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

† Author for correspondence: Hsu, C.-C., Department of Economics, National Central University, Jhongli 320, Taiwan; Tel: +886 3 4227151 ext. 66308; Fax: +886 3 4222876; Email: cchsu@mgt.ncu.edu.tw

Estimation of a Panel Stochastic Frontier Model with Unobserved Common Shocks

Abstract

This paper develops a panel stochastic frontier model with unobserved common shocks to capture cross-section dependence among individual firms. The novel feature of our model is to separate technical inefficiency from the effects induced by unobserved common shocks and individual heterogeneity. We propose a modified maximum likelihood method that does not require estimating unobserved common correlated effects and discuss the asymptotic properties of the proposed estimation procedure. The basic idea of our approach is similar to that in Pesaran (2006) for the linear panel regression. We show that the proposed method can control the common correlated effects and obtain consistent estimates of parameters 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-section dependence for relatively small T. The proposed method is also illustrated in applications based on a comparison of the efficiency of savings and commercial banking industries in US.

JEL classification: C23

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

1 Introduction

The use of panel data has been increasingly popular in stochastic frontier models to analyze technical or cost inefficiencies of production units and financial institutions. There are two approaches that have been employed for controlling unobservable cross-sectional heterogeneity and estimating time-varying technical inefficiency. The first is to consider the linear panel models with fixed or random effects but without imposing distributional assumptions on technical inefficiency; see Cornwell et al. (1990), Han et al. (2005), Lee (2006), Ahn et al. (2001, 2007), among others. The generalized method of moments (GMM) is adapted in these studies to estimate stochastic frontier models with time-varying technical inefficiency. The second approach is to assume technical inefficiency to be random and specific distributional assumptions are required; see Kumbhakar (1990), Wang and Schmidt (2002), Greene (2003, 2005a, b), Wang and Ho (2010), among others. The maximum likelihood (ML) method, based on suitable distributional assumptions, is suggested to estimate the effects of firm characteristics on technical efficiency levels. Both approaches, however, do not provide a tractable way to model unobserved common shocks and their heterogeneous impacts on cross-sectional production units.

Unobserved common shocks (e.g., financial crises, technological innovations, free trade agreements etc.) are a likely source of cross-section dependence which is a prevalent feature in panel data. Ignoring cross-section dependence induced by unobserved common shocks can be problematic in the estimation of cross-section and panel regressions; see Andrews (2005), Pesaran (2006), Bai (2009) for further discussion. Conventional panel stochastic frontier models do not distinguish between unobserved common shocks and technical inefficiency. Ahn et al. (2007, hereafter ALS) assume that firms' inefficiencies consist of unobserved factors each of which changes over time in a temporal pattern common to all individual firms. This modeling allows cross-sectional dependence among individual firms, but the assumption might be quite strong because all the time-varying effects induced by unobserved common factors are attributed to technical inefficiency. For example, it is hard to conclude that local and small banks suffer less from global financial shocks are in general more efficient than multinational banks. Another feature of the ALS approach is that they simply estimate firms' inefficiency components instead of analyzing the effects of exogenous variables on inefficiency levels. It would be likely to limit their applicability in empirical studies.

In this paper we develop a panel stochastic frontier model with unobserved common shocks to capture cross-sectional dependence among individual firms. The novel feature of our model is to separate technical inefficiency from the effects induced by unobserved common shocks[1] and individual heterogeneity. We propose a likelihood-based method to estimate parameters in the stochastic frontier model and discuss the asymptotic properties of the proposed estimation procedure. The basic idea of our approach is similar to that in Pesaran (2006) for the linear panel regression. We first transform the model by regressing cross-section averages of the dependent and independent variables to filter out common correlated and fixed effects, and then maximize the marginal log-likelihood function of the transformed model to yield parameter estimates. It is shown that the proposed ML estimator has consistency and asymptotic normality when $(T, N) \to \infty$ jointly and $T/N \to 0$.

A few additional comments are in order. First, the proposed model possesses the scaling property proposed by Wang and Schmidt (2002) and Wang and Ho (2010). In contrast with ALS, the use of scaling-property model will enable us to investigate how firms' efficiency levels vary with exogenous variables. Second, we show that the estimates will be biased if there exists common correlated effects and we wrongly use the within-transformation. Third, our approach can be applied to estimate the cost function and cost inefficiency. We also conduct some Monte Carlo simulations to investigate the finite sample properties of the proposed method. Simulation results show that the proposed estimator has quite smaller biases and MSEs than of the within-transformation estimator when the model exists unobserved common shocks and cross-sectional dependence.

To illustrate the relevance of our approach, the proposed approach is applied for analyzing cost inefficiency of savings and commercial banking industry in U.S.. Recent researches in bank efficiency do not deal with the effects of unobserved common shocks; see, for example, Lensink et al. (2008) and Sun and Chang (2010). The empirical results show that bank

[1] These effects are usually referred to as common correlated effects (Pesaran, 2006) or interactive effects (Bai, 2009).

efficiency improves before 2006 and the estimated inefficiency index might bias if we do not take account of unobserved common shocks.

The remainder of this paper is organized as follows. Section 2 describes the panel stochastic frontier model with a multifactor error structure and discusses the asymptotic properties of the proposed estimation procedure. Section 3 conducts some Monte Carlo simulations to investigates the small-sample properties of the proposed estimator. An empirical application is discussed in Section 4. Section 5 concludes this paper. All mathematic proofs are provided in the Appendix.

2 Panel Stochastic Frontier Model

2.1 The Model

Consider a panel stochastic frontier model with the following specifications:

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + v_{it} - u_{it}, \qquad i = 1, \ldots, N,\ t = 1, \ldots, T, \tag{1}$$

$$\boldsymbol{x}_{it} = \mathbf{A}_i + \boldsymbol{\tau}_i'\boldsymbol{f}_t + \boldsymbol{e}_{it} \tag{2}$$

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

where y_{it} is the logarithm of output of firm i in period t, $\boldsymbol{x}_{it}$ is a $k \times 1$ vector of the logarithm of inputs in this production system, α_i denotes individual fixed effects, and v_{it} is a zero-mean idiosyncratic error. Let $\boldsymbol{f}_t$ be a $r \times 1$ vector of unobserved common shocks or common correlated effects, $\boldsymbol{\lambda}_i$ be the heterogeneous impact of common shocks on firm i, and u_{it} is the term used to measure inefficiency. The regressors are also affected by individual fixed effects, $\mathbf{A}_i$, and common shocks, where $\mathbf{A}_i$ is a $k \times 1$ vector which is correlated with α_i, and $\boldsymbol{\tau}_i$ denotes a $r \times k$ vector of factor loadings. The model specification not only allows for cross-sectional dependence through a multifactor error structure but makes for correlation between common factors and regressors.[2] The idiosyncratic error $\boldsymbol{e}_{it}$ is independent of all

[2] Ackerberg, Caves and Frazer (2006) have mentioned that although some shocks are unobserved for econometrician, they are potentially predictable by firms when they are are making input decision, such as expected defect rates, expected down-time due to machine breakdown, or expected government policies. This will be

observations on v_{it} and u_{it}. Finally, let h_{it} be a positive function of firms' inefficiency determinants $\boldsymbol{z}_{it}$, $u_i^* \sim N^+(\mu, \sigma_u^2)$, where the distribution is truncated from below at zero such that $u_i^* > 0$. This specification is referred to as the scaling property, which allows us to estimate coefficients and inefficiency in a one-step procedure.[3] The scaling property also allows the inefficiency u_{it} to be correlated over time for a given individual.

A number of features in these specifications are of interest. Firstly, in contrast with the conventional stochastic frontier literature, our model can distinguish the common correlated effects, $\boldsymbol{\lambda}_i' \boldsymbol{f}_t$, from technical inefficiency, u_{it}. The common correlated effects are used to capture the heterogeneous impacts of unobservable common shocks, such as a sharp global economic downturn. Secondly, an endogeneity problem may arise because unobserved common shocks may affect both firms' input decisions, $\boldsymbol{x}_{it}$, and their outputs, y_{it}.[4] Thirdly, the conventional fixed-effect stochastic frontier models proposed by Greene (2005a, b) and Wang and Ho (2010) are special cases of our specification with $\boldsymbol{f}_t = 1$. Fourthly, compared with ALS, our specification enables us to directly investigate the effects of observed variables $\boldsymbol{z}_{it}$ on inefficiency and then obtains meaningful policy inferences to improve efficiency.[5]

2.2 Estimation

In this section we propose a transformation to control for common correlated effects (referred to as CCE transformation), and then apply the maximum likelihood method to consistently estimate parameters in the stochastic frontier model (1) − (3).

Define

$$\bar{\mathbf{M}}_0 = \mathbf{I}_T - \bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0'\bar{\mathbf{H}}_0)^{-1}\bar{\mathbf{H}}_0',$$

classic endogeneity problem that the firm's optimal choice of inputs will generally be correlated with these unobserved shocks.

[3]Conditional on z_{it}, the scaling property means that technical inefficiency equals some function of exogenous variables times a one-sided error distributed independently of z_{it}; see Wang and Schmidt (2002).

[4]To solve the endogeneity problem, Olley and Pakes (1996) and Levinsohn and Petrin (2003) show 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.

[5]Notice that $\boldsymbol{z}_{it}$ is allowed to include unobserved common shocks, $\boldsymbol{f}_t$.

where

$$\bar{\mathbf{H}}_0 = (\mathbf{D}, \bar{\mathbf{Y}}, \bar{\boldsymbol{h}}_0 \mu^+), \qquad \mu^+ = \left(\mu + \frac{\phi(\frac{-\mu}{\sigma_u})}{1 - \Phi(\frac{-\mu}{\sigma_u})} \sigma_u \right),$$

$\mathbf{D} = (d_1, ..., d_T)' = (1, ..., 1)'$ is a $T \times 1$ vector of ones, $\bar{\mathbf{Y}} = (\bar{\mathbf{y}}, \bar{\mathbf{X}})$ is the cross-sectional average of $(\mathbf{y}_i, \mathbf{X}_i)$, $\bar{\boldsymbol{h}}_0$ denotes the cross-sectional average of $\boldsymbol{h}_i$ evaluated at $\boldsymbol{\delta}_0$, where the subscript "0" is used to denote the parameter is evaluated at true value. μ^+ 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 $\bar{\mathbf{M}}_0$, which depends on the dimension of $\bar{\mathbf{H}}_0 = (\mathbf{D}, \bar{\mathbf{Y}}, \bar{\boldsymbol{h}}_0 \mu_+)$, is $T - \dim(\bar{\mathbf{H}}_0) = T - s$.

Transform (1) by multiplying $\bar{\mathbf{M}}_0$,

$$\bar{\mathbf{M}}_0 \mathbf{y}_i = \bar{\mathbf{M}}_0 \mathbf{X}_i \boldsymbol{\beta} + \bar{\mathbf{M}}_0 \boldsymbol{\varepsilon}_i + \bar{\mathbf{M}}_0 \mathbf{F} \boldsymbol{\lambda}_i, \tag{4}$$

where $\bar{\mathbf{M}}_0 \boldsymbol{\varepsilon}_i = \bar{\mathbf{M}}_0 \boldsymbol{v}_i - \bar{\mathbf{M}}_0 \boldsymbol{u}_i$. In particular, $\boldsymbol{v}_i = (v_{i1}, ..., v_{iT})'$ and $\boldsymbol{u}_i = (u_{i1}, ..., u_{iT})'$, thus, $\bar{\mathbf{M}}_0 \boldsymbol{v}_i \sim N(0, \Pi_0)$, $\Pi_0 = \sigma_v^2 \bar{\mathbf{M}}_0$, and $\bar{\mathbf{M}}_0 \boldsymbol{u}_i = \bar{\mathbf{M}}_0 h\left(\mathbf{z}_i' \boldsymbol{\delta}\right) u_i^*$. Further, $\mathbf{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, ..., \boldsymbol{f}_T)'$ is a $T \times r$ matrix. Since $\bar{\mathbf{M}}_0$ is an idempotent matrix, we solve the non-invertible problem of $\bar{\mathbf{M}}_0$ based on the method of Khatri (1968). In addition, following Wang and Ho (2010), we obtain the conditional log-likelihood function for each i as

$$\begin{aligned} \ln L_i(\boldsymbol{\theta}) = & -\frac{1}{2}(T - s)\left(\ln(2\pi) + \ln \sigma_v^2\right) - \frac{1}{2}\left(\boldsymbol{\varepsilon}_i + \mathbf{F}\boldsymbol{\lambda}_i\right)' \bar{\mathbf{M}}_0 \Pi_0^- \bar{\mathbf{M}}_0 \left(\boldsymbol{\varepsilon}_i + \mathbf{F}\boldsymbol{\lambda}_i\right) \\ & + \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) + \ln\left(\sigma_* \Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \ln\left(\sigma_u \Phi\left(\frac{\mu}{\sigma_u}\right)\right), \end{aligned} \tag{5}$$

where

$$\mu_* = \frac{\mu/\sigma_u^2 - \left(\boldsymbol{\varepsilon}_i + \mathbf{F}\boldsymbol{\lambda}_i\right)' \bar{\mathbf{M}}_0 \Pi_0^- \bar{\mathbf{M}}_0 \boldsymbol{h}_i}{\boldsymbol{h}_i' \bar{\mathbf{M}}_0 \Pi_0^- \bar{\mathbf{M}}_0 \boldsymbol{h}_i + 1/\sigma_u^2} \tag{6}$$

$$\sigma_*^2 = \frac{1}{\boldsymbol{h}_i' \bar{\mathbf{M}}_0 \Pi_0^- \bar{\mathbf{M}}_0 \boldsymbol{h}_i + 1/\sigma_u^2}. \tag{7}$$

The model parameters can be estimated numerically by maximizing the objective function, $\tilde{Q}_{NT}(\theta) = (NT)^{-1} \sum_{i=1}^{N} \ln L_i(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$ is an unknown parameter vector, where d is the number of parameters.

Notice that the above estimation procedure is designed for the production system. For the cost function, the model should be modified as

$$y_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i \boldsymbol{f}_t + v_{it} + u_{it}, \tag{8}$$

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

$$\mu_* = \frac{\mu/\sigma_u^2 + (\boldsymbol{\varepsilon}_i + \mathbf{F}\boldsymbol{\lambda}_i)' \bar{\mathbf{M}}_0 \Pi_0^- \bar{\mathbf{M}}_0 \boldsymbol{h}_i}{\boldsymbol{h}'_i \bar{\mathbf{M}}_0 \Pi_0^- \bar{\mathbf{M}}_0 \boldsymbol{h}_i + 1/\sigma_u^2}.$$

2.3 The Properties of the Proposed Method

By an analogous argument to Pesaran (2006), it is shown that $\bar{\mathbf{M}}_0$ can filter out the common correlated effects. To complete the inferences of consistency and asymptotic normality of the proposed estimator. The following assumptions are used throughout this paper.

Assumptions:

1. The error structure contains v_{it}, $\boldsymbol{e}_{it}$ and u_i^*, which are distributed independently of each other and of the regressors $\boldsymbol{x}_{it}, z_{it}$, $\forall\ i, t$. 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 $\boldsymbol{f}_t$ are covariance stationary with absolute summable autocovariances, distributed independently of v_{it}, $\boldsymbol{e}_{it}$ and u_i^*, $\forall\ i, t$.

3. The unobserved factor loadings $\boldsymbol{\lambda}_i$ with mean λ and $\boldsymbol{\tau}_i$ with mean τ are mutually independent and of v_{it}, $\boldsymbol{e}_{it}$, u_i^*, and the common factors d_t, $\boldsymbol{f}_t$, $\forall\ i, t$. In particular, $\|\boldsymbol{\lambda}_i\|$ and $\|\boldsymbol{\tau}_i\|$ are bounded with finite second moment.

4. The function of the determinants $h(z'_{it}\delta)$ should be assumed to have finite first second, and fourth moments and to be distributed independently of v_{it}, $\boldsymbol{e}_{it}$ and u_i^* $\forall\ i, t$.

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

We rewrite the stochastic frontier model (1) − (3) as

$$\begin{bmatrix} y_{it} \\ \boldsymbol{x}_{it} \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \alpha_i \\ \mathbf{A}_i \end{bmatrix} d_t + \begin{bmatrix} 1 & \boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_i' \\ \boldsymbol{\tau}_i' \end{bmatrix} \boldsymbol{f}_t - \begin{bmatrix} u_{it} \\ \mathbf{0}_{(k\times 1)} \end{bmatrix} + \begin{bmatrix} v_{it} + \boldsymbol{\beta}'\boldsymbol{e}_{it} \\ \boldsymbol{e}_{it} \end{bmatrix}$$

or

$$\mathbf{Y}_{it} = \mathbf{B}_i' d_t + \mathbf{C}_i' \boldsymbol{f}_t - \mathbf{U}_{it} + \boldsymbol{\xi}_{it};$$

here $d_t = 1$. After taking the cross-sectional average under the equal weight, we have

$$\bar{\mathbf{Y}}_t = \bar{\mathbf{B}}' d_t + \bar{\mathbf{C}}' \boldsymbol{f}_t - \bar{\mathbf{U}}_t + \bar{\boldsymbol{\xi}}_t, \tag{9}$$

where $\bar{\mathbf{U}}_t = (\bar{u}_t, \mathbf{0}')'$. In the light of Pesaran (2006), we obtain $\bar{\boldsymbol{\xi}}_t \xrightarrow{\mathbb{P}} 0$ and $\bar{\mathbf{C}} \xrightarrow{\mathbb{P}} \mathbf{C}$ as $N \to \infty$, where $\mathbf{C} = \begin{bmatrix} \boldsymbol{\lambda} & \boldsymbol{\tau} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}' \\ \boldsymbol{\beta} & \mathbf{I}_k \end{bmatrix}$. Under the assumption Rank$(\bar{\mathbf{C}}) = r \leq k+1$, it can be shown that

$$\boldsymbol{f}_t - (\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}(\bar{\mathbf{Y}}_t - \bar{\mathbf{B}}' d_t + \bar{\mathbf{U}}_t) \xrightarrow{\mathbb{P}} 0. \tag{10}$$

Thus, the set $\{\mathbf{D}, \bar{\mathbf{y}}, \bar{\mathbf{X}}, \bar{\mathbf{U}}\}$ can be regarded as the proxy of the factor structure. Based on Pesaran (2006), to proxy the common factors in our model, we could use

$$\bar{\mathbf{H}}^* = [\ \mathbf{D} \quad \bar{\mathbf{y}} \quad \bar{\mathbf{X}} \quad \bar{\boldsymbol{u}}\ ].$$

Notice that u_i^* is not observed in data. To overcome this problem, we propose using $\bar{\boldsymbol{h}}_0 \mu^+$ as a proxy of $\bar{\boldsymbol{u}}$. Under Assumptions 1 and 4, we have

$$\bar{u}_t - \bar{h}_{t,0}\mu^+ \xrightarrow{\mathbb{P}} 0$$

as $N \to \infty$, it follows that

$$\boldsymbol{f}_t - (\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}\left(\bar{\mathbf{Y}}_t - \bar{\mathbf{B}}' d_t + \begin{bmatrix} \bar{h}_{t,0}\mu^+ \\ \mathbf{0} \end{bmatrix}\right) \xrightarrow{\mathbb{P}} 0. \tag{11}$$

By substituting $\bar{\boldsymbol{h}}_0\mu^+$ in $\bar{\mathbf{H}}^*$, we obtain

$$\bar{\mathbf{H}}_0 = [\ \mathbf{D} \quad \bar{\mathbf{y}} \quad \bar{\mathbf{X}} \quad \bar{\boldsymbol{h}}_0\mu^+\ ].$$

The transformed matrix which consists of $\bar{\mathbf{H}}_0$ as we mentioned earlier could work because we construct this matrix by using the true value of $\boldsymbol{\delta}$ and μ^+. However, it is not reasonable to assume that we know these values *ex ante*. Therefore, we shall prove that the deviation of $\boldsymbol{\delta}$ and μ^+ should lead the transformed log-likelihood function not converge to the correctly specified log-likelihood function and less than it with probability one when this deviation is not vanish as the sample size increases. To show this property, we define two log-likelihood functions after transformation by using the transformed matrix $\bar{\mathbf{M}}$. In contrast to $\bar{\mathbf{M}}_0$, here, $\bar{\mathbf{M}}$ denotes the transformed matrix which is evaluated at estimated $\boldsymbol{\delta}$ and μ^+. The first of these two functions is the correctly specified log-likelihood function considering the common correlated effects,

$$Q_{NT}(\boldsymbol{\theta}) = (NT)^{-1}\sum_{i=1}^{N}\left\{-\frac{1}{2}(T-s)\left(\ln(2\pi)+\ln\sigma_v^2\right)-\frac{1}{2}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{F}_0\boldsymbol{\lambda}_{i,0})'\times\right.$$
$$\left.\bar{\mathbf{M}}\Pi^-\bar{\mathbf{M}}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{F}_0\boldsymbol{\lambda}_{i,0})+\frac{1}{2}\left(\frac{\mu_c^2}{\sigma_*^2}-\frac{\mu^2}{\sigma_u^2}\right)+\ln\left(\sigma_*\Phi\left(\frac{\mu_c}{\sigma_*}\right)\right)-\ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right)\right\}, \quad (12)$$

where $\mu_c = \frac{\mu/\sigma_u^2+\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}\Pi^-\bar{\mathbf{M}}\boldsymbol{h}_i}{\boldsymbol{h}_i'\bar{\mathbf{M}}\Pi^-\bar{\mathbf{M}}\boldsymbol{h}_i+1/\sigma_u^2}$. The second one is the "*feasible*" log-likelihood function ignoring those common shocks,

$$\tilde{Q}_{NT}(\boldsymbol{\theta}) = (NT)^{-1}\sum_{i=1}^{N}\left\{-\frac{1}{2}(T-s)\left(\ln(2\pi)+\ln\sigma_v^2\right)-\frac{1}{2}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})'\times\right.$$
$$\left.\bar{\mathbf{M}}\Pi^-\bar{\mathbf{M}}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})+\frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2}-\frac{\mu^2}{\sigma_u^2}\right)+\ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right)-\ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right)\right\}. \quad (13)$$

The main differences of these two functions can be disclosed by focusing on two parts. First, in the mean equation of (12), since we assume that this function is correctly specified by treating the factor structure as an observed structure. As a result, we can put it into the mean equation. Second, because we do not ignore effects from the factor structure, these effects will not enter the error term. Hence, we replace μ_* as we defined before by μ_c. The only difference is that we drop the factor structure. On the other hand, the "*feasible*" log-likelihood function defined in (13), however, is more realistic because we usually cannot observe these factors and their corresponding effects. Thus, as we specified in (13), the factor structure enter the error term instead of entering the mean equation. This log-likelihood function is "*feasible*" because after the transformation, we can prove that the difference between (12) and (13) can be ignored under some assumptions as we mentioned before. We state the main properties of these two functions in the following proposition.

Proposition 1. *Under Assumptions 1-4 and let* $\mathbb{B} = \{\boldsymbol{\theta}_0 + b_{NT}\boldsymbol{d} : \|\boldsymbol{d}\| \leq K\}$*, where* b_{NT} *converges to 0 as* $N, T \to \infty$*. The* "feasible" *log-likelihood function by using the transformed matrix* $\bar{\mathbf{M}}$ *has the following properties:*

1. $|Q_{NT}(\boldsymbol{\theta}) - \tilde{Q}_{NT}(\boldsymbol{\theta})| \xrightarrow{\mathbb{P}} 0$ *when* $\boldsymbol{\theta} \in \mathbb{B}$.

2. $\mathbb{P}[Q_{NT}(\boldsymbol{\theta}_0) - \tilde{Q}_{NT}(\boldsymbol{\theta}) > 0] = 1$*, when* $\boldsymbol{\theta} \in \mathbb{B}^c \cap \boldsymbol{\Theta}$,

as $N, T \to \infty$ *jointly.*

The first result of this proposition indicates that if we can construct an open ball, $\mathbb{B}$, which includes the true value of $\boldsymbol{\theta}$ and the distance between $\boldsymbol{\theta}_0$ and any element in this ball converging to zero, we can show that the "*feasible*" log-likelihood function is uniformly close to the correctly specified likelihood function when $\boldsymbol{\theta} \in \mathbb{B}$. In addition, the second result implies that, with probability one, there is a positive difference between $Q_{NT}(\boldsymbol{\theta}_0)$ and $\tilde{Q}_{NT}(\boldsymbol{\theta})$, and it does not vanish as $N, T \to \infty$. This means that if we consider a candidate solution of $\boldsymbol{\theta} \in \mathbb{B}^c \cap \boldsymbol{\Theta}$, named $\boldsymbol{\theta}'$, we have $Q_{NT}(\boldsymbol{\theta}_0) > \tilde{Q}_{NT}(\boldsymbol{\theta}')$ in probability one. Roughly speaking, we can conclude that $\boldsymbol{\theta}'$ is not maximizing $\tilde{Q}_{NT}$, in other words, $\boldsymbol{\theta}'$ is not the solution of the "*feasible*" likelihood function because we can always find another solution $\boldsymbol{\theta}'' \in \mathbb{B}$ which is more closer to $\boldsymbol{\theta}_0$ to make $\tilde{Q}_{NT}(\boldsymbol{\theta}'')$ close to $Q_{NT}(\boldsymbol{\theta}_0)$. Consequently, these results give the following theorem about the consistency of our "*feasible*" log-likelihood function.

Theorem 1. *Under the Assumptions 1 − 4, and the following conditions (i)* $Q_0(\boldsymbol{\theta})$ *is uniquely maximized at* $\boldsymbol{\theta}_0$*; (ii)* $\boldsymbol{\Theta}$ *is compact; (iii)* Q_0 *and* $\tilde{Q}_0$ *are continuous at* $\boldsymbol{\theta}$ *and (iv)* $Q_{NT}(\boldsymbol{\theta})$ *and* $\tilde{Q}_{NT}(\boldsymbol{\theta})$ *converge uniformly in probability to* $Q_0(\boldsymbol{\theta})$ *and* $\tilde{Q}_0(\boldsymbol{\theta})$ *respectively, then* $\tilde{\boldsymbol{\theta}} \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0$ *as* $N, T \to \infty$ *jointly, where* $\tilde{\boldsymbol{\theta}}$ *is obtained from maximizing the objective function* $\tilde{Q}_{NT}(\boldsymbol{\theta})$.

Theorem 1 shows that, instead of maximizing the correctly specified log-likelihood function, if we maximize the "*feasible*" log-likelihood function, then we can obtain a consistent estimator of $\boldsymbol{\theta}_0$. To investigate the asymptotic behavior of this estimator from maximizing the "*feasible*" log-likelihood function, we should further investigate the behavior of this "*feasible*" function. Since the "*feasible*" function is an approximate function of the true one, we can not apply the traditional method, for example, the mean value theorem to obtain the asymptotic behavior of its estimator. Instead of the well-known method, we first show the difference between $Q_{NT}(\boldsymbol{\theta})$ and $\tilde{Q}_{NT}(\boldsymbol{\theta})$ after multiplying root-NT still converges to zero under certain requirement. Then we apply Lemma 1 which has been proved by Kristensen and Shin (2012) to prove the asymptotic normality of our proposed estimator. We summary the result of the requirement to ensure the stronger convergence of $Q_{NT}(\boldsymbol{\theta})$ and $\tilde{Q}_{NT}(\boldsymbol{\theta})$ as follows:

Proposition 2. *Using assumptions in Theorem 1 we have the following result:*

$\sqrt{NT}|Q_{NT}(\boldsymbol{\theta}) - \tilde{Q}_{NT}(\boldsymbol{\theta})| \xrightarrow{\mathbb{P}} 0$ *when* $\boldsymbol{\theta} \in \mathbb{B}$ *and* $b_{NT} = o_p(C_{NT})$*, where* $C_{NT} = \min\{N^{-1/2}, (NT)^{-1/4}\}$*, as* $N, T \to \infty$ *jointly and* $T/N \to 0$*.*

This result shows the minimum requirement of the converge rate of b_{NT} to guarantee the stronger convergence property of $Q_{NT}(\boldsymbol{\theta})$ and $\tilde{Q}_{NT}(\boldsymbol{\theta})$. Furthermore, according to the results of Caner (2006), we have $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p((NT)^{-1/2})$ by the smoothness of $\tilde{Q}_{NT}(\boldsymbol{\theta})$. Thus, it satisfies the minimum requirement of the converge rate of the ball $\mathbb{B}$ which is supposed to be to have the property of Proposition 2. That is, the difference between $Q_{NT}(\boldsymbol{\theta})$ and $\tilde{Q}_{NT}(\boldsymbol{\theta})$ converges to 0 as $N, T \to \infty$ jointly and $T/N \to 0$. This stronger result is important because it can be used to show that the asymptotic behavior of $\tilde{\boldsymbol{\theta}}$ is asymptotically equivalent to $\hat{\boldsymbol{\theta}}$ obtained from $Q_{NT}(\boldsymbol{\theta})$. We state the above result as the following theorem.

Theorem 2. *Using assumptions in Theorem 1 and additional assumption (L1),* $Q_0(\boldsymbol{\theta})$ *is three times continuously differentiable with its derivatives satisfying, together with* $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p((NT)^{-1/2})$*, which is smaller than* C_{NT}*, we have the following result:*

$\sqrt{NT}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \{\mathbb{E}[-\tilde{H}(\boldsymbol{\theta}_0)]\}^{-1})$*, and* $\mathbb{E}[\tilde{H}(\boldsymbol{\theta}_0)] \xrightarrow{\mathbb{P}} \mathbb{E}[H(\boldsymbol{\theta}_0)]$*,* *as $N, T \to \infty$ jointly and $T/N \to 0$. Here, $\tilde{H}(\boldsymbol{\theta}_0)$ is the Hessian matrix of $\tilde{Q}_0(\boldsymbol{\theta})$ and $H(\boldsymbol{\theta}_0)$ is the Hessian matrix of $Q_0(\boldsymbol{\theta})$at θ_0, respectively.*

Compared with ALS, our estimation allows us to focus on z_{it} that is concerned with measuring inefficiency and treat other unobserved inefficiencies as part of the common correlated effects which can be filtered out by our transformation. According the above asymptotic properties, our estimation still have asymptotic normality and is asymptotically equivalent to the function which treats the factor structure as observed structure. Furthermore, the GMM method can not distinguish between inefficiency and common correlated effects.

2.4 The Inefficiency Index

It is important to measure the inefficiency index in applications. How the inefficiency index can be estimated after the proposed transformation? We follow Wang and Ho (2010), who use the conditional expectation estimator proposed by Jondrow et al. (1982), namely, $E(\boldsymbol{u}_i|\boldsymbol{\varepsilon}_i)$ evaluated at $\boldsymbol{\varepsilon}_i = \hat{\boldsymbol{\varepsilon}}_i$, to construct the inefficiency index. In the same manner, the inefficiency index in our estimation is the conditional expectation of u_{it} on the vector of the transformed $\boldsymbol{\varepsilon}_i = \boldsymbol{v}_i - \boldsymbol{u}_i$, i.e., $\bar{\mathbf{M}}\boldsymbol{\varepsilon}_i$. Note that $\bar{\mathbf{M}}\boldsymbol{\varepsilon}_i$ is evaluated at $\widehat{\bar{\mathbf{M}}\boldsymbol{\varepsilon}_i}$, and following Wang and Ho (2010), the conditional inefficiency index is

$$E\left(\boldsymbol{u}_i|\bar{\mathbf{M}}\boldsymbol{\varepsilon}_i\right) = h(\boldsymbol{z}_i'\boldsymbol{\delta})\left[\mu_* + \frac{\phi\left(\frac{\mu_*}{\sigma_*}\right)\sigma_*}{\Phi\left(\frac{\mu_*}{\sigma_*}\right)}\right] \tag{14}$$

3 Monte Carlo Simulations

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

$$y_{it} = \alpha_i + x_{it}\beta + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + v_{it} - \exp(\boldsymbol{z}_{it}'\boldsymbol{\delta})u_i^* \tag{15}$$

$$x_{it} = \mathbf{A}_i + \boldsymbol{\tau}_i'\boldsymbol{f}_t + e_{it}, \tag{16}$$

where $\alpha_i \sim U(0,1)$, x_{it} is a regressor, $\boldsymbol{f}_t \sim N(0, \sigma_f)$ is a common factor, $\sigma_f^2 = 0.2$, factor loadings $\boldsymbol{\lambda}_i$ and $\boldsymbol{\tau}_i$ follow $N(1, 0.2)$, z_{it} consists of $z_{it,1} \sim N(0,1)$ and $z_{it,2} = t$, which implies the inefficiency is time-varying, $v_{it} \sim N(0, \sigma_v^2)$, $u_i^* \sim N^+(\mu, \sigma_u^2)$, 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 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. We find that 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 three settings with $\sigma_f^2 = 0.1$, 1 and 0, 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. The last case implies the model which has only fixed effects. 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 $T = 5$. 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 $\sigma_f^2 = 0.1$. However, CCE estimator tends to be less effeicienct when the model only contains fixed effects.

We next consider the experiment that both x_{it} and z_{it} are correlated with an unobservable common factor. We set $u_{it} = \exp(\mathbf{z}_{it}'\boldsymbol{\delta})u_i^*$ to ensure that u_{it} is positive. Let

$$\mathbf{z}_{it} = \boldsymbol{\gamma}_i'\boldsymbol{f}_t + \boldsymbol{e}_{z,it}, \tag{17}$$

and z_{it} is correlated with $\boldsymbol{f}_t$. 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, $\boldsymbol{f}_t \sim N(0, 0.6)$ to let factor is important in this model, and each of $\boldsymbol{e}_{z,it}$ 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 $1-4$.

4 Empirical Study

In the years leading up to the 2008 financial crisis, banks in U.S. have been suffered from severe environment. Given this crisis was induced by a rise in subprime mortgage delinquencies and foreclosures, key question is generated concerning banks' performance before said crisis. Among the banks in U.S., two basic types co-exist in the banking market, namely savings and commercial banks. These two types are generally characterized by their ownership structure and the service they provid. In U.S., savings institution could be owed by shareholders (stock), or by their depositors and borrowers (mutual). Based on the agency theory and the property rights theory addressed by the seminal works of Jensen and Meckling (1976) and Fama and Jensen (1983). In contrast to commercial banks who are generally stock corporations, saving banks may not appear to engage in skimping behavior. Particularly in the period before the crisis. As we know, savings banks must hold a certain proportion of their loan portfolio in housing-related assets to preserve their charter. Therefore, these savings banks faced the overbuilding during the boom period, increasing loans and inappropriate government regulation before the financial crisis. In particular, more and more loans to higher-risk borrowers had offered from lenders, thus it may reveals inappropriate managerial behavior of savings banks before the crisis.

Another aim of this paper is to examine the change of efficiency from the baking consolidation. According to the data from Federal Deposit Insurance Corporation(FDIC), the number of commercial banks had fallen to 6,279 at the end of 2011, a drop about 49.1%

from 1990. Similarly, the number of savings institution fell from 2,815 to 1,067 over the same period. It is still a debate between the efficiency and the baking consolidation. In general, the consolidation will increase the market power, and therefore have a decline of competition. From the viewpoint of competitive efficiency, the efficiency of banks should be lower in this scenario. Put differently, an increase in competition will wear bank's pricing power away, increase bank's risk taking behavior, see Berger et al. (2009b) and Beck, Jonghe and Schepens (2013). Hence, increase in competition could lead lower profit and higher cost under the same allocation of input, in other words, cost inefficiency. To explore the relationship between baking consolidation and efficiency, we focus on the banks which are not failure or merged from other banks, in other words, we collect the banks exist over the whole sample period we considered. Build on this situation, we can show, on average, the effects of consolidation without the failure banks .

4.1 Data

We evaluate the cost efficiency of commercial and savings banks in U.S. 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. (2009a) 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 non-interest expenses to total fixed assets and the ratio of interest expenses to total deposits, 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

$$\ln\left(\frac{\mathrm{TC}}{\mathrm{PF}}\right)_{it} = \beta_0 \ln\left(\frac{\mathrm{PC}}{\mathrm{PF}}\right)_{it} + \beta_1 \ln \mathrm{TL}_{it} + \beta_2 \ln \mathrm{OEA}_{it} + \beta_3 \ln \mathrm{TD}_{it} + \beta_4 \ln \mathrm{LA}_{it} + \boldsymbol{\lambda}_i \boldsymbol{f}_t + v_{it} + u_{it}. \tag{18}$$

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

$$h(z_{it}'\delta) = \exp(\delta_1 \ln \mathrm{TA}_{it} + \delta_2 \mathrm{ETA}_{it} + \delta_3 \mathrm{ROAA}_{it} + \mathrm{Type}), \quad (19)$$

where TA denotes the total assets subtracts liquid assests, 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 type dummy variable is also included to capture the effect of different of type of banks.

We consider a balanced panel data set covering 1994-2007 with 223 banks in U.S.. 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 in the left panel of Table 6.[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

[6] We also consider the trend effects while we implementing the Within approach by adding t and t^2 along with intercept to form the idempotent matrix $\mathbf{M}$.

coefficient of OEA is negative, it plays a slightly low effects in contrast to other variables. The empirical results from the Within approach are qualitatively similar to those based on our CCE approach. However, the former tends to deliver smaller estimated coefficients of TL, TD and LA than our approach.

Next, we turn our focus to the coefficients of the inefficiency equation. The coefficient for TA, equal to -0.202, 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 et al. (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.[7] 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. Furthermore, the negative relationship between ROAA and inefficiency is also in line with Lensink et al. (2008).

Though the ROAA should exhibit a negative relationship with inefficiency as pointed out by Lensink et al. (2008), we can not find a strong evidence to link ROAA and efficiency, even the sign is negative and the play a very slight effect.

Furthermore, the type dummy variable for identify the different performance shows the positive effects on commercial banks. The effect is not only statically, but also economically large. The result in Table is equal to -0.263, which provides a strong evidence to show that savings banks is less efficiency than commercial counterparts. It supports that savings banks

[7] Following Berger et al. (2009a), 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.

had poorly managerial behavior before the crisis when they faced overbuilding during the boom period, increasing loans and inappropriate government regulation and did not tend to minimize their cost. On the contrary, commercial banks are more efficient.

Comparing the results from different approaches further reflects the importance of controlling the common correlated effects in the frontier model. The second column of Table from the alternative approach which only take account of the fixed effects provides different results. It shows that the effects of ETA, ROAA and type dummy are completely opposite compared with our results. Despite of the ETA, it is uncanny to explain the relationship between ROAA and efficiency is negative. [8] Moreover, the result implies that the saving banks are efficient which opposes the traditional concept. 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 different estimated value based on the Within approach appears to reflect the fact that ignoring the common correlated effects.

Finally, we further compare the pattern of cost efficiency of savings and commercial banks. Figure 1 plots the average cost efficiency of each group over the 1994-2007 period. Both Within and CCE approaches have upward trend for savings and commercial banks, which implies the bank industry operated more efficient under consolidation. This result may support the evidence that most U.S. banks faced increasing returns recently discussed by Wheelock and Wilson (2012). However, the pattern further shows the difference between savings and commercial banks is relatively small by using Within approach rather than CCE approach. As the figure illustrates, savings banks even more efficient than commercial from Within estimation. As we discussed before, the efficiency may affected by ignoring the common correlated effects, and which gives biased estimated efficiency.

[8] This result is the same as Sun and Chang (2010), while it might be caused by endogeneity.

5 Concluding Remarks

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. In this paper, a stochastic frontier model with unobserved common shocks is develpoed to capture cross-section dependence among individual firms. The novel feature of our model is to separate technical inefficiency from the effects induced by unobserved common shocks and individual heterogeneity. We propose a maximum likelihood method by model transformation that does not require estimating unobserved common correlated effects. With the CCE transformation, we can control the common correlated effects and obtain consistent estimates of parameters 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-section dependence for relatively small T. The desirable results and computational ease should appeal to empirical researchers.

Table 1: Simulation results with cross-section dependence

	$T=5$					$T=10$					$T=20$				
	Within		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)

	$T=5$					$T=10$					$T=20$				
	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-section dependence under different σ_f

(T=10)	$\sigma_f^2 = 0$(only fixed effects)					$\sigma_f^2 = 0.1$					$\sigma_f^2 = 1$				
$N = 50$	Within		CCE			Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.000	0.015	0.000	0.020	0.776	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	0.031	0.004	0.066	0.478	0.001	0.075	0.006	0.077	0.971	0.014	0.126	0.002	0.074	1.694
$\hat{\delta}_2$	0.000	0.007	0.000	0.016	0.446	0.001	0.200	-0.003	0.232	0.862	0.015	0.280	0.004	0.216	1.299
$\hat{\sigma}_v^2$	0.000	0.007	-0.002	0.009	0.763	0.109	0.119	-0.001	0.009	13.655	0.592	0.604	-0.001	0.009	67.373
$\hat{\sigma}_u^2$	0.012	0.103	0.012	0.129	0.801	0.017	0.152	0.009	0.151	1.008	0.077	0.258	0.017	0.158	1.629
$\hat{\mu}$	-0.020	0.150	-0.009	0.172	0.870	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			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.000	0.011	0.000	0.014	0.781	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	0.022	0.005	0.049	0.458	0.001	0.052	0.001	0.050	1.037	0.002	0.096	-0.001	0.051	1.871
$\hat{\delta}_2$	0.000	0.005	-0.001	0.012	0.420	0.009	0.130	0.001	0.162	0.805	0.002	0.182	0.004	0.162	1.125
$\hat{\sigma}_v^2$	0.000	0.005	-0.001	0.006	0.816	0.112	0.123	-0.000	0.006	19.624	0.596	0.607	-0.000	0.006	98.686
$\hat{\sigma}_u^2$	0.003	0.064	0.000	0.081	0.786	0.009	0.103	0.009	0.105	0.982	0.082	0.229	0.011	0.107	2.143
$\hat{\mu}$	0.001	0.092	0.001	0.111	0.833	-0.009	0.128	-0.001	0.125	1.026	-0.040	0.192	-0.004	0.136	1.412

(continued)

(T=10)	$\sigma_f^2 = 0$(only fixed effects)					$\sigma_f^2 = 0.1$					$\sigma_f^2 = 1$				
$N = 200$	Within		CCE			Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.000	0.008	0.000	0.010	0.789	0.089	0.096	-0.000	0.009	10.234	0.430	0.441	0.000	0.010	45.506
$\hat{\delta}_1$	0.000	0.015	-0.001	0.035	0.430	0.002	0.037	0.001	0.037	0.995	0.001	0.068	-0.002	0.037	1.845
$\hat{\delta}_2$	0.000	0.003	0.001	0.010	0.353	0.003	0.094	0.008	0.114	0.824	-0.007	0.133	-0.003	0.115	1.154
$\hat{\sigma}_v^2$	0.000	0.004	-0.001	0.005	0.765	0.111	0.121	0.000	0.004	28.167	0.597	0.608	-0.000	0.004	135.998
$\hat{\sigma}_u^2$	0.002	0.043	0.004	0.059	0.736	0.005	0.068	0.002	0.070	0.978	0.083	0.195	0.009	0.076	2.560
$\hat{\mu}$	-0.001	0.064	0.001	0.081	0.801	-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			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.000	0.006	0.000	0.007	0.803	0.088	0.094	-0.000	0.007	13.411	0.426	0.438	0.000	0.007	65.428
$\hat{\delta}_1$	-0.001	0.011	0.001	0.026	0.416	0.003	0.026	0.001	0.024	1.068	-0.003	0.049	0.002	0.025	1.980
$\hat{\delta}_2$	0.000	0.002	0.000	0.009	0.283	0.002	0.067	-0.000	0.079	0.843	-0.007	0.094	0.004	0.077	1.229
$\hat{\sigma}_v^2$	0.000	0.003	0.000	0.003	0.808	0.109	0.119	-0.000	0.003	38.680	0.594	0.606	-0.000	0.003	193.373
$\hat{\sigma}_u^2$	0.001	0.032	-0.001	0.042	0.755	0.002	0.049	0.000	0.043	1.124	0.086	0.178	0.003	0.046	3.831
$\hat{\mu}$	0.001	0.043	0.000	0.056	0.761	-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-section dependence under different σ_f

(T=20)	$\sigma_f^2 = 0$(only fixed effects)					$\sigma_f^2 = 0.1$					$\sigma_f^2 = 1$				
$N = 50$	Within		CCE			Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.000	0.011	0.001	0.012	0.912	0.089	0.094	0.000	0.011	8.272	0.447	0.453	0.000	0.012	38.062
$\hat{\delta}_1$	0.000	0.007	0.000	0.014	0.466	-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	0.001	0.000	0.002	0.759	0.000	0.171	-0.002	0.194	0.885	0.002	0.209	-0.006	0.193	1.084
$\hat{\sigma}_v^2$	0.000	0.005	-0.001	0.005	0.923	0.110	0.116	-0.000	0.005	22.063	0.626	0.631	-0.000	0.005	122.391
$\hat{\sigma}_u^2$	0.006	0.079	0.004	0.080	0.988	0.010	0.110	-0.001	0.101	1.096	0.054	0.207	-0.000	0.102	2.030
$\hat{\mu}$	-0.015	0.122	-0.014	0.124	0.988	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			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.000	0.008	0.000	0.009	0.909	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	0.004	0.000	0.010	0.437	0.000	0.033	0.000	0.025	1.299	0.000	0.062	0.001	0.025	2.512
$\hat{\delta}_2$	0.000	0.001	0.000	0.001	0.778	0.002	0.118	-0.003	0.135	0.875	-0.008	0.140	-0.006	0.133	1.055
$\hat{\sigma}_v^2$	0.000	0.003	0.000	0.004	0.908	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	0.053	0.003	0.054	0.973	0.003	0.078	-0.002	0.069	1.127	0.041	0.148	-0.002	0.069	2.133
$\hat{\mu}$	-0.009	0.082	-0.008	0.085	0.974	-0.004	0.094	0.003	0.091	1.037	-0.018	0.125	0.004	0.087	1.428

(continued)

(T=20)	$\sigma_f^2 = 0$(only fixed effects)					$\sigma_f^2 = 0.1$					$\sigma_f^2 = 1$				
$N = 200$	Within		CCE			Within		CCE			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.000	0.005	0.000	0.006	0.906	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	0.003	0.000	0.007	0.437	-0.000	0.023	-0.000	0.017	1.314	-0.001	0.044	-0.000	0.017	2.558
$\hat{\delta}_2$	0.000	0.001	0.000	0.001	0.681	0.003	0.085	0.004	0.090	0.940	-0.003	0.101	-0.001	0.094	1.081
$\hat{\sigma}_v^2$	0.000	0.002	0.000	0.003	0.902	0.110	0.115	-0.000	0.003	44.603	0.631	0.636	0.000	0.003	244.367
$\hat{\sigma}_u^2$	0.000	0.035	0.000	0.037	0.966	0.005	0.052	-0.001	0.045	1.165	0.040	0.116	0.000	0.049	2.383
$\hat{\mu}$	-0.003	0.053	-0.003	0.054	0.977	-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			Within		CCE		
	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ	Bias	RMSE	Bias	RMSE	ψ
$\hat{\beta}$	0.000	0.004	0.000	0.004	0.906	0.088	0.091	-0.000	0.004	22.551	0.444	0.449	-0.000	0.004	107.029
$\hat{\delta}_1$	0.000	0.002	0.000	0.005	0.441	0.001	0.015	-0.000	0.012	1.291	0.001	0.030	0.000	0.013	2.374
$\hat{\delta}_2$	0.000	0.001	0.000	0.001	0.593	0.007	0.059	0.004	0.065	0.919	-0.003	0.075	-0.002	0.066	1.126
$\hat{\sigma}_v^2$	0.000	0.002	0.000	0.002	0.911	0.109	0.114	-0.000	0.002	62.508	0.635	0.639	0.000	0.002	354.291
$\hat{\sigma}_u^2$	0.001	0.025	0.001	0.026	0.956	0.002	0.035	0.001	0.032	1.097	0.031	0.082	0.001	0.033	2.524
$\hat{\mu}$	-0.002	0.040	-0.002	0.040	0.978	-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: x and z are affected by an unobservable common shock (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	1.11×10^3	4.60×10^3	4.10	8.08×10^4
Output quantities				
Total loans	1.06×10^4	4.29×10^4	42.60	6.77×10^5
Other earning assets	5.77×10^3	3.30×10^4	0.50	6.92×10^5
Total deposits	1.20×10^4	5.21×10^4	1.80	7.94×10^5
Liquid assets	3.28×10^3	2.59×10^4	0.10	6.48×10^6
Input prices				
Price of capital	0.04	0.05	1.99×10^{-3}	1.24
Price of funds	5.29	1.48×10^3	0.34	7.56×10^4
Other variables' quantity and ratios				
Total assets	1.86×10^4	8.31×10^4	62.00	1.32×10^6
Return on average assets	1.32	1.20	-3.18	24.04
Equity to assets	9.77	5.39	4	82.36

[1] The variables in total cost and output quantities are measured in U.S. $ millions.

[2] There are a total of 3,122 bank-year observations.

Table 6: Estimation results of the cost frontier

		Within		CCE	
	Exp. Sign	$\hat{\theta}$	Std. Dev.	$\hat{\theta}$	Std. Dev.
Effects on cost function					
ln(PC/PF)	(+)	0.371 ***	0.018	0.184 ***	0.006
ln(TL)	(+)	0.033 *	0.019	0.216 ***	0.015
ln(OEA)	(+)	0.002	0.018	-0.012 **	0.005
ln(TD)	(+)	0.696 ***	0.020	0.861 ***	0.014
ln(LA)	(+)	0.048 ***	0.018	0.027 ***	0.003
Effects on inefficiency					
ln(TA)	(?)	-0.347 ***	0.022	-0.202 ***	0.018
ETA	(+)	-0.166 **	0.018	0.068 ***	0.010
ROAA	(-)	0.470 ***	0.024	-0.005 ***	0.002
TYPE	(-)	0.117 ***	0.018	-0.263 ***	0.085
σ_v^2		0.153		0.004	
σ_u^2		518.796		38.127	

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

[2] Exp. Sign explains the expected relationship between inefficiency and variables.

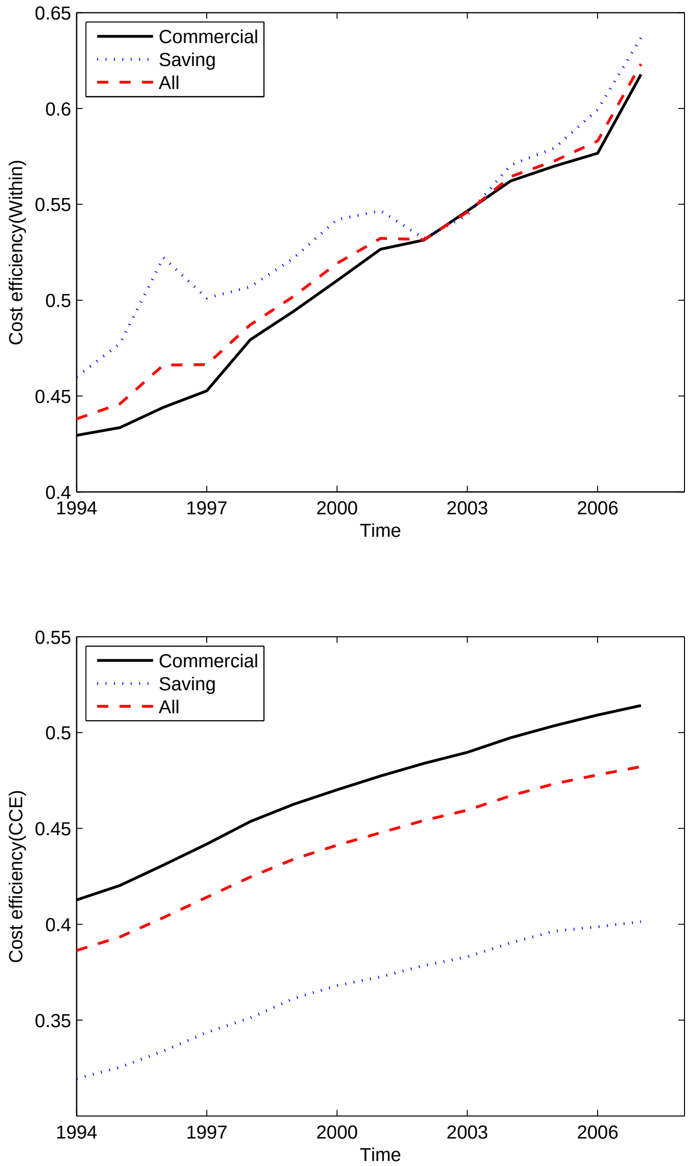

Figure 1: Average Cost Efficiency in All Banks

Useful Lemmas

Below we introduce some useful lemmas for proving the main results in our paper. The proof can be founded in the Supplementary Material.

Assumption for Lemma 1:

(L1) $Q_0(\boldsymbol{\theta})$ is three times continuously differentiable with its derivatives satisfying

$$\sqrt{NT}S(\boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \{E[-H(\boldsymbol{\theta}_0)]\}^{-1}),$$
$$H(\boldsymbol{\theta}_0) \xrightarrow{\mathbb{P}} E[H(\boldsymbol{\theta}_0)],$$
$$\max_{j=1,\dots,d} \sup_{\boldsymbol{\theta}} || \frac{\partial Q_0(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}_j} || = O_p(1),$$

where $S(\boldsymbol{\theta}_0) = \frac{\partial Q_0(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}_0}$ and $H(\boldsymbol{\theta}_0) = \frac{\partial Q_0(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}|_{\boldsymbol{\theta}_0}$.

Lemma 1. *As assumption (L1) holds with Θ which is compact, and $\sqrt{NT}\sup_{\boldsymbol{\theta}} |\tilde{Q}_{NT}(\boldsymbol{\theta}) - Q_{NT}(\boldsymbol{\theta})| = o_p(1)$ as $N, T \to \infty$. Then $\sqrt{NT}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \{E[-H(\boldsymbol{\theta}_0)]\}^{-1})$.*

Throughout the following lemmas, we use the following notations: $\bar{\boldsymbol{\xi}} = (\bar{\boldsymbol{\xi}}_1, ..., \bar{\boldsymbol{\xi}}_T)'$, $\boldsymbol{\xi}_i = (\boldsymbol{\xi}_{i1}, ..., \boldsymbol{\xi}_{iT})'$, $\boldsymbol{v}_i = (v_{i1}, ..., v_{iT})'$, $\bar{\boldsymbol{u}} = (\bar{\boldsymbol{u}}_1, ..., \bar{\boldsymbol{u}}_T)$, $\bar{\boldsymbol{\zeta}} = \bar{\boldsymbol{h}}_0\mu^+ - \bar{\boldsymbol{u}}$, $\bar{u}_i = T^{-1}\sum_{t=1}^T u_{it}$, $\bar{h}_i = T^{-1}\sum_{t=1}^T h_{it}$ and $\mathbf{G} = [\ \mathbf{D} \quad \mathbf{F} \quad \bar{\boldsymbol{u}}\ ]$. Recall that $\bar{\mathbf{H}}_0 = \left[\mathbf{D}, \bar{\mathbf{y}}, \bar{\mathbf{X}}, \bar{\boldsymbol{h}}_0\mu^+\right]$, together with equation (9), $\bar{\mathbf{H}}_0$ can be rewritten as $\bar{\mathbf{H}}_0 = \left[\mathbf{G}\bar{\mathbf{P}} + \bar{\boldsymbol{\psi}} + \bar{\boldsymbol{\xi}}^*\right]$, where

$$\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{1} & \bar{\mathbf{B}} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{C}} & \mathbf{0} \\ 0 & \mathbf{1}_{1\times(k+1)} & 1 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{D} & \mathbf{F} & \bar{\boldsymbol{u}} \end{bmatrix}$$

and

$$\bar{\boldsymbol{\xi}}^* = \begin{bmatrix} \mathbf{0}_{(T\times 1)} & \bar{\boldsymbol{\xi}} & \mathbf{0}_{(T\times 1)} \end{bmatrix}, \quad \bar{\boldsymbol{\psi}} = \begin{bmatrix} \mathbf{0}_{(T\times 1)} & \mathbf{0}_{(T\times(k+1))} & \bar{\boldsymbol{\zeta}} \end{bmatrix}.$$

Lemma 2. *As assumptions 1–4 hold, we have*

(B1) $T^{-1}\bar{\boldsymbol{\xi}}'\bar{\boldsymbol{\xi}} = O_p(N^{-1})$;

(B2) $T^{-1}\boldsymbol{\xi}_i'\bar{\boldsymbol{\xi}} = O_p(N^{-1}) + O_p((NT)^{-1/2})$;

(B3) $T^{-1}\mathbf{D}'\bar{\boldsymbol{\xi}} = O_p((NT)^{-1/2})$;

(B4) $T^{-1}\mathbf{F}'\bar{\boldsymbol{\xi}} = O_p((NT)^{-1/2})$;

(B5) $T^{-1}\mathbf{D}'\boldsymbol{v}_i = O_p(T^{-1/2})$;

(B6) $T^{-1}\mathbf{F}'\boldsymbol{v}_i = O_p(T^{-1/2})$.

Lemma 3. *As assumptions 1–4 hold, we have*

(C1) $T^{-1}\left(\bar{\boldsymbol{u}}'\bar{\boldsymbol{\xi}}\right) = O_p\left((NT)^{-1/2}\right)$;

(C2) $T^{-1}\left(\bar{\boldsymbol{\xi}}'\bar{\boldsymbol{\zeta}}\right) = O_P(N^{-1}T^{-1/2})$;

(C3) $T^{-1}\left(\mathbf{G}'\bar{\boldsymbol{\zeta}}\right) = O_p(N^{-1/2})$;

(C4) $T^{-1}\left(\bar{\boldsymbol{\zeta}}'\bar{\boldsymbol{\zeta}}\right) = O_p(N^{-1})$;

(C5) $T^{-1}\left(\boldsymbol{\xi}_i\bar{\boldsymbol{\zeta}}\right) = O_p((NT)^{-1/2})$;

(C6) $T^{-1}\left(\boldsymbol{\xi}_i'\mathbf{G}\right) = O_p(T^{-1/2})$;

(C7) $T^{-1}\left((\boldsymbol{u}_i - \bar{u}_i)'\bar{\boldsymbol{\xi}}\right) = O_p((NT)^{-1/2})$;

(C8) $T^{-1}\left((\boldsymbol{u}_i - \bar{u}_i)'\bar{\boldsymbol{\zeta}}\right) = O_p(N^{-1}) + O_p((NT)^{-1/2})$;

(C9) $T^{-1}\left((\boldsymbol{u}_i - \bar{u}_i)'\mathbf{G}\right) = O_p(N^{-1}) + O_p(T^{-1/2})$;

(C10) $T^{-1}(\boldsymbol{h}_i - \bar{h}_i)'N^{-1}\sum_{j=1}^{N}(\boldsymbol{h}_j - \bar{h}_j)(u_j^* - \mu^+) = O_p(N^{-1}) + O_p((NT)^{-1/2})$.

Proof of Main Propositions and Theorems

Recall the transformed log-likelihood functions of (12) and (13),

$$\begin{aligned} Q_{NT}(\boldsymbol{\theta}) = (NT)^{-1}\sum_{i=1}^{N}\Bigg\{ -\frac{1}{2}(T-s)\left(\ln(2\pi)+\ln\sigma_v^2\right) - \frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{F}_0\boldsymbol{\lambda}_{i,0})' \times \\ \bar{\mathbf{M}}\Pi^{-}\bar{\mathbf{M}}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{F}_0\boldsymbol{\lambda}_{i,0}) + \frac{1}{2}\left(\frac{\mu_c^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) + \ln\left(\sigma_*\Phi\left(\frac{\mu_c}{\sigma_*}\right)\right) - \ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right)\Bigg\}, \end{aligned} \tag{12}$$

and

$$\begin{aligned} \tilde{Q}_{NT}(\boldsymbol{\theta}) = (NT)^{-1}\sum_{i=1}^{N}\Bigg\{ -\frac{1}{2}(T-s)\left(\ln(2\pi)+\ln\sigma_v^2\right) - \frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})' \times \\ \bar{\mathbf{M}}\Pi^{-}\bar{\mathbf{M}}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) + \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) + \ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right)\Bigg\}. \end{aligned} \tag{13}$$

Proof of Proposition 1. To complete the proof of Proposition 1, we separate (13) into five parts:

$$\begin{aligned} P1 &= (NT)^{-1}\sum_{i=1}^{N}\left\{-\frac{1}{2}(T-s)\left(\ln(2\pi)+\ln\sigma_v^2\right)\right\}, \\ P2 &= (NT)^{-1}\sum_{i=1}^{N}\left\{-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\bar{\mathbf{M}}\Pi^{-}\bar{\mathbf{M}}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right\}, \\ P3 &= (NT)^{-1}\sum_{i=1}^{N}\left\{\frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right)\right\}, \\ P4 &= (NT)^{-1}\sum_{i=1}^{N}\left\{\ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right)\right\}, \\ P5 &= (NT)^{-1}\sum_{i=1}^{N}\left\{-\ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right)\right\}. \end{aligned}$$

Since $P1$ and $P5$ are the same as part of (12), we only need to investigate the differences of $P2$, $P3$ and $P4$ between (12) and (13).

Consider $P2$. By the facts that $\mathbf{y}_i = \mathbf{D}\alpha_i + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i$, $\bar{\mathbf{M}}\mathbf{D}\alpha_i = \mathbf{0}$ and $\bar{\mathbf{M}}\Pi^{-}\bar{\mathbf{M}} = \sigma_v^{-2}\bar{\mathbf{M}}$,

$P2$ can be rewritten as,

$$
\begin{aligned}
&(NT)^{-1}\sum_{i=1}^{N}\left\{-\frac{1}{2}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})'\bar{\mathbf{M}}\Pi^{-}\bar{\mathbf{M}}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\right\}\\
&=\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left\{-\frac{1}{2}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})'\bar{\mathbf{M}}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\right\}\\
&=\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left\{-\frac{1}{2}\left(\mathbf{F}_0\boldsymbol{\lambda}_{i0}+\boldsymbol{\varepsilon}_i+\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}\left(\mathbf{F}_0\boldsymbol{\lambda}_{i0}+\boldsymbol{\varepsilon}_i+\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)\right\}\\
&=\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left\{-\frac{1}{2}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)\right\}\\
&-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0}\right\}-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}\boldsymbol{\varepsilon}_i\right\}\\
&-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0}-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\frac{1}{2}\boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0}-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\frac{1}{2}\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}\boldsymbol{\varepsilon}_i\\
&=: A_1(\boldsymbol{\theta})+A_2(\boldsymbol{\theta})+A_3(\boldsymbol{\theta})+A_4(\boldsymbol{\theta})+A_5(\boldsymbol{\theta})+A_6(\boldsymbol{\theta}).
\end{aligned}
$$

Particularly, $A_1(\boldsymbol{\theta})$, $A_3(\boldsymbol{\theta})$ and $A_6(\boldsymbol{\theta})$ do not affected by the factor structure, therefore we will focus on the properties of $A_2(\boldsymbol{\theta})$, $A_4(\boldsymbol{\theta})$ and $A_5(\boldsymbol{\theta})$ respectively. For $A_2(\boldsymbol{\theta})$,

$$
\begin{aligned}
A_2(\boldsymbol{\theta})=&-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0}\\
=&-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}\mathbf{F}_0(\bar{\boldsymbol{\lambda}}-\bar{\boldsymbol{\eta}})-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\eta}_{i0}\\
=&-\sigma_v^{-2}T^{-1}(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\bar{\mathbf{X}}'\bar{\mathbf{M}}\mathbf{F}_0(\bar{\boldsymbol{\lambda}}-\bar{\boldsymbol{\eta}})-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\eta}_{i0}\\
=&0--\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}_0\mathbf{F}_0\boldsymbol{\eta}_{i0}-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT}\\
=&: A_{2,1}(\boldsymbol{\theta})+A_{2,2}(\boldsymbol{\theta}).
\end{aligned}
$$

Since $\boldsymbol{\lambda}_{i0}=\boldsymbol{\eta}+\boldsymbol{\eta}_{i0}$, after taking cross-sectional average of $\boldsymbol{\lambda}_{i0}$, we have $\bar{\boldsymbol{\lambda}}=\boldsymbol{\eta}+\bar{\boldsymbol{\eta}}$. The second equality holds by replacing $\boldsymbol{\lambda}_{i0}$ by $\bar{\boldsymbol{\lambda}}-\bar{\boldsymbol{\eta}}+\boldsymbol{\eta}_{i0}$. The fourth equality holds because $\bar{\mathbf{M}}\bar{\mathbf{X}}=0$ and the fifth equality holds because $\bar{\mathbf{M}}=\bar{\mathbf{M}}_0+\kappa_{NT}$, where $\kappa_{NT}=O(b_{NT})$ by $\boldsymbol{\theta}\in\mathbb{B}$. Note that for easy to state, we use $A_{2,1}(\boldsymbol{\theta})$ and $A_{2,2}(\boldsymbol{\theta})$ to denote the rest of terms we need to discuss.

Consider $A_{2,1}(\boldsymbol{\theta})$, because of the fact that $\mathbf{F}_0=-(\bar{\boldsymbol{\xi}}+\bar{\mathbf{U}})\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}$ from equation (10) and $\bar{\mathbf{U}}=(\bar{\mathbf{U}}_1,...,\bar{\mathbf{U}}_T)'$, we have

$$
\begin{aligned}
A_{2,1}(\boldsymbol{\theta}) =& \sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}_0(\bar{\boldsymbol{\xi}}+\bar{\mathbf{U}})\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\boldsymbol{\eta}_{i0}\\
=& \sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}_0\bar{\boldsymbol{\xi}}\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\boldsymbol{\eta}_{i0}\\
&+\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}_0\bar{\mathbf{U}}\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\boldsymbol{\eta}_{i0}.
\end{aligned}
$$

The property of the first term can be obtained from the fact that $\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}$ is bounded and the result that $\frac{(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\mathbf{X}_i'\bar{\mathbf{M}}_0\sqrt{N}\bar{\boldsymbol{\xi}}}{\sqrt{T}} = O_P(1)$ proved by Pesaran (2006). Therefore, with $\boldsymbol{\eta}_{i0}$ which is distributed independently of $\mathbf{X}_i$, $\bar{\boldsymbol{\xi}}$ and elements in $\bar{\mathbf{M}}_0$, we have

$$
N^{-1}\sum_{i=1}^{N}\frac{(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\mathbf{X}_i'\bar{\mathbf{M}}_0\sqrt{N}\bar{\boldsymbol{\xi}}}{\sqrt{T}}\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\boldsymbol{\eta}_{i0} = O_p(N^{-1/2}),
$$

that is $\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right)'\bar{\mathbf{M}}_0\bar{\boldsymbol{\xi}}\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\boldsymbol{\eta}_{i0} = O_p(N^{-1}T^{-1/2})$. We can prove the second term in $A_{2,1}(\boldsymbol{\theta})$ in the similar way because $\bar{\mathbf{M}}_0\bar{\mathbf{U}} = \bar{\mathbf{M}}_0(\bar{\mathbf{U}}-[\bar{\boldsymbol{h}}_0\mu^+,\mathbf{0}]) = [\bar{\mathbf{M}}_0\bar{\boldsymbol{\zeta}},\mathbf{0}]$, and

$$
N^{-1}\sum_{i=1}^{N}\frac{(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\mathbf{X}_i'\bar{\mathbf{M}}_0\sqrt{N}\bar{\boldsymbol{\zeta}}}{\sqrt{T}}\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\boldsymbol{\eta}_{i0} = O_p(N^{-1/2}).
$$

Thus, we have $A_{2,1}(\boldsymbol{\theta}) = O_p(N^{-1}T^{-1/2})$.

Next, consider $A_{2,2}(\boldsymbol{\theta})$. We have

$$
\begin{aligned}
A_{2,2}(\boldsymbol{\theta}) =& -\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\mathbf{X}_i'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT}\\
=& -\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\beta}_0-\boldsymbol{\beta})'(\mathbf{F}_0\boldsymbol{\tau}_{i0}+\boldsymbol{e}_i)'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT}\\
=& -\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\boldsymbol{\tau}_{i0}'\mathbf{F}_0'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT}-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\boldsymbol{e}_i'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT},
\end{aligned}
$$

where the first equality comes from facts that $\mathbf{X}_i = \mathbf{D}\mathbf{A}_i'+\mathbf{F}\boldsymbol{\tau}_i+\boldsymbol{e}_i$ and $\mathbf{D}\mathbf{A}_i'$ has been removed by $\bar{\mathbf{M}}$. The first term of the last equation can be rearranged as $-\sigma_v^{-2}(N)^{-1}\sum_{i=1}^{N}(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\boldsymbol{\tau}_{i0}'\frac{\mathbf{F}_0'\mathbf{F}_0}{T}\boldsymbol{\eta}_{i0}\kappa_{NT}$. Since $\frac{\mathbf{F}_0'\mathbf{F}_0}{T} = O_p(1)$ and $\boldsymbol{\eta}_{i0}$ is distributed independently of $\boldsymbol{\tau}_{i0}$ and $\mathbf{F}_0$, we have $-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\boldsymbol{\tau}_{i0}'\mathbf{F}_0'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT} = O_p(N^{-1/2}b_{NT})$. Further, according to the result of $\frac{\mathbf{e}_i'\mathbf{F}_0}{T} = O_p(T^{-1/2})$ and the property of $\boldsymbol{\eta}_{i0}$ we used before. We can show that $-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\boldsymbol{e}_i'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT} =$

$O_p((NT)^{-1/2}b_{NT})$. Combining these results, we have $A_{2,2}(\boldsymbol{\theta}) = O_p(N^{-1/2}b_{NT})$. Therefore, $A_2(\boldsymbol{\theta}) = O_P(N^{-1}T^{-1/2}) + O_p(N^{-1/2}b_{NT})$.

For $A_5(\boldsymbol{\theta})$, using the same fact that $\boldsymbol{\lambda}_{i0} = \bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}} + \boldsymbol{\eta}_{i0}$, we have

$$\begin{aligned}
A_5(\boldsymbol{\theta}) &= -\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0} \\
&= -\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}}) - \sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\eta}_{i0} \\
&= -\sigma_v^{-2}T^{-1}\bar{\boldsymbol{\varepsilon}}'\bar{\mathbf{M}}\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}}) - \sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\eta}_{i0} \\
&=: A_{5,1}(\boldsymbol{\theta}) + A_{5,2}(\boldsymbol{\theta}).
\end{aligned}$$

In particular,

$$\begin{aligned}
A_{5,1}(\boldsymbol{\theta}) &= -\sigma_v^{-2}T^{-1}\bar{\boldsymbol{\varepsilon}}'\bar{\mathbf{M}}_0\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}}) - \sigma_v^{-2}T^{-1}\bar{\boldsymbol{\varepsilon}}'\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}})\kappa_{NT} \\
&= -\sigma_v^{-2}T^{-1}(\bar{\boldsymbol{v}} + (\bar{\boldsymbol{u}} - \bar{\boldsymbol{h}}_0\mu^+))'\bar{\mathbf{M}}_0\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}}) - \sigma_v^{-2}T^{-1}(\bar{\boldsymbol{v}} + \bar{\boldsymbol{u}})'\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}})\kappa_{NT}.
\end{aligned}$$

We can rewrite the first term of the above equation as

$$\begin{aligned}
&-\sigma_v^{-2}T^{-1}(\bar{\boldsymbol{v}} + (\bar{\boldsymbol{u}} - \bar{\boldsymbol{h}}_0\mu^+))'\bar{\mathbf{M}}_0\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}}) \\
&= -\sigma_v^{-2}T^{-1}\bar{\boldsymbol{v}}'\bar{\mathbf{M}}_0\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}}) - \sigma_v^{-2}T^{-1}(\bar{\boldsymbol{u}} - \bar{\boldsymbol{h}}_0\mu^+)'\bar{\mathbf{M}}_0\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}}).
\end{aligned}$$

Using the fact $\frac{1}{T}\bar{\boldsymbol{v}}'\bar{\mathbf{M}}_0\mathbf{F}_0 = \frac{1}{T}\bar{\boldsymbol{v}}'\mathbf{F}_0 - \frac{1}{T}\bar{\boldsymbol{v}}\bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0')^{-1}\bar{\mathbf{H}}_0'\mathbf{F}_0$ and the results from lemmas (B1), (B3), (B4), (C1)-(C4) and the fact $\bar{\boldsymbol{v}} = O_p(N^{-1/2})$, we have

$$\begin{aligned}
&\frac{1}{T}\bar{\boldsymbol{v}}'\bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0')^{-1}\bar{\mathbf{H}}_0'\mathbf{F}_0 = \frac{\bar{\boldsymbol{v}}'\bar{\mathbf{H}}_0}{T}\left(\frac{\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0'}{T}\right)^{-1}\frac{\bar{\mathbf{H}}_0'\mathbf{F}_0}{T} \\
&= \left(\underbrace{\frac{\bar{\boldsymbol{v}}'\mathbf{G}}{T}\bar{\mathbf{P}}}_{O_p(N^{-1/2})} + \underbrace{\frac{\bar{\boldsymbol{v}}'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T}}_{O_p(N^{-1})}\right)\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}} + \underbrace{\bar{\mathbf{P}}'\frac{\mathbf{G}'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T} + \frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'\mathbf{G}}{T}\bar{\mathbf{P}}}_{O_p(N^{-1/2})} + \underbrace{\frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T}}_{O_p(N^{-1})}\right)^{-1} \\
&\quad\times\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}_0}{T} + \underbrace{\frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'\mathbf{F}_0}{T}}_{O_p((NT)^{-1/2})}\right) \\
&= = \frac{\bar{\boldsymbol{v}}'\mathbf{G}}{T}\bar{\mathbf{P}}\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}}\right)^{-1}\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}_0}{T} + O_p(N^{-1}).
\end{aligned}$$

Notice that we keep the first term of the above equation to illustrate the fact that $\frac{1}{T}\bar{\boldsymbol{v}}'\mathbf{F}_0 - \frac{\bar{\boldsymbol{v}}'\mathbf{G}}{T}\bar{\mathbf{P}}\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}}\right)^{-1}\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}_0}{T} = \frac{1}{T}\bar{\boldsymbol{v}}'\bar{\mathbf{M}}_{\mathbf{G}}\mathbf{F}_0 = \mathbf{0}$ because $\mathbf{F}_0 \in \mathbf{G}$. Combining these results, we have $\frac{1}{T}\bar{\boldsymbol{v}}'\bar{\mathbf{M}}_0\mathbf{F}_0 = O_p(N^{-1})$. In the same manner, we have $\frac{1}{T}(\bar{\boldsymbol{u}} - \bar{\boldsymbol{h}}_0\mu^+)'\bar{\mathbf{M}}_0\mathbf{F}_0 = O_p(N^{-1})$. In addition, the term, $\sigma_v^{-2}T^{-1}(\bar{\boldsymbol{v}} + \bar{\boldsymbol{u}})'\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}})\kappa_{NT}$ is needed to investigate. The property of this term can be obtained by using $\frac{1}{T}\bar{\boldsymbol{v}}'\mathbf{F}_0 = O_p((NT)^{-1/2})$ and $\frac{1}{T}\bar{\boldsymbol{u}}'\mathbf{F}_0 = O_p(T^{-1/2})$. Thus, $\sigma_v^{-2}T^{-1}(\bar{\boldsymbol{v}} + \bar{\boldsymbol{u}})'\mathbf{F}_0(\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\eta}})\kappa_{NT} = O_p(T^{-1/2}b_{NT})$. These give $A_{5,1}(\boldsymbol{\theta}) = O_p(N^{-1}) + O_p(T^{-1/2}b_{NT})$.

Next, consider $A_{5,2}(\boldsymbol{\theta})$,

$$A_{5,2}(\boldsymbol{\theta}) = -\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\bar{\mathbf{M}}_0\mathbf{F}_0\boldsymbol{\eta}_{i0} - \sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT}.$$

The first term of $A_{5,2}(\boldsymbol{\theta})$ can be decomposed into $\frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\bar{\mathbf{M}}_0\mathbf{F}_0 = \frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{F}_0 - \frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0')^{-1}\bar{\mathbf{H}}_0'\mathbf{F}_0$ and using lemmas (B2), (C5) and (C7)-(C8), with $\frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{G} = O_p(N^{-1}) + O_p(T^{-1/2})$ by lemmas (C6) and (C9), we have

$$\begin{aligned}
&\frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0')^{-1}\bar{\mathbf{H}}_0'\mathbf{F}_0 \\
&= \Bigg(\underbrace{\frac{(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{G}}{T}\bar{\mathbf{P}}}_{O_p(N^{-1})+O_p(T^{-1/2})} + \underbrace{\frac{(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T}}_{O_p(N^{-1})+O_p((NT)^{-1/2})}\Bigg)\Bigg(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}} + \underbrace{\bar{\mathbf{P}}'\frac{\mathbf{G}'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T} + \frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'\mathbf{G}}{T}\bar{\mathbf{P}}}_{O_p(N^{-1/2})} + \underbrace{\frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T}}_{O_p(N^{-1})}\Bigg)^{-1} \\
&\times \Bigg(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}_0}{T} + \underbrace{\frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'\mathbf{F}_0}{T}}_{O_p((NT)^{-1/2})}\Bigg) \\
&= \frac{(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{G}}{T}\bar{\mathbf{P}}\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}}\right)^{-1}\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}_0}{T} + O_p(N^{-3/2}) + O_p((NT)^{-1/2}).
\end{aligned}$$

Similarly, we keep the first interaction term, together with $\frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{F}_0$, then we have $\frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{F}_0 - \frac{(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{G}}{T}\bar{\mathbf{P}}\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}}\right)^{-1}\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}_0}{T} = \frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\bar{\mathbf{M}}_{\mathbf{G}}\mathbf{F}_0 = \mathbf{0}$. Thus, $\frac{1}{T}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\bar{\mathbf{M}}_0\mathbf{F}_0 = O_p(N^{-3/2}) + O_p((NT)^{-1/2})$. Since $\boldsymbol{\eta}_{i0}$ is distributed independently of $\mathbf{F}_0$, $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$, we can conclude that $-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\bar{\mathbf{M}}_0\mathbf{F}_0\boldsymbol{\eta}_{i0} = O_p(N^{-2}) + O_p(N^{-1}T^{-1/2})$. Finally, $\frac{1}{T}\boldsymbol{v}_i'\mathbf{F}_0 = \frac{1}{T}(\boldsymbol{u}_i - \bar{u}_i)'\mathbf{F}_0 = O_p(T^{-1/2})$, we therefore have $-\sigma_v^{-2}(NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{\varepsilon}_i - \bar{\varepsilon}_i)'\mathbf{F}_0\boldsymbol{\eta}_{i0}\kappa_{NT} = O_p((NT)^{-1/2}b_{NT})$. Taking these results from $A_{5,1}(\boldsymbol{\theta})$ and $A_{5,2}(\boldsymbol{\theta})$, we have $A_5(\boldsymbol{\theta}) = O_p(N^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p((NT)^{-1/2}b_{NT})$.

Now, consider $A_4(\boldsymbol{\theta})$. By using the following inequality

$$\|\frac{1}{T}\boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0}\| = \|\frac{1}{T}\boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\bar{\mathbf{M}}\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0}\| \leq \frac{1}{T}\sum_{t=1}^{T}\|\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0(t)}\|^2,$$

where $\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0(t)}$ denotes the $t-$th element of $\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0}$. Since

$$
\begin{aligned}
\boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\bar{\mathbf{M}} &= \boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\kappa_{NT} + \boldsymbol{\lambda}_{i0}'\mathbf{F}_0' - \boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0')^{-1}\bar{\mathbf{H}}_0' \\
&= \boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\kappa_{NT} + \boldsymbol{\lambda}_{i0}'\mathbf{F}_0' - \boldsymbol{\lambda}_{i0}'\left(\frac{\mathbf{F}_0'\mathbf{G}}{T}\bar{\mathbf{P}} + \underbrace{\frac{\mathbf{F}_0'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T}}_{O_p((NT)^{-1/2})}\right) \\
&\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}} + \underbrace{\bar{\mathbf{P}}'\frac{\mathbf{G}'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T} + \frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'\mathbf{G}}{T}\bar{\mathbf{P}}}_{O_p(N^{-1/2})} + \underbrace{\frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T}}_{O_p(N^{-1})}\right)^{-1} \times \left(\bar{\mathbf{P}}'\mathbf{G}' + \underbrace{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}_{O_P(N^{-1/2})}\right) \\
&= \boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\kappa_{NT} + \boldsymbol{\lambda}_{i0}'\mathbf{F}_0' - \boldsymbol{\lambda}_{i0}'\frac{\mathbf{F}_0'\mathbf{G}}{T}\bar{\mathbf{P}}\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}}\right)^{-1}\bar{\mathbf{P}}'\mathbf{G}' + O_p(N^{-1/2}) \\
&= \boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\kappa_{NT} + \boldsymbol{\lambda}_{i0}'\mathbf{F}_0'\bar{\mathbf{M}}_{\mathbf{G}} + O_p(N^{-1/2}) \\
&= O_p(b_{NT}) + O_p(N^{-1/2}),
\end{aligned}
$$

we have $\frac{1}{T}\sum_{t=1}^{T}\|\bar{\mathbf{M}}\mathbf{F}_0\boldsymbol{\lambda}_{i0(t)}\|^2 = O_p(b_{NT}^2) + O_p(N^{-2}) + O_p(N^{-1}b_{NT})$ and $A_4(\boldsymbol{\theta}) = O_p(b_{NT}^2) + O_p(N^{-2}) + O_p(N^{-1}b_{NT})$. Combining the above results of $A_2(\boldsymbol{\theta})$, $A_4(\boldsymbol{\theta})$ and $A_5(\boldsymbol{\theta})$, we have $P2 = O_p(N^{-1}) + O_p(N^{-1}b_{NT}) + O_p(b_{NT}^2)$.

So far, we still need to examine $P3$ and $P4$. First, we define

$$
(NT)^{-1}\sum_{i=1}^{N}\ln\Phi\left(\frac{\mu_c}{\sigma_*}\right) =: (NT)^{-1}\sum_{i=1}^{N} f\left(\frac{\mu_c}{\sigma_*}\right),
$$

and by the first order of Taylor expansion at $\frac{\mu_*}{\sigma_*}$, we have

$$
(NT)^{-1}\sum_{i=1}^{N} f\left(\frac{\mu_c}{\sigma_*}\right) \approx (NT)^{-1}\sum_{i=1}^{N}\left[f\left(\frac{\mu_*}{\sigma_*}\right) + f'\left(\frac{\mu_*}{\sigma_*}\right)\frac{(\boldsymbol{h}_i - \bar{h}_i)'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{\left(\boldsymbol{h}_i'\bar{\mathbf{M}}\boldsymbol{h}_i/\sigma_v^2 + 1/\sigma_u^2\right)^{1/2}}\right],
$$

where $\bar{h}_i = T^{-1}\sum_{t=1}^{T} h_{it}$. Rewrite the second term in the brackets of right hand side as

$$
\begin{aligned}
&(NT)^{-1}\sum_{i=1}^{N}\left[f'\left(\frac{\mu_*}{\sigma_*}\right)\frac{(\boldsymbol{h}_i - \bar{h}_i)'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{\left(\boldsymbol{h}_i'\bar{\mathbf{M}}\boldsymbol{h}_i/\sigma_v^2 + 1/\sigma_u^2\right)^{1/2}}\right] \\
&= N^{-1}T^{-1/2}\sum_{i=1}^{N}\left[\underbrace{f'\left(\frac{\mu_*}{\sigma_*}\right)\left(\frac{\boldsymbol{h}_i'\bar{\mathbf{M}}\boldsymbol{h}_i/\sigma_v^2 + 1/\sigma_u^2}{T}\right)^{-1/2}}_{O_p(1)}\left(\frac{(\boldsymbol{h}_i - \bar{h}_i)'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{T}\right)\right].
\end{aligned}
$$

Here,

$$
\frac{(\boldsymbol{h}_i - \bar{h}_i)'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{T} = \frac{(\boldsymbol{h}_i - \bar{h}_i)'\bar{\mathbf{M}}_0\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{T} + \frac{(\boldsymbol{h}_i - \bar{h}_i)'\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{T} \times \kappa_{NT}.
$$

The first term can be decomposed into $\frac{1}{T}(\boldsymbol{h}_i-\bar{h}_i)'\bar{\mathbf{M}}_0\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2 = \frac{1}{T}(\boldsymbol{h}_i-\bar{h}_i)'\mathbf{F}_0-\frac{1}{T}(\boldsymbol{h}_i-\bar{h}_i)'\bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0')^{-1}\bar{\mathbf{H}}_0'\mathbf{F}_0$. We use the results similar to Lemmas (C7)-(C8) and obtain

$$
\begin{aligned}
&\frac{1}{T}(\boldsymbol{h}_i - \bar{h}_i)'\bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0')^{-1}\bar{\mathbf{H}}_0'\mathbf{F}_0 \\
&= \left(\frac{(\boldsymbol{h}_i - \bar{h}_i)'\mathbf{G}}{T}\bar{\mathbf{P}} + \underbrace{\frac{(\boldsymbol{h}_i - \bar{h}_i)'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T}}_{O_p(N^{-1})+O_p((NT)^{-1/2})}\right)\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}} + \underbrace{\bar{\mathbf{P}}'\frac{\mathbf{G}'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T} + \frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'\mathbf{G}}{T}\bar{\mathbf{P}}}_{O_P(N^{-1/2})} + \underbrace{\frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})}{T}}_{O_p(N^{-1})}\right)^{-1} \\
&\times \left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}_0}{T} + \underbrace{\frac{(\bar{\boldsymbol{\xi}}^* + \bar{\boldsymbol{\psi}})'\mathbf{F}_0}{T}}_{O_P((NT)^{-1/2})}\right).
\end{aligned}
$$

Notice that $\frac{1}{T}(\boldsymbol{h}_i - \bar{h}_i)'\mathbf{G} = O_p(N^{-1}) + O_p(T^{-1/2})$ because of lemma (C10) and a similar argument of (C9), thus

$$
\begin{aligned}
\frac{1}{T}(\boldsymbol{h}_i - \bar{h}_i)'\bar{\mathbf{H}}_0(\bar{\mathbf{H}}_0\bar{\mathbf{H}}_0')^{-1}\bar{\mathbf{H}}_0'\mathbf{F}_0 =& \frac{(\boldsymbol{h}_i - \bar{h}_i)'\mathbf{G}}{T}\bar{\mathbf{P}}\left(\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{G}}{T}\bar{\mathbf{P}}\right)^{-1}\bar{\mathbf{P}}'\frac{\mathbf{G}'\mathbf{F}_0}{T} \\
&+ O_p(N^{-1}) + O_p((NT)^{-1/2}).
\end{aligned}
$$

Further, together with a similar argument of (C9), the second term $\frac{(\boldsymbol{h}_i-\bar{h}_i)'\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{T}\times\kappa_{NT} = O_p(T^{-1/2}b_{NT})$. Thus

$$
\frac{(\boldsymbol{h}_i - \bar{h}_i)'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{T} = O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1/2}b_{NT}).
$$

Using this result, the term $f'\left(\frac{\mu_*}{\sigma_*}\right)\left(\frac{\boldsymbol{h}_i'\bar{\mathbf{M}}\boldsymbol{h}_i/\sigma_v^2+1/\sigma_u^2}{T}\right)^{-1/2}\left(\frac{(\boldsymbol{h}_i-\bar{h}_i)'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\lambda}_i/\sigma_v^2}{T}\right)$ should be $O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1/2}b_{NT})$. It implies that the difference between $(NT)^{-1}\sum_{i=1}^N f\left(\frac{\mu_c}{\sigma_*}\right)$ and $(NT)^{-1}\sum_{i=1}^N f\left(\frac{\mu_*}{\sigma_*}\right)$ is $O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-1/2}b_{NT})$. The results of $P3$ and $P4$ are readily obtained.

Taking results from $P2$, $P3$ and $P4$, we have

$$
P2+P3+P4 = O_p(N^{-1})+O_p(N^{-1/2}T^{-1})+O_p(T^{-1/2}b_{NT})+O_p(N^{-1}b_{NT})+O_p(b_{NT}^2). \quad \text{(M.1)}
$$

The first result of Proposition 1 can be proved because when $b_{NT}\to 0$, $P2+P3+P4 \xrightarrow{\mathbb{P}} 0$. The second result about $\boldsymbol{\theta}\in\mathbb{B}^c\cap\boldsymbol{\Theta}$ can be proved by assuming b_{NT} does not converge to zero. In this case, it implies that the difference of $P2+P3+P4$ will be dominated by the term $O_p(b_{NT}^2)$ which comes from the quadratic term of $A_4(\boldsymbol{\theta})$. Thus the difference between $Q_{NT}(\boldsymbol{\theta}_0)$ and $\tilde{Q}_{NT}(\boldsymbol{\theta})$ is greater than zero in probability one when $\boldsymbol{\theta}\in\mathbb{B}^c\cap\boldsymbol{\Theta}$. $\square$

Proof of Theorem 1. For any $\epsilon > 0$, we have (a) $\tilde{Q}_{NT}(\tilde{\boldsymbol{\theta}}) > \tilde{Q}_{NT}(\boldsymbol{\theta}_0) - \frac{\epsilon}{3}$; (b) $\tilde{Q}_{NT}(\boldsymbol{\theta}_0) > Q_0(\boldsymbol{\theta}_0) - \frac{\epsilon}{3}$ and (c) $\tilde{Q}_0(\boldsymbol{\theta}) > \tilde{Q}_{NT}(\boldsymbol{\theta}) - \frac{\epsilon}{3}$. (a) holds because $\tilde{\boldsymbol{\theta}}$ maximizes $\tilde{Q}_{NT}$, (b) holds because the result 1 from Proposition 1 by letting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and (c) holds because (iv). Therefore, we have

$$\tilde{Q}_0(\tilde{\boldsymbol{\theta}}) > \tilde{Q}_{NT}(\tilde{\boldsymbol{\theta}}) - \frac{\epsilon}{3} > \tilde{Q}_{NT}(\boldsymbol{\theta}_0) - \frac{2\epsilon}{3} > Q_0(\boldsymbol{\theta}_0) - \epsilon.$$

Using the same definitions of b_{NT} and $\mathbb{B}$, we have $Q_{NT}(\boldsymbol{\theta}_0) - \tilde{Q}_{NT}(\boldsymbol{\theta}) > 0$ with probability 1 for all $\boldsymbol{\theta} \in \mathbb{B}^c \cap \boldsymbol{\Theta}$ from the first result of Proposition 1. Taking this result with regular conditions (iii) and (iv), for any given $\epsilon > 0$, there is a constant $K > 0$ such that

$$\mathbb{P}[|Q_0(\boldsymbol{\theta}_0) - \tilde{Q}_0(\boldsymbol{\theta}) > K] \geq 1 - \epsilon,$$

for all $\boldsymbol{\theta} \in \mathbb{B}^c \cap \boldsymbol{\Theta}$. Also $Q_0(\boldsymbol{\theta}_0) = \tilde{Q}_0(\boldsymbol{\theta}_0)$ if and only if $Q_0 = \tilde{Q}_0$ and $Q_0(\boldsymbol{\theta}_0) > \tilde{Q}_0(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathbb{B}^c \cap \boldsymbol{\Theta}$. Therefore, by $\mathbb{B}^c \cap \boldsymbol{\Theta}$ is compact, (i) and (iii), $\sup_{\boldsymbol{\theta} \in \mathbb{B}^c \cap \boldsymbol{\Theta}} \tilde{Q}_0(\boldsymbol{\theta}) = Q_0(\boldsymbol{\theta}^*) < Q_0(\boldsymbol{\theta}_0)$ for some $\boldsymbol{\theta}^* \in \mathbb{B}^c \cap \boldsymbol{\Theta}$. Thus, choosing $\epsilon = Q_0(\boldsymbol{\theta}_0) - \sup_{\boldsymbol{\theta} \in \mathbb{B}^c \cap \boldsymbol{\Theta}} \tilde{Q}_0(\boldsymbol{\theta})$, it follows that

$$\tilde{Q}_0(\tilde{\boldsymbol{\theta}}) > \sup_{\boldsymbol{\theta} \in \mathbb{B}^c \cap \boldsymbol{\Theta}} \tilde{Q}_0(\boldsymbol{\theta}).$$

with probability one, and hence $\tilde{\boldsymbol{\theta}} \in \mathbb{B}$. □

Proof of Proposition 2. It can be proved immediately by multiplying $\sqrt{NT}$ and equation (M.1) from Proposition 1. Notice that we drop $N^{1/2}T^{-1/2}$ form $C_{NT} = \min\{N^{-1/2}, (NT)^{-1/4}\}$ because it will explode when $T/N \to 0$. □

Proof of Theorem 2. Since the result from Proposition 2 satisfies the requirement of Lemma 1, we can prove the asymptotic normality of our proposed estimator immediately. □

Supplementary Material

Proof of Lemma 1. See Theorem A.5 of Kristensen and Shin (2012). □

Proof of Lemma 2. It can be shown based on Lemma 2 of Pesaran (2006). □

Proof of Lemma (C1). Let $\bar{\boldsymbol{\xi}}_l = (\bar{\xi}_{1,l}, \bar{\xi}_{2,l}, ..., \bar{\xi}_{T,l})'$ denotes the l-th element of $\bar{\boldsymbol{\xi}}$. Since h_{it}, u_i^*, ε_{it} and $\boldsymbol{e}_{it}$ are mutually independent and note that $E(h_{it}) < K$ and $E(u_i^*) < K$, $\forall i, j$. We have

$$E\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{h}_i' u_i^* \bar{\boldsymbol{\xi}}_l\right) = 0 \tag{S.1}$$

and

$$\begin{aligned}
E(\bar{u}_t^2) &= E\left[\left(N^{-1}\sum_{i=1}^{N} h_{it}u_i^*\right)^2\right] \\
&= N^{-2}E\left(\sum_{i=1}^{N} h_{it}^2 u_i^{*2} + \sum_{i=1}^{N}\sum_{j\neq i} h_{it}h_{jt}u_i^* u_j^*\right) \\
&= N^{-2}\sum_{i=1}^{N} E(h_{it}^2)E(u_i^{*2}) + N^{-2}\sum_{i=1}^{N}\sum_{j\neq i} E(h_{it})E(h_{jt})E(u_i^*)E(u_j^*) = O(1).
\end{aligned}$$

Thus,

$$\begin{aligned}
\text{Var}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{h}_i' u_i^* \bar{\boldsymbol{\xi}}_l\right) &= \text{Var}\left(\sum_{t=1}^{T}\bar{u}_t\bar{\xi}_{t,l}\right) = \sum_{t=1}^{T}\text{Var}\left(\bar{u}_t\bar{\xi}_{t,l}\right) \\
&= \sum_{t=1}^{T} E(\bar{u}_t^2)E(\bar{\xi}_{t,l}^2) = O\left(TN^{-1}\right),
\end{aligned} \tag{S.2}$$

the second equality comes from the fact,

$$\begin{aligned}
&\text{Cov}\left(\bar{u}_t\bar{\xi}_{t,l}, \bar{u}_s\bar{\xi}_{s,l}\right) = E\left(\bar{u}_t\bar{\xi}_{t,l}\bar{u}_s\bar{\xi}_{s,l}\right) - E\left(\bar{u}_t\bar{\xi}_{t,l}\right)E\left(\bar{u}_s\bar{\xi}_{s,l}\right) \\
&\quad = E\left(\bar{\xi}_{t,l}\right)E\left(\bar{\xi}_{s,l}\right)E\left(\bar{u}_t\bar{u}_s\right) - E\left(\bar{\xi}_{t,l}\right)E\left(\bar{u}_t\right)E\left(\bar{\xi}_{s,l}\right)E\left(\bar{u}_s\right) = 0,
\end{aligned}$$

where the last equality holds by $E(v_{it}v_{is}) = 0$ and $E(\boldsymbol{e}_{it}\boldsymbol{e}_{is}') = \mathbf{0}$ for all i, j, and the last equality of (S.2) holds by $E(\bar{\xi}_{t,l}^2) = O(N^{-1})$. Together with (S.1) and (S.2), we obtain

$$\text{Var}\left(T^{-1}\bar{\boldsymbol{u}}'\bar{\boldsymbol{\xi}}\right) = O\left((NT)^{-1}\right),$$

hence, $T^{-1}\bar{\boldsymbol{u}}'\bar{\boldsymbol{\xi}} = O_p\left((NT)^{-1/2}\right)$. □

Proof of Lemma (C2): Since $\bar{\boldsymbol{\zeta}} = \begin{bmatrix} \bar{\boldsymbol{u}} - \bar{\boldsymbol{h}}\mu^+ & \mathbf{0} \end{bmatrix}$, we focus our analysis on the term $\bar{\boldsymbol{u}} - \bar{\boldsymbol{h}}\mu^+$. Notice that the mean is equal to 0 by the fact that h_{it}, u_i^*, ε_{it} and $\boldsymbol{e}_{it}$ are mutually independent. Furthermore,

$$\begin{aligned} &\text{Var}\left[(NT)^{-1}\sum_{t=1}^{T}\bar{\boldsymbol{\xi}}_t\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right] \\ &\quad = (NT)^{-2}\sum_{t=1}^{T}E\left(\bar{\boldsymbol{\xi}}_t\bar{\boldsymbol{\xi}}_t'\right)E\left[\left(\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\right] \\ &\quad = (NT)^{-2}\sum_{t=1}^{T}E\left(\bar{\boldsymbol{\xi}}_t\bar{\boldsymbol{\xi}}_t'\right)\sum_{i=1}^{N}E(h_{it}^2)E\left[(u_i^*-\mu^+)^2\right], \end{aligned}$$

in particular, the second and third equalities hold by $E(\bar{\boldsymbol{\xi}}_t\bar{\boldsymbol{\xi}}_s') = 0\ \forall t \neq s$ and $E[(u_i^*-\mu^+)(u_j^*-\mu^+)] = 0\ \forall i \neq j$, respectively. Moreover, $E(h_{it}^2) < K$, $E\left[(u_i^*-\mu^+)^2\right] < K$ and $\left(\bar{\boldsymbol{\xi}}_t\bar{\boldsymbol{\xi}}_t'\right) = O_p(N^{-1})$, thus,

$$\text{Var}\left[(NT)^{-1}\sum_{t=1}^{T}\bar{\boldsymbol{\xi}}_t\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right] = O(N^{-2}T^{-1}).$$

We therefore have $(NT)^{-1}\sum_{t=1}^{T}\bar{\boldsymbol{\xi}}_t\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+) = O_p(N^{-1}T^{-1/2})$. □

Proof of Lemma (C3). Recall $\mathbf{G} = [\ \mathbf{D}\quad \mathbf{F}\quad \bar{\boldsymbol{u}}\ ]$, we prove (C3) for each element of G, first, we turn our focus on $(NT)^{-1}\sum_{t=1}^{T}D_t\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)$. Notice that the mean is equal to 0 by u_i^* and h_it are mutually independent, and

$$\begin{aligned} &\text{Var}\left[(NT)^{-1}\sum_{t=1}^{T}D_t\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right] \\ &=(NT)^{-2}E\left[\sum_{t=1}^{T}D_t\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right]^2 \\ &=(NT)^{-2}E\left[\sum_{t=1}^{T}D_t^2\left(\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\right. \\ &\quad\left.+\sum_{t=1}^{T}\sum_{s\neq t}D_tD_s\left(\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)\left(\sum_{j=1}^{N}h_{js}(u_j^*-\mu^+)\right)\right]. \end{aligned}$$

For each t, the first term can be written as

$$
\begin{aligned}
&E\left[(NT)^{-2}\sum_{t=1}^{T}D_t^2\left(\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\right]\\
=&(NT)^{-2}\sum_{t=1}^{T}D_t^2\sum_{i=1}^{N}E(h_{it}^2)E\left[(u_i^*-\mu^+)^2\right]\\
=&O((NT)^{-1}),
\end{aligned}
\tag{S.3}
$$

where the second equality comes from the fact that $E[(u_i^*-\mu^+)(u_j^*-\mu^+)]=0\ \forall i\neq j$, and u_i is independent of h_{jt} for all i,j. The last equality holds by $E(h_{it}^2)<K$, $E\left[(u_i^*-\mu^+)^2\right]<K$. The second term,

$$
\begin{aligned}
&E\left[(NT)^{-2}\sum_{t=1}^{T}\sum_{s\neq t}D_tD_s\left(\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)\left(\sum_{j=1}^{N}h_{js}(u_j^*-\mu^+)\right)\right]\\
=&(NT)^{-2}\sum_{t=1}^{T}\sum_{s\neq t}D_tD_s\sum_{i=1}^{N}E(h_{it})E(h_{is})E(u_i^*-\mu^+)^2=O(N^{-1}),
\end{aligned}
\tag{S.4}
$$

the second equality holds for the same reason that $(u_i^*-\mu^+)E(u_j^*-\mu^+)]=0$. The result comes from the assumption of finite first moment of h_{it}. To sum up (S.3) and (S.4), we obtain

$$
\text{Var}\left((NT)^{-1}\sum_{t=1}^{T}D_t\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)=O(N^{-1}),
$$

and which implies $(NT)^{-1}\sum_{t=1}^{T}D_t\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)=O_p(N^{-1/2})$.

Next, consider the l-th row of $T^{-1}\left[\mathbf{F}'N^{-1}\sum_{i=1}^{N}\boldsymbol{h}_i(u_i^*-\mu^+)\right]$, which can be written as $T^{-1}[\sum_{t=1}^{T} f_{lt}N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)]$. Notice that its mean is equal to 0 by the similar argument in the previous case, and the variance,

$$
\begin{aligned}
&\mathrm{Var}\left[T^{-1}\left(\sum_{t=1}^{T} f_{lt} N^{-1} \sum_{i=1}^{N} h_{it}(u_i^* - \mu^+)\right)\right] \\
=&(NT)^{-2} E\left[\sum_{t=1}^{T} f_{lt} \sum_{i=1}^{N} h_{it}(u_i^* - \mu^+)\right]^2 \\
=&(NT)^{-2} E\left[\sum_{t=1}^{T} f_{lt}^2 \left(\sum_{i=1}^{N} h_{it}(u_i^* - \mu^+)\right)^2\right. \\
&\left. + \sum_{t=1}^{T}\sum_{s\neq t} f_{lt} f_{ls}\left(\sum_{i=1}^{N} h_{it}(u_i^* - \mu^+)\right)\left(\sum_{j=1}^{N} h_{js}(u_j^* - \mu^+)\right)\right] \\
=&(NT)^{-2}\left[\sum_{t=1}^{T} E(f_{lt}^2) \sum_{i=1}^{N} E(h_{it}^2) E(u_i^* - \mu^+)^2\right. \\
&\left. + \sum_{t=1}^{T}\sum_{s\neq t} E(f_{lt} f_{ls})\left(\sum_{i=1}^{N} E(h_{it}h_{is}) E(u_i^* - \mu^+)^2\right)\right],
\end{aligned}
$$

the third equality holds by $(u_i^* - \mu^+)E(u_j^* - \mu^+)] = 0$. Furthermore, because $\mathbf{F}$, h_{it} are covariance stationary process distributed independently of u_i^*, the autocovariance function decays exponentially in $|t-s|$. By these assumptions,

$$
\begin{aligned}
&\mathrm{Var}\left[T^{-1}\left(\sum_{t=1}^{T} f_{lt} N^{-1} \sum_{i=1}^{N} h_{it}(u_i^* - \mu^+)\right)\right] \\
=&(NT)^{-2}\left[\sum_{t=1}^{T} E(f_{lt}^2) \sum_{i=1}^{N} E(h_{it}^2) E(u_i^* - \mu^+)^2\right. \\
&\left. + \sum_{t=1}^{T}\sum_{s\neq t} \Gamma_{fl}(|t-s|)\left(\sum_{i=1}^{N} E(h_{it}h_{is}) E(u_i^* - \mu^+)^2\right)\right] = O((NT^{-1})),
\end{aligned}
$$

where Γ_{fl} is the autocovariance function of f_{lt}, and the last equality holds by $E(f_{lt}^2) < K$, $E(h_{it}^2) < K$, $E(u_i^* - \mu^+)^2 < K$ and $E(h_{it}h_{is}) < K$, which establishes $T^{-1}\left[\mathbf{F}' N^{-1} \sum_{i=1}^{N} \boldsymbol{h}_i(u_i^* - \mu^+)\right] = O_p((NT)^{-1/2})$.

Finally, we analyze the last term. Notice that

$$
\begin{aligned}
&E\left[T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N} h_{it}(u_i^* - \mu^+)\right)\left(N^{-1}\sum_{j=1}^{N} h_{jt}u_j^*\right)\right] \\
=&N^{-2}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} E(h_{it}^2) E(u_i^* - \mu^+)u_i^* \\
=&O(N^{-1}), \qquad \text{(S.5)}
\end{aligned}
$$

the first equality holds by the assumption that $E[(u_i^* - \mu^+)u_j^*] = 0$, $\forall i \neq j$, and the last is true by $E(h_{it}^2) < K$, and $E(u_i^{*2}) < K$. The variance,

$$
\begin{aligned}
&\text{Var}\left[T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)\left(N^{-1}\sum_{j=1}^{N}h_{jt}u_j^*\right)\right] \\
=&E\left[T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)\left(N^{-1}\sum_{j=1}^{N}h_{jt}u_j^*\right)\right]^2 \\
&-\left[E\left(T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)\left(N^{-1}\sum_{j=1}^{N}h_{jt}u_j^*\right)\right)\right]^2,
\end{aligned}
$$

where the first term can be rearranged as

$$
\begin{aligned}
&T^{-2}E\left[\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\left(N^{-1}\sum_{j=1}^{N}h_{jt}u_j^*\right)^2\right] \\
&+T^{-2}E\left[\sum_{t=1}^{T}\sum_{s\neq t}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)\left(N^{-1}\sum_{j=1}^{N}h_{jt}u_j^*\right)\right. \\
&\times\left.\left(N^{-1}\sum_{j=1}^{N}h_{js}(u_j^*-\mu^+)\right)\left(N^{-1}\sum_{j=1}^{N}h_{js}u_j^*\right)\right] \\
=&:A_1+A_2.
\end{aligned}
$$

Consider A_1,

$$
A_1 = N^{-4}T^{-2}E\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}h_{it}h_{jt}h_{kt}h_{lt}(u_i^*-\mu^+)(u_j^*-\mu^+)u_k^*u_l^*\right],
$$

in which expectation is non-zero only in the following six case: (i) $i=j=k=l$, (ii) $i=j$, (iii) $i=k$ (iv) $i=l$ (v) $j=k$ and (vi) $j=l$ by assuming the forth moment of h_{it} exists. It follows that

$$
\begin{aligned}
A_1 =&N^{-4}T^{-2}E\sum_{t=1}^{T}\left[\sum_{i=1}^{N}h_{it}^4(u_i^*-\mu^+)^2u_i^{*2}+\sum_{i=1}^{N}\sum_{k\neq i}\sum_{l\neq i}h_{it}^2h_{kt}h_{lt}(u_i^*-\mu^+)^2u_k^*u_l^*\right. \\
&\left.+\sum_{i=1}^{N}\sum_{j\neq i}\sum_{l\neq i}h_{it}^2h_{jt}h_{lt}(u_i^*-\mu^+)(u_j^*-\mu^+)u_i^*u_l^*\right] = O((NT)^{-1}).
\end{aligned}
$$

Furthermore, A_2 has the similar result except that we have to sum up the terms for all $t \neq s$, $t, s = 1, ..., T$. Thus, we have $A_2 = O((N)^{-1})$. Taking A_1,A_2 and (S.5) together, we have

$$
\text{Var}\left[T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)\left(N^{-1}\sum_{j=1}^{N}h_{jt}u_j^*\right)\right] = O(N^{-1}),
$$

which implies $T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)\left(N^{-1}\sum_{j=1}^{N}h_{jt}u_j^*\right)=O(N^{-1/2})$. Therefore

$$T^{-1}\left[G'N^{-1}\sum_{i=1}^{N}\boldsymbol{h}_i(u_i^*-\mu^+)\right]=O_p(N^{-1/2})$$

as required. □

Proof of Lemma (C4). Write

$$\begin{aligned}
&E\left[T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\right]\\
=&N^{-2}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}E(h_{it}^2)E(u_i^*-\mu^+)^2=O(N^{-1}),
\end{aligned}\tag{S.6}$$

which holds by the assumption that $E\left[(u_i^*-\mu^+)(u_j^*-\mu^+)\right]=0$, $E(h_{it}^2)<K$ and $E(u_i^*-\mu^+)^2<K$. Furthermore,

$$\begin{aligned}
&E\left[T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\right]^2\\
=&T^{-2}E\left[\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^4\right]\\
&+T^{-2}E\left[\sum_{t=1}^{T}\sum_{s\neq t}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\left(N^{-1}\sum_{j=1}^{N}h_{js}(u_j^*-\mu^+)\right)^2\right]\\
=:&A_1+A_2.
\end{aligned}$$

Consider A_1, in which expectation is non-zero only in the following case

$$\begin{aligned}
A_1=&N^{-4}T^{-2}E\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{N}h_{it}^4(u_i^*-\mu^+)^4\right)\right]\\
&+N^{-4}T^{-2}E\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\sum_{j\neq i}h_{it}^2h_{jt}^2(u_i^*-\mu^+)^2(u_j^*-\mu^+)^2\right]=O(N^{-2}T^{-1}),
\end{aligned}$$

where the result comes from assuming h_{it} and u_i^* are independently distributed with finite fourth moment, and the fact that u_i^*'s are cross-sectional independent. Now consider A_2,

$$\begin{aligned}
A_2=&N^{-4}T^{-2}E\sum_{t=1}^{T}\sum_{s\neq t}\left[\sum_{i=1}^{N}\sum_{k=1}^{N}h_{it}h_{kt}(u_i^*-\mu^+)(u_k^*-\mu^+)\right]\\
&\times\left[\sum_{j=1}^{N}\sum_{l=1}^{N}h_{js}h_{ls}(u_j^*-\mu^+)(u_l^*-\mu^+)\right],
\end{aligned}$$

in which expectation is non-zero only in the following cases: (i) $i = j = k = l$, (ii) $i = k$, $j = l$ (iii) $i = j$, $k = l$, it follows that

$$A_2 = N^{-4}T^{-2}E\sum_{t=1}^{T}\sum_{s\neq t}\left[\sum_{i=1}^{N}h_{it}^2h_{is}^2(u_i^*-\mu^+)^4+\sum_{i=1}^{N}\sum_{j\neq i}h_{it}^2h_{js}^2(u_i^*-\mu^+)^2(u_j^*-\mu^+)^2\right.$$
$$\left.+\sum_{i=1}^{N}\sum_{k\neq i}h_{it}h_{is}h_{kt}h_{ks}(u_i^*-\mu^+)^2(u_k^*-\mu^+)^2\right] = O(N^{-2}).$$

Taking A_1, A_2 and (S.6) together, we have

$$\mathrm{Var}\left[T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\right] = O(N^{-2}),$$

which implies $T^{-1}\sum_{t=1}^{T}\left(N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2 = O_p(N^{-1})$. □

Proof of Lemma (C5). Recall that $\boldsymbol{\xi}_{it} = \begin{bmatrix} v_{it}+\boldsymbol{\beta}'\boldsymbol{e}_{it} \\ \boldsymbol{e}_{it}\end{bmatrix}$, it is easy to show that its expectation is 0 by the fact that v_{it}, $\boldsymbol{e}_{it}$, h_{it} and u_i^* are distributed independently. So we can write the variance as

$$\begin{aligned}
&\mathrm{Var}\left[T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{it}N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right]\\
=&(NT)^{-2}E\left[\sum_{t=1}^{T}\boldsymbol{\xi}_{it}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right]^2\\
=&(NT)^{-2}\left[\sum_{t=1}^{T}E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}')E\left(\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right)^2\right]\\
=&(NT)^{-2}\left[\sum_{t=1}^{T}E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}')E\left(\sum_{i=1}^{N}h_{it}^2(u_i^*-\mu^+)^2\right)\right],
\end{aligned} \tag{S.7}$$

the second equality holds by the fact that v_{it} and $\boldsymbol{e}_{it}$ are serially uncorrelated, and the third equality holds by u_i^* are cross-sectionally independent. Furthermore, the term $E||\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}'|| < K$ by v_{it} and $\boldsymbol{e}_{it}$ have finite variance, together with $E(h_{it}^2) < K$ and $E(u_i^*-\mu^+)^2 < K$, we can obtain $\mathrm{Var}\left[T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{it}N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right] = O((NT)^{-1})$. Therefore $T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{it}N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+) = O_p((NT)^{-1/2})$. □

Proof of Lemma (C6). Given Lemmas (B5) and (B6), we already discussed two of three elements in $\mathbf{G}$. It remains to show the rate of $T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{it}N^{-1}\sum_{i=1}^{N}h_{it}u_i^*$. Consider its mean. Again, given the fact that v_{it}, $\boldsymbol{e}_{it}$, h_{it} and u_i^* are distributed independently, it can be show that the mean is 0. The variance,

$$\begin{aligned}
&\text{Var}\left[T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{it}N^{-1}\sum_{i=1}^{N}h_{it}u_i^*\right]\\
=&(NT)^{-2}E\left[\sum_{t=1}^{T}\boldsymbol{\xi}_{it}\sum_{i=1}^{N}h_{it}u_i^*\right]^2\\
=&(NT)^{-2}\left[\sum_{t=1}^{T}E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}')E\left(\sum_{i=1}^{N}h_{it}u_i^*\right)^2\right]\\
=&(NT)^{-2}\left[\sum_{t=1}^{T}E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}')\left(\sum_{i=1}^{N}\sum_{j=1}^{N}E(h_{it}h_{jt})E(u_i^*u_j^*)\right)\right],
\end{aligned}$$

where the second equality holds as the same as preceding discuss that v_{it} and $\boldsymbol{e}_{it}$ are serially uncorrelated. However, by expanding $\left(\sum_{i=1}^{N}h_{it}u_i^*\right)^2$, it is $O_p(N^2)$ by the assumptions that $E(h_{it}h_{jt}) < K$ and $E(u_i^*u_j^*) < K$ for all i, j. Together with $E||\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}'|| < K$, we get

$$\text{Var}\left[T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{it}N^{-1}\sum_{i=1}^{N}h_{it}u_i^*\right] = O(T^{-1}),$$

which implies $T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{it}N^{-1}\sum_{i=1}^{N}h_{it}u_i^* = O_p(T^{-1/2})$. □

Proof of Lemma (C7). Consider the mean. Because v_{it}, $\boldsymbol{e}_{it}$, h_{it} and u_i^* are mutually independent, we can obtain the mean is 0 easily. Next, the variance,

$$\begin{aligned}
\text{Var}\left[T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)\bar{\boldsymbol{\xi}}_t\right] =&T^{-2}\sum_{t=1}^{T}E\left(u_i^*(h_{it}-\bar{h}_i)\right)^2 E(\bar{\boldsymbol{\xi}}_t\bar{\boldsymbol{\xi}}_t')\\
=&T^{-2}\sum_{t=1}^{T}E(u_i^{*2})E(h_{it}-\bar{h}_i)^2E(\bar{\boldsymbol{\xi}}_t\bar{\boldsymbol{\xi}}_t').
\end{aligned}$$

Notice that the above holds by the fact v_{it} and $\boldsymbol{e}_{it}$ are serially uncorrelated and assumptions we used in the mean. Because we have $E(u_i^{*2}) < K$, $E(h_{it}-\bar{h}_i)^2 < K$ and the order of $E(\bar{\boldsymbol{\xi}}_t\bar{\boldsymbol{\xi}}_t')$ is $O(N^{-1})$. Thus, we have $\text{Var}\left[T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)\bar{\boldsymbol{\xi}}_t\right] = O((NT)^{-1})$, and it follows that $T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)\bar{\boldsymbol{\xi}}_t = O_p((NT)^{-1/2})$. □

Proof of Lemma (C8). We first consider its mean. Write,

$$
\begin{aligned}
&E\left[T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)N^{-1}\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+)\right] \\
=&(NT)^{-1}\left[E\sum_{t=1}^{T}(u_{it}-\bar{u}_i)h_{it}(u_i^*-\mu^+)+E\sum_{t=1}^{T}(u_{it}-\bar{u}_i)\sum_{j\neq i}^{N}h_{jt}(u_j^*-\mu^+)\right],
\end{aligned}
$$

where the second term inside the square brackets is 0 by the assumption that u_i^* is cross-sectional independent. Further, since $u_{it} = h_{it}u_i^*$ and using the assumptions that h_{it} and u_i^* are mutually independent with finite mean and variance, we get

$$
\begin{aligned}
&E\left[T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)N^{-1}\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+)\right] \\
=&(NT)^{-1}\sum_{t=1}^{T}\left[E(h_{it}^2-h_{it}\bar{h}_i)E(u_i^{*2}-u_i^*\mu^+)\right]=O(N^{-1}).
\end{aligned}
\tag{S.8}
$$

Consider the variance, we first evaluate the term

$$
\begin{aligned}
&(NT)^{-2}E\left[\sum_{t=1}^{T}(u_{it}-\bar{u}_i)\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+)\right]^2 \\
&\quad=(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)^2u_i^{*2}\left(\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+)\right)\left(\sum_{k=1}^{N}h_{kt}(u_k^*-\mu^+)\right)\right. \\
&\qquad\left.+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}-\bar{h}_i)(h_{is}-\bar{h}_i)u_i^{*2}\left(\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+)\right)\left(\sum_{k=1}^{N}h_{ks}(u_k^*-\mu^+)\right)\right].
\end{aligned}
$$

Note that the expected value of above equation is non-zero only in the case that $j = k$, so we can rewrite as

$$
\begin{aligned}
&(NT)^{-2}E\left[\sum_{t=1}^{T}(u_{it}-\bar{u}_i)\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+)\right]^2\\
&=(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)^2u_i^{*2}\sum_{j=1}^{N}h_{jt}^2(u_j^*-\mu^+)^2\right.\\
&\quad\left.+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}-\bar{h}_i)(h_{is}-\bar{h}_i)u_i^{*2}\left(\sum_{j=1}^{N}h_{jt}h_{js}(u_j^*-\mu^+)^2\right)\right]\\
&=(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}^2-h_{it}\bar{h}_i)^2(u_i^{*2}-u_i^*\mu^+)^2+\sum_{t=1}^{T}(h_{it}-\bar{h}_i)u_i^{*2}\sum_{j\neq i}h_{jt}^2(u_j^*-\mu^+)^2\right.\\
&\quad+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}^2-h_{it}\bar{h}_i)(h_{is}^2-h_{is}\bar{h}_i)(u_i^{*2}-u_i^*\mu^+)^2\\
&\quad\left.+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}-\bar{h}_i)(h_{is}-\bar{h}_i)u_i^{*2}\left(\sum_{j\neq i}^{N}h_{jt}h_{js}(u_j^*-\mu^+)^2\right)\right].
\end{aligned}
\tag{S.9}
$$

Given the assumptions that h_{it} and u_i^* are mutually independent with finite fourth moment, the first term inside square brackets divided by $(NT)^2$ is $O(N^{-2}T^{-1})$. Using the similar argument, the third term divided by $(NT)^2$ is $O(N^{-2})$. Further, since u_i^* is cross-sectional independent and h_{it} is covariance stationary process, the second and fourth terms divided by $(NT)^2$ are $(NT)^{-1}$. Thus, by summarizing (S.8) and (S.9), we have Var $\left[T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)N^{-1}\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+)\right]=O(N^{-2})+O((NT)^{-1})$. Therefore, we obtain $T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)N^{-1}\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+)=O_p(N^{-1})+O_p((NT)^{-1/2})$. □

Proof of Lemma (C9). Since h_{it}, D_t and $\boldsymbol{f}_t$ are independent stationary process, it is easy to obtain $T^{-1}(\boldsymbol{h}_i'D)=O_p(T^{-1/2})$ and $T^{-1}(\boldsymbol{h}_i'\mathbf{F})=O_p(T^{-1/2})$. The remains can be denote as $T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)N^{-1}\sum_{i=1}^{N}h_{it}u_i^*$, and using the similar arguments in Lemma (C8), the mean,

$$
\begin{aligned}
&E\left[T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)N^{-1}\sum_{i=1}^{N}h_{it}u_i^*\right]\\
&=(NT)^{-1}\left[E\sum_{t=1}^{T}(h_{it}^2-h_{it}\bar{h}_i)u_i^{*2}+E\sum_{t=1}^{T}(h_{it}-\bar{h}_i)u_i^*\sum_{j\neq i}^{N}h_{jt}u_j^*\right]\\
&=(NT)^{-1}\left[\sum_{t=1}^{T}E(h_{it}^2-h_{it}\bar{h}_i)E(u_i^{*2})\right]=O(N^{-1}).
\end{aligned}
\tag{S.10}
$$

The second equality holds by the fact that h_{it} is cross-sectional independent with $E\left(\sum_{t=1}^{T} h_{it} - \bar{h}_i\right) = 0$. The result holds by h_{it} and u_i^* are mutually independent with finite mean and variance. Next, we consider

$$\begin{aligned}
&(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)u_i^*\sum_{j=1}^{N}h_{jt}u_j^*\right]^2\\
&=(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)^2u_i^{*2}\left(\sum_{j=1}^{N}h_{jt}u_j^*\right)\left(\sum_{k=1}^{N}h_{kt}u_k^*\right)\right.\\
&\quad\left.+\sum_{t=1}^{T}\sum_{s\neq t}^{T}(h_{it}-\bar{h}_i)(h_{is}-\bar{h}_i)u_i^{*2}\left(\sum_{j=1}^{N}h_{jt}u_j^*\right)\left(\sum_{k=1}^{N}h_{ks}u_k^*\right)\right]\\
&=(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}^2-h_{it}\bar{h}_i)^2u_i^{*4}+\sum_{t=1}^{T}(h_{it}-\bar{h}_i)^2u_i^{*2}\left(\sum_{j\neq i}^{N}\sum_{k\neq i}^{N}h_{jt}h_{kt}u_j^*u_k^*\right)\right.\\
&\quad+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}^2-h_{it}\bar{h}_i)(h_{is}^2-h_{is}\bar{h}_i)u_i^{*4}\\
&\quad\left.+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}-\bar{h}_i)(h_{is}-\bar{h}_i)\left(\sum_{j\neq i}^{N}\sum_{k\neq i}^{N}h_{jt}h_{ks}u_j^*u_k^*\right)\right]. \qquad \text{(S.11)}
\end{aligned}$$

The above expressions are quite similar with (C8), the assumptions that h_{it} and u_i^* are mutually independent with finite fourth moment imply the first and third terms divided by $(NT)^2$ are $O(N^{-2}T^{-1})$ and $O(N^{-2})$. The difference is that the case $j \neq k$ is non-zero here, thus the second and forth terms divided by $(NT)^2$ are $O(T^{-1})$. Taking (S.10) and(S.11) together, we have $\text{Var}\left[T^{-1}\sum_{t=1}^{T}(h_{it}-\bar{h}_i)u_i^*N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+)\right] = O(N^{-2})+O(T^{-1})$, which implies $T^{-1}\sum_{t=1}^{T}(h_{it}-\bar{h}_i)u_i^*N^{-1}\sum_{i=1}^{N}h_{it}(u_i^*-\mu^+) = O_p(N^{-1})+O_p(T^{-1/2})$. □

Proof of Lemma (C10). The proof of (C10) is quite similar to (C8), we first consider the mean,

$$\begin{aligned}
&E\left[T^{-1}\sum_{t=1}^{T}(h_{it}-\bar{h}_i)N^{-1}\sum_{j=1}^{N}(h_{jt}-\bar{h}_j)(u_j^*-\mu^+)\right]\\
&=(NT)^{-1}\left[E\sum_{t=1}^{T}(h_{it}-\bar{h}_i)\sum_{j=1}^{N}(h_{jt}-\bar{h}_j)(u_j^*-\mu^+)\right]=0, \qquad \text{(S.12)}
\end{aligned}$$

the above result holds by $E(u_i^*-\mu^+)=0$ and the fact u_i^* is independent of h_{it}, for all i, t. Consider the variance,

$$(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)\sum_{j=1}^{N}(h_{jt}-\bar{h}_j)(u_j^*-\mu^+)\right]^2$$
$$=(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)^2\left(\sum_{j=1}^{N}(h_{jt}-\bar{h}_j)(u_j^*-\mu^+)\right)\left(\sum_{k=1}^{N}(h_{kt}-\bar{h}_k)(u_k^*-\mu^+)\right)\right.$$
$$\left.+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}-\bar{h}_i)(h_{is}-\bar{h}_i)\left(\sum_{j=1}^{N}(h_{jt}-\bar{h}_j)(u_j^*-\mu^+)\right)\left(\sum_{k=1}^{N}(h_{ks}-\bar{h}_k)(u_k^*-\mu^+)\right)\right].$$

Note that the expected value of above equation is non-zero only in the case that $j=k$, so we can rewrite as

$$(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)\sum_{j=1}^{N}(h_{jt}-\bar{h}_j)(u_j^*-\mu^+)\right]^2$$
$$=(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)^2\sum_{j=1}^{N}(h_{jt}-\bar{h}_j)^2(u_j^*-\mu^+)^2\right.$$
$$\left.+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}-\bar{h}_i)(h_{is}-\bar{h}_i)\left(\sum_{j=1}^{N}(h_{jt}-\bar{h}_j)(h_{js}-\bar{h}_j)(u_j^*-\mu^+)^2\right)\right]$$
$$=(NT)^{-2}E\left[\sum_{t=1}^{T}(h_{it}-\bar{h}_i)^4(u_i^*-\mu^+)^2+\sum_{t=1}^{T}(h_{it}-\bar{h}_i)\sum_{j\neq i}(h_{jt}-\bar{h}_j)^2(u_j^*-\mu^+)^2\right.$$
$$+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}-\bar{h}_i)^2(h_{is}-\bar{h}_i)^2(u_i^*-\mu^+)^2$$
$$\left.+\sum_{t=1}^{T}\sum_{s\neq t}(h_{it}-\bar{h}_i)(h_{is}-\bar{h}_i)\left(\sum_{j\neq i}^{N}(h_{jt}-\bar{h}_j)(h_{js}-\bar{h}_j)(u_j^*-\mu^+)^2\right)\right]. \quad \text{(S.13)}$$

Given the assumptions that h_{it} and u_i^* are mutually independent with finite fourth and second moment respectively, the first term inside square brackets divided by $(NT)^2$ is $O(N^{-2}T^{-1})$. Similarly, the third term divided by $(NT)^2$ is $O(N^{-2})$. Further, since u_i^*, h_{it} are cross-sectional independent and h_{it} is also covariance stationary process, the second and fourth terms divided by (S.13)$= O(N^{-2})+O((NT)^{-1})$. Therefore, we obtain $T^{-1}\sum_{t=1}^{T}(u_{it}-\bar{u}_i)N^{-1}\sum_{j=1}^{N}h_{jt}(u_j^*-\mu^+) = O_p(N^{-1})+O_p((NT)^{-1/2})$. □

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, 2001. GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, **101**, 219–255.

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.

Andrews, D. W. K. 2005. Cross-section regression with common shocks, *Econometrica*, **73**, 1551–1585.

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

Beck, T. , O. D. Jonghe and G. Schepens, 2013. Bank competition and stability: Cross-country heterogeneity, *Journal of Financial Intermediation*, **22**, 218–244.

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, 2009a. Bank ownership and efficiency in China: What will happen in the world's largest nation? *Journal of Banking and Finance*, **33**, 113–130.

Berger, A. N., L. F. Klapper and R. Turk-Ariss, 2009b. Bank Competition and Financial Stability, *Journal of Financial Services Research*, **35**, 99–118.

Caner, M., 2006. M-estimators with non-standard rates of convergence and weakly dependent data, *Journal of Statistical Planning and Inference*, 136, 1207-1219.

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.

Fama, E. F. and M. C. Jensen, 1983. Separation of Ownership and Control, *Journal of Law and Economics*, **26**, 301–325.

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., 2005a. Fixed and random effects in stochastic frontier models, *Journal of Productivity Analysis*, **23**, 7–32.

Greene, W., 2005b. Reconsidering heterogeneity and inefficiency: Alternative estimators for stochastic frontier models, *Journal of Econometrics*, **126**, 269–303.

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.

Jensen, M. C. and W. H. Meckling, 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure, *Journal of Financial Economics*, **3**, 305–360.

Jondrow, J., C. A. K. Lovell, I. S. Materov and P. Schmidt, 1984. On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics*, **19**, 233–238.

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

Kristensen, D. and Y. Shin, 2012. Estimation of dynamic models with nonparametric simulated maximum likelihood, *Journal of Econometrics*, **167**, 76-9-4.

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

Lee, L. F., 1995. Asymptotic bias in simulated maximum likelihood estimation of discrete choice models, *Econometric Theory*, **11**, 437–483.

Lee, L. F., 1999. Statistical inference with simulated likelihood functions, *Econometric Theory*, **15**, 337–360.

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.

Wheelock, D. C. and P. W. Wilson, 2012. Do Large Banks Have Lower Costs? New Estimates of Returns to Scale for U.S. Banks, *Journal of Money, Credit and Banking*, **44**, 171–

199.