(1). \tag{24}$$

Thus, by (21)–(24) and Lemma2 from Kiviet and Phillips (1994), we have

$$\begin{aligned}
\frac{1}{T}\sum_{i=1}^{N} w_i\varepsilon_i' M_w \varepsilon_i &= \frac{1}{T}\sum_{i=1}^{N} w_i\varepsilon_i'\varepsilon_i - \sum_{i=1}^{N} w_i \frac{\varepsilon_i' H_w}{T}\left(\frac{H_w' H_w}{T}\right)^{-1}\frac{H_w'\varepsilon_i}{T} \\
&= \frac{1}{T}\sum_{i=1}^{N} w_i\varepsilon_i'\varepsilon_i - \sum_{i=1}^{N} w_i\left(\frac{\varepsilon_i' H_w^*}{T} + O_p\left(\frac{1}{\sqrt{N}}\right)\right)\times \\
&\left(\left(\frac{H_w^{*'}H_w^*}{T}\right)^{-1} + O_p\left(\frac{1}{\sqrt{N}}\right)\right)\left(\frac{H_w^{*'}\varepsilon_i}{T} + O_p\left(\frac{1}{\sqrt{N}}\right)\right) \\
&= \frac{1}{T}\sum_{i=1}^{N} w_i\varepsilon_i' M_w^*\varepsilon_i + O_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

To proof (*D2*), let $\mathbf{F}_l = (f_{l1}, \ldots, f_{lT})'$ denotes the l-th column of $\mathbf{F}$. Since $\mathbf{F}$ is covariance stationary, we have

$$E\left(\frac{1}{T}\sum_{t=1}^{T} f_{lt}\right) = O(1), \qquad Var\left(\frac{1}{T}\sum_{t=1}^{T} f_{lt}\right) = \frac{1}{T^2}Var\left(\sum_{t=1}^{T} f_{lt}\right) = O\left(\frac{1}{T}\right).$$

Therefore, we have

$$\frac{\mathbf{F}_l' H_w}{T} = \frac{\mathbf{F}_l' H_w^*}{T} + \frac{1}{T}\sum_{t=1}^{T} f_{lt} O_p\left(\frac{1}{\sqrt{N}}\right) = \frac{\mathbf{F}_l' H_w^*}{T} + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{25}$$

By (22), (23), (25), and the fact that $\frac{H_w^{*\prime} H_w^*}{T} = O_P(1)$ due to $\mathbf{G}$ is covariance stationary and $\bar{\mathbf{P}}_w$ is bounded, we have

$$\frac{1}{T}\sum_{i=1}^{N} w_i \lambda_i' \mathbf{F}' M_w \varepsilon_i = \frac{1}{T}\sum_{i=1}^{N} w_i \lambda_i' \mathbf{F}' \varepsilon_i - \sum_{i=1}^{N} w_i \lambda_i' \left(\frac{\mathbf{F}' H_w^*}{T} + O_p\left(\frac{1}{\sqrt{N}}\right)\right) \times$$
$$\left(\left(\frac{H_w^{*\prime} H_w^*}{T}\right)^{-1} + O_p\left(\frac{1}{\sqrt{N}}\right)\right)\left(\frac{H_w^{*\prime}\varepsilon_i}{T} + O_p\left(\frac{1}{\sqrt{N}}\right)\right)$$
$$= \frac{1}{T}\sum_{i=1}^{N} w_i \lambda_i' \mathbf{F}' \left(I_T - H_w^*\left(H_w^{*\prime} H_w^*\right)^{-1} H_w^{*\prime}\right)\varepsilon_i + O_p\left(\frac{1}{\sqrt{N}}\right)$$
$$= \frac{1}{T}\sum_{i=1}^{N} w_i \lambda_i' \mathbf{F}' M_w^* \varepsilon_i + O_p\left(\frac{1}{\sqrt{N}}\right)$$

By Lemma 2(*B10*), we have

$$\frac{H_w^{*\prime}\mathbf{F}}{T} = \frac{\bar{\mathbf{P}}_w'\mathbf{G}'\mathbf{F}}{T} + \frac{\bar{\xi}_w^{*\prime}\mathbf{F}}{T} = \frac{\bar{\mathbf{P}}_w'\mathbf{G}'\mathbf{F}}{T} + O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{26}$$

Similarly, by Lemma 2(*B1–6*) and Lemma 3(*C2*), we have

$$\frac{H_w^{*\prime}\varepsilon_i}{T} = \frac{\bar{\mathbf{P}}_w'\mathbf{G}'\varepsilon_i}{T} + \frac{\bar{\xi}_w^{*\prime}\varepsilon_i}{T} = \frac{\bar{\mathbf{P}}_w'\mathbf{G}'\varepsilon_i}{T} + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{27}$$

and by Lemma 2(*B10–12*) and Lemma 3(*C4*)

$$\frac{H_w^{*\prime} H_w^*}{T} = \frac{\bar{\mathbf{P}}_w'\mathbf{G}'\mathbf{G}\bar{\mathbf{P}}_w}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{28}$$

Notice that if the rank condition hold, $I_T - \mathbf{G}\bar{\mathbf{P}}_w\left(\bar{\mathbf{P}}_w'\mathbf{G}'\mathbf{G}\bar{\mathbf{P}}_w\right)^{-}\bar{\mathbf{P}}_w'\mathbf{G}' = I_T - \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-}\mathbf{G}' = M_g$. Since $F \subset G$, we have $M_g F = 0$. Together with (26)–(28), we have

$$\frac{1}{T}\sum_{i=1}^{N} w_i \lambda_i' \mathbf{F}' M_w^* \varepsilon_i$$
$$= \frac{1}{T}\sum_{i=1}^{N} w_i \lambda_i' \mathbf{F}' \left(I_T - \mathbf{G}\bar{\mathbf{P}}_w\left(\bar{\mathbf{P}}_w'\mathbf{G}'\mathbf{G}\bar{\mathbf{P}}_w\right)^{-}\bar{\mathbf{P}}_w'\mathbf{G}'\right)\varepsilon_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$$
$$= \frac{1}{T}\sum_{i=1}^{N} w_i \lambda_i' \mathbf{F}' M_g \varepsilon_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$$

Therefore, we obtain

$$\frac{1}{T}\sum_{i=1}^{N} w_i\lambda_i'\mathbf{F}'M_w\varepsilon_i = \frac{1}{T}\sum_{i=1}^{N} w_i\lambda_i'\mathbf{F}'M_w^*\varepsilon_i + O_p\left(\frac{1}{\sqrt{N}}\right) = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Similarly, we can obtain (*D3*). Finally, Consider (*C4*) and (*C5*). Since h_{it} is covariance stationary, we have

$$\begin{aligned}
E\left(\frac{1}{T}\sum_{t=1}^{T} h_{it}\right) &= \frac{1}{T}E\left(\sum_{t=1}^{T} h_{it}\right) = O(1) \\
Var\left(\frac{1}{T}\sum_{t=1}^{T} h_{it}\right) &= \frac{1}{T^2}Var\left(\sum_{t=1}^{T} h_{it}\right) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} Cov(h_{it}, h_{is}) \\
&= O\left(\frac{1}{NT^2}\right)\sum_{t=1}^{T}\sum_{s=1}^{T} (\Gamma_h|t-s|) = O\left(\frac{1}{NT}\right)
\end{aligned}$$

where $\Gamma_h|t-s|$ is the autocovariance of h_{it}. Therefore, we have $\frac{1}{T}\sum_{t=1}^{T} h_{it} = O_p(1)$ and

$$\frac{\mathbf{h}_i'H_w}{T} = \frac{\mathbf{h}_i'H_w^*}{T} + \frac{1}{T}\sum_{t=1}^{T} h_{it}O_p\left(\frac{1}{\sqrt{N}}\right) = \frac{\mathbf{h}_i'H_w^*}{T} + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Then, similar to the proof of (*D1*), we can further obtain (*D4*) and (*D5*). □

Lemma 5. *Under Assumptions 1-5,*

(E1) $\ln(\sigma_*) = \ln(\sigma_{**}) + O_p\left(\frac{1}{\sqrt{N}}\right)$.

(E2) $\frac{\mu_*}{\sigma_*} = \frac{\mu_{**}}{\sigma_{**}} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right)$.

(E3) $\frac{\mu_*^2}{\sigma_*^2} = \frac{\mu_{**}^2}{\sigma_{**}^2} + O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{T}{\sqrt{N}}\right)$.

Proof of Lemma 5: First, consider (*E1*). Notice that

$$\begin{aligned}
\ln(\sigma_*) &= -\frac{1}{2}\ln(h_i'M_w\Pi^-M_wh_i + 1/\sigma_u^2) = -\frac{1}{2}\ln\left(\frac{h_i'M_wh_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}\right) - \ln T, \\
\ln(\sigma_{**}) &= -\frac{1}{2}\ln\left(h_i'M_w^*\Pi^-M_w^*h_i + 1/\sigma_u^2\right) = -\frac{1}{2}\ln\left(\frac{h_i'M_w^*h_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}\right) - \ln T.
\end{aligned}$$

By lemma 4(*D4*), we have

$$\frac{h_i' M_w^* h_i}{T} = \frac{h_i' M_w^* h_i}{T} + O_p\left(\frac{1}{\sqrt{N}}\right),$$

by using Taylor expansion at $\Theta_1^* = \frac{h_i' M_w^* h_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}$ and the fact that $\Theta_1 = \frac{h_i' M_w^* h_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}$ is strictly positive, we obtain

$$\ln(\Theta_1) = \ln(\Theta_1^*) + \frac{1}{\Theta_1^*}\left(O_p\left(\frac{1}{\sqrt{N}}\right)\right) = \ln(\Theta_1^*) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Next, consider (*E2*). Notice that

$$\frac{\mu_*}{\sigma_*} = \left(\frac{\mu}{\sigma_u^2} - \varepsilon_i' M_w \Pi^{-1} M_w \mathbf{h}_i\right)\left(\mathbf{h}_i' M_w \Pi^{-1} M_w \mathbf{h}_i + \frac{1}{\sigma_u^2}\right)^{-\frac{1}{2}}. \tag{29}$$

The denominator can be rearranged as

$$\left(\mathbf{h}_i' M_w \Pi^{-1} M_w \mathbf{h}_i + \frac{1}{\sigma_u^2}\right)^{-\frac{1}{2}} = T^{-\frac{1}{2}}\left(\frac{\mathbf{h}_i' M_w \mathbf{h}_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}\right)^{-\frac{1}{2}}.$$

Let $\Theta_2 = \frac{\mathbf{h}_i' M_w \mathbf{h}_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}$ and $\Theta_2^* = \frac{\mathbf{h}_i' M_w^* \mathbf{h}_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}$. By using Taylor expansion, we have

$$\Theta_2^{-\frac{1}{2}} = (\Theta_2^*)^{-\frac{1}{2}} - \frac{1}{2}(\Theta_2^*)^{-\frac{3}{2}}\left(O_p\left(\frac{1}{\sqrt{N}}\right)\right)$$

Then (*E2*) follows if $(\Theta_2^*)^{-\frac{3}{2}} = O_p(1)$.

Since Θ_2^* is a scalar, it is either $(\Theta_2^*)^{-1} \leq (\Theta_2^*)^{-\frac{3}{2}} \leq (\Theta_2^*)^{-2}$ or $(\Theta_2^*)^{-1} \geq (\Theta_2^*)^{-\frac{3}{2}} \geq (\Theta_2^*)^{-2}$ by Lemma 2 from Kiviet and Phillips (1994). Also,

$$(\Theta_2^*)^{-1} = \left(\frac{\mathbf{h}_i' M_w^* \mathbf{h}_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}\right)^{-1} = O_p(1) > 0$$

and

$$(\Theta_2^*)^{-2} = \left(\left(\frac{\mathbf{h}_i' M_w^* \mathbf{h}_i}{T\sigma_v^2}\right)^{-1} + O\left(\frac{1}{T^2}\right)\right)^2 = O_p(1) \tag{30}$$

Thus, $(\Theta_2^*)^{-\frac{3}{2}} = O_p(1)$ and, therefore,

$$\begin{aligned} &T^{\frac{-1}{2}}\left(\frac{\mathbf{h}_i' M_w \mathbf{h}_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}\right)^{\frac{-1}{2}} \\ &\quad = T^{\frac{-1}{2}}\left(\left(\frac{\mathbf{h}_i' M_w^* \mathbf{h}_i}{T\sigma_v^2} + \frac{1}{T\sigma_u^2}\right)^{\frac{-1}{2}} + O_p\left(\frac{1}{\sqrt{N}}\right)\right) \end{aligned} \tag{31}$$

Similarly, by using lemma 4(*D5*), it can be shown that

$$T\left(\frac{\mu}{T\sigma_u^2}-\frac{\varepsilon_i' M_w \mathbf{h}_i}{T\sigma_v^2}\right)=T\left(\left(\frac{\mu}{T\sigma_u^2}-\frac{\varepsilon_i' M_w^* \mathbf{h}_i}{T\sigma_v^2}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)\right) \tag{32}$$

Together with (31) and (32), we obtain (*E2*).

The proof of (*E3*) is similar. First, by (30) that $(\Theta_2^*)^{-2}=\left(\frac{\mathbf{h}_i' M_w^* \mathbf{h}_i}{T\sigma_v^2}+\frac{1}{T\sigma_u^2}\right)^{-2}=O_p\left(1\right)$, we have

$$T^{-1}\left(\frac{\mathbf{h}_i' M_w^* \mathbf{h}_i}{T\sigma_v^2}+\frac{1}{T\sigma_u^2}\right)^{-1}=T^{-1}\left(\left(\frac{\mathbf{h}_i' M_w^* \mathbf{h}_i}{T\sigma_v^2}+\frac{1}{T\sigma_u^2}\right)^{-1}+O_p\left(\frac{1}{\sqrt{N}}\right)\right)$$

Similarly,

$$T^2\left(\frac{\mu}{T\sigma_u^2}-\frac{\varepsilon_i' M_w \mathbf{h}_i}{T\sigma_v^2}\right)^2=T^2\left(\left(\frac{\mu}{T\sigma_u^2}-\frac{\varepsilon_i' M_w^* \mathbf{h}_i}{T\sigma_v^2}\right)^2+O_p\left(\frac{1}{\sqrt{N}}\right)\right)$$

Combining the above two equations, we obtain (*E3*). □

Proof of Proposition 1 Recall that

$$\begin{aligned}\frac{1}{T}\sum_{i=1}^N w_i \ln L_i =& \frac{1}{T}\sum_{i=1}^N w_i\left[-\frac{1}{2}\left(T-s\right)\left(\ln\left(2\pi\right)+\ln\sigma_v^2\right)-\frac{1}{2\sigma_v^2}\left(\varepsilon_i+\mathbf{F}\lambda_i\right)' M_w\left(\varepsilon_i+\mathbf{F}\lambda_i\right)\right.\\ &\left.+\frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2}-\frac{\mu^2}{\sigma_u^2}\right)+\ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right)-\ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right)\right].\end{aligned}$$

By Lemmas 3(*D1*)–(*D3*), the second term becomes

$$\begin{aligned}&\frac{\left(\varepsilon_i+\mathbf{F}\lambda_i\right)' M_w \Pi^- M_w\left(\varepsilon_i+\mathbf{F}\lambda_i\right)}{T}=\frac{\left(\varepsilon_i+\mathbf{F}\lambda_i\right)' M_w\left(\varepsilon_i+\mathbf{F}\lambda_i\right)}{T\sigma_v^2}\\ &=\frac{\varepsilon_i' M_w^*\varepsilon_i}{T\sigma_v^2}+O_p\left(\frac{1}{\sqrt{N}}\right).\end{aligned} \tag{33}$$

Next, consider the third term. Define $f(\cdot)=\ln\left(\Phi(\cdot)\right)$. By Lemma 5(*E2*) and using Taylor expansion,

$$\ln\left(\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right)=\ln\left(\Phi\left(\left(\frac{\mu_{**}}{\sigma_{**}}\right)\right)\right)+O_p\left(\sqrt{\frac{T}{N}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right).$$

Together with Lemma 5(*E1*), we obtain

$$\ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right)=\ln\left(\sigma_{**}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)+\ln\left(\Phi\left(\frac{\mu_{**}}{\sigma_{**}}\right)\right)+O_p\left(\sqrt{\frac{T}{N}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)$$

and, therefore,

$$T^{-1}\ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right)=T^{-1}\ln\left(\sigma_{**}\Phi\left(\frac{\mu_{**}}{\sigma_{**}}\right)\right)+O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{34}$$

Similarly, by Lemma 5(*E3*),

$$T^{-1}\left(\frac{\mu_*}{\sigma_*}\right)^2=T^{-1}\left(\frac{\mu_{**}}{\sigma_{**}}\right)^2+O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{35}$$

Combining (33)–(35), we obtain the desired result. □

References

Ackerberg, D. A., K. Caves, and G. Frazer, 2006. Structural identification of production functions, mimeo, UCLA Department of Economics.

Ahn, S. G., Y. H. Lee, and P. Schmidt, 2006. Panel data models with multiple time-varying effects, Mimeo, Arizona State University.

Ahn, S. G., Y. H. Lee, and P. Schmidt, 2007. Stochastic frontier models with multiple time-varying individual effects, *Journal of Productivity Analysis*, **27**, 1-12.

Aigner, D.J., C.A.K. Lovell, and P. Schmidt, 1977. Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics*, **6**, 21-37.

Bai, J., 2009. Panel data models with interactive fixed effects, *Econometrica*, **77**, 1229-1279.

Bai, J. and S. Ng, 2002. Determining the number of factors in approximate factor models, *Econometrica*, **70**, 191-221.

Battese, G. E. and T. J. Coelli, 1988. Prediction of firm-level technical effeiciencies with a generalized frontier production function and panel data, *Journal of Econometrics*, **38**, 387-399.

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

Berger, A. N., I. Hasan and M. Zhou, 2009. Bank ownership and efficiency in China: What will happen in the world's largest nation? *Journal of Banking and Finance*, **33**, 113-130.

Cornwell, C., P. Schmidt and R. Sickles, 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels, *Journal of Econometrics*, 46, 185-200.

Delis, M. D. and N. I. Papanikolaou, 2009. Determinants of bank efficiency: evidence from a semi-parametric methodology, *Managerial Finance*, **35**, 260-275.

Greene, W., 2003. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. Working Paper 03-10, Department of Economics, Stern School of Business, New York University.

Greene, W., 2005. Fixed and random effects in stochastic frontier models, *Journal of Productivity Analysis*, **23**, 7-32.

Han, C., L. Orea and P. Schmidt, 2005, Estimation of a panel data model with parametric temporal variation in individual effects, *Journal of Econometrics*, **126**, 241-267.

Khatri, C. G., 1968. Some results for the singular normal multivariate regression models, *Sankhya*, **30**, 267-280.

Kim, Y. and P. Schmidt, 2000. A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data, *Journal of Productivity Analysis*, **14**, 91-118.

Kiviet, J. F. and G. D. A. Phillips, 1994, Bias assessment and reduction in linear error-correction models, *Journal of Econometrics*, **63**, 215-243.

Kumbhakar, S., 1990. Production frontiers, panel data, and time-varying technical inefficiency, *Journal of Econometrics*, 46, 201-212.

Lee, Y. H., 2006. A stochastic production frontier model with group-specific temporal variation in technical efficiency, *European Journal of Operational Research*, **174**, 1616-1630.

Lensink, R., A. Meesters and I. Naaborg, 2008. Bank efficiency and foreign ownership: Do good institutions matter? *Journal of Banking and Finance*, **32**, 834-844.

Levinsohn, J. and A. Petrin, 2003. Estimating production functions using inputs to control for unobservables, *Review of Economic Studies*, **70**, 317-341.

Olley, S. and A. Pakes, 1996. The dynamics of productivity in the telecommunications equipment industry, *Econometrica*, **64**, 1263-1298.

Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica*, **74**, 967-1012.

Sun, L. and T.P. Chang, 2010. A comprehensive analysis of the effects of risk measures on bank efficiency: Evidence from emerging Asian countries, *Journal of Banking and Finance*, **35**, 1727-1735.

Wang, H. J. and C. W. Ho, 2010. Estimating fixed-effect panel stochastic frontier models by model transformation, *Journal of Econometrics*, **157**, 286-296.

Wang, H. J. and P. Schmidt, 2002. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels, *Journal of Productivity Analysis*, **18**, 129-144.