Panel Nonparametric Estimation of Conditional Heteroskedastic Frontiers

Jun Cai∗ William C. Horrace† Yoonseok Lee‡

September 16, 2019

Abstract

We study a nonparametric panel stochastic frontier model that (a) contains a nonparametric production/cost function with non-separable unobserved heterogeneity, (b) accommodates time-varying, conditional heteroskedastic variance components, and (c) does not require distributional assumptions on the noise term except conditional symmetry. We use conditional characteristic functions to derive new moment conditions that allow identification of the heteroskedastic variance of inefficiency and noise. Identification only requires a panel with two time periods. We develop a nonparametric estimation procedure for the conditional variance of inefficiency and random noise and derive polynomial convergence rates. Monte Carlo simulation shows that the estimator has good finite sample properties for various designs. We apply the proposed nonparametric estimator to a panel of US banks and compare the results to the parametric counterparts.

∗Department of Economics, Syracuse University, jcai106@syr.edu.
†Department of Economics, Syracuse University, whorrace@syr.edu.
‡Department of Economics, Syracuse University, ylee41@syr.edu.
1 Introduction

Since Aigner et al. (1977), stochastic frontier analysis (SFA) has been an important tool in the analysis of productive efficiency. The original model includes a log-linear production or cost function with an additively separable noise term and an additively separable, time-invariant, inefficiency term, which decreases output in the production function or increases costs in the cost function. It is a leading case of the “composed error model” with two independent components, and it has been the workhorse for countless empirical investigations of firm-level cost or productive inefficiency. While estimation of the linear production or cost function is important in these models, the primary concern is often characterization of the error components. In particular, interest centers on characterizing the inefficiency term in a meaningful way. That is, most regression-based empirical studies treat errors or error components as nuisance parameters and focus on the marginal effects in the conditional mean function. However, in the stochastic frontier literature estimation of the error components or features of the error component distributions is an important aspect of the model’s specification: one can not specify a production or cost function with inefficiency without trying to understand the statistical and economic significance of that inefficiency.

While early models were designed for cross sectional data and were fully parametric, proliferation of panel data has facilitated relaxation of parametric assumptions. Pitt and Lee (1981), Schmidt and Sickles (1984), and Battese and Coelli (1988) consider fixed-effect and random-effect estimation of the stochastic frontier model with additively separable time-invariant technical or cost inefficiency. Here, the fixed- or random-effect embodies inefficiency. Cornwell et al. (1990), Kumbhakar (1990), Lee and Schmidt (1993), Han et al. (2005) and Ahn et al. (2007, 2013) propose time-varying inefficiency versions of the model. These models are more sensible because they allow firms to decrease their inefficiency over time. That is, any reasonable model specification should allow firms to move closer to the
efficient frontier over time. More recently, Greene (2005a, 2005b) introduces a “true” fixed-effect/random-effect stochastic frontier which includes the additively separable noise and inefficiency terms but also features an additively separable, time-invariant fixed or random effect. That is, each firm in the sample has an idiosyncratic and persistent effect in addition to a time-varying effect and a draw from the noise distribution. Wang and Ho (2009), Chen, Schmidt and Wang (2014) and Wikstrom (2015) study estimation techniques for these models. This paper presents a nonparametric version of the “true” fixed-effect/random-effect model, where (a) the production or cost function is a general function of the inputs, (b) inefficiency is time-varying, and (c) the fixed or random effect is non-separable from the production or cost function. To the best of our knowledge we are the first to consider panel nonparametric estimation of the stochastic frontier model with nonseparable fixed/random effects.

Studies on nonparametric stochastic frontier models for cross-sectional data are Fan et al. (1996), Kumbhakar et al. (2007), Parmeter and Racine (2012), Noh (2014), Martins-Filho and Yao (2015), Parmeter et al. (2017), Horrace and Parmeter (2011), and Cai et al. (2019). The last two papers explore deconvolution-type nonparametric estimation of the cross-sectional stochastic frontier model. For panel data, there are only a couple nonparametric stochastic frontier models. Kneip and Simar (1996) employ a Nadaraya-Watson estimator to estimate a nonparametric version of the Schmidt and Sickles (1984) model with time-invariant inefficiency. Yao et al. (2018) investigate a semi-parametric smooth coefficient stochastic frontier model for panel data. It should be noted that, however, these studies do not consider (non-separable) fixed/random effects.

In this paper, we identify and consistently estimate the variance parameters associated with the noise and inefficiency components in the panel stochastic frontier model with nonseparable fixed/random effects, using the results of Kotlarski (1967) and Evdokimov and
These variance components are often the primary focus of the stochastic frontier literature and of this paper. In particular, we consider

\[ Y_{it} = m(X_{it}, \alpha_i) + \varepsilon_{it} \]  

(1)

\[ \varepsilon_{it} = U_{it} + V_{it}, \quad i = 1, ..., n, \quad t = 1, ..., T, \]  

(2)

where \( m(X_{it}, \alpha_i) \) is the unspecified cost function or production function which allows nonseparable unobserved heterogeneity \( \alpha_i \); \( X_{it} \in \mathbb{R}^p \), \( \alpha_i \) is the random effects (RE) or fixed effects (FE). \( U_{it} \) is the inefficiency term which could be conditionally heteroskedastic in some exogenous \( X_{it} \). It is constrained positive in a cost function and negative in a production function. The \( V_{it} \) is the random noise or disturbance term from unspecified distributional family. We derive moment conditions, based on which the model can be identified and consistently estimated with two time periods.

The nearest neighbors to our model and contribution is the aforementioned approach of Evdokimov (2010) and Evdokimov and White (2012), which are concerned with estimation of the model in equation (1) with the restriction \( \varepsilon_{it} = V_{it} \). Like Evdokimov (2010) our model is quite general and only requires that the noise component \( V \) be from a zero-mean, conditionally symmetric distribution with finite second moment. However, unlike Evdokimov (2010), our contribution is identification and estimation of the variances of the error components \( (U, V) \) in equation (2), instead of the distribution of \( \alpha_i \) or \( m(x, \alpha) \) itself, though the latter can be obtained with a straightforward application of Evdokimov (2010) or more recently Ju, Gan and Li (2017). Since the model is identified when inefficiency equals zero (i.e., \( U = 0 \)), we show that the model with non-zero inefficiency is still identified when its distribution is known up to its variance (i.e., half-normal or exponential inefficiency).

---

1The results of Kotlarski (1967) have been applied in a variety of economic settings. See, for example, Li and Vuong (1998), Schennach (2004), Li, Perrigne, Vuong (2000), Krasnokutskaya (2011), Arellano and Bonhomme (2009), Bonhomme and Robin (2010) and Kennan and Walker (2011).
Showing this requires a “common support” assumption on $X_{i1}$ and $X_{i2}$, which will often be met in empirical applications where inputs are relatively stable across (short) periods and which allows us to remove the $m$ function with time-differencing of equation (1). It also requires independence between $V_{it}$ and $(\alpha_i, V_{is}, X_{is})$ for any $t \neq s$ conditional on $X_{it}$ which is standard in non-linear panel data setting.

The paper is organized as follows. In Section 2 we introduce the assumptions needed to identify the model in equation (1) and derive moment conditions for estimation of the variance components. Section E.3 discusses estimation issues and bandwidth selection methods. Section 4 derives the large sample properties of the estimator. We study the finite sample properties of the estimator with Monte Carlo simulations in Section 5. In Section 6 we apply the model to a large US bank dataset and compare the nonparametric results to those from parametric models. Directions for future research and conclusions are in Section 7.

2 Identification

For stochastic frontier analysis, the inefficiency term $U_{it}$ in equation (2) is of primary interest. We focus on the identification of the distribution of $U_{it}$, especially its conditional variance. In what follows, we consider $X_{it} \in \mathbb{R}^1$ (i.e., $p = 1$) for simplicity of presentation and assume that $m$ is a cost function so that $U_{it} \geq 0$. We let $T = 2$ and assume the following conditions:

Assumption 1 (Identification). (i) $\{X_{it}, U_{it}, V_{it}, \alpha_i\}$ are i.i.d. across $i$ and stationary over $t$.

(ii) The inefficiency satisfies $U_{it}|X_{it} = x$ is distributed as either half-normal, $|N(0, \sigma_u^2(x))|$; or Exponential, $\text{Exp}(\sigma_u(x))$, where $0 < \sigma_u^2(x) < \infty$ is time-invariant.

(iii) Given $X_{it}$, the random noise $V_{it}$ is independent of $U_{it}$ and its distribution is symmetric with $E(V_{it}|X_{it} = x) = 0$ and $\text{Var}(V_{it}|X_{it} = x) = \sigma_v^2(x) < \infty$ for all $x$ and $t = 1, 2$.

(iv) (Conditional Independence) $f_{V_{it}|X_{it}, \alpha_i, X_{it}, V_{is}, \alpha_s}(v|x, \alpha, \tilde{x}, \tilde{v}) = f_{V_{it}|X_{it}}(v|x)$ for all $(v, x, \alpha, \tilde{x}, \tilde{v})$.
and \( t \neq \tau \), where \( f_{V_{it}} \) is the conditional density function of \( V_{it} \).

(v)(Common Support) The joint density of \( X_i = (X_{i1}, X_{i2}) \) satisfies \( f_{X_{i1},X_{i2}}(x,x) > 0 \) for all \( x \in \chi \), where \( \chi \) is the common support of \( X_{i1} \) and \( X_{i2} \).

(vi) The conditional characteristic function \( \phi_{V_{it}|X_{it}}(s|x) \) does not vanish for all \( s, x \) and \( t = 1, 2 \).

Assumptions 1-(i) and (ii) are standard assumptions for the stochastic frontier model for panel data. We allow for conditional heteroskedastic variances, \( \sigma^2_v(x) \) and \( \sigma^2_u(x) \), which may also be a function of other (environmental) variables, say \( Z_{it} \). In that case, the model allows for the “scaling property” specification \( U_{it} \sim G(Z_{it})|N(0,\sigma^2_u)| \) of Wang and Ho (2010).\(^2\) In addition to the half-normal and exponential distributions, Assumption 1-(ii) can be generalized to include single-parameter distributional families with bounded second moments. In Assumption 1-(iii), both \( U_{it} \) and \( V_{it} \) are related to \( X_{it} \) through the variance terms, but they are independent once \( X_{it} \) is given.

Assumption 1-(iv) is crucial for identification. It implies that conditional on the contemporaneous covariate \( X_{it} \), the disturbance term \( V_{it} \) is independent of \( \alpha_i, V_{i\tau}, \) and \( X_{i\tau} \) for any \( t \neq \tau \). For example, we let \( V_{it} = \sigma_v(X_{it})\eta_{it} \), where \( \sigma_v(x) \) is a bounded positive function and \( \eta_{it} \) are i.i.d \( N(0,1) \) that are independent of \( (\alpha_i, X_{i\tau}) \). It rules out lagged dependent variables in \( X_{it} \) and serially correlated disturbances. The assumption is strong but necessary for the deconvolution techniques that follow. For the case with serially correlated \( V_{it} \), we need at least three periods to identify the model as Evdokimov(2010). This case is summarized in the Appendix.

Assumption 1-(v) is also important for identification, but it generally holds in the stochastic frontier models since \( X_{it} \) includes input elements, prices, and some environment variables that are continuous. Assumption 1-(vi) is satisfied for most of the distributions, including

---

\(^2\)The scaling property specification possesses some appealing features. See Wang and Ho (2010), Wang and Schmidt (2002), Alvarez et al. (2006) and some others.
normal, log-normal, Cauchy, Laplace, $\chi^2$ and Student-t distributions.

If we only care about characterizing the inefficiency term $U_{it}$ (which may be reasonable in frontier analysis), then only Assumption 1 is needed for the identification. Under Assumption 1, we first introduce the following theorem, which is the key identification result of this paper. Proof of Theorem 1 is in the Appendix, which is based on proof of Lemma 1 in Evdokimov and White (2012).

**Theorem 1.** Suppose Assumption 1 holds. Then the distribution of inefficiency $U_{it}$ and the elasticity of the mean inefficiency $\mu_E := E[U_{it}]$ with respect to covariates $X_{it}$ are identified. That is, $\sigma^2_u(x)$ and $\xi_{\mu X} := \frac{\partial \mu_E}{\partial x} \frac{x}{\mu_E}$ are identified for all $x \in \chi$ and $t = 1, 2$.

Identification of the distribution of $U_{it}$ is achieved by identifying $\sigma^2_u(x)$ under Assumption 1-(ii). In particular, it is identified based on the following two moment conditions, which is obtained when $U_{it}$ is half-normal for instance:

$$E_x[Y_{it}(Y_{it} - Y_{i\tau})] - E_x[Y_{it}]E_x[(Y_{it} - Y_{i\tau})] = \left(1 - \frac{2}{\pi}\right) \sigma^2_u(x) + \sigma^2_v(x), \quad (3)$$

and

$$E_x[Y_{it}(Y_{it} - Y_{i\tau})^2] - E_x[Y_{it}]E_x[(Y_{it} - Y_{i\tau})^2] = \frac{(4 - \pi)\sqrt{2}}{\pi \sqrt{\pi}} \sigma^3_u(x), \quad (4)$$

where $E_x$ denotes expectation conditional on $X_{it} = X_{i\tau} = x$. These moment conditions are calculated similarly to the moment conditions in Simar et al. (2017).\(^3\) When $U_{it}$ has exponential distribution, we can obtain the same moment conditions by just replacing the scaling terms $(1 - 2/\pi)$ and $(4 - \pi)\sqrt{2}/\pi \sqrt{\pi}$ by 1 and 2 in equations (3) and (4), respectively.

In addition to the inefficiency term $U_{it}$, if the cost or production function $m$ is also of interest, we need additional conditions as follows, similarly as Evdokimov (2010).\(^3\)

\(^3\)If one would like to assume that higher central moments exist for $U$ and $V$, more odd moment conditions can be used to identify the unknown variance parameters $\sigma^2_u$ in a similar way.
**Assumption 2** (Identification of \( m \)). (i) \( m(x, \alpha) \) is weakly increasing in \( \alpha \) for all \( x \in \chi \).
(ii) \( \alpha_i | X_i \) is continuously distributed for all \( X_i \in \chi \times \chi \).
(iii) \( m(x, \alpha) \) and conditional densities \( f_{V \mid X_{it}}(v \mid x) \), \( f_{\alpha_i \mid X_{it}}(\alpha \mid x) \), and \( f_{\alpha_i \mid X_{i1}, X_{i2}}(\alpha \mid x_1, x_2) \) are almost everywhere continuous in the continuously distributed components of \( x, x_1, x_2 \) for all \( \alpha \) and \( v \).

Assumption 2-(i) is not too restrictive in the stochastic frontier models. It can be weakened to “monotonic in \( \alpha \)”. Assumption 2-(ii) and Assumption 2-(iii) are standard. For further discussions, see Evdokimov (2010).

Finally, in order to identify the idiosyncratic term \( \alpha_i \), we need to decide if it will be treated as a random-effect (RE) or a fixed-effect (FE) and use either set of the following assumptions.

**Assumption 3** (RE). (i) \( \alpha_i \) and \( X_i \) are independent. (ii) \( \alpha_i \) is Uniform on \([0, 1]\).

Assumption 3-(i) defines the random effects model, while Assumption 3-(ii) is a standard normalization. This normalization is necessary because the function \( m(x, \alpha) \) is modeled nonparametrically.

**Assumption 4** (FE). (i) \( m(x, \alpha) \) is strictly increasing in \( \alpha \) for all \( x \).
(ii) For some \( \bar{x} \in \chi \), \( m(\bar{x}, \alpha) = \alpha \) for all \( \alpha \).
(iii) \( \text{Supp}\{\alpha_i | (X_{it}, X_{i\tau}) = (x, \bar{x})\} = \text{Supp}\{\alpha_i | X_{it} = x\} \) for all \( x \in \chi \) and \( t \neq \tau \), where \( \text{Supp}\{\alpha_i | \vartheta\} \) denotes the support of \( \alpha_i \) conditional on an event \( \vartheta \).

Assumption 4-(i) is standard and guarantees invertibility of function \( m(x, \alpha) \) in \( \alpha \). Assumption 4-(ii) is a normalization given Assumptions 2-(ii) and 4-(i). Assumption 4-(iii) requires that the “extra” conditioning on \( X_{i\tau} = \bar{x} \) does not reduce the support of \( \alpha_i \). A conceptually similar support assumption is made by Altonji and Matzkin (2005).

Direct application of the results in Evdokimov (2010) leads to the following two theorems. We let \( F_{\alpha_i} \) denote the distribution function of \( \alpha_i \).
Theorem 2. Suppose Assumptions 1, 2, and 3 hold. Then \( m(x, \alpha), \sigma_u^2(x) \) and \( \sigma_v^2(x) \) are identified for all \( x \in \chi, \alpha \in (0,1) \) and \( t = 1, 2 \).

Theorem 3. Suppose Assumptions 1, 2, and 4-(i), (ii) hold. Then \( m(x, \alpha), \sigma_u^2(x), \sigma_v^2(x), F_{\alpha_i}(\alpha|X_{it} = x) \) and \( F_{\alpha_i}(\alpha) \) are identified for all \( x \in \chi, \alpha \in \text{Supp}\{\alpha_i|(X_{it}, X_{i(\sim t)}) = (x, \bar{x})\} \) and \( t = 1, 2 \). If in addition Assumption 4-(iii) is satisfied, \( \sigma_u^2(x), \sigma_v^2(x), F_{\alpha_i}(\alpha|X_{it} = x) \) and \( F_{\alpha_i}(\alpha) \) are identified for all \( x \in \chi, \alpha \in \text{Supp}\{\alpha_i|X_{it} = x\} \) and \( t = 1, 2 \).

The proofs of Theorems 2 and 3 directly follow from Evdokimov (2010), which we sketch in the Appendix. Specifically, we consider a cost stochastic frontier model with fixed effects \( \alpha_i \), which is a common case in the literature. The proof can be easily extend to the random effects setting.

3 Estimation

Since we focus on nonparametric identification and estimation of the variance components in equation (2), we now derive their estimation strategy under \( T = 2 \). As a by-product, the elasticity of mean inefficiency with respect to the covariates \( X_{it}, \xi_{\mu X} \), is also estimated. Specifically, consistent estimators for the heterogeneous variance of the inefficiency and random noise (i.e., \( \sigma_u^2 \) and \( \sigma_v^2 \)) are obtained by taking advantage of the conditional covariance structure of the panel model. This is consistent with the literature on nonparametric panel model estimation as well as the panel stochastic frontier model. For instance, Wang (2003) proposes a novel method for estimating nonparametric panel data models that utilizes the information contained in the covariance structure of the model’s disturbances, as do Henderson et al. (2008) and some others (see Wang et al., 2005).

Note that the variance of the inefficiency term and the random noise are important in the stochastic frontier model, both theoretically and empirically. Theoretically, technical or cost
inefficiency, which is usually defined as $E(U_{it}|\varepsilon_{it})$, is a function of $\sigma_u^2$ and $\sigma_v^2$, and they are closely related to one another. Here, we allow $E(U_{it}|\varepsilon_{it})$ to be not only a function of $\sigma_u^2$ and $\sigma_v^2$ but also a function of $X_{it}$ though the variance components. Wang and Schmidt (2009) point out that the variance of random noise matters and the technical efficiency estimate $E(U_{it}|\varepsilon_{it})$ is a shrinkage of $U_{it}$ toward its mean in a probabilistic sense. Empirically, the expectation of time-varying inefficiency $E(U_{it})$ equals $\sqrt{2/\pi}\sigma_u$ under the half-normal assumption and $\sigma_u$ under the exponential distributional assumption. Both are monotonic functions of the inefficiency variance. Alternatively, instead of using the formula for $E(U_{it}|\varepsilon_{it})$, we can use the best linear predictor of $U_{it}$ given $\varepsilon_{it}$ (i.e., $a + b\varepsilon_{it}$ for some finite constant $a$ and $b$), which was analyzed in detail in Waldman (1984). In particular, a simple calculation leads to $b = Var(U_{it})/(Vau(U_{it}) + Var(V_{it}))$ and $a = E(U_{it})(1 + b)$.

To estimate $\sigma_u^2(x)$ and $\sigma_v^2(x)$, we first define $A(x)$ as the conditional covariance between $Y_{it}$ and its first difference ($Y_{it} - Y_{i\tau}$) and $B(x)$ as the conditional covariance between $Y_{it}$ and $(Y_{it} - Y_{i\tau})^2$. It follows that, for the half-normal $U_{it}$ case, the moment conditions in (3) and (4) are written as

\[
A(x) = \left(1 - \frac{2}{\pi}\right)\sigma_u^2(x) + \sigma_v^2(x)
\]
\[
B(x) = \frac{(4 - \pi)\sqrt{2}}{\pi\sqrt{\pi}} \left\{\sigma_u^2(x)\right\}^{3/2},
\]

from which $\sigma_u^2(x)$ and $\sigma_v^2(x)$ are identified as

\[
\sigma_u^2(x) = \left\{\frac{\pi\sqrt{\pi}}{(4 - \pi)\sqrt{2}} B(x)\right\}^{2/3}
\]
\[
\sigma_v^2(x) = A(x) - \left(1 - \frac{2}{\pi}\right) \left\{\frac{\pi\sqrt{\pi}}{(4 - \pi)\sqrt{2}} B(x)\right\}^{2/3}.
\]
For the exponential case, we similarly have

\[ \sigma_n^2(x) = \left\{ \frac{1}{2}B(x) \right\}^{2/3} \]

\[ \sigma_p^2(x) = A(x) - \left\{ \frac{1}{2}B(x) \right\}^{2/3}. \]

Then, \( \sigma_n^2(x) \) and \( \sigma_p^2(x) \) can be estimated using the following conditional expectation estimators:

\[
\hat{A}(x) = \frac{1}{2} \left\{ \sum_{i=1}^{n} Y_{i1}(Y_{i1} - Y_{i2})\omega_{i,A1}(x) - \sum_{i=1}^{n} Y_{i1}\omega_{i,A1}(x)\sum_{i=1}^{n}(Y_{i1} - Y_{i2})\omega_{i,A1}(x) \right\} \quad (5)
\]

\[
+ \frac{1}{2} \left\{ \sum_{i=1}^{n} Y_{i2}(Y_{i2} - Y_{i1})\omega_{i,A2}(x) - \sum_{i=1}^{n} Y_{i2}\omega_{i,A2}(x)\sum_{i=1}^{n}(Y_{i2} - Y_{i1})\omega_{i,A2}(x) \right\},
\]

\[
\hat{B}(x) = \frac{1}{2} \left\{ \sum_{i=1}^{n} Y_{i1}(Y_{i1} - Y_{i2})^2\omega_{i,B1}(x) - \sum_{i=1}^{n} Y_{i1}\omega_{i,B1}(x)\sum_{i=1}^{n}(Y_{i1} - Y_{i2})^2\omega_{i,B1}(x) \right\} \quad (6)
\]

\[
+ \frac{1}{2} \left\{ \sum_{i=1}^{n} Y_{i2}(Y_{i2} - Y_{i1})^2\omega_{i,B2}(x) - \sum_{i=1}^{n} Y_{i2}\omega_{i,B2}(x)\sum_{i=1}^{n}(Y_{i2} - Y_{i1})^2\omega_{i,B2}(x) \right\},
\]

where

\[ \omega_{i,j}(x) = \frac{K \left( (X_{i1} - x)/h_j \right) K \left( (X_{i2} - x)/h_j \right)}{\sum_{i=1}^{n} K \left( (X_{i1} - x)/h_j \right) K \left( (X_{i2} - x)/h_j \right)} \]

for \( j = A1, A2, B1, B2 \). The \( A1 \) and \( A2 \) denote the bandwidth index for \( \omega_{i,A1} \) and \( \omega_{i,A2} \) in \( A(x) \) respectively and \( B1 \) and \( B2 \) stand similar for \( B(x) \). Note that \( K \) is a non-negative kernel function and \( h_j \) is a bandwidth that is common for \( t = 1, 2 \).\(^4\) In that case, for selecting

\(^4\)For the multivariate case (i.e., \( p > 1 \)), we let

\[ \omega_{i,j}(x) = \frac{K \left( H^{-1}_j(X_{i1} - x) \right) K \left( H^{-1}_j(X_{i2} - x) \right)}{\sum_{i=1}^{n} K \left( H^{-1}_j(X_{i1} - x) \right) K \left( H^{-1}_j(X_{i2} - x) \right)} \]

for \( j = A, B \), where \( K \) is a non-negative \( p \)-variate kernel function, and \( H_j \) is a \( p \times p \) bandwidth matrix that is symmetric and positive definite. The rest of the discussion holds if we simply consider the product kernel \( K(r) = \prod_{t=1}^{p} k(r_t) \) and \( H_j = h_j I_p \) for some bandwidth parameter \( h_j \), where \( I_p \) is the identity matrix of rank.
and \( h_B \), we can use nonparametric covariance estimation approaches, which we summarize two existing methods as follows.\(^5\)

The first approach is the adapted rule of thumb method:

\[
h = C \cdot X_{sd}n^{-1/(4+2p)}
\]

where \( p \) is the dimension of \( X \); \( X_{sd} \) is the sample standard deviation of \( \{X_{it}\}_{i=1}^n \); \( C \) is a constant and in practice we choose \( C = 1.06 \). Note that though \( X \in R^1 \), we use two dimensions based on the conditional argument \( E_x(\cdot) := E(\cdot|X_{it} = X_{i\tau} = x) \). Hence, the power of \( n \) is \(-1/(4+2p)\) rather than \(-1/(4+p)\).

Alternatively, one can directly apply the least squares leave-one-out cross validation method to the covariance, following Li et al. (2007) and Diggle and Verbyla (1998). For choosing \( h_A \), for instance, we suggest minimizing the following cross-validation criterion

\[
CV_{LS}(h) = \sum_{i=1}^n (A^0(X_i) - \hat{A}_h(X_{-i}))^2 W(X_i),
\]

where \( A^0(X_i) = Y^0_{it}(Y^0_{it} - Y^0_{i\tau}) \) with \( Y^0_{it} = Y_{it} - n^{-1} \sum_{i=1}^n Y_{it} \) and \( \hat{A}_h(X_{-i}) \) is the leave-one-out estimator of \( A(x) \) with the bandwidth \( h \), defined in equation (5). The \( W(x) \) is a weight function, such as the probability density function of \( X \). If \( n \) is large, above cross validation evaluation could be performed on a random subsample of \( m \) units, where \( m < n \) to reduce the computational burden. The \( h_B \) can be derived similarly by replacing \( (Y_{it} - Y_{i\tau}) \) with \( (Y_{it} - Y_{i\tau})^2 \).

Though both the identification theorem in Section 2 and the estimation strategy in Section 3 are derived based on the basic setting with \( T = 2 \), they can be extended to more general cases with \( T > 2 \) accordingly. A sequential conditional independence assumption

\(^5\)Here \( A \) and \( B \) are generic symbols for \( A_1, A_2 \) and \( B_1, B_2 \).
and a corresponding common support assumption are needed to identify the models and a consecutive-period estimation strategy is also proposed. Moreover, we consider the serial correlations within $V$ and $U$ respectively in the cases with multiple periods. For details, please refer to the Appendix (E).

4 Asymptotics

In this section, we drive the asymptotic properties of $\hat{\sigma}_u^2(x)$ and $\hat{\sigma}_v^2(x)$. For any $t \neq \tau (\tau = 1, 2)$ and $X_{it} \in R^1$, we let

$$\tilde{Y}_{[i,t,\tau]} = \begin{pmatrix} Y_{it} \\ Y_{it} - Y_{i\tau} \\ (Y_{it} - Y_{i\tau})^2 \end{pmatrix}$$

and denote $\tilde{Y}_{[i,t,\tau],j}$ as the $j$th element of $\tilde{Y}_{[i,t,\tau]}$, $m_j(x) = E(\tilde{Y}_{[i,t,\tau],j} | X_{it} = X_{i\tau} = x)$ for $j = 1, 2, 3$. We assume the following regularity conditions from Yin et al. (2010).

**Assumption 5.**

(i) $X_{it}$ has compact support and probability density $f(x)$, which is bounded away from zero and has two continuous derivatives.

(ii) For any $1 \leq j_1, j_2 \leq 3$, there exists $\delta \in [0, 1)$ such that $\sup_x E[|\tilde{Y}_{[i,t,\tau],j_1} \tilde{Y}_{[i,t,\tau],j_2}|^{2+\delta} | X_{it} = X_{i\tau} = x] < \infty$.

(iii) The conditional mean $E[\tilde{Y}_{[i,t,\tau],j} | X_{it} = X_{i\tau} = x]$ has two continuous derivatives.

(iv) $E[\tilde{Y}_{[i,t,\tau],j_1}^{k_1} \tilde{Y}_{[i,t,\tau],j_2}^{k_2} \tilde{Y}_{[i,t,\tau],j_3}^{k_3} \tilde{Y}_{[i,t,\tau],j_4}^{k_4} | X_{it} = X_{i\tau} = x]$ has two continuous derivatives in $x$ for $k_1, k_2, k_3, k_4 \in \{0, 1\}$, where $1 \leq j_1, j_2, j_3, j_4 \leq 3$, and $j_1, j_2, j_3, j_4$ are not necessarily different.

(v) The Bandwidth satisfies $h \to 0$ and $nh^5 \to c > 0$ for some $0 < c < \infty$.

(vi) Kernel function $K(v)$ is a bounded probability density function symmetric about 0. For the $\delta$ in (ii), $\int K^{2+\delta}(v)v^j dv < \infty$ for $j = 0, 1, 2$. For two arbitrary indices $v_1$ and $v_2$, $|K(v_1) - K(v_2)| \leq K_c|v_1 - v_2|$ for some $K_c > 0$. 

13
(vii) \(|A(x)| > 1/M_1 > 0, |B(x)| > 1/M_2 > 0\) for all \(x \in \chi\), where \(\chi\) is the support of \(x\) defined in Assumption 1-(v) and \(M_1\) and \(M_2\) are some positive constants.

Assumption 5-(i) restricts the density of the covariates \(X_{it}\) and it is slightly stronger than Assumptions 1-(v) and 2-(iii). Assumption 5-(ii) is a moment requirement for the dependent variables. Assumptions 5-(iii) and (iv) are smoothness constraints on the conditional mean and conditional variance, respectively (Fan, 1993 and Yao and Tong, 1998). Assumption 5-(v) holds with the optimal bandwidth choice, which yields the optimal convergence rate as optimal rate of convergence \(n^{-1/5}\). When \(p > 1\), we suppose \(nh^{4+p} \rightarrow c > 0\). Assumption 5-(vi) is a standard requirement for the kernel function (Li and Racine, 2007), which is trivially satisfied by the Gaussian and Epanechnikov kernels. Assumption 5-(vii) ensures the reciprocal of \(A(x)\) and \(B(x)\) are bounded which is necessary for the convergence proof of target parameters \(\sigma^2_u(x)\) and \(\sigma^2_v(x)\).

Define \(\sigma_{j_1j_2}(x)\) as the \((j_1, j_2)\)th element of the \(3 \times 3\) matrix \(\Sigma(x) = \text{Var}[\tilde{Y}_{[i,t,\tau]}|X_{it} = X_{ir} = x]\) for \(1 \leq j_1, j_2 \leq 3\), and \(\hat{\sigma}_{j_1j_2}(x)\) as its consistent estimator. Then, \(A(x) = \sigma_{12}(x)\) and \(B(x) = \sigma_{13}(x)\). The following lemma characterizes the joint asymptotic distribution of \((\hat{A}(x), \hat{B}(x))' = (\hat{\sigma}_{12}(x), \hat{\sigma}_{13}(x))'\), which extends Theorem 1 of Yin et al. (2010). We let \(\nu_0 = \int K^2(u)du\) and \(\mu_2 = \int u^2K(u)du\). For any continuously differentiable function \(g(x)\), we denote \(\dot{g}_s(x)\) and \(\ddot{g}_{ss}(x)\) be the first and the second derivative of \(g(x)\) with respect to the \(s\)th dimensional element of \(x\), respectively.

**Lemma 1.** Under Assumption 5,

\[
\sqrt{nh^2} \begin{pmatrix} 
\hat{\sigma}_{12}(x) - \sigma_{12}(x) - h^2 \gamma_{12}(x) \\
\hat{\sigma}_{13}(x) - \sigma_{13}(x) - h^2 \gamma_{13}(x)
\end{pmatrix} \Rightarrow_d \mathcal{N} \left( 0, \frac{\nu_0}{f(x)} \begin{pmatrix}
\phi^2_{12}(x) & \phi_{12,13}(x) \\
\phi_{12,13}(x) & \phi^2_{13}(x)
\end{pmatrix} \right)
\]

\(^6\)Recall the target distribution parameter \(\sigma_u(x)\) and \(\sigma_v(x)\) is a power function of the nonparametric covariance \(A(x)\) and \(B(x)\). This is also necessary for the uniform convergence of the target parameter.
as \( n \to \infty \), where \( \gamma_{j_{1},j_{2}}(x) = \frac{\mu_{x}}{2} \{ (\hat{\sigma}_{j_{1},j_{2}})_{11}(x) + 2(\hat{\sigma}_{j_{1},j_{2}})_{12}(x) \} + \frac{\mu_{x}}{2} \{ (\hat{\sigma}_{j_{1},j_{2}})_{22}(x) + 2(\hat{\sigma}_{j_{1},j_{2}})_{23}(x) \} \), \( \phi_{j_{1},j_{2}}(x) = \text{Var}[\varepsilon_{j_{1},j_{2}}(x)|X_{it} = X_{i\tau} = x] \), \( \varphi_{j_{1},j_{2},j_{3},j_{4}}(x) = \text{Cov}[\varepsilon_{j_{1},j_{2}}(x), \varepsilon_{j_{3},j_{4}}(x)|X_{it} = X_{i\tau} = x] \) and \( \varepsilon_{j_{1},j_{2}}(X) = \{ \tilde{Y}_{[i,\tau]} - m_{j_{1}}(X) \} \{ \tilde{Y}_{[i,\tau]} - m_{j_{2}}(X) \} - \sigma_{j_{1},j_{2}}(X). \)

From Lemma 1 and by the delta method, we derive asymptotic distributions of \( \hat{\sigma}_{u}^{2}(x) \) and \( \hat{\sigma}_{v}^{2}(x) \). The proof is in the Appendix.

**Theorem 4.** Suppose Assumptions 1 and 5 are satisfied. Then, as \( n \to \infty \),

\[
\sqrt{nh^{2}} \{ \hat{\sigma}_{u}^{2}(x) - \sigma_{u}^{2}(x) - h^{2}b(x)\gamma_{12}(x) \} \rightarrow_{d} \mathcal{N} \left( 0, \frac{\nu_{0}b^{2}(x)\phi_{13}^{2}(x)}{f(x)} \right)
\]

and

\[
\sqrt{nh^{2}} \{ \hat{\sigma}_{v}^{2}(x) - \sigma_{v}^{2}(x) - h^{2}(\gamma_{13}(x) - ab(x)\gamma_{12}(x)) \} \rightarrow_{d} \mathcal{N} \left( 0, \frac{\nu_{0}\{\phi_{12}^{2}(x) - 2ab(x)\phi_{12,13}(x) + a^{2}b^{2}(x)\phi_{13}^{2}(x)\} + \sigma_{13}^{-1/3}(x)}{f(x)} \right)
\]

where

\[
a = 1 - \frac{2}{\pi} \text{ and } b(x) = \frac{2}{3} \left( \frac{\pi \sqrt{\pi}}{(4 - \pi)\sqrt{2}} \right)^{2/3} \sigma_{13}^{-1/3}(x)
\]

for half-normal \( U_{it} \); or \( a = 1 \) and \( b(x) = \frac{2}{3}(1/2)^{2/3} \sigma_{13}^{-1/3}(x) \) for exponential \( U_{it} \).

Under the optimal bandwidth of \( h \sim n^{-1/6} \), the optimal convergence rate of the conditional variance estimators \( \hat{\sigma}_{u}^{2}(x) \) and \( \hat{\sigma}_{v}^{2}(x) \) is obtained as \( n^{-2/6} \), which is the standard result of the kernel estimator.\(^7\) It hence can be conjectured that the uniform rate of convergence of these two estimators is also the standard one, say \( O_{p}((\ln n/(nh^{2}))^{1/2} + 2h^{2}) \). Note that, however, this result does not extend to the regression function estimator \( \hat{m}(x, \alpha) \). For instance, Evdokimov (2010) shows that, depending on either ordinary smooth or super smooth \( U_{it} \), the

\(^7\)Recall that we have two kernels for a univariate \( X_{it} \) as we consider two consecutive periods.
convergence rate of \( \hat{m}(x, \alpha) \) is of order \((\ln n/n)^c\) and \((\ln n)^c\) for some \( c > 0 \).

The main reason for this difference lies in the different strategies for identification and estimation. Identification in the present paper uses properties of (conditional) characteristic functions and their (conditional) moments, while identification in Evdokimov (2010) is based on nonparametric deconvolution techniques.

## 5 Simulation

This section presents a Monte Carlo study of the finite sample properties of the proposed estimators \( \hat{\sigma}_u^2 \) and \( \hat{\sigma}_v^2 \) in the stochastic cost frontier model for both the fixed effects and random effects specifications (Greene 2005a, 2005b). We consider the following non-separable panel data model:

\[
Y_{it} = m(X_{it}, \alpha_i) + U_{it} + V_{it} \quad \text{for } i = 1, ..., n, \quad t = 1, ..., T = 2
\]

\[
m(x, \alpha) = \alpha + (1 + 0.5\alpha)(2x - 1)^3
\]

\[
U_{it} \sim |\mathcal{N}(0, \sigma_u^2(X_{it}))| \quad \text{or} \quad Exp(\sigma_u(X_{it}))
\]

\[
V_{it} \sim \mathcal{N}(0, \sigma_v^2(X_{it}))
\]

\[
X_{it} \sim iid \mathcal{U}[0, 1]
\]

\[
\alpha_i = \begin{cases} 
(p/T) \sum_{t=1}^{T} \sqrt{T^2(X_{it} - 0.5) + \sqrt{1 - \rho^2}\phi_i} & \text{for FE} \\
\sqrt{1 - \rho^2}\phi_i & \text{for RE}
\end{cases}
\]

with \( \rho = 0.5 \) and \( \phi_i \sim iid \mathcal{N}(0, 1) \)

---

8Here \( c \) is a function of \( d_1, d_2 \) and \( p \) where \( d_1 \) is the maximum continuous derivative of conditional cumulative distribution function \( F_{m}(t|x) \), \( d_2 \) is the maximum continuous derivative of the joint density \( f(x, x) \) and \( p \) is the dimension of \( X \). For details, see the Theorem 4-5, Theorem 7-8 in Evdokimov (2010).
for \( n = 2500 \) and 10000. The following specifications for the variance of inefficiency and noise are considered:

Specification I: \( \sigma_u^2 = 2, \sigma_v^2 = 1; \)

Specification II: \( \sigma_u^2 = 2X_{it}^2, \sigma_v^2 = 1; \)

Specification III: \( \sigma_u^2 = 2, \sigma_v^2 = X_{it}^2; \)

Specification IV: \( \sigma_u^2 = 2X_{it}^2, \sigma_v^2 = X_{it}^2; \)

Since we focus on identification and estimation of the variance components which hinges on the first three conditional moments of the compound error term \( \varepsilon_{it} = U_{it} + V_{it} \) in equation (2), the signal to noise ratio defined by \( \frac{\text{Var}(\varepsilon)}{\text{Var}(\varepsilon) + \text{Var}(m(X))} \) is important. In particular, for Specification I (with constant variance) the signal-to-noise ratio is \( 1.73/2.73 \approx 0.63 \) in the half-normal case and \( 3/4 = 0.75 \) in the exponential case. For the remaining specifications, the average signal-to-noise ratios are between 0.2 and 0.75 for different realizations of \( X_{it} \). Each Monte Carlo experiment is based on 1,000 replications. We use the rule of thumb bandwidth \( h = 1.06 \times stdv(X_{it})n^{-1/6} \) for simplicity and consistency. We can also use the proposed maximum likelihood or leave-one-out cross validation method to choose the unknown bandwidths. Recall that for univariate \( X_{it} \in \mathbb{R}^1 \) with the special conditional argument \( X_{i1} = X_{i2} = x \), we need to choose two bandwidths for each of \( \omega_{i,A1}, \omega_{i,A2}, \omega_{i,B1}, \omega_{i,B2} \) in equation (5) and (6) respectively.\(^9\)

We report the root integrated mean squared error (\( RIMSE \)), the root integrated squared

---

\(^9\)In the case of \( N=2500, T=2 \), one Monte Carlo simulation is less than 1 second with rule of thumb bandwidth but about one hour implementing the leave-one-out cross validation method to choose the bandwidth. In the application section, we use leave-one-out cross validation to choose the bandwidths.
bias ($RIBIAS^2$), and the root integrated variance ($RIVAR$), which are calculated as

$$RIMSE = \sqrt{\frac{1}{100} \sum_{k=1}^{100} \frac{1}{R} \sum_{r=1}^{R} [\hat{\sigma}_{ur}^2(x_k) - \sigma_u^2(x_k)]^2},$$

$$RIBIAS^2 = \sqrt{\frac{1}{100} \sum_{k=1}^{100} \left[ \frac{1}{R} \sum_{r=1}^{R} \hat{\sigma}_{ur}^2(x_k) - \sigma_u^2(x_k) \right]^2},$$

$$RIVAR = \sqrt{\frac{1}{100} \sum_{k=1}^{100} \left[ \frac{1}{R} \sum_{r=1}^{R} (\hat{\sigma}_{ur}^2(x_k))^2 - \left\{ \frac{1}{R} \sum_{r=1}^{R} \hat{\sigma}_{ur}^2(x_k) \right\}^2 \right]},$$

where $x_k = 0.1 + 0.008k$ for $k = 1, 2, ..., 100$ is the $k$th grid point between the 10th and 90th percentiles of $x$; $\hat{\sigma}_{ur}^2$ is the estimate of the cost inefficiency variance function in the $r$th Monte Carlo replication with $R = 1000$.\(^\text{10}\)

For comparison, we also present an infeasible estimator of the variance components: the “true” fixed effects and random effects ($TFRE$) models of Greene (2005a). Here, “Infeasible” means that the nonlinear model is fully known with correct specification of the distributions of the inefficiency and noise. We calculated the infeasible estimates using the Stata package: $sfpanel$ by Belotti, Daidone, Ilardi and Atella (2013). For the heteroskedastic specifications, $sfpanel$ reports the average of the estimated heterogeneous variance estimates. For computational parsimony, only 10 simulations are run for each specification and each case.\(^\text{11}\) We report the $RIMSE$ of the $TFRE$ estimator based on the Stata results. It is of interest to see how the feasible estimators, $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$, compare to the $TFRE$ estimator.

Table 1 contains the results for the design $N = 2500$ and $T = 2$. The rows of the table are divided into four panels for each of our four specifications: I, II, III and IV, respectively. For example, the first panel contains the results for Specification I ($\sigma_u^2 = 2$ and $\sigma_v^2 = 1$).

\(^{10}\)The support trimming procedure ensures that the joint density $f_{X_1, X_2}(x, x) > 0$, which is a key identification assumption in the nonparametric panel setting.

\(^{11}\)It takes one whole day to get one simulation with $sfpanel$ in Stata with a sample size of 10000 for one specification and one case.
The first three rows of each panel contain the \textit{RIMSE}, the \textit{RIBIAS}^2 and the \textit{RIVAR} (respectively) for the proposed estimator. The last row contains the \textit{RIMSE} of the \textit{TFRE} model, and is a useful benchmark for comparison. The columns of results are (left to right) for fixed effects with half-normal inefficiency, random effects with half-normal inefficiency, fixed effects with exponential inefficiency, random effects with exponential inefficiency.

Table 1 suggests that both proposed estimators of the variance components perform well when compared to the \textit{TFRE} variance estimators. For Specification I (first panel), the proposed estimators always outperform the true fixed or random effects models when the error components are homoskedastic, regardless of the parametric assumption on the inefficiency distribution. For example, 0.847 vs 0.887 for the half-normal fixed effects model and 0.532 vs 1.219 for the exponential fixed effects model. Under homoskedasticity, the proposed estimate of the noise variance is always more accurate (in terms of \textit{RIMSE}) than the inefficiency variance, and this result is consistent in most of the heteroskedastic specifications (second, third and fourth panel). Generally speaking, the infeasible estimator performs better in the heteroskedastic inefficiency and homoskedastic noise case (second panel). However, the proposed estimators outperform the infeasible estimators when inefficiency is homoskedastic (first panel and third panel). For instance, 0.648 vs 0.924 for the half-normal fixed effects model in the panel III. Moreover, the proposed estimators are comparable to the infeasible estimators when both error components are heteroskedastic (fourth panel).

Continuing with the results in Table 1, the proposed method works better (compared to itself) for heteroskedastic \(v\) (third and fourth panel) than for homoskedastic \(v\) (first and second panel). The rule of thumb bandwidth choice may be driving this result since the bandwidth should go to infinity in the homoskedastic design in light of the irrelevance of the covariates. The feasible estimators perform better (compared to itself) when inefficiency is exponentially distributed than when it is half-normal (first and second column VS third and fourth column). This is probably due to the fact that in all specifications the noise
$u_t$ is normally distributed, and disentangling moments from the same distributional family is always harder than from different families.\footnote{Half-normal and normal distribution are both in the super smooth distributional family while exponential distribution is in the ordinary smooth distributional family. It is always harder to disentangle inefficiency from the random noise in the former case than in the latter.} Another interesting pattern is that under the proposed method the fixed effects models and the random effects models yield similar results. This corresponds to Theorem 1 in which the identification and estimation of $\sigma_u^2$ and $\sigma_v^2$ does not hinge on the fixed effects or random effects assumptions. The slight difference between them is an artifact of finite sampling variability.

Similar findings can be found when we increase the sample size to 10000. Table 2 reports the results for the larger sample size design $N = 10000$ and $T = 2$. The proposed estimator works better in all specifications with either heteroskedastic inefficiency or heteroskedastic noise than the case with none of them (second, third and fourth panel vs first panel). The proposed estimator outperforms the infeasible estimator even in the most heteroskedastic case (fourth panel). Exponential stochastic frontier model outperforms the half-normal counterpart. A different finding is that the proposed estimator outperforms the infeasible estimator in the heteroskedastic inefficiency and homoskedastic noise case (second panel). For example, 0.677 vs 1.052 for the half-norm random effects model in the second panel. This is due to the fact that the proposed estimator captures the heteroskedasticity better than the infeasible estimator which researcher usually assume homoskedasticity for that. Another main finding is that in all specifications, the proposed estimators perform better in Table 2 than in Table 1, which demonstrates the consistency and decent rate of convergence of the proposed estimators.
6 Application

We apply the proposed method to a panel of US banks. We consider a cost frontier with either fixed or random effects, as well as time-varying inefficiency in equations (1) and (2). The data are from Feng and Serletis (2009) and are obtained from “Reports of Income and Condition” (Call Reports).\textsuperscript{13} The data are a sample of US banks, covering the period from 1998 to 2005. After deleting those with negative or zero input prices, there remains a balanced panel of 6010 observations for the 8-year period. In order to make the key identification assumption ID 5 more plausible, observations below the 1\textsuperscript{st} percentile or above the 99\textsuperscript{th} percentile are eliminated, leaving a balanced panel of \( n = 5,030 \) individual banks.

To make the analysis better reflect the \( T = 2 \) estimation described in the paper, we only use the last two years of data, 2004 and 2005. We have data on three output quantities and three input prices. The three outputs are consumer loans, \( Y_1 \); commercial loans, \( Y_2 \); and securities, \( Y_3 \). The commercial loans are non-consumer loans and include commercial real estate loans and industrial loans. Securities include all non-loan financial assets minus the sum of consumer loans, commercial loans, securities and equity. All outputs are deflated by the Consumer Price Index (CPI) to the base year of 1988. The three input prices are: the “per worker” wage bill, \( P_1 \); the interest rate for borrowed funds, \( P_2 \); and the price of physical capital, \( P_3 \).\textsuperscript{14} The total cost, \( C \), is the sum of three corresponding input costs: total salaries and benefits, expenses on premises and equipment, and total interest expenses. A more detailed description can be found in Feng and Serletis (2009). The means and standard deviations (std. dev.) for total cost, the three input prices, and the three output quantities are in Table 3.

Empiricists typically specify a log-linear or translog specification for the stochastic cost frontier model (see for example, Greene, 2005a; Kumbhakar and Tsionas, 2005; Kumbhakar,\textsuperscript{13}The data are publicly available on the Journal of Applied Econometrics website.\textsuperscript{14}The wage bill here is equal to total salaries and benefits divided by the number of full-time employees.
\[ c_{it} = \alpha_i + x_{it}'\beta + u_{it} + v_{it} \quad i = 1, 2, ..., n; \quad t = 1, 2, ..., T, \tag{8} \]

where \( c_{it} = \ln C_{it} \); \( \alpha_i \) is the fixed or random effect; \( u_{it} > 0 \) is time-varying cost inefficiency; \( v_{it} \) is statistical noise; and \( x_{it} = \ln X_{it} \) where \( X_{it} \) includes functions of the three output quantities and three input prices: \( Y_1, Y_2, Y_3, P_1, P_2, P_3 \), depending on the specification. However, the log-linear or translog specifications are just convenient first-order or second-order approximations of the “true” cost function.\(^{15}\) Therefore, it could be misspecified, and a nonparametric version may be preferred (see Wilson and Carey, 2004). Consequently, we relax the assumptions of log-linear or translog specification and estimate the following cost frontier model instead:

\[ c_{it} = m(x_{it}, \alpha_i) + \varepsilon_{it}, \quad \varepsilon_{it} = u_{it} + v_{it} \quad i = 1, ..., n, \quad t = 1, ..., T \tag{9} \]

with inefficiency either \( |N(0, \sigma_u^2)| \) or \( \text{Exp}(\sigma_u) \). As previously stated, the model is flexible in the following respects: the function \( m \) may be a separable or nonseparable function of \( \alpha_i \); both \( u_{it} \) and \( v_{it} \) may be heteroskedastic; we do not need a specific parametric assumption on the distribution of \( v_{it} \) except conditional symmetry.

Based on the summary statistics in Table 3, some interesting patterns can be observed. The average total cost rose significantly from 2004 ($6,559,520) to 2005 ($8,037,230). The wage bill increased slightly (from $48,960 to $50,830), while the interest rate and capital price were relatively stable over the period. In terms of output, consumer loans decreased while commercial loans and securities increased. Another point worth mentioning is that the supports of the covariates are highly variable. For example, in Table 3 the mean of the

\(^{15}\)Here, “true” means given output and input prices and assuming a Cobb-Douglas production function, minimizing the cost function for a given quantity of a single output.
interest rate is 0.01 for the full sample, while the mean of the wage bill (total wage bill divided by the number of employees) is $49,890. To facilitate discussion of the variables and their associated bandwidths, we transform all covariates into their logarithm forms with proper scales. Specifically, we multiply the wage bill (in thousands of dollars), the interest rate and the capital price by 100 before taking logarithms. For consumer loans, commercial loans and securities (all in thousands of dollars), we take logarithms directly, as they are relatively large.\footnote{In the figures that follow, we use ln(Wage), ln(Interest), ln(Capital), ln(Conloans), ln(Comloans) and ln(Securities) for consistency.}

To facilitate illustration we never run the regression with all six covariates at one time. Instead, we choose to run the regression with various pairs of covariates ($p = 2$): (wage bill, commercial loans), (wage bill, securities), (interest rate, commercial loans) and (interest rate, securities) and assume that $u_{it} \sim |N(0, \sigma_u^2)|$. For the case of $u_{it} \sim Exp(\sigma_u)$, we get very similar results.\footnote{Recall what we estimate for half-normal inefficiency term is the $\sigma_u^2$ while the effective variance is $(1 - 2/\pi)\sigma_u^2$. For exponential inefficiency, we are estimating the variance, $\sigma_u^2$.} These covariate pairs allow us to produce convenient 3D plots of the heteroskedastic variance parameters, as we shall see.

Figure A indicates the common support assumption hold for selected covariates across year 2004 and year 2005. Specifically, the support of covariates are almost identical except for interest rate. After transformation, the supports of the covariates are of the same magnitude which facilitates our illustration with tables and graphs. As illustrated in equation (7), the regression conditions on the covariates being the same across different periods (i.e, $X_{i1} = X_{i2} = x$, where $X_{it}$ is a $2 \times 1$ vector of each of our covariate pairs).\footnote{In practice they can be different, and this creates no problems as we use kernel smoothing.} Therefore, we allow for different bandwidths for each covariate in each pair in each of two years (i.e., four bandwidths). Table 4 reports the selected bandwidth by least square leave-one-out Cross Validation. Specifically, we allow the search grid to be 0.2 to 4 times the rule of thumb ($rot$) bandwidth in each period: $h_{rot1}$ and $h_{rot2}$ in the table. As we choose one pair of covariates

23
at a time, we set the order of the bandwidth as \( n^{-1/8} \) in Table 4 for each covariate pair. For example, in the first row of Table 4 for \( \ln(\text{Wage Bill}) \) the \( \text{rot} \) bandwidth in period 1 \((h_{\text{rot}1})\) is 0.074, and in period 2 \((h_{\text{rot}2})\) it is 0.075. These rule-of-thumb bandwidths (times 0.2) are used as starting values for the cross-validation procedure, which produces the optimal bandwidths \( \{h_{A1}, h_{A2}\}, \{h_{B1}, h_{B2}\} \) for \( A(x) \) and \( B(x) \), respectively, as previously defined.

As previously stated our focus is estimation of the variance components, which are very important in determining firm-level inefficiency scores, \( E(u_{it}|\varepsilon_{it}) \), in the stochastic frontier literature (see, Horrace, 2005, or Wang and Schmidt, 2009, for example). These inefficiency scores are functions of the variance component \( \sigma_u^2 \), and are (therefore) the focus of our empirical analysis. Assuming heteroskedastic variance allows us to uncover the determinants of inefficiency. Therefore, Figures 1-4 show the estimated 3D surface of \( \sigma_u^2(X_{it}) \) as a function of the four aforementioned covariate pairs. Figure 5-12 show the corresponding 2D curves of \( \sigma_u^2(X_{it}) \) as a function of each covariate, fixing the other covariate at its median (i.e., a single slice in each dimension for the 3D plots in Figures 1-4). The dash lines indicate the 95% confidence intervals derived by 199 bootstrap simulations with undersmoothed bandwidths.

For the “wage bill/commercial loans” pair in Figure 1, there is a peak in the wage bill (around 8) when \( \ln(\text{Comloans}) \) is at the 90th percentile (around 14). For the rest of the figure, the 3D surface is rather smooth and there are some small peaks as the wage bill changes and as commercial loans increase. We can see this point more clearly when we look at the 2D curves in Figure 5 and 6. In Figure 5, when fixing commercial loans at its median, the variance of inefficiency decreases as the logarithm of the wage bill increases from 8.0 to 8.3 and then becomes flat. In Figure 6 when the wage bill is fixed at its median, the variance of inefficiency increases monotonically as the logarithm of commercial loans increase from 8.0 to 14.5, then sharply decreases. Both covariates are highly nonlinearly correlated with the variance of the variance of the cost inefficiency.

For the “wage bill/securities” pair in Figure 2, the 3D surface shows three local peaks
when the wage bill is high, or securities are high, or both. There seems to be several local minimum points when the wage bill is at its 20 ∼ 40 percentiles and securities is at its 25 ∼ 50 percentiles. This pattern can be seen clearly in the 2D curves in Figure 7 and 8. When fixing the securities at its median, the variance of inefficiency decreases monotonically until it reaches about 8.15 and then it shows an inverse U shape as the wage bill rises in Figure 7. The maximum point is reached at around 8.7. There is a significant drop at around 8.15, i.e., wage ≈ $34.63 in thousand dollars. In Figure 8, when we fix the wage bill at its median, \( \sigma_u^2 \) shows a highly non-monotonic curve: increase at beginning, remaining flat, and then decreasing to its minimum point (nearly zero) when securities increase to around 13.2, and then increasing again. Clearly there is a local minimum point at \( \ln(\text{Securities}) \approx 13.2 \), i.e., securities = $540,365 in thousand dollars. Wage bill and securities seem to be both very important in determining \( \sigma_u^2 \) and (hence) cost inefficiency.

For the “interest rate/commercial loans” pair in Figure 3, the 3D surface is very similar to that of “wage bill/commercial loans” pair. There is a ‘peak at a low interest rate and a relatively high commercial loan level. The rest of the 3D surface is very smooth with a few small humps. For the 2D curves, Figure 10 shows similar monotonically increasing as that in the “wage bill/commercial loans” pair when the interest rate is fixed at its median. Figure 9 shows a very intuitive graph: at the beginning when the interest rate increases, the variance of the inefficiency increases as the total cost may increase; then it reach the highest inefficiency variance at around \( \ln(100\text{interest}) = 0.48 \), and after that it monotonically decreases as the interest rate continues to rise.

For the “interest rate/securities” pair, there are two obvious peaks in the 3D surface in Figure 4. The first one happens when interest rate is low and securities is relatively high (85 percentiles) while the other one happens at a point where the interest rate is relatively high (65 percentiles) and the securities are high. We can see this more clearly in the 2D curves in Figure 11 and 12. Figure 11 shows that the inefficiency variance is low at both high and low
values of the interest rate. The middle range of interest rate renders lower cost inefficiency when we fixed the securities at its median. On the other hand, the variance of inefficiency almost monotonically increases as the securities increase, when the interest rate is fixed at its median. This makes sense since security markets are more risky than other banks products, such as consumer loans and commercial loans. The more securities the bank sells, the more potential loss it may occur which increases cost. In the pairwise analysis, securities seems to contribute more monotonically to the cost inefficiency than the interest rate does.

For comparison, we also apply the “true” fixed/random effects stochastic frontier model by Greene (2005a,2005b) on the same data set, assuming half normal and exponential distribution for the cost inefficiency respectively. The results are reported in Table 5. We estimate the models by MLE and simulated moment of methods with the help of the Stata package *sfpanel*, assuming homoskedasticity for both the inefficiency and the noise variances. The “true” fixed effects model with exponential inefficiency does not converge.\textsuperscript{19} For the remaining three specifications, the results are robust. Two input prices and two outputs quantities are most important to total cost: wage and interest rate; commercial loans and securities. This justifies the pairwise covariates choice in the proposed nonparametric estimation from another perspective. For variance of inefficiency, they are all very significant and in the range of 0.004 $\sim$ 0.018. This is smaller than what we get in most of the 3D surface in Figure 1-4 and 2D graphs in Figure 5-12 which is at around 0.1. They are still around the same magnitude though sometimes the nonparametric estimation is more volatile as we allow the variance of inefficiency to be a function of covariates.\textsuperscript{20}

Finally, we estimate the elasticity of mean inefficiency with respect to each covariate (fixing the other at its median) in the covariate pairs: $\xi_{\mu x}$. Due to the limitation of space, we just

\textsuperscript{19}This always happens when we estimate a complicated model with MLE or simulated moments of methods.

\textsuperscript{20}We also tried to include the heteroskedasticity case in the “true” fixed/random stochastic frontier model using the *sfpanel* package but it didn’t converge even for the case with only two covariates, e.g., $\sigma_u^2 = f(x_{i11}, x_{i12})$. 

26
show the two elasticities from the “wage bill/securities” pair and “interest rate/securities” pair respectively. As it is known, the wage bill reflects the price for the labor which stands for one of the key input in the production process. The interest rate denotes the price for the other key input: capital. We want to see how the mean (cost) inefficiency changes with respect to the two input prices (i.e., wage bill and interest rate) fixing the output securities at its median. Figure 13-14 show these. In Figure 13, the elasticity decreases monotonically from positive to negative as wage bill increases. This indicates that mean cost inefficiency increases at the beginning and then decreases as wage bill rises. However, the elasticity with respect to interest rate shows a very different graph in Figure 14. It fluctuates up and down several times around the zero line and but it remains positive in most of the interest rate range, especially when interest rate goes high. This means that the mean cost inefficiency increases as the interest rate rises. All these heterogeneous variation of the cost inefficiency are of primary interest for the researchers as well as the bank management.

7 Conclusion

We propose a new nonparametric methodology to estimate the variance parameters of the time-varying inefficiency and time-varying noise in the stochastic production/cost frontier model for panel data. Compared with existing methods, the proposed methodology (a) doesn’t impose log-linearity for the cost/production function; (b) allows for heteroskedastic inefficiency and noise which may be a function of environmental variables; and (c) relaxes the distribution assumption for the random noise which is typically assumed to be normal or Laplace in the literature. Identification and estimation of the unknown production/cost function is built on the novel deconvolution methodology of Evdokimov (2010). Identification and estimation of the variance parameters does not require specific information about the cost/production function as they are built on the conditional first-difference transformation
of the model. The proposed method for estimating the variance parameters is straightforward and easy to implement as it requires no deconvolution techniques.

One interesting result is that once we implement the conditional first-difference transformation of the model, the fixed effects (corrected with $X$) or random effects (independent of $X$) doesn’t matter. Distributional assumptions on the inefficiency term only changes the scale of the heterogeneous variance and the relative ranks of firms are stable for different distributional assumptions.

For future research, some refinements of the nonparametric estimation can be pursued. Hall and Horowitz (2013) propose a new bootstrap method to construct more precise nonparametric confidence bands for functions which can be directly applied on our proposed estimators. Also more interesting applications can be explored and compared with the typical parametric panel stochastic frontier model.
References


33


Appendices

A Proof of Theorem 1

Proof. 1. Rewrite the model as:

\[ Y_{it} = \tilde{m}(X_{it}, \alpha_i) + \tilde{\varepsilon}_{it} \quad (10) \]

\[ \tilde{\varepsilon}_{it} = U_{it} + V_{it} - E[U_{it}], \quad i = 1, \ldots, n, \quad t = 1, \ldots, T \quad (11) \]

Observe that \( \tilde{m}(X_{i1}, \alpha_i) = \tilde{m}(X_{i2}, \alpha_i) \) when \( X_{i1} = X_{i2} = x \). For any \( x \in \chi \),

\[ \left( \begin{array}{c} Y_{i1} \\ Y_{i2} \end{array} \right) |\{X_{i1} = X_{i2} = x\} = \left( \begin{array}{c} \tilde{m}(x, \alpha) + \tilde{\varepsilon}_{i1} \\ \tilde{m}(x, \alpha) + \tilde{\varepsilon}_{i2} \end{array} \right) |\{X_{i1} = X_{i2} = x\}. \quad (12) \]

There are four conditions (which come from the original Kotlarski’s Lemma) need to check before applying Lemma 1 on Evdokimov and White (2012): (1) \( \tilde{m}, \tilde{\varepsilon}_{i1} \) and \( \tilde{\varepsilon}_{i2} \) are mutually (conditional) independent; (2) \( \tilde{m}, \tilde{\varepsilon}_{i1} \) and \( \tilde{\varepsilon}_{i2} \) have at least one absolute moment; (3) \( E(\tilde{\varepsilon}_{i1}) = 0 \); (4) characteristic function \( \phi_{\tilde{\varepsilon}_{it}}(s) \neq 0 \) for all \( s \) and \( t \in \{1, 2\} \).

For condition (1), condition on \( X_{i1} = X_{i2} = x \), \( \tilde{m} \) is independent with \( \tilde{\varepsilon}_{i1} \) and \( \tilde{\varepsilon}_{i2} \) respectively. The crucial part is to show \( \tilde{\varepsilon}_{i1} \) are conditionally independent with \( \tilde{\varepsilon}_{i2} \). Recall that \( \varepsilon_{it} = U_{it} + V_{it} \). \( V_{i1} \) is conditional independent with \( V_{i2} \) by Assumption ID 4. Based on Assumption ID 2 \( U_{it} \) can be represented as \( U_{it} = \sigma_u(X_{it})\eta_{it} \) where \( \sigma_u(x) \) is a bounded positive function and \( \eta_{it} \) are i.i.d \( |N(0, 1)| \) which is independent of \( (\alpha_i, X_i(-t)) \) where \( -t \) stands for other period. So \( U_{it} \) also satisfy the conditional independence in Assumption ID 4. As \( U_{it} \) and \( V_{it} \) are conditional independent with each other by Assumption ID 3, \( \varepsilon_{it} \) is also conditional independent with \( \varepsilon_{i(-t)} \), so as its demeaned version \( \bar{\varepsilon}_{it} = U_{it} + V_{it} - E(U_{it}) \) since \( E(U_{it}) \) is just a constant conditional on \( X_{i1} = X_{i2} = x \).

For Condition (2), it is trivially satisfied since \( m(\cdot|X_{i1} = X_{i2} = x) \) is a bounded function, \( U_{it} \sim |N(0, \sigma^2_u(X_{it}))| \), and \( V_{it} \) is conditionally symmetrically distributed with finite variance. For condition (3), obviously \( E(\tilde{\varepsilon}_{i1}|X_{i1} = X_{i2} = x) = 0 \). For condition (4), it holds since \( U_{it} \sim |N(0, \sigma^2_u(X_{it}))| \) and conditional characteristic function \( \phi_{\tilde{V}_{it}|X_{it}}(s|X_{it} = x) \) does not vanish for all
s, x and t = 1, 2 by Assumption ID 6.

Assumption ID 1-6 ensures that the Lemma 1 on Evdokimov and White (2012) applies to (12), conditional on the event $X_{i1} = X_{i2} = x$ and identifies the conditional distributions (or characteristic functions) of $m(x, \alpha)$, $\tilde{\varepsilon}_{i1}$ and $\tilde{\varepsilon}_{i2}$, given that $X_{i1} = X_{i2} = x$, for all $x \in \chi$. By the conditional independence Assumption ID 4 and its above discussion, $f_{\tilde{\varepsilon}_{it} | X_{it}, \alpha_i, X_{i(t-t)}, \tilde{\varepsilon}_{i(-t)}(\varepsilon_t | x, x(\tau), \tilde{\varepsilon}_{(-t)}) = f_{\tilde{\varepsilon}_{it} | X_{it}}(\tilde{\varepsilon}_i | x)$ for $t \in \{1, 2\}$. That is the conditional density $f_{\tilde{\varepsilon}_{it} | X_{it}}(\tilde{\varepsilon}_i | x)$ is identified for all $x \in \chi$, $\tilde{\varepsilon} \in R$ and $t \in \{1, 2\}$, as is the conditional characteristic function $\phi_{\tilde{\varepsilon}_{it} | X_{it}}(s | x)$ for all $s$.

\[
\phi_{\tilde{\varepsilon}_{it} | X_{it}}(s | x) = \exp \left( \int_0^s i E[Y_{it} \exp(i \varphi(Y_{it} - Y_{i\tau})) | X_{it} = X_{i\tau} = x] d\xi - is E(Y_{it} | X_{it} = X_{i\tau} = x) \right) \tag{13}
\]

where $i = \sqrt{-1}$ and $\tau \in \{1, 2\}$ and $\tau \neq t$.\(^{21}\)

2. Note that, given $X_{it} = x$, $U_{it} \sim |N(0, \sigma_u^2(x))|$ by Assumption ID 2 which is a one-parameter asymmetric distribution and $V_{it}$ is conditionally independent of $U_{it}$ and symmetric with $E(V_{it} | X_{it} = x) = 0$ and finite variance $\sigma_v^2(x)$ by Assumption ID 3. Therefore, given $X_{it} = x$, for the first three moments of the demeaned disturbance $\tilde{\varepsilon}_{it}$ can be written as

\[
E(\tilde{\varepsilon}_{it} | X_{it} = x) = 0
\]

\[
E(\tilde{\varepsilon}_{it}^2 | X_{it} = x) = (1 - \frac{2}{\pi}) \sigma_u^2(x) + \sigma_v^2(x)
\]

\[
E(\tilde{\varepsilon}_{it}^3 | X_{it} = x) = \frac{(4 - \pi)\sqrt{2}}{\pi \sqrt{\pi}} \sigma_u^3(x)
\]

3. We can also calculate the first three moments of the demeaned disturbance $\tilde{\varepsilon}_{it}$ conditional on $X_{it} = x$ by taking derivative of the conditional characteristic function and setting $s = 0$:

\[
E(\tilde{\varepsilon}^k) = (-i)^k \frac{\partial \phi_{\tilde{\varepsilon} | X} | s | x)}{\partial s} |_{s=0} \quad \text{for } k = 1, 2, 3.
\]

Plugging this into equation (13) and rearranging,

\(^{21}\)Note the slight notational abuse: the subscript $i$ is an index, while the bold $i$ is the imaginary number.
\[-i \cdot e^{0} \left( \int_{0}^{s} \frac{iE_{x}[Y_{it}\exp(i\xi(Y_{it} - Y_{i\tau}))]}{E_{x}[\exp(i\xi(Y_{it} - Y_{i\tau}))]} d\xi - i\cdot sE_{x}(Y_{it}))|_{s=0} = 0 \right) \right. \\
E_{x}[Y_{it}(Y_{it} - Y_{i\tau})] - E_{x}[Y_{it}]E_{x}[(Y_{it} - Y_{i\tau})] = (1 - \frac{2}{\pi})\sigma_{u}^{2}(x) + \sigma_{v}^{2}(x) \quad (14)

\[E_{x}[Y_{it}(Y_{it} - Y_{i\tau})^{2}] - E_{x}[Y_{it}]E_{x}[(Y_{it} - Y_{i\tau})^{2}] - 2E_{x}[(Y_{it} - Y_{i\tau})](E_{x}[Y_{it}(Y_{it} - Y_{i\tau})] - \\
E_{x}[Y_{it}]E_{x}[(Y_{it} - Y_{i\tau})]) = \frac{(4 - \pi)\sqrt{2}}{\pi \sqrt{\pi}} \sigma_{u}^{3}(x) \quad (16)\]

where \(E_{x}(.) = E(.)|X_{it} = X_{i\tau} = x) and \tau \in \{1, 2\} and \tau \neq t.\)

Given \(X_{it} = X_{i\tau} = x,\) the two unknowns \(\sigma_{u}^{2}(x)\) and \(\sigma_{v}^{2}(x)\) can be uniquely solved out by equation (15) and equation (35). The third term in equation (35) is zero as \(E_{x}(Y_{it} - Y_{i\tau}) = 0.\)

Varying \(x \in \chi,\) we can identify the nonparametric function of \(\sigma_{u}^{2}(.)\). Therefore, the distribution of inefficiency, \(U_{it},\) is identified.

Consequently, the elasticity of the mean efficiency with respect to the covariate \(X_{it}\) can be identified following Simar et al. (2017). Define the elasticity as \(\xi_{\mu X} = \frac{\partial \mu_{E}}{\partial x},\) note that \(\mu_{E}(x) := E(U_{it}(x)) = \frac{\sqrt{2}\sigma_{u}(x)}{\sqrt{\pi}}\) for the half-normal distribution (\(\sigma_{u}(x)\) for exponential distribution), so we can easily derive:

\[\xi_{\mu X} = \frac{\partial \sigma_{u}(x)}{\partial x} \frac{x}{\sigma_{u}(x)} = \frac{1}{3} \frac{\partial E(\tilde{\varepsilon}_{it}^{3}|X_{it} = x)}{\partial x} \frac{x}{E(\tilde{\varepsilon}_{it}^{3}|X_{it} = x)} \quad (17)\]

in which \(E(\tilde{\varepsilon}_{it}^{3}|X_{it} = x)\) is identified in equation (35).

For \(U_{it} \sim \text{Exp}(\sigma_{u})\) where variance is \(\sigma_{u}^{2}(x),\) just replace \((1 - \frac{2}{\pi})\) and \((\frac{(4 - \pi)\sqrt{2}}{\pi \sqrt{\pi}})\) by 1 and 2 in equation (15) and (35).\(^{22}\), and the result follows.

\(^{22}\) \(E(u) = \sigma_{u}, Var(u) = \sigma_{u}^{2}, Skewness(u) = 2.\)
B Sketch Proof of Theorem 2 and 3

Proof. We consider a cost stochastic frontier model with fixed effects $\alpha_i$ which is a common case in the literature. The proof can be easily extend to the random effects setting. With assumption ID 1-9 and FE 1-3, we can sketch a procedure for identifying the $m(X_{it}, \alpha_i)$ which is the production function (or profit function) in the SFA context.

(I) Step one: Identifying the conditional distribution of $\tilde{\epsilon}_{it}$ given $X_{it}$ exactly following the first step of the proof of Theorem 1. In particular, the conditional characteristic functions $\phi_{\tilde{\epsilon}_{it}|X_{it}}(s|x)$ are identified for all $t \in \{1, 2\}$.

(II) Step two: Identifying the distribution of $\tilde{m}(x, \alpha)|\{X_{i1} = x, X_{i2} = \bar{x}\}$ and $\alpha|\{X_{i1} = x, X_{i2} = \bar{x}\}$.

Conditional on the event $\{(X_{i1}, X_{i2}) = (x, \bar{x})\}$, by the conditional independence Assumption ID 4 and the normalization Assumption FE 2, we have

$$\phi_{Y_{i1}}(s|X_{i1} = x, X_{i2} = \bar{x}) = \phi_{\tilde{m}(X_{i1}, \alpha_i)}(s|X_{i1} = x, X_{i2} = \bar{x})\phi_{\tilde{\epsilon}_{i1}}(s|X_{i1} = x),$$

$$\phi_{Y_{i2}}(s|X_{i1} = x, X_{i2} = \bar{x}) = \phi_{\alpha_i}(s|X_{i1} = x, X_{i2} = \bar{x})\phi_{\tilde{\epsilon}_{i2}}(s|X_{i2} = \bar{x}).$$

Then,

$$\phi_{\tilde{m}(X_{i1}, \alpha_i)}(s|X_{i1} = x, X_{i2} = \bar{x}) = \frac{\phi_{Y_{i1}}(s|X_{i1} = x, X_{i2} = \bar{x})}{\phi_{\tilde{\epsilon}_{i1}}(s|X_{i1} = x)},$$

$$\phi_{\alpha_i}(s|X_{i1} = x, X_{i2} = \bar{x}) = \frac{\phi_{Y_{i2}}(s|X_{i1} = x, X_{i2} = \bar{x})}{\phi_{\tilde{\epsilon}_{i2}}(s|X_{i2} = \bar{x})}. \quad (18)$$

The left-hand side of equation (18) and equation (19) can be identified since the numerators can be identified from the data and the denominators is already identified from the previous step. The conditional CDFs of $F_{\tilde{m}(x, \alpha_i)|X_{i1}, X_{i2}}(w|x, \bar{x})$ and $F_{\alpha_i|X_{i1}, X_{i2}}(a|x, \bar{x})$ can be obtained following (Gil-Pelaez 1951; Evdokimov 2010):

$$F_{\tilde{m}(x, \alpha_i)|X_{i1}, X_{i2}}(w|x, \bar{x}) = \frac{1}{2} - \lim_{\chi \to \infty} \int_{-\chi}^{\chi} e^{-isw} \phi_{\tilde{m}(X_{i1}, \alpha_i)|X_{it}, X_{ir}}(s|x, \bar{x})ds, t, \tau = 1, 2, t \neq \tau.$$
\begin{equation*}
F_{\alpha|X_{i1},X_{i2}}(a|x, \bar{x}) = \frac{1}{2} - \lim_{\chi \to \infty} \int_{-\chi}^{\chi} e^{-is\alpha} \phi_{\alpha|X_{it},X_{it}}(s|x, \bar{x}) ds, t, \tau = 1, 2, t \neq \tau
\end{equation*}

(III) Step three: Identifying the functional \( m(x,.) \).

Inverting the conditional CDF \( F_{\tilde{m}(x,\alpha)|X_{i1},X_{i2}}(a|x, \bar{x}) \), we can obtain the conditional quantile function

\begin{equation*}
Q_{\tilde{m}(x,\alpha)|X_{i1},X_{i2}}(q|x, \bar{x}) = \inf \{ w : F_{\tilde{m}(x,\alpha)|X_{i1},X_{i2}}(w|x, \bar{x}) \geq q \}, q \in (0,1)
\end{equation*}

According to property of quantiles, we have

\begin{equation*}
\tilde{m}(x, a) = Q_{\tilde{m}(x,\alpha)|X_{i1},X_{i2}}(F_{\alpha|X_{i1},X_{i2}}(a|x, \bar{x})|x, \bar{x})
\end{equation*}

for all \( x \) and \( a \). And \( m(x, \alpha) = \tilde{m}(x, \alpha) - E(u) \) where \( E(u) \) is a function of \( \sigma_u(x) \).

(IV) Step four: identifying \( F_{\alpha_i}(a|X_{it} = x) \).

Similar to step 2, function \( \phi_{Y_{it}}(s|X_{it} = x) \) is identified from the data and hence

\begin{equation*}
\phi_{\tilde{m}(X_{it},\alpha_i)}(s|X_{it} = x) = \phi_{Y_{it}}(s|X_{it} = x)/\phi_{\tilde{\varepsilon}_{it}}(s|X_{it} = x)
\end{equation*}

is identified. Hence, the CDF \( F_{\tilde{m}(x,\alpha_i)|X_{it}}(w|x) \) and the quantile function \( Q_{\tilde{m}(x,\alpha_i)|X_{it}}(q|x) \) can be identified. By assumption FE 1, \( \tilde{m}(x, \alpha) \) is strictly increasing in \( \alpha \), by the property of quantiles,

\begin{equation*}
Q_{\alpha|X_{it}}(q|x) = \tilde{m}^{-1}(x, Q_{\tilde{m}(X_{it},\alpha_i)|X_{it}}(q|x))
\end{equation*}

Finally, one can identify the conditional cumulative distribution function \( F_{\alpha_i}(a|X_{it} = x) \) by inverting the quantile function \( Q_{\alpha_i|X_{it}}(q|x) \).

\[ \Box \]

C Proof of Lemma 1

Proof. Under Assumption 5, Theorem 1 in Yin et al. (2010) holds. That is

\begin{equation*}
\sqrt{n}\{ \hat{\sigma}_{j_1j_2}(x) - \sigma_{j_1j_2}(x) - \theta_n \} \to_d N(0, f^{-1}(x)\nu_0w_{j_1j_2}(x))
\end{equation*}
As $n \to \infty$, where $\theta_n = \frac{h^2}{2} \{ \hat{\sigma}_{j_1,j_2}(x) + 2\hat{\sigma}_{j_1,j_2}(x) \frac{f(j)}{f(x)} \}, \nu_0 = \int K^2(u)du, \mu_2 = \int u^2 K(u)du, f(x)$ is the probability density function of $X$ evaluated at $X = x$; $w_{j_1,j_2} \equiv \text{Var}(\varepsilon_{j_1,j_2}(i)|X_i)$, where $\varepsilon_{j_1,j_2}(i) = \{ \tilde{Y}_{[i,t],j_1} - m_{j_1}(X_i) \}\{ \tilde{Y}_{[i,t],j_2} - m_{j_2}(X_i) \} - \sigma_{j_1,j_2}(X_i)$. $\tilde{Y}_{[i,t],r}$ is defined in equation (34) and $m_j(x) = \mathbb{E}(\tilde{Y}_{[i,t],j}|X_{it} = x)$.

As we consider a panel model with $T=2$ periods, there are two kernels for the special conditional argument $E_x(.) = E(.|X_{it} = x)$ with univariate $X_{it}$. Assume the same bandwidth are chosen for the two kernels and let $(\hat{\sigma}_{j_1,j_2})_x(x)$ (or $(\hat{\sigma}_{j_1,j_2})_{xx}(x)$) denote the first (or second) order derivative with respect to the sth dimensional element of $X$, then we have

$$\sqrt{n}h^2\{ \hat{\sigma}_{j_1,j_2}(x) - \sigma_{j_1,j_2}(x) - \theta_n \} \to_d \mathcal{N}(0, f^{-1}(x)\nu_0 w_{j_1,j_2}(x))$$

as $n \to \infty$, where $\theta_n = \frac{h^2}{2} \{ (\hat{\sigma}_{j_1,j_2})_{11}(x) + 2(\hat{\sigma}_{j_1,j_2})_{1}(x) \frac{f(j)}{f(x)} \} + \frac{h^2}{2} \{ (\hat{\sigma}_{j_1,j_2})_{22}(x) + 2(\hat{\sigma}_{j_1,j_2})_{2}(x) \frac{f(j)}{f(x)} \}, \nu_0 = \int K^2(u)du, \mu_2 = \int u^2 K(u)du, f(x)$ is the probability density function of $X$ evaluated at $X = x$ ($X \in \mathbb{R}^2$ here), $f_s(x)$ denote the first (or second) order derivative with respect to the sth dimensional element of $X$; $w_{j_1,j_2} \equiv \text{Var}(\varepsilon_{j_1,j_2}(i)|X_i)$, where $\varepsilon_{j_1,j_2}(i) = \{ \tilde{Y}_{[i,t],j_1} - m_{j_1}(X_i) \}\{ \tilde{Y}_{[i,t],j_2} - m_{j_2}(X_i) \} - \sigma_{j_1,j_2}(X_i)$. $\tilde{Y}_{[i,t],r}$ is defined in equation (34) and $m_j(x) = \mathbb{E}(\tilde{Y}_{[i,t],j}|X_{it} = x)$.

Specifically, we have

$$\sqrt{n}h^2\{ \hat{\sigma}_{12}(x) - \sigma_{12}(x) - h^2 \gamma_{12} \} \to_d \mathcal{N}(0, f^{-1}(x)\nu_0 w_{12}(x))$$

where $\gamma_{12} = \frac{\mu_2}{2} \{ (\hat{\sigma}_{12})_{11}(x) + 2(\hat{\sigma}_{12})_{1}(x) \frac{f(j)}{f(x)} \} + \frac{\mu_2}{2} \{ (\hat{\sigma}_{12})_{22}(x) + 2(\hat{\sigma}_{12})_{2}(x) \frac{f(j)}{f(x)} \}$ and $w_{12} = \text{Var}(\varepsilon_{12}(i)|X_{it} = X_{i\tau} = x)$, and

$$\sqrt{n}h\{ \hat{\sigma}_{13}(x) - \sigma_{13}(x) - h^2 \gamma_{13} \} \to_d \mathcal{N}(0, f^{-1}(x)\nu_0 w_{13}(x))$$

where $\gamma_{13} = \frac{\mu_2}{2} \{ (\hat{\sigma}_{13})_{11}(x) + 2(\hat{\sigma}_{13})_{1}(x) \frac{f(j)}{f(x)} \} + \frac{\mu_2}{2} \{ (\hat{\sigma}_{13})_{22}(x) + 2(\hat{\sigma}_{13})_{2}(x) \frac{f(j)}{f(x)} \}$ and $w_{13} = \text{Var}(\varepsilon_{13}(i)|X_{it} = X_{i\tau} = x)$.

Jointly, we have
where $\phi_{12}^2(x) = w_{12}(x) = Var(\varepsilon_{12}(i)|X_{it} = X_{itr} = x)$, $\phi_{13}^2(x) = w_{13}(x) = Var(\varepsilon_{13}(i)|X_{it} = X_{itr} = x)$ and accordingly $\phi_{12,13}(x) = Cov(\varepsilon_{12}, \varepsilon_{13}|X_{it} = X_{itr} = x)$ with $\varepsilon_{j1j2}(i) = \{\tilde{Y}_{i,t,r,j1} - m_{j1}(X_i)\}{\tilde{Y}_{i,t,r,j2} - m_{j2}(X_i)} - \sigma_{j1j2}(X_i)$.

Then the conclusion follows.

\[\]  

D Proof of Theorem 4

Proof. By Lemma (1),

\[
\sqrt{n h^2} \left( \begin{array}{c} \hat{\sigma}_{12}(x) - \sigma_{12}(x) - h^2 \gamma_{12}(x) \\ \hat{\sigma}_{13}(x) - \sigma_{13}(x) - h^2 \gamma_{13}(x) \end{array} \right) \xrightarrow{d} \mathcal{N} \left( \begin{array}{cc} 0, \frac{\nu_0}{f(x)} & \phi_{12}^2(x) \\ \phi_{12,13}(x) & \phi_{13}^2(x) \end{array} \right)
\]

where $\phi_{12}^2(x) = w_{12}(x) = Var(\varepsilon_{12}(i)|X_{it} = X_{itr} = x)$, $\phi_{13}^2(x) = w_{13}(x) = Var(\varepsilon_{13}(i)|X_{it} = X_{itr} = x)$ and accordingly $\phi_{12,13}(x) = Cov(\varepsilon_{12}, \varepsilon_{13}|X_{it} = X_{itr} = x)$.

As $A(x) = \sigma_{12}(x)$, $B(x) = \sigma_{13}(x)$ by definition, the identification strategy simplifies to

\[
\sigma_u^2(x) = c^{-2/3} \sigma_{13}(x)^{2/3}
\]
\[
\sigma_v^2(x) = \sigma_{12}(x) - ac^{-2/3} \sigma_{13}(x)^{2/3}
\]

where $a$ and $c$ are constants. For example, if the inefficiency term $U_{it} \sim |N(0, \sigma_u^2(X_{it}))|$, $a = 1 - \frac{2}{\pi}$, $c = \frac{(4-\pi)\sqrt{2}}{\pi \sqrt{\pi}}$; if $U_{it} \sim \text{Exp}(b)$ where $Var(U_{it}) = \sigma_u^2(X_{it})$, $a = 1$ and $c = 2$.

By delta method, the asymptotic distribution of $\sigma_u^2(x)$ and $\sigma_v^2(x)$ follows

\[\]
\[
\sqrt{n \hat{h}^2} \left\{ \hat{\sigma}^2_u(x) - \sigma^2_u(x) - h^2 b(x) \gamma_{12}(x) \right\} \rightarrow_d \mathcal{N} \left( 0, \frac{\nu_0 b^2(x) \phi_{13}^2(x)}{f(x)} \right)
\]

where \( b(x) = \frac{2}{3} \sigma^{-1/3}_{13}(x) c^{-2/3} \) and

\[
\sqrt{n \hat{h}^2} \left\{ \hat{\sigma}^2_v(x) - \sigma^2_v(x) - h^2 (\gamma_{13}(x) - ab(x) \gamma_{12}(x)) \right\} \rightarrow_d \mathcal{N} \left( 0, D' \Sigma D \right)
\]

where \( D = \frac{\partial \sigma^2_v(x)}{\partial \sigma_{12,13}} = \begin{pmatrix} 1 \\ -b(x) \end{pmatrix} \) where \( \sigma_{12,13} = \left( \sigma_{12}, \sigma_{13} \right), b(x) = \frac{2}{3} \sigma^{-1/3}_{13} c^{-2/3} \) which is bounded constant by (vii) in Assumption 5;

\[
\Sigma = \frac{\nu_0}{f(x)} \begin{pmatrix} \phi_{12}^2(x) & \phi_{12,13}(x) \\ \phi_{12,13}(x) & \phi_{13}^2(x) \end{pmatrix}
\]

Specifically, we have

\[
\sqrt{n \hat{h}^2} \left\{ \hat{\sigma}^2_v(x) - \sigma^2_v(x) - h^2 (\gamma_{13}(x) - ab(x) \gamma_{12}(x)) \right\} \\
\rightarrow_d \mathcal{N} \left( 0, \frac{\nu_0 \left\{ \phi_{12}^2(x) - 2ab(x) \phi_{12,13}(x) + a^2b^2(x) \phi_{13}^2(x) \right\}}{f(x)} \right)
\]

The results follow.

\[\square\]

### E General Models with \( T > 2 \)

In this section, we consider the nonlinear panel model with more than two periods. Most of the identification results hold and the estimation techniques apply readily. However, there are some minor but crucial points worth mentioning explicitly.
E.1 When both U and V are Serially Uncorrelated

The first difference is the conditional independence assumption (i.e., Assumption 1 (iv)) and the common support assumption (i.e., Assumption 1 (v)) change accordingly with multiple time periods. When \( T > 2 \), especially when \( T \) is large, the original assumptions become restrictive and they imply conditional serial independence between \( V_{it} \) and \( X_{is} \), \( V_{is} \) where \( s \) stands for all the other time periods except \( t \) and there exists some common support among \((X_{i1}, X_{i2}, ..., X_{iT})\). So we proposed a sequential version of these two assumptions for \( T > 2 \) which is much weaker than the original ones as follows:

**Assumption 1 (iv'). (Sequential Conditional Independence)** \( f_{V_{it+1}|X_{it+1},\alpha_{it}}(v|x,\tilde{x},\tilde{v}) = f_{V_{it+1}|X_{it+1}}(v|x) \) for all \((v,x,\alpha,\tilde{x},\tilde{v})\) and \( t = 1, ..., T - 1 \), where \( f_{V_{it+1}} \) is the conditional density function of \( V_{it+1} \).

**Assumption 1 (v'). (Common Support)** The joint density of \((X_{it}, X_{it+1})\) satisfies \( f_{X_{it},X_{it+1}}(x,x) > 0 \) for all \( x \in \chi \) and \( t = 1, ..., T - 1 \), where \( \chi \) is the common support of \( X_{it} \) and \( X_{it+1} \). That is, \((X_{it}, X_{it+1}) \in \chi \times \chi \).

Assumption (iv') requires that conditional on contemporaneous \( X_{it} \), \( V_{it} \) is independent with \( X_{it-1} \) and \( V_{it-1} \) as well as the unobserved heterogeneity \( \alpha_{it} \). This is a much weaker assumption. For instance, in a production frontier model, conditional on current period’s inputs, the current period’s measurement error (or random noise) is very likely to be independence with the previous period’s inputs and measurement errors, and also the unobserved firm heterogeneity. Accordingly, Assumption (v') is a sequential common support assumption which only require the common support of covariates in two consecutive periods. For firms in a stable market, this is a very reasonable assumption.

All the identification results hold under Assumption 1 (i)-(iii), (iv'), (v') and (vi). We have similar identification theorem as follows: Suppose Assumption 1 (i)-(iii), (iv'), (v') and (vi) are satisfied. Then the distribution of inefficiency \( U_{it} \) and the elasticity of the mean inefficiency \( \mu_E := E[U_{it}] \) with respect to covariates \( X_{it} \) are identified. That is, \( \sigma^2_u(x) \) and \( \xi_{\mu X} := \frac{\partial \mu_E}{\partial x} \frac{x}{\mu_E} \) are identified.
for all $x \in \chi$ and $t \in 1, 2$. The proof is very similar to that for Theorem 1 for $T = 2$. We omit it here. Details are upon request.

The second difference is that the nonparametric covariance estimation should be changed accordingly as we have more than two periods now. A possible solution is to apply the proposed estimation techniques for each of the consecutive two periods (with large $n$ for each period) and average them. \footnote{Actually with multiple periods, we can let $\sigma_u(\cdot)$ be time varying. The benefit is we can look at the trace of the inefficiency variance to understand how the inefficiency evolve across time periods.} Under the weaker Assumption ID 4’-5’, we only leverage the covariance moment conditions between the $Y_{it}$ and $(Y_{it} - Y_{it+1})$, the $Y_{it}$ and $(Y_{it} - Y_{it+1})^2$, for $t = 1, ..., T - 1$ respectively.

Define $A(x)$ as the conditional covariance between $Y_{it}$ and its first difference $(Y_{it} - Y_{it+1})$ and $B(x)$ as the conditional covariance between $Y_{it}$ and $(Y_{it} - Y_{it+1})^2$ as in the main text. The moment conditions in (3) and (4) can be written as

$$A(x) = (1 - \frac{2}{\pi})\sigma_u^2(x) + \sigma_v^2(x)$$  \hspace{1cm} (20)

$$B(x) = (4 - \pi)\sqrt{\frac{2}{\pi}}\sigma_u^3(x).$$  \hspace{1cm} (21)

Since we assume $\sigma_u(\cdot)$ and $\sigma_v(\cdot)$ are time-invariant functions, we can flip around with $t$ and $\tau$ for the covariance estimation. They are both valid sample analogy for the $A(x)$ and $B(x)$. So they can be estimated as

\[
\hat{A}(x) = \frac{1}{T - 1} \sum_{t=1, \tau=t+1}^{T-1} \left\{ \sum_{i=1}^{n} Y_{it}(Y_{it} - Y_{it+1})\omega_{i,At}(x) - \sum_{i=1}^{n} Y_{it}\omega_{i,At}(x) \sum_{i=1}^{n} (Y_{it} - Y_{it+1})\omega_{i,At}(x) \right\} \quad (22)
\]

\[
+ \frac{1}{T - 1} \sum_{t=1, \tau=t+1}^{T-1} \left\{ \sum_{i=1}^{n} Y_{it}(Y_{it} - Y_{it+1})\omega_{i,At}(x) - \sum_{i=1}^{n} Y_{it}\omega_{i,At}(x) \sum_{i=1}^{n} (Y_{it} - Y_{it+1})\omega_{i,At}(x) \right\},
\]

\[
\hat{B}(x) = \frac{1}{T - 1} \sum_{t=1, \tau=t+1}^{T-1} \left\{ \sum_{i=1}^{n} Y_{it}(Y_{it} - Y_{it+1})^2\omega_{i,Bt}(x) - \sum_{i=1}^{n} Y_{it}\omega_{i,Bt}(x) \sum_{i=1}^{n} (Y_{it} - Y_{it+1})^2\omega_{i,Bt}(x) \right\} \quad (23)
\]

\[
+ \frac{1}{T - 1} \sum_{t=1, \tau=t+1}^{T-1} \left\{ \sum_{i=1}^{n} Y_{it}(Y_{it} - Y_{it+1})^2\omega_{i,Bt}(x) - \sum_{i=1}^{n} Y_{it}\omega_{i,Bt}(x) \sum_{i=1}^{n} (Y_{it} - Y_{it+1})^2\omega_{i,Bt}(x) \right\},
\]
where

$$
\omega_{i,j}(x) = \frac{K((X_{i1} - x)/h_j) K((X_{i2} - x)/h_j)}{\sum_{j=1}^{n_i} K((X_{i1} - x)/h_j) K((X_{i2} - x)/h_j)}
$$

for $j = \{At, A\tau, Bt, B\tau\}$ ($t = 1, ..., T - 1, \tau = t + 1$). The $At$ and $A\tau$ denote the bandwidth index for $\omega_{i,At}$ and $\omega_{i,A\tau}$ in $A(x)$ respectively and $Bt$ and $B\tau$ stand similar for $B(x)$. Note that $K$ is a non-negative kernel function and $h_j$ is a bandwidth that is common for $t = 1, 2$.

Let

$$
\tilde{Y}_{[i,t]} = \begin{pmatrix} Y_{it} \\ Y_{it} - Y_{it+1} \\ (Y_{it} - Y_{it+1})^2 \end{pmatrix}
$$

(24)

For each $t \in \{1, ..., T - 1\}$, we have a three by one matrix $\tilde{Y}_{[i,t]}$. Stacking it across individuals leads to a 3 by $n$ matrix $\tilde{Y}_{[t]}$. We suggest for each of these $\tilde{Y}_{[t]}$, choose the bandwidth $h_{At}$, $h_{At+1}$, $h_{Bt}$, $h_{Bt+1}$ accordingly to the base setting with $T = 2$. And then averaging them across $T - 1$ periods gives us the estimator $\hat{A}(x)$ and $\hat{B}(x)$ which could be used to estimate the target variance functions in equations (20) and (21).

### E.2 When V are Serially Correlated

In the basic setting with $T = 2$, $U_{it} \sim |N(0, \sigma^2_{u}(X_{it}))|$, and $V_{it}$ is conditional symmetrically distributed with finite variance which excludes serially correlation disturbances. When there are multiple periods, this condition is hard to satisfied. In this section, we consider the case disturbance $V_{it}$ in model (1) is serially correlated, specifically, $V_{it}$ follows an autoregressive process of order one (AR(1)): $V_{it} = \rho V_{it-1} + e_{it}$ for some constant $\rho$ with $0 < |\rho| < 1$.

In general, the model with serially correlated $V_{it}$ can be identified with three periods of time ($T = 3$). Consider the following modification of Assumption 1 in the main text:

**Assumption 1’ (Identification).** (i) $\{X_i, U_i, V_i, \alpha_i\}_{i=1}^{n_i}$ is a random sample, where $X_i \equiv (X_{i1}, X_{i2}, X_{i3})$, $U_i \equiv (U_{i1}, U_{i2}, U_{i3})$ and $V_i \equiv (V_{i1}, V_{i2}, V_{i3})$. Suppose $V_{it} = \rho V_{it-1} + e_{it}$ for all $t \geq 2$ with $0 < |\rho| < 1$.

(ii) The inefficiency term $U_{it}|X_{it} = x \sim |N(0, \sigma^2_{u}(x))|$ where $\sigma^2_{u}(.)$ is a time-invariant function; or
\[ U_{it} | X_{it} = x \sim \exp(\sigma_u(x)) \] where the variance is \( \sigma^2_u(x) \).

(iii) Given \( X_{it} = x \), random innovation noise \( e_{it} \) is independent of \( U_{it} \) and random noise \( V_{it} \) is symmetric with \( E(V_{it} | X_{it} = x) = 0 \) and finite conditional variance \( \text{Var}(V_{it} | X_{it} = x) = \sigma^2_v(x) < \infty \) for all \( x \) and \( t = 1, 2, 3 \).

(iv) (Conditional Independence 1) \( f_{e_{it} | X_{it}, \alpha_i, X_{i(-t)}}(e_{it} | x, \alpha, \bar{x}, \tilde{e}) = f_{e_{it} | X_{it}}(e | x) \) for all \( e, x, \alpha, \bar{x}, \tilde{e} \) and \( (t) = \{ \tau : \tau \neq t \text{ with } t, \tau = 1, 2, 3 \} \), where \( f_{e_{it}} \) is the conditional density function of \( e_{it} \).

(v) (Conditional Independence 2) \( f_{\varepsilon_{i1} | X_{i1}, \alpha_i, X_{i(-1)}}(\varepsilon_{i1} | x_1, \alpha, \bar{x}_(-1), e) = f_{\varepsilon_{i1} | X_{i1}}(\varepsilon | x) \) for all \( (x_1, x_1, \alpha, \bar{x}_(-1), e) \) and \( (-t) = \{ \tau : \tau = 2, 3 \} \), where \( f_{\varepsilon_{i1}} \) is the conditional density function of \( \varepsilon_{i1} \).

(vi) (Common Support) For each \( x \in \chi \) there is a \( x_1(x) \in \chi \) such that \( f_{X_{i1}, X_{i2}, X_{i3}}(x_1(x), x, x) > 0 \); also, \( f_{X_{i1}, X_{i2}}(x, x) > 0 \) for all \( x \in \chi \).

(vii) The conditional characteristic function \( \phi_{\varepsilon_{i1} | X_{i1}}(s | X_{i1} = x) \), \( \phi_{e_{it} | X_{it}}(s | X_{it} = x) \) does not vanish for all \( s \in R, x \in \chi \) and \( t = 2, 3 \).

Assumption 1'(i) incorporates the first order autoregressive relation for disturbance \( V_{it} \) and assume \( T = 3 \). Assumption 1'(ii) and 1'(iii) assume time-invariant variance functions for inefficiency \( U_{it} \) and innovation term \( e_{it} \) following the basic setting when \( T = 2 \) (i.e., Assumption 1(ii) and 1(iii)). As \( V_{it} = \sum_{s=0}^{\infty} \rho^t e_{i,t-s} \), the conditional independence between \( U_{it} \) and \( e_{it} \) still holds and \( e_{it} \) is conditional symmetric with \( \text{Var}(V_{it}) = \frac{1}{1-\rho^2} \sigma^2_e \) where \( \sigma^2_e \) is the variance of the innovation \( e_{it} \). These assumptions can be relax to allow time-varying variance functions, i.e., \( \sigma^2_{u,t} \) and \( \sigma^2_{v,t} \) when \( T > 2 \). Assumption 1'(iv) and 1'(v) are similar to Assumption 1(iv) and imply that \( \alpha_i, e_{it}, e_{is} \) and \( \varepsilon_{i1} \) are mutually independent for all \( t, s \geq 2, t \neq s \). Assumption 1'(v) is about the initial value problem which assumes \( \alpha_i \perp \varepsilon_{i1} \) conditional on \( (X_{i1}, X_{i2}, X_{i3}) = (x_1, x_2, x_3) \). Note that Assumption 1'(vi) is a very weak extension of the common support assumption in Assumption 1(v). Assumption 1'(vii) is the same as Assumption 1(vi) with an additional period \( t = 3 \).

Similar identification result holds: Suppose Assumption 1’ is satisfied. Then the distribution of inefficiency \( U_{it} \) and the elasticity of the mean inefficiency \( \mu_E := E[U_{it}] \) with respect to covariates \( X_{it} \) are identified. That is, \( \sigma^2_u(x) \) and \( \xi_{\mu X} := \frac{\partial \mu_E}{\partial x} \mu_E \) are identified for all \( x \in \chi \) and \( t = 1, 2, 3 \).

\(^{25}\)In that case, large number of observation in each period is needed to derive decent convergence rate for the target variance parameters.
The proof is as follows: use the ratio between conditional covariance of $Y_{i3}$ and $Y_{i3} - Y_{i2}$ and conditional covariance of $Y_{i2}$ and $Y_{i2} - Y_{i1}$ to pin down $\rho$, then identify the variance function $\sigma_u^2(x)$ and $\sigma_v^2(x)$ based on similar moment conditions as those in $T = 2$ case. Explicitly, we have

$$
\rho = E[(Y_{i3} - Y_{i2})Y_{i1}|X_i = (x_1(x), x, x)]/E[(Y_{i2} - Y_{i1})Y_{i1}|X_i = X_{i2} = x]
$$

$$
E_x[Y_{i1}(Y_{i1} - Y_{i2})] - E_x[Y_{i1}]E_x[(Y_{i1} - Y_{i2})] = (1 - \frac{2}{\pi})\sigma_u^2(x) + (1 - \rho)\sigma_v^2(x)
$$

$$
E_x[Y_{i1}(Y_{i1} - Y_{i2})^2] - E_x[Y_{i1}]E_x[(Y_{i1} - Y_{i2})^2] = \frac{(4 - \pi)\sqrt{2}}{\pi\sqrt{\pi}}\sigma_u^3(x)
$$

where $E_x = E(\cdot|X_{i1} = X_{i2} = x)$. Note the right hand side of the second equation is different with that in the basic setting with $T = 2$ and no serial correlation. Details of the derivation are in Appendix (E.4).

The estimation procedure is very standard as the basic setting with $T = 2$ or as the case $T > 2$ with no serial correlation when there are more than three periods.

### E.3 When U are Serially Correlated

In this section, we consider another serial correlated case when there are multiple periods: the disturbance $U_{it}$ in model (1) is serially correlated, specifically, $U_{it}$ follows an autoregressive process of order one (AR(1)): $U_{it} = \rho U_{it-1} + e_{it}$ for some constant $\rho$ with $0 < |\rho| < 1$.

Similarly, the model can be identified with three periods of time ($T = 3$). Consider the following modification of Assumption 1:

**Assumption 1" (Identification).** (i) $\{X_i, U_i, V_i, \alpha_i\}_{i=1}^n$ is a random sample, where $X_i \equiv (X_{i1}, X_{i2}, X_{i3})$, $U_i \equiv (U_{i1}, U_{i2}, U_{i3})$ and $V_i \equiv (V_{i1}, V_{i2}, V_{i3})$. Suppose $U_{it} = \rho U_{it-1} + e_{it}$ for all $t \geq 2$ with $0 < |\rho| < 1$.

(ii) The inefficiency term $U_{it}|X_{it} = x \sim N(0, \sigma_u^2(x))$ where $\sigma_u^2(.)$ is a time-invariant function; or $U_{it}|X_{it} = x \sim Exp(\sigma_u(x))$ where the variance is $\sigma_u^2(x)$.

(iii) Assume $V_{it}$ is symmetric conditional on $X_{it} = x$ with $E(V_{it}|X_{it} = x) = 0$ and finite conditional variance $\text{Var}(V_{it}|X_{it} = x) = \sigma_v^2(x) < \infty$ for all $x$ and $t = 1, 2, 3$. Given $X_{it} = x$, random
inefficiency innovation $e_{it}$ is independent of $V_{it}$ with finite conditional variance $\operatorname{Var}(e_{it}|X_{it} = x) = \sigma^2_{e}(x) < \infty$ for all $x$ and $t = 2, 3$.\footnote{The distribution of $e_{it}$ may be \textit{asymmetric} and since $E(U_{it}|X_{it} = x) \neq 0$ we do not have much to say about $E(e_{it}|X_{it} = x)$.}

(iv) (Conditional Independence 1) $f_{e_{it}|X_{it}, \alpha_{i}, X_{i(-1)}, e_{i(-1)}|\varepsilon_{i1}|x, \alpha, \bar{x}, \varepsilon_1, e_1) = f_{e_{it}|X_{it}}(e|x)$ for all $(e, x, \alpha, \bar{x}, \varepsilon_1)$ and $(-t) = \{\tau : \tau \neq t$ with $t, \tau = 1, 2, 3\}$, where $f_{e_{it}}$ is the conditional density function of $e_{it}$.

(v) (Conditional Independence 2) $f_{\varepsilon_{i1}|X_{it}, \alpha_{i}, X_{i(-1)}, e_{i}|\varepsilon_{1}|x_1, \alpha, \bar{x}(-1), e) = f_{\varepsilon_{i1}|X_{it}}(\varepsilon|x)$ for all $(\varepsilon_1, x_1, \alpha, \bar{x}(-1), e) \in \chi \times \chi \times \chi^2 \times \mathbb{R}^2$ and $(-t) = \{\tau : \tau = 2, 3\}$, where $f_{\varepsilon_{i1}}$ is the conditional density function of $\varepsilon_{i1}$.

(vi) (Common Support) For each $x \in \chi$ there is a $x_1(x) \in \chi$ such that $f_{X_{i1}, X_{i2}, X_{i3}}(x_1(x), x, x) > 0$; also, $f_{X_{i1}, X_{i2}}(x, x) > 0$ for all $x \in \chi$.

(vii) The conditional characteristic function $\phi_{\varepsilon_{i1}|X_{it}}(s|X_{i1} = x)$, $\phi_{e_{it}|X_{it}}(s|X_{it} = x)$ does not vanish for all $s \in \mathbb{R}$, $x \in \chi$ and $t = 2, 3$.

Assumption 1"(i) incorporates the first order autoregressive relation for inefficiency $U_{it}$ and assume $T = 3$. Assumption 1"(ii) and 1"(iii) assume time-invariant variance functions for inefficiency $U_{it}$, $V_{it}$ and innovation term $e_{it}$ following the basic setting when $T = 2$. As $U_{it} = \sum_{s=0}^{\infty} \rho^i e_{i,t-s}$, the conditional independence between $U_{it}$ and $V_{it}$ still holds as $e_{it}$ is conditional independent with $V_{it}$. Again, these assumptions can be relax to allow time-varying variance functions, i.e., $\sigma^2_{u,i,t}$ and $\sigma^2_{e,i,t}$ when $T > 2$. Assumption 1"(iv) and 1"(v) are similar to Assumption 1(iv) and imply that $\alpha_{i}, e_{it}, e_{is}$ and $\varepsilon_{i1}$ are mutually independent for all $t, s \geq 2, t \neq s$. Assumption 1"(v) is about the initial value problem which assumes $\alpha_{i} \perp \varepsilon_{i1}$ conditional on $(X_{i1}, X_{i2}, X_{i3}) = (x_1, x_2, x_3)$. Note that Assumption 1"(vi) and 1"(vii) are weak extensions of Assumption 1(v) and 1(vi), respectively.

similar identification theorem holds: Suppose Assumption 1" is satisfied. Then the distribution of inefficiency $U_{it}$ and the elasticity of the mean inefficiency $\mu_{E} := E[U_{it}]$ with respect to covariates $X_{it}$ are identified. That is, $\sigma^2_{u}(x)$ and $\xi_{\mu X} := \frac{\partial \mu_{E}}{\partial x} \frac{x}{\mu_{E}}$ are identified for all $x \in \chi$ and $t \in 1, 2, 3$.

The proof is very similar to the case with serial correlated disturbance $V_{it}$. First we use the ratio between conditional covariance of $Y_{i3}$ and $Y_{i3} - Y_{i2}$ and conditional covariance of $Y_{i2}$ and $Y_{i2} - Y_{i1}$ to pin down the serial correlation coefficient $\rho$, then identify the variance function $\sigma^2_{u}(x)$.
and $\sigma^2_v(x)$ based on similar moment conditions as those in $T = 2$ case. Explicitly, we have

$$\rho = E[(Y_{i3} - Y_{i2})Y_{i1}|X_i = (x_1(x), x_2)]/E[(Y_{i2} - Y_{i1})Y_{i1}|X_{i2} = x_2]$$  \hspace{1cm} (28)$$

$$E_x[Y_{i1}(Y_{i1} - Y_{i2})] - E_x[Y_{i1}]E_x[(Y_{i1} - Y_{i2})] = (1 - \rho)(1 - \frac{2}{\pi})\sigma^2_u(x) + \sigma^2_v(x)$$  \hspace{1cm} (29)$$

$$E_x[Y_{i1}(Y_{i1} - Y_{i2})^2] - E_x[Y_{i1}]E_x[(Y_{i1} - Y_{i2})^2] = (1 - \rho)^2 \frac{4 - \pi}{\pi \sqrt{\pi}} \sigma^3_u(x)$$  \hspace{1cm} (30)$$

where $E_x = E(\cdot | X_{i1} = X_{i2} = x)$. Note the right hand sides of the second equation and the third equation are different from those in the basic setting with $T = 2$ and no serial correlation. The serial correlation with $\{U_{it}\}$ has impact on both the second moment and the third moment. Details of the derivation are in Appendix (E.5).

The estimation procedure is very standard as the basic setting with $T = 2$ or as the case $T > 2$ with no serial correlation when there are more than three periods. We omit it here for brevity.

**E.4 Derivation of Moment Conditions (25)-(27)**

Let $X_i := (X_{i1}, X_{i2}, X_{i3})$. Recall that $\varepsilon_{it} := U_{it} + V_{it}$ and its demeaned version $\bar{\varepsilon}_{it} = U_{it} - E[U_{it}] + V_{it}$. Consider

$$E((Y_{i2} - Y_{i1})Y_{i1}|X_{i1} = X_{i2} = x) = E((\bar{\varepsilon}_{i2} - \bar{\varepsilon}_{i1})\bar{\varepsilon}_{i1}|X_{i1} = X_{i2} = x)$$

$$= E((V_{i2} - V_{i1})V_{i1}|X_{i1} = X_{i2} = x)$$

$$= E[(\rho - 1)V_{i1} + e_{i2}]V_{i1}|X_{i1} = X_{i2} = x]$$

The first equality holds since $\bar{m}(\alpha_i, X_{i1})$ is conditional independent with $\bar{\varepsilon}_{i1}$ and $\bar{\varepsilon}_{i2}$. The second equality holds as $U_{it} - E(U_{it})$ is conditional independent with $V_{it}$. The last equality holds by $V_{it} = \rho V_{it-1} + e_{it}$ for all $t \geq 2$ and some constant $\rho$ with $0 < |\rho| < 1$.  

49
Similarly, 

\[ E((Y_{i3} - Y_{i2})Y_{i1}|X_i = (x_1(x), x, x)) = E((\tilde{\varepsilon}_{i3} - \tilde{\varepsilon}_{i2})\tilde{\varepsilon}_{i1}|X_i = (x_1(x), x, x)) \]

\[ = E((V_{i3} - V_{i2})V_{i1}|X_i = (x_1(x), x, x)) \]

\[ = E(\rho(V_{i2} - V_{i1})V_{i1}|X_i = (x_1(x), x, x)) \]

\[ = E[\rho((\rho - 1)V_{i1} + e_{i2})V_{i1}|X_i = (x_1(x), x, x)] \]

So for any \( x \in \chi \), we can obtain \( \rho \) by taking ratio of above two equations

\[ \rho = E[(Y_{i3} - Y_{i2})Y_{i1}|X_i = (x_1(x), x, x)]/E[(Y_{i2} - Y_{i1})Y_{i1}|X_i = X_{i2} = x] \] (31)

Then consider the similar moment conditions as the basic setting with \( T = 2 \)

\[ Cov(Y_{i1}, (Y_{i1} - Y_{i2})|X_{i1} = X_{i2} = x) \]

\[ = Cov(\tilde{\varepsilon}_{i1}, \tilde{\varepsilon}_{i1} - \tilde{\varepsilon}_{i2}|X_{i1} = X_{i2} = x) \]

\[ = Var(\tilde{U}_{i1}|X_{i1} = X_{i2} = x) + Cov(V_{i1}, V_{i1} - V_{i2}|X_{i1} = X_{i2} = x) \]

\[ = (1 - \frac{2}{\pi})\sigma^2_u(x) + (1 - \rho)Var(V_{i1}|X_{i1} = X_{i2} = x) \]

\[ = (1 - \frac{2}{\pi})\sigma^2_u(x) + (1 - \rho)\sigma^2_v(x) \]

where \( \tilde{U}_{i1} = U_{i1} - E(U_{i1}) \). The second equality holds as \( V_{i1} \) is independent with \( e_{i2} \).

Similarly,

\[ Cov(Y_{i1}, (Y_{i1} - Y_{i2})^2|X_{i1} = X_{i2} = x) \]

\[ = Cov(\tilde{\varepsilon}_{i1}, (\tilde{\varepsilon}_{i1} - \tilde{\varepsilon}_{i2})^2|X_{i1} = X_{i2} = x) \]

\[ = E(\tilde{U}_{i1}^3|X_{i1} = X_{i2} = x) + Cov(V_{i1}, (V_{i1} - V_{i2})^2|X_{i1} = X_{i2} = x) \]

\[ = E(\tilde{U}_{i1}^3|X_{i1} = X_{i2} = x) + Cov(V_{i1}, ((1 - \rho)V_{i1} + e_{i2})^2|X_{i1} = X_{i2} = x) \]

\[ = \frac{(4 - \pi)\sqrt{2}}{\pi\sqrt{\pi}}\sigma^3_v(x) \]
Where the last equality holds since \( U_{it} \mid X_{it} = x \sim N(0, \sigma_u^2(x)) \) and \( E(V_{i1}^2 \mid X_{i1} = X_{i2} = x) = 0, \) \( V_{i1} \)

is independent with \( e_{i2}. \)

So we can identify the variance function by following moments:

\[
E_x[Y_i(Y_i - Y_{i2})] - E_x[Y_i]E_x[(Y_i - Y_{i2})] = (1 - \frac{2}{\pi})\sigma_u^2(x) + (1 - \rho)\sigma_v^2(x)
\]

(32)

\[
E_x[Y_i(Y_i - Y_{i2})^2] - E_x[Y_i]E_x[(Y_i - Y_{i2})^2] = \frac{(4 - \pi)\sqrt{2}}{\pi \sqrt{\pi}}\sigma_u^3(x)
\]

(33)

where \( E_x = E(\cdot | X_{i1} = X_{i2} = x) \). Once we identify the \( \sigma_v^2(x) \), by the same token of proof in

Theorem 1, we can identify the elasticity of mean inefficiency with respect to the covariates \( \xi_{\mu X}. \)

**E.5 Derivation of Moment Conditions (28)-(30)**

Let \( X_i := (X_{i1}, X_{i2}, X_{i3}). \) Recall that \( \varepsilon_{it} := U_{it} + V_{it} \) and its demeaned version \( \bar{\varepsilon}_{it} = U_{it} - E[U_{it}] + V_{it}. \)

Consider

\[
E((Y_{i2} - Y_{i1})Y_{i1} \mid X_{i1} = X_{i2} = x) = E((\bar{\varepsilon}_{i2} - \bar{\varepsilon}_{i1})\bar{\varepsilon}_{i1} \mid X_{i1} = X_{i2} = x)
= E((U_{i2} - U_{i1})U_{i1} \mid X_{i1} = X_{i2} = x)
= E[\{(\rho - 1)U_{i1} + e_{i2}\}U_{i1} \mid X_{i1} = X_{i2} = x]
\]

The first equality holds since \( \bar{\varepsilon}_{i1}(\alpha_i, X_{i1}) \) is conditional independent with \( \bar{\varepsilon}_{i1} \) and \( \bar{\varepsilon}_{i2}. \) The second equality holds as \( U_{it} - E(U_{it}) \) is conditional independent with \( V_{it}. \) The last equality holds by \( U_{it} = \rho U_{it-1} + e_{it} \) for all \( t \geq 2 \) and some constant \( \rho \) with \( 0 < |\rho| < 1. \)

Similarly,

\[
E((Y_{i3} - Y_{i2})Y_{i2} \mid X_i = (x_1(x), x, x)) = E((\bar{\varepsilon}_{i3} - \bar{\varepsilon}_{i2})\bar{\varepsilon}_{i2} \mid X_i = (x_1(x), x, x))
= E((U_{i3} - U_{i2})U_{i2} \mid X_i = (x_1(x), x, x))
= E(\rho(U_{i2} - U_{i1})U_{i1} \mid X_i = (x_1(x), x, x))
= E[\rho\{(\rho - 1)U_{i1} + e_{i2}\}U_{i1} \mid X_i = (x_1(x), x, x)]
\]
So for any \( x \in \chi \), we can obtain \( \rho \) by taking ratio of above two equations

\[
\rho = E[(Y_{i3} - Y_{i2})Y_{i1}|X_i = (x_1(x), x, x)]/E[(Y_{i2} - Y_{i1})Y_{i1}|X_{i1} = X_{i2} = x]
\]  \hspace{1cm} (34)

Then consider the similar moment conditions as the basic setting with \( T = 2 \)

\[
\text{Cov}(Y_{i1}, (Y_{i1} - Y_{i2})|X_{i1} = X_{i2} = x) = \text{Cov}(\tilde{\epsilon}_{i1}, \tilde{\epsilon}_{i1} - \tilde{\epsilon}_{i2}|X_{i1} = X_{i2} = x)
\]

\[
\text{Cov}(\tilde{U}_{i1}, \tilde{U}_{i1} - \tilde{U}_{i2}|X_{i1} = X_{i2} = x) + \text{Var}(V_{i1}|X_{i1} = X_{i2} = x)
\]

\[
= (1 - \rho)\text{Var}(U_{i1}|X_{i1} = X_{i2} = x) + \sigma^2_u(x)
\]

\[
= (1 - \rho)(1 - \frac{2}{\pi})\sigma^2_u(x) + \sigma^2_v(x)
\]

where \( \tilde{U}_{i1} = U_{i1} - E(U_{i1}), \tilde{U}_{i2} = U_{i2} - E(U_{i2}), \text{Var}(\tilde{U}_{i1}|X_{i1} = X_{i2} = x) = (1 - \frac{2}{\pi})\sigma^2_u \) as \( U_{i1}|X_{i1} = x \sim |N(0, \sigma^2_u(x))| \) by assumption.

Similarly,

\[
\text{Cov}(Y_{i1}, (Y_{i1} - Y_{i2})^2|X_{i1} = X_{i2} = x)
\]

\[
= \text{Cov}(\tilde{\epsilon}_{i1}, (\tilde{\epsilon}_{i1} - \tilde{\epsilon}_{i2})^2|X_{i1} = X_{i2} = x)
\]

\[
= \text{Cov}(\tilde{U}_{i1}, (\tilde{U}_{i1} - \tilde{U}_{i2})^2|X_{i1} = X_{i2} = x) + \text{Cov}(V_{i1}, (V_{i1} - V_{i2})^2|X_{i1} = X_{i2} = x)
\]

\[
= E((1 - \rho)^2\tilde{U}_{i1}^3|X_{i1} = X_{i2} = x) + \text{Cov}(V_{i1}, (V_{i1}^2 - 2V_{i1}V_{i2} + V_{i2}^2)|X_{i1} = X_{i2} = x)
\]

\[
= (1 - \rho)^2\frac{(4 - \pi)\sqrt{2}}{\pi\sqrt{\pi}}\sigma^3_u(x)
\]

Where the last equality holds since \( U_{it}|X_{it} = x \sim |N(0, \sigma^2_u(x))| \) and \( E(V_{i1}^3|X_{i1} = X_{i2} = x) = 0, V_{i1} \) is independent with \( V_{i2} \).

So we can identify the variance function by following moments:

\[
E_x[Y_{i1}(Y_{i1} - Y_{i2})] - E_x[Y_{i1}]E_x[(Y_{i1} - Y_{i2})] = (1 - \rho)(1 - \frac{2}{\pi})\sigma^2_u(x) + \sigma^2_v(x)
\]  \hspace{1cm} (35)
\[ E_x[Y_{i1}(Y_{i1} - Y_{i2})^2] - E_x[Y_{i1}]E_x[(Y_{i1} - Y_{i2})^2] = (1 - \rho)^2 \frac{(4 - \pi\sqrt{2})}{\pi\sqrt{\pi}} \sigma^2_u(x) \]

where \( E_x = E(\cdot | X_{i1} = X_{i2} = x) \). Once we identify the \( \sigma^2_u(x) \), by the same token of proof in Theorem 1, we can identify the elasticity of mean inefficiency with respect to the covariates \( \xi_{\mu X} \).
| $\sigma_u^2$ | $\sigma_v^2$ | $U_{it} \sim |N(0, \sigma_u^2)|$ | $U_{it} \sim \text{Exp}(\sigma_u)$ |
|---------|---------|-------------------------------|-------------------------------|
|         |         | Fixed Effects | Random Effects | Fixed Effects | Random Effects |
|         |         | ($\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$) | ($\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$) | ($\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$) | ($\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$) |
| $2^2$   | $1^2$  | $\text{RIMSE}$ | (0.847, 0.302) | (0.854, 0.298) | (0.532, 0.383) | (0.537, 0.385) |
|         |         | $\text{RIBIAS}^2$ | (0.229, 0.063) | (0.227, 0.079) | (0.071, 0.072) | (0.246, 0.050) |
|         |         | $\text{RIVAR}$ | (0.819, 0.297) | (0.831, 0.293) | (0.535, 0.370) | (0.538, 0.372) |
| $2x^2$  | $1^2$  | $\text{RIMSE}$ | (2.000, 0.127) | (0.169, 0.086) | (1.853, 0.703) | (0.056, 0.050) |
|         |         | $\text{RIBIAS}^2$ | (0.834, 0.251) | (0.827, 0.266) | (0.122, 0.102) | (0.117, 0.088) |
|         |         | $\text{RIVAR}$ | (0.592, 0.217) | (0.574, 0.213) | (0.274, 0.232) | (0.259, 0.224) |
| $2^2$   | $x^2$  | $\text{RIMSE}$ | (2.000, 0.332) | (2.000, 0.988) | (1.867, 0.565) | (0.572, 0.817) |
|         |         | $\text{RIBIAS}^2$ | (0.648, 0.209) | (0.614, 0.204) | (0.487, 0.313) | (0.484, 0.314) |
|         |         | $\text{RIVAR}$ | (0.613, 0.204) | (0.598, 0.199) | (0.494, 0.302) | (0.491, 0.304) |
| $2x^2$  | $x^2$  | $\text{RIMSE}$ | (2.000, 0.332) | (0.673, 0.193) | (0.575, 0.170) | (0.250, 0.173) | (0.235, 0.165) |
|         |         | $\text{RIBIAS}^2$ | (0.523, 0.126) | (0.411, 0.094) | (0.092, 0.034) | (0.074, 0.016) |
|         |         | $\text{RIVAR}$ | (0.423, 0.147) | (0.401, 0.141) | (0.233, 0.169) | (0.223, 0.164) |
| $2x^2$  | $x^2$  | $\text{RIMSE}$ | (0.667, 0.231) | (0.667, 1.092) | (1.085, 0.413) | (0.386, 0.917) |

Notes: $\text{RIMSE}$, $\text{RIBIAS}^2$, and $\text{RIVAR}$ are “Root of the Integrated Mean Squared Error,” “Root of the Integrated Squared Bias,” and “Root of the Integrated Variance,” respectively. “$\text{Exp}(b)$” is the exponential pdf: $f(x) = \frac{1}{b}e^{-\frac{x}{b}}$ for $x \geq 0$. $\text{TFRE}$ is the “True Fixed or Random [effects] Estimator” of Greene (2005a,b) with correct functional and distributional specifications.
Table 2: DESIGN II, $n=10,000$, $T=2$

<table>
<thead>
<tr>
<th>$\sigma^2_u$</th>
<th>$\sigma^2_v$</th>
<th>$U_{it} \sim [N(0, \sigma^2_u)]$</th>
<th>$U_{it} \sim \text{Exp}(\sigma_u)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed Effects</td>
<td>Random Effects</td>
<td>Fixed Effects</td>
</tr>
<tr>
<td></td>
<td>$(\hat{\sigma}^2_u, \hat{\sigma}^2_v)$</td>
<td>$(\hat{\sigma}^2_u, \hat{\sigma}^2_v)$</td>
<td>$(\hat{\sigma}^2_u, \hat{\sigma}^2_v)$</td>
</tr>
<tr>
<td>2 1</td>
<td>RIMSE</td>
<td>(0.609, 0.218)</td>
<td>(0.606, 0.220)</td>
</tr>
<tr>
<td></td>
<td>RIBIAS$^2$</td>
<td>(0.078, 0.016)</td>
<td>(0.049, 0.015)</td>
</tr>
<tr>
<td></td>
<td>RIVAR</td>
<td>(0.604, 0.218)</td>
<td>(0.605, 0.220)</td>
</tr>
<tr>
<td></td>
<td>TFRE</td>
<td>(2.000, 0.131)</td>
<td>(0.241, 0.098)</td>
</tr>
<tr>
<td>2$x^2$ 1</td>
<td>RIMSE</td>
<td>(0.688, 0.232)</td>
<td>(0.677, 0.226)</td>
</tr>
<tr>
<td></td>
<td>RIBIAS$^2$</td>
<td>(0.543, 0.167)</td>
<td>(0.527, 0.173)</td>
</tr>
<tr>
<td></td>
<td>RIVAR</td>
<td>(0.421, 0.159)</td>
<td>(0.425, 0.158)</td>
</tr>
<tr>
<td></td>
<td>TFRE</td>
<td>(0.667, 0.346)</td>
<td>(1.052, 0.331)</td>
</tr>
<tr>
<td>2 $x^2$</td>
<td>RIMSE</td>
<td>(0.417, 0.135)</td>
<td>(0.404, 0.136)</td>
</tr>
<tr>
<td></td>
<td>RIBIAS$^2$</td>
<td>(0.132, 0.021)</td>
<td>(0.076, 0.017)</td>
</tr>
<tr>
<td></td>
<td>RIVAR</td>
<td>(0.397, 0.134)</td>
<td>(0.400, 0.135)</td>
</tr>
<tr>
<td></td>
<td>TFRE</td>
<td>(2.000, 0.341)</td>
<td>(1.999, 0.947)</td>
</tr>
<tr>
<td>2$x^2$ $x^2$</td>
<td>RIMSE</td>
<td>(0.452, 0.129)</td>
<td>(0.400, 0.120)</td>
</tr>
<tr>
<td></td>
<td>RIBIAS$^2$</td>
<td>(0.351, 0.080)</td>
<td>(0.280, 0.061)</td>
</tr>
<tr>
<td></td>
<td>RIVAR</td>
<td>(0.285, 0.102)</td>
<td>(0.286, 0.103)</td>
</tr>
<tr>
<td></td>
<td>TFRE</td>
<td>(0.665, 0.240)</td>
<td>(0.667, 0.944)</td>
</tr>
</tbody>
</table>

Notes: RIMSE, RIBIAS$^2$, and RIVAR are “Root of the Integrated Mean Squared Error,” “Root of the Integrated Squared Bias,” and “Root of the Integrated Variance,” respectively. “Exp(b)” is the exponential pdf: $f(x) = \frac{1}{b} e^{-\frac{x}{b}}$ for $x \geq 0$. TFRE is the “True Fixed or Random [effects] Estimator” of Greene (2005a,b) with correct functional and distributional specifications.
Table 3: Summary Statistics for US Bank Data, 2004-2005

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>Total Cost</td>
<td>7,298</td>
<td>16,125</td>
<td>6,559</td>
</tr>
<tr>
<td>Wage Bill</td>
<td>49.89</td>
<td>10.79</td>
<td>48.96</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>0.0142</td>
<td>0.0064</td>
<td>0.0120</td>
</tr>
<tr>
<td>Capital Price</td>
<td>0.2785</td>
<td>0.2326</td>
<td>0.2837</td>
</tr>
<tr>
<td>Consumer Loans</td>
<td>9,376</td>
<td>25,350</td>
<td>9,646</td>
</tr>
<tr>
<td>Commercial Loans</td>
<td>130,735</td>
<td>264,113</td>
<td>126,691</td>
</tr>
<tr>
<td>Securities</td>
<td>78,302</td>
<td>175,853</td>
<td>77,346</td>
</tr>
<tr>
<td>(n)</td>
<td>10,600</td>
<td>5,300</td>
<td>5,300</td>
</tr>
</tbody>
</table>

Notes: Total Cost, Wage Bill and the three output quantities (Consumer and Commercial Loans and Securities) are in thousands of dollars. Interest Rate and Capital Price are real values.

Figure A. Common Support of Selected Covariates
Table 4: Bandwidths for Nonparametric Covariance Estimators

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(Wage Bill)</td>
<td>(0.206, 0.208)</td>
<td>0.059</td>
<td>0.074</td>
<td>0.074</td>
<td>0.074</td>
<td>0.074</td>
<td>0.075</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Commercial Loans)</td>
<td>(1.220, 1.249)</td>
<td>1.771</td>
<td>0.531</td>
<td>0.453</td>
<td>0.443</td>
<td>0.453</td>
<td>0.453</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Wage Bill)</td>
<td>(0.206, 0.208)</td>
<td>0.089</td>
<td>0.295</td>
<td>0.302</td>
<td>0.074</td>
<td>0.075</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Securities)</td>
<td>(1.098, 1.119)</td>
<td>0.319</td>
<td>0.398</td>
<td>1.625</td>
<td>0.398</td>
<td>0.406</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Interest Rate)</td>
<td>(0.605, 0.440)</td>
<td>0.220</td>
<td>0.264</td>
<td>0.639</td>
<td>0.220</td>
<td>0.160</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Commercial Loans)</td>
<td>(1.220, 1.249)</td>
<td>1.771</td>
<td>1.771</td>
<td>0.453</td>
<td>0.443</td>
<td>0.453</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Interest Rate)</td>
<td>(0.605, 0.440)</td>
<td>0.220</td>
<td>0.264</td>
<td>0.639</td>
<td>0.220</td>
<td>0.160</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Securities)</td>
<td>(1.098, 1.119)</td>
<td>0.239</td>
<td>1.593</td>
<td>0.812</td>
<td>0.398</td>
<td>0.406</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: $h_{A1}, h_{A2}$ are bandwidths for covariance $A(x)$ for period $t = 2004$ and $t = 2005$, respectively. $h_{B1}$ and $h_{B2}$ are similarly defined for the covariance $B(x)$ in each year. $C(x)$ is negligible by Lemma 2. The standard deviations in each year (2004, 2005) are calculated after the log transformation.

Table 5: US Bank ln(Total Cost) Stochastic Frontier Models - $T=2$

<table>
<thead>
<tr>
<th></th>
<th>Fixed Effects</th>
<th>Random Effects</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Half-normal</td>
<td>Exponential</td>
<td>Half-normal</td>
<td>Exponential</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Wage Bill)</td>
<td>0.3637***</td>
<td>0.3279</td>
<td>0.3097***</td>
<td>0.2925***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Interest Rate)</td>
<td>0.2763***</td>
<td>0.3013</td>
<td>0.2621***</td>
<td>0.2669***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Capital Price)</td>
<td>0.0167***</td>
<td>0.0151</td>
<td>0.0150***</td>
<td>0.0039</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Consumer Loans)</td>
<td>0.0521***</td>
<td>0.0051</td>
<td>0.0941***</td>
<td>0.0745***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Commercial Loans)</td>
<td>0.4896***</td>
<td>0.4853</td>
<td>0.5708***</td>
<td>0.5912***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Securities)</td>
<td>0.2051***</td>
<td>0.1793</td>
<td>0.2891***</td>
<td>0.2781***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-</td>
<td>-</td>
<td>-2.0251***</td>
<td>-1.8655***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.0695***</td>
<td>0.0330</td>
<td>0.1329***</td>
<td>0.0640***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>$3.67e^{-8}$</td>
<td>$1.65e^{-6}$</td>
<td>$0.0205***$</td>
<td>$0.0340***$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-Likelihood</td>
<td>20,569.2</td>
<td>25,573.1</td>
<td>5,563.5</td>
<td>6,308.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>5,300</td>
<td>5,300</td>
<td>5,300</td>
<td>5,300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: * = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$. Half-normal and Exponential are the distributions assumed for cost inefficiency. FE-Exponential model did not converge.
Figure 1: $\sigma_u^2$ as a Function of the Wage Bill and Commercial Loans.

Figure 2: $\sigma_u^2$ as a Function of the Wage Bill and Securities.
Figure 3: $\sigma_u^2$ as a Function of the Interest Rate and Commercial Loans

Figure 4: $\sigma_u^2$ as a Function of the Interest Rate and Securities.
Figure 5: $\sigma^2_u$ as a Function of the Wage Bill at the Median of Commercial Loans

![Graph of $\sigma^2_u$ as a Function of the Wage Bill at the Median of Commercial Loans]

Figure 6: $\sigma^2_u$ as a Function of Commercial Loans at the Median of the Wage Bill

![Graph of $\sigma^2_u$ as a Function of Commercial Loans at the Median of the Wage Bill]
Figure 7: $\sigma_u^2$ as a Function of the Wage Bill at the Median of securities

Figure 8: $\sigma_u^2$ as a Function of Securities at the Median of Wage Bill
Figure 9: $\sigma_u^2$ as a Function of the Interest Rate at the Median of Commercial Loans

Figure 10: $\sigma_u^2$ as a Function of Commercial Loans at the Median of the Interest Rate
Figure 11: $\sigma_u^2$ as a Function of the Interest Rate at the Median of Securities

Figure 12: $\sigma_u^2$ as a Function of Securities at the Median of the Interest Rate
Figure 13: $\xi_{\mu x}$ as a Function of the Wage Bill at the Median of Securities

Figure 14: $\xi_{\mu x}$ as a Function of Interest Rate at the Median of Securities