Asymptotic properties of a spatial autoregressive stochastic frontier model

Fei Jin*[1] and Lung-fei Lee[2]

[1]School of Economics, Fudan University, and Shanghai Institute of International Finance and Economics, Shanghai 200433, China

[2]Department of Economics, The Ohio State University, Columbus, OH 43210, USA

May 11, 2020

Abstract

This paper considers asymptotic properties of a spatial autoregressive stochastic frontier model. Relying on the asymptotic theory for nonlinear spatial NED processes, we prove the consistency and asymptotic distribution of the maximum likelihood estimator under regularity conditions. When inefficiency exists, all parameter estimators have the $\sqrt{n}$-rate of convergence and are asymptotically normal. However, when there is no inefficiency, only some parameter estimators have the $\sqrt{n}$-rate of convergence, and the rest have slower convergence rates. We also investigate a corrected two stage least squares estimator that is computationally simple, and derive the asymptotic distributions of the score and likelihood ratio test statistics that test for the existence of inefficiency.

Keywords: Stochastic frontier, spatial autoregression, maximum likelihood, asymptotic property, test

JEL classification: C12, C13, C21, C51, R32

1 Introduction

In spatial econometrics, there are several popular modeling strategies to take into account cross sectional dependence: in a spatial autoregressive (SAR) or spatial lag model (Cliff and Ord, 1973, 1981), the outcome of a spatial unit is specified as a weighted sum of neighbors' outcomes, i.e., a spatial lag of the dependent variable; in a spatial error model, the SAR process is specified on the error terms; in a spatial Durbin model, weighted sums of neighbors' characteristics are included

*Corresponding author. E-mail addresses: jin.fei@live.com (F. Jin), lee.1777@osu.edu (L.-F. Lee).

as explanatory variables. We may also consider spatial dependence in the dependent variable, exogenous variables and/or error terms simultaneously, e.g., a SAR model with SAR disturbances. The SAR model captures global spillovers which can have a structural economic interpretation, the spatial Durbin model captures in addition local spillovers, and the spatial error model reflects spillovers in unobserved variables. Ignoring spatial dependence can lead to inconsistent estimation and/or incorrect inference. This is also the case for stochastic frontier (SF) models. This paper considers large sample properties of a SAR SF (SARSF) model which contains a spatial lag of the dependent variable and a half normal inefficiency term.

Our research in this paper is motivated by some existing papers in the literature on SF models with spatial dependence. Druska and Horrace (2004) consider an SF model for panel data with fixed effects and spatial error dependence, and calculate efficiency using fixed effects. Glass et al. (2013, 2014) use a similar strategy for a SARSF panel data model with fixed effects. Papers on SF models with spatial dependence in error terms include, among others, Schmidt et al. (2009), Pavlyuk (2011), Areal et al. (2012), Fusco and Vidoli (2013), Tsionas and Michaelides (2016), Vidolia et al. (2016) and Carvalho (2018).[1] Brehm (2013) and Adetutu et al. (2015) consider SF models with local spatial dependence. Pavlyuk (2013) and Glass et al. (2016) study SARSF models.

We notice that the above papers have not considered large sample properties of SF models with spatial dependence. The asymptotic theory for such models is of interest as they are nonlinear SAR models, which cannot be analyzed by laws of large numbers (LLN) and central limit theorems (CLT) designed for linear processes. But they might be studied by recently developed asymptotic theories on nonlinear spatial models. For the consistency of the maximum likelihood estimator (MLE) of our SARSF model with a half normal inefficiency term, due to the composite error term in the model, the usual LLN for linear-quadratic forms of disturbances (Kelejian and Prucha, 2001) for a linear SAR model would not be applicable.

We provide a first rigorous analysis on asymptotic properties of this SARSF model in this paper. For nonlinear spatial econometrics, Jenish and Prucha (2012) introduce the near-epoch dependence (NED) concept of spatial processes and develop a useful LLN. We use their LLN to prove the consistency of the MLE under regularity conditions. For the general case with technical inefficiency, the asymptotic distribution of the MLE can be derived as usual by expanding the first order condition and applying an NED CLT. However, there might be a specific case that there is no inefficiency (but unknown). For such an irregular case, the asymptotic distribution might be different. For the half-normal SF model with no spatial dependence, there is an irregular feature that the information matrix is singular when there is no inefficiency (Lee, 1993). This is also the case for the SARSF model. We shall show that the presence of spatial dependence will generally not create extra irregularity. We derive the asymptotic distribution of the MLE in both the cases

[1] Some papers consider SF models with cross sectional dependence in error terms using a factor-based approach, e.g., Mastromarco et al. (2013, 2016).

with and without inefficiency. If inefficiency exists, the information matrix is nonsingular and all parameter estimators have the $\sqrt{n}$-rate of convergence and asymptotic normal distribution. But if there is no inefficiency, only some parameter estimators have the $\sqrt{n}$-rate of convergence, and the rest of parameters can have slower rates of convergence. The asymptotic distribution of the MLE in the irregular case with no inefficiency is derived by reparameterizing the model into one with a nonsingular information matrix, and the analysis essentially relies on higher order Taylor expansions of the original log likelihood function (Lee, 1993; Rotnitzky et al., 2000).

We also investigate the score and likelihood ratio (LR) tests that test for the existence of inefficiency. All the analysis takes into account spatial correlation of observed dependent variables. These tests are useful since the asymptotic distribution of the MLE depends on whether inefficiency exists or not. Because the inefficiency parameter is nonnegative, the score test is left sided and its test statistic is asymptotically normal, similar to the SF model with no spatial dependence (Lee and Chesher, 1986). But the asymptotic distribution of the LR test statistic is a mixture of a chi-square distribution with one degree of freedom and a degenerate distribution with a unit mass at 0, in accordance with the result in Lee (1993).

It is possible to consider other distributions of the inefficiency term, e.g., the exponential distribution (Meeusen and van Den Broeck, 1977), the truncated normal distribution (Stevenson, 1980) or the Gamma distribution (Greene, 1990), but the half-normal distribution in Aigner et al. (1977) is arguably most popular in empirical applications. We may also consider spatial Durbin terms and/or spatial error dependence. Spatial Durbin terms and spatial lags or spatial moving averages of disturbances are linear spatial dependence processes, so the analysis would be similar. Kumbhakar et al. (2013) consider a subgroup approach that can allow for a mixture of both fully efficient and inefficient firms, which is useful for empirical research. Large sample properties of models with alternative specifications are of interest in future research.

The rest of this paper is organized as follows. Section 2 studies large sample properties of the MLE. A computationally simple corrected two stage least squares estimator is also investigated. The score and likelihood ratio tests for frontier functions to be possibly efficient are proposed. Section 3 reports Monte Carlo results for the estimators and test statistics. Section 4 concludes. Proofs of propositions are collected in an appendix.

2 MLE

Consider the following SARSF model:

$$y_{ni} = \lambda_0 w_{n,i\cdot} Y_n + x'_{ni}\beta_0 + \epsilon_{ni}, \quad \epsilon_{ni} = v_{ni} - u_{ni}, \quad i = 1, \ldots, n, \tag{2.1}$$

where y_{ni} is the logged value of a dependent variable for the ith unit, $w_{n,i\cdot}$ is the ith row of an $n \times n$ spatial weights matrix $W_n = [w_{n,ij}]$, $Y_n = [y_{n1}, \ldots, y_{nn}]'$, $x_{ni} = [x_{ni,1}, \ldots, x_{ni,k_x}]'$ is a $k_x \times 1$

vector of exogenous variables in logarithm, λ_0 is a scalar spatial dependence parameter, β_0 is a $k_x \times 1$ parameter vector, v_{ni} follows the normal distribution $N(0, \sigma_{v0}^2)$, u_{ni} follows the nonnegative half normal distribution $|N(0, \sigma_{u0}^2)|$, u_{ni} and v_{ni} are independent, and $[u_{ni}, v_{ni}]$'s are i.i.d. for all i. The x_{ni} typically includes an intercept term, so we let $x_{ni} = [1, x_{2ni}']'$ and $\beta_0 = [\beta_{10}, \beta_{20}']'$.

With a nonnegative inefficiency term u_{ni}, model (2.1) can be for production, revenue, profit frontiers and so on. For cost distance frontiers, $v_{ni} - u_{ni}$ can be replaced by $v_{ni} + u_{ni}$ to capture cost inefficiency, but the analysis is similar. Such a model has been introduced in the empirical literature of frontier functions, e.g., Glass et al. (2016), where the maximum likelihood estimation is described and various efficiency measures such as direct, indirect and total relative efficiencies are proposed. Model (2.1) can be extended for a panel data set by introducing a subscript t, as in Glass et al. (2016). Without loss of generality, we consider model (2.1) for cross sectional data. This model is similar to the SAR model except for the composite error term ϵ_{ni} with a nonzero mean. However, due to the half normal distribution of u_{ni}, estimates of the model parameters would in general not depend on linear and quadratic moments of independent disturbances. So asymptotic analysis and results for the linear SAR model would not be applicable. One has to sort for nonlinear spatial asymptotic theories for estimation and testing.[2]

Let $[\lambda, \beta', \sigma_u^2, \sigma_v^2]$ be an arbitrary parameter vector and the corresponding true parameter vector be $[\lambda_0, \beta_0', \sigma_{u0}^2, \sigma_{v0}^2]$. Denote $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\delta = \sigma_u / \sigma_v$, and $\theta = [\lambda, \beta', \sigma^2, \delta]'$. The log likelihood function of θ for model (2.1) is

$$
\begin{aligned}
\ln L_n(\theta) = n \ln 2 - \frac{n}{2} \ln(2\pi\sigma^2) + \ln |I_n - \lambda W_n| - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_{ni} - \lambda w_{n,i\cdot} Y_n - x_{ni}'\beta)^2 \\
+ \sum_{i=1}^{n} \ln \Phi\big(-\frac{\delta}{\sigma}(y_{ni} - \lambda w_{n,i\cdot} Y_n - x_{ni}'\beta)\big),
\end{aligned}
\tag{2.2}
$$

where $\Phi(\cdot)$ is the distribution function of a standard normal random variable, whose presence is due to the stochastic frontier disturbance u_{ni}. The log likelihood function $\ln L_n(\theta)$ involves the log determinant $|I_n - \lambda W_n|$, which can be computed once eigenvalues of W_n are computed (Ord, 1975), or by a Taylor series approximation of the log determinant as suggested in LeSage and Pace (2009), even when the sample size is large. Note that $\Phi\big(-\frac{\delta}{\sigma}(y_{ni} - \lambda w_{n,i\cdot} Y_n - x_{ni}'\beta)\big)$ is a nonlinear function of $y_{n1}, \ldots, y_{nn}$, so the LLN for linear-quadratic forms of disturbances in spatial econometrics, originated in Kelejian and Prucha (2001), is not applicable. However, we investigate the asymptotic theory in Jenish and Prucha (2012) for near-epoch dependent (NED) random fields, which are generalized from the time series literature.[3] The spatial NED property is preserved under certain transformations such as summation, multiplication, and Lipschitz transformation and some of its generalizations. We note that some terms in (2.2) are similar to those in the log likelihood

[2] In the existing frontier function literature with interactions, there are several papers on model specification and empirical estimation but there are no rigorous asymptotic studies.

[3] The definition of an NED random field is given in Appendix B.

function of a SAR Tobit model in Xu and Lee (2015), so some of their analysis on NED properties of relevant terms can be adapted to investigate large sample properties of model (2.1).

The following assumptions are maintained for model (2.1).

Assumption 1. *Individual units in the economy are located or living in a region* $D_n \subset D \subset \mathbb{R}^d$*, where* D *is a (possibly) unevenly spaced lattice, the cardinality* $|D_n|$ *of a finite set* D_n *satisfies* $\lim_{n\to\infty} |D_n| = \infty$*. The distance* $d(i, j)$ *between any two different individuals* i *and* j *is larger than or equal to a positive constant, which will be assumed to be* 1 *for convenience.*

Assumption 2. $c_1 \equiv \lambda_m \sup_n \|W_n\|_\infty < 1$*, and* $[-\lambda_m, \lambda_m]$ *is the compact parameter space of* λ *on the real line.*

Assumption 3. *In addition to the diagonal elements of* W_n*,* $w_{n,ii} = 0$ *for all* i*, the elements of* W_n *satisfy at least one of the following two conditions:*

(a) Only individuals whose distances are less than or equal to some positive constant d_0 *may affect each other directly, i.e.,* $w_{n,ij} \neq 0$ *only if* $d(i, j) \leq d_0$*.*

(b) (i) For every n*, the number of columns* $w_{n,\cdot j}$ *of* W_n *with* $|\lambda_0| \sum_{i=1}^n |w_{n,ij}| > c_1$ *is less than or equal to some fixed nonnegative integer that does not depend on* n*;*[4] *(ii) there exists an* $\alpha > d$ *and a constant* c_2 *such that* $|w_{n,ij}| \leq c_2/d(i, j)^\alpha$*.*

Assumption 4. *(a)* $v_{ni} \sim N(0, \sigma_{v0}^2)$ *and* $u_{ni} \sim |N(0, \sigma_{u0}^2)|$ *a half normal random variable; (b)* x_{ni}*,* v_{ni} *and* u_{ni} *are mutually independent; (c)* $[v_{ni}, u_{ni}]$*'s are i.i.d.*

Assumption 5. *(a)* $\sup_{1\leq k\leq k_x,i,n} \mathrm{E}[|x_{ni,k}|^{4+\iota}] < \infty$ *for some* $\iota > 0$*; (b)* $\{x_{ni}\}_{i=1}^n$ *is an* α*-mixing random field with* α*-mixing coefficient* $\alpha(u, v, s) \leq (u + v)^{c_3}\hat{\alpha}(s)$ *for some* $c_3 \geq 0$*, where* $\hat{\alpha}(s)$ *satisfies* $\sum_{s=1}^\infty s^{d-1}\hat{\alpha}(s) < \infty$*.*

Assumption 6. $\limsup_{n\to\infty} \frac{1}{n}[\mathrm{E} \ln L_n(\theta) - \mathrm{E} \ln L_n(\theta_0)] < 0$ *for any* $\theta \neq \theta_0$*.*

Assumption 7. *The parameter space of* $[\beta', \sigma^2, \delta]'$ *is a compact subset of* $\mathbb{R}^{k_x+2}$ *and* $\delta \geq 0$*.*

Assumption 1 is introduced by Jenish and Prucha (2009, 2012) for spatial mixing and NED processes. As the distance between two units can be a geometrical distance or an economic distance or a mixture of both, the space D is allowed to be high dimensional as a subset of $\mathbb{R}^d$ and the distance can be induced from any norm in $\mathbb{R}^d$. The increasing domain asymptotics imposed in Assumption 1 are natural for a regional study, and are usually needed for regular asymptotic properties of estimators. Since the distance between any two different individuals is larger than or equal to some positive constant, the sample region must expand as the sample size increases. Another asymptotic method is the so-called infill asymptotics, where the growth of the sample

[4]The same c_1 as in Assumption 2 is used for simplicity. It can be any positive number smaller than 1.

size can be achieved by sampling points arbitrarily dense in a fixed sample region. Under infill asymptotics, even some popular estimators, such as the least squares and the method of moments estimators, may not be consistent (see, e.g., Lahiri, 1996). Assumptions 2–3 are from Xu and Lee (2015). Assumption 2 is used in Xu and Lee (2015) to establish the NED property of a term similar to $\Phi\big(-\frac{\delta}{\sigma}(y_{ni} - \lambda w_{n,i\cdot} Y_n - x'_{ni}\beta)\big)$ in (2.2), so we also impose it, although it is stronger than the condition that λ is in a compact subset of $(1/\mu_{\min}, 1/\mu_{\max})$, where $\mu_{\min}$ and $\mu_{\max}$ are, respectively, the smallest and largest real eigenvalues of the spatial weights matrix for a linear SAR model, as discussed in, e.g., LeSage and Pace (2009) and Kelejian and Prucha (2010). Assumption 2 also implies the existence of the reduced form of Y_n in (2.1) and the Neuman series expansion $(I_n - \lambda W_n)^{-1} = I_n + \lambda W_n + \lambda^2 W_n^2 + \dots$ for any λ in that parameter space. The compactness of the parameter space for λ in Assumption 2 and that for the rest of parameters in Assumption 7 are typically maintained for extremum estimators. Assumption 3 avoids self-influence, i.e., $w_{n,ii} = 0$ for all i, and also requires the interaction of units i and j in terms of $w_{n,ij}$ to decline fast enough. While Assumption 3(a) requires no direct interaction for any two units when they are far enough from each other, Assumption 3$(b)(ii)$ possibly allows all non-diagonal elements of W_n to be nonzero but their interactions decline geometrically fast. Assumption 3$(b)(i)$ is a condition on the column sums of W_n in absolute value, i.e., the total effects of each spatial unit on those who are connected to (or nominate) him/her. Only a fixed number of spatial units are allowed to have large aggregated effects on other units. In a network setting, the units with large aggregated effects on other units are referred to as stars. Assumptions 4 and 5 summarize the exogeneity of explanatory variables and distributional assumptions of disturbances. The conditions in Assumption 5 are needed for the NED properties of relevant terms. The mixing coefficient for the random field $\{x_{ni}\}_{i=1}^n$ in Assumption 5(b) does not only depend on the distance between two separate subsets of spatial units but also their sizes.[5] Assumption 6 is an identification condition for the model.[6] As a ratio of two standard deviations, δ is necessarily nonnegative, that is stated in Assumption 7. Under the above assumptions, pointwise and uniform LLNs can be applied to prove the uniform convergence of the sample average log likelihood function. With identification uniqueness of the true parameters and equicontinuity of the limiting expected log likelihood function in parameters, the MLE $\hat{\theta}$ will be consistent (White, 1994). The detailed proof is in Appendix B.

Proposition 2.1. *Under Assumptions 1–7, the MLE $\hat{\theta}$ of model (2.1) is consistent.*

We next investigate the asymptotic distribution of the MLE. Let $G_n(\lambda) = W_n(I_n - \lambda W_n)^{-1}$, $\epsilon_{ni}(\lambda, \beta) = y_{ni} - \lambda w_{n,i\cdot} Y_n - x'_{ni}\beta$, and $f(t) = \phi(t)/\Phi(t)$ be the inverse Mills ratio, where $\phi(t)$ is the density function of the standard normal distribution. The first order derivatives of the log

[5]See Jenish and Prucha (2012) for the detailed definition.
[6]Due to the nonlinearity of model (2.1), a primitive identification condition is not obvious.

likelihood function on its parameters are

$$\frac{\partial \ln L_n(\theta)}{\partial \lambda} = -\operatorname{tr}[G_n(\lambda)] + \frac{1}{\sigma^2}\sum_{i=1}^{n} w_{n,i\cdot} Y_n \epsilon_{ni}(\lambda,\beta) + \frac{\delta}{\sigma}\sum_{i=1}^{n} w_{n,i\cdot} Y_n f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{2.3}$$

$$\frac{\partial \ln L_n(\theta)}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_{ni}\epsilon_{ni}(\lambda,\beta) + \frac{\delta}{\sigma}\sum_{i=1}^{n} x_{ni} f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{2.4}$$

$$\frac{\partial \ln L_n(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} \epsilon_{ni}^2(\lambda,\beta) + \frac{\delta}{2\sigma^3}\sum_{i=1}^{n} f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big)\epsilon_{ni}(\lambda,\beta), \tag{2.5}$$

$$\frac{\partial \ln L_n(\theta)}{\partial \delta} = -\frac{1}{\sigma}\sum_{i=1}^{n} f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big)\epsilon_{ni}(\lambda,\beta). \tag{2.6}$$

These scores and the second order derivatives used to construct the information matrix will be regular when $\delta_0 > 0$. However, there are some irregularities for the case with δ_0 happened to be zero. Here we consider both situations. If $\delta_0 = 0$ but unknown to the investigator, then this constraint would not be imposed for estimation. For this case, we see that $\frac{\partial \ln L_n(\eta,0)}{\partial \delta} = -\sigma\sqrt{\frac{2}{\pi}}\frac{\partial \ln L_n(\eta,0)}{\partial \beta_1}$, where $\eta = [\lambda, \beta_1, \beta_2', \sigma^2]'$ and β_1 is the first component of $\beta = [\beta_1, \beta_2']'$, because $f(0) = \frac{\phi(0)}{\Phi(0)} = \sqrt{\frac{2}{\pi}}$. Thus, when the true value δ_0 is zero, the scores of model (2.1) are linearly dependent and the information matrix is singular, which is similar to the SF model with no spatial dependence. The SARSF model has an additional term $\frac{\partial \ln L_n(\eta,0)}{\partial \lambda}$, but it would not create additional linear dependence on other derivatives because $\frac{\partial \ln L_n(\eta,0)}{\partial \lambda} = \frac{1}{\sigma^2}[G_n(\lambda)X_n\beta]'\epsilon_n(\lambda,\beta) + \frac{1}{\sigma^2}\epsilon_n'(\lambda,\beta)G_n(\lambda)\epsilon_n(\lambda,\beta) - \operatorname{tr}[G_n(\lambda)]$ is linear-quadratic in $\epsilon_n(\lambda,\beta)$, where $\epsilon_n(\lambda,\beta) = [\epsilon_{n1}(\lambda,\beta),\ldots,\epsilon_{nn}(\lambda,\beta)]'$ and $X_n = [x_{n1},\ldots,x_{nn}]'$, so the additional score $\frac{\partial \ln L_n(\eta_0,0)}{\partial \lambda}$ due to the presence of spatial dependence will not be linearly dependent on other scores, which do not have a quadratic term. With $\delta_0 = 0$, however the asymptotic distribution of the MLE can be derived by reparameterizing the model into one with a nonsingular information matrix, as in the usual SF model without spatial interactions (Lee, 1993). When $\delta_0 \neq 0$, the scores are generally not linearly dependent, so the asymptotic distribution of the MLE can be derived as usual by a mean value theorem expansion. In the following, we shall first consider the regular case with $\delta_0 \neq 0$ and then the irregular one under $\delta_0 = 0$.

2.1 Asymptotic distribution under $\delta_0 \neq 0$

When $\delta_0 \neq 0$, the information matrix $\mathrm{E}(\frac{\partial \ln L_n(\theta_0)}{\partial \theta}\frac{\partial \ln L_n(\theta_0)}{\partial \theta'}) = -\mathrm{E}(\frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'})$ is generally nonsingular and we assume that it is so in the limit.

Assumption 8. $\lim_{n\to\infty}\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'})$ *is positive definite.*

The asymptotic distribution of $\hat{\theta}$ follows by a mean value theorem expansion of its first order condition. The following regularity conditions are needed in the analysis.

Assumption 9. *(a) When $\delta_0 > 0$, the true value of θ is in the interior of its parameter space. (b) For the case with $\delta_0 = 0$, the true values of all remaining parameters are in the interior of their parameter subspace.*

Assumption 10. *(a) $\sup_{1\le k\le k_x,i,n} \mathrm{E}[|x_{ni,k}|^6] < \infty$; (b) $\alpha > 5d$, where α is that one in Assumption 3(b)(ii); (c) for the α-mixing coefficients of $\{x_{ni}\}_{i=1}^n$ in Assumption 5, $\breve{\alpha}(s)$ satisfies $\sum_{s=1}^{\infty} s^{d[1+c_3\iota^*/(2+\iota^*)]-1}[\hat{\alpha}(s)]^{\iota^*/(4+2\iota^*)} < \infty$ for some $0 < \iota^* < 1$, , where the constant c_3 is the one in Assumption 5(b).*

Assumption 9 is a familiar condition required to derive asymptotic distributions of estimators. With $\delta_0 > 0$, all components of the true value θ_0 are in the interior of the parameter space. On the other hand, if $\delta_0 = 0$, the remaining true parameters are not subject to boundary constraints so they are in the interior of their parameter subspace. The moment condition in Assumption 10(a) is used to have the convergence of the second order derivatives of $\frac{1}{n}\ln L_n(\theta)$ at a consistent estimator of θ_0. In addition to $\sup_{1\le k\le k_x,i,n} \mathrm{E}[|x_{ni,k}|^6] < \infty$, Assumption 10($b$)–($c$) modify the declining rate α of $w_{n,ij}$ in Assumption 3 and that on $\breve{\alpha}(s)$ accordingly for the applicability of the CLT for an NED process in Jenish and Prucha (2012).

Proposition 2.2. *Under Assumptions 1–10 and $\delta_0 \neq 0$,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\Big(0, \lim_{n\to\infty}\big(-\frac{1}{n}\,\mathrm{E}\,\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\big)^{-1}\Big).$$

This proposition gives the asymptotic distribution of the MLE $\hat{\theta}$ of θ_0 when there is inefficiency in the stochastic frontier function. This is a regular situation of the model. It remains to consider the irregular case when the production function of each firm is efficient. For that case, the analysis is relatively complicated and will be presented in the next subsection.

2.2 The boundary case with $\delta_0 = 0$

When $\delta_0 = 0$, it is on the boundary of its parameter space and the scores are linearly dependent as mentioned above. Then the usual analysis on asymptotic distributions does not work, but we can provide an analysis based on reparameterizations.

Let $\beta_1^\dagger = \beta_1 - \delta\sigma\sqrt{\frac{2}{\pi}}$ be a reparameterization. Then the log likelihood function in terms of $[\lambda, \beta_1^\dagger, \beta_2', \sigma^2, \delta]'$ is

$$\ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta) \equiv \ln L_n\Big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\Big).$$

Denote $\eta^\dagger = [\lambda, \beta_1^\dagger, \beta_2', \sigma^2]'$. The derivatives of $\ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta)$ are

$$\frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta)}{\partial\lambda} = \frac{\partial \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial\lambda},$$

$$\frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta)}{\partial \beta_1^\dagger} = \frac{\partial \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \beta_1},$$

$$\frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta)}{\partial \beta_2} = \frac{\partial \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \beta_2},$$

$$\frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta)}{\partial \sigma^2} = \frac{\delta}{2\sigma}\sqrt{\frac{2}{\pi}}\frac{\partial \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \beta_1} + \frac{\partial \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \sigma^2},$$

$$\frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta)}{\partial \delta} = \sigma\sqrt{\frac{2}{\pi}}\frac{\partial \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \beta_1} + \frac{\partial \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \delta}.$$

At $\delta = 0$, because $\beta_1^\dagger = \beta_1$ and $\eta^\dagger = \eta$, we have $\frac{\partial \ln L_{2n}(\lambda,\beta_1,\beta_2,\sigma^2,0)}{\partial \eta^\dagger} = \frac{\partial \ln L_n(\eta,0)}{\partial \eta}$, but

$$\frac{\partial \ln L_{2n}(\lambda, \beta_1, \beta_2, \sigma^2, 0)}{\partial \delta} = 0 \tag{2.7}$$

identically for all possible values of η due to the linear dependence of $\frac{\partial \ln L_n(\lambda,\beta_1,\beta_2,\sigma^2,0)}{\partial \theta}$ in (2.3)–(2.6). Thus, as $\frac{\partial \ln L_{2n}(\eta_0,0)}{\partial \delta}$ is not the leading order term of a Taylor expansion in deriving the asymptotic distribution of the MLE that maximizes $\ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta)$, we need to investigate the second order derivative $\frac{\partial^2 \ln L_{2n}(\eta_0,0)}{\partial \delta^2}$. Since

$$\begin{aligned}\frac{\partial^2 \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^2, \delta)}{\partial \delta^2} = {}& \frac{2\sigma^2}{\pi}\frac{\partial^2 \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \beta_1^2} \\ & + 2\sigma\sqrt{\frac{2}{\pi}}\frac{\partial^2 \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \beta_1 \partial \delta} \\ & + \frac{\partial^2 \ln L_n\big(\lambda, \beta_1^\dagger + \delta\sigma\sqrt{\frac{2}{\pi}}, \beta_2, \sigma^2, \delta\big)}{\partial \delta^2},\end{aligned}$$

by the second order derivatives of $\ln L_n(\theta)$ in Appendix A, $\frac{\partial^2 \ln L_{2n}(\lambda,\beta_1,\beta_2,\sigma^2,0)}{\partial \delta^2} = -\frac{2}{\pi\sigma^2}\sum_{i=1}^n[\epsilon_{ni}^2(\lambda,\beta) - \sigma^2]$. Then

$$\frac{\partial^2 \ln L_{2n}(\lambda, \beta_1, \beta_2, \sigma^2, 0)}{\partial \delta^2} = -\frac{4\sigma^2}{\pi}\frac{\partial \ln L_{2n}(\lambda, \beta_1, \beta_2, \sigma^2, 0)}{\partial \sigma^2} \tag{2.8}$$

is linearly dependent on the score with respect to σ^2. Let $\sigma^{\dagger 2} = \sigma^2/(1 + \frac{2}{\pi}\delta^2)$ be another reparameterization, and

$$\begin{aligned}\ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \delta) &\equiv \ln L_{2n}\Big(\lambda, \beta_1^\dagger, \beta_2, \big(1 + \frac{2}{\pi}\delta^2\big)\sigma^{\dagger 2}, \delta\Big) \\ &= \ln L_n\Big(\lambda, \beta_1^\dagger + \delta\sigma^\dagger\big(\frac{2}{\pi} + \frac{4}{\pi^2}\delta^2\big)^{1/2}, \beta_2, \big(1 + \frac{2}{\pi}\delta^2\big)\sigma^{\dagger 2}, \delta\Big).\end{aligned}$$

Then,

$$\frac{\partial \ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \delta)}{\partial \lambda} = \frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \lambda}, \tag{2.9}$$

$$\frac{\partial \ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \delta)}{\partial \beta_1^\dagger} = \frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \beta_1^\dagger}, \tag{2.10}$$

$$\frac{\partial \ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \delta)}{\partial \beta_2} = \frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \beta_2}, \tag{2.11}$$

$$\frac{\partial \ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \delta)}{\partial \sigma^{\dagger 2}} = \big(1+\frac{2}{\pi}\delta^2\big)\frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \sigma^2}, \tag{2.12}$$

$$\begin{aligned}\frac{\partial \ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \delta)}{\partial \delta} &= \frac{4\delta\sigma^{\dagger 2}}{\pi}\frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \sigma^2} \\ &\quad + \frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \delta};\end{aligned} \tag{2.13}$$

and hence,

$$\begin{aligned}&\frac{\partial^2 \ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \delta)}{\partial \delta^2} \\ &= \frac{4\sigma^{\dagger 2}}{\pi}\frac{\partial \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \sigma^2} + \frac{16\delta^2\sigma^{\dagger 4}}{\pi^2}\frac{\partial^2 \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \sigma^4} \\ &\quad + \frac{8\delta\sigma^{\dagger 2}}{\pi}\frac{\partial^2 \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \sigma^2 \partial \delta} + \frac{\partial^2 \ln L_{2n}(\lambda, \beta_1^\dagger, \beta_2, (1+\frac{2}{\pi}\delta^2)\sigma^{\dagger 2}, \delta)}{\partial \delta^2}.\end{aligned} \tag{2.14}$$

It follows from (2.8) and (2.14) that

$$\frac{\partial^2 \ln L_{3n}(\lambda, \beta_1, \beta_2, \sigma^2, 0)}{\partial \delta^2} = 0.$$

Also from (2.7)–(2.12) at $\delta_0 = 0$,

$$\frac{\partial \ln L_{3n}(\lambda, \beta_1, \beta_2, \sigma^2, 0)}{\partial \eta^\ddagger} = \frac{\partial \ln L_n(\eta, 0)}{\partial \eta},$$

where $\eta^\ddagger = [\lambda, \beta_1^\dagger, \beta_2', \sigma^{\dagger 2}]'$. Furthermore, by (2.7) and (2.13),

$$\frac{\partial \ln L_{3n}(\lambda, \beta_1, \beta_2, \sigma^2, 0)}{\partial \delta} = 0.$$

Thus, neither $\frac{\partial \ln L_{3n}(\eta_0,0)}{\partial \delta}$ nor $\frac{\partial^2 \ln L_{3n}(\eta_0,0)}{\partial \delta^2}$ is the leading order term of a Taylor expansion in deriving the asymptotic distribution of the MLE that maximizes $\ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \delta)$, and the third order derivative $\frac{\partial^3 \ln L_{3n}(\eta_0,0)}{\partial \delta^3}$ need be examined.[7] It follows that, by one more reparameterization, the model can be transformed to be one with a nonsingular information matrix so that the asymptotic

[7] See also Rotnitzky et al. (2000) for such an analysis on models with i.i.d. data.

distribution of the MLE for the reparameterized coefficients can be derived as usual (Lee, 1993). Let $\tau = \delta^3$ and

$$\ln L_{4n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \tau) \equiv \ln L_{3n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \tau^{1/3}) \\ = \ln L_n\Big(\lambda, \beta_1^\dagger + \tau^{1/3}\sigma^\dagger\big(\frac{2}{\pi} + \frac{4}{\pi^2}\tau^{2/3}\big)^{1/2}, \beta_2, \big(1 + \frac{2}{\pi}\tau^{2/3}\big)\sigma^{\dagger 2}, \tau^{1/3}\Big). \tag{2.15}$$

Then by Proposition 3 in Lee (1993), $\frac{\partial \ln L_{4n}(\lambda,\beta_1,\beta_2,\sigma^2,0)}{\partial \tau} = \frac{1}{6}\frac{\partial^3 \ln L_{3n}(\lambda,\beta_1,\beta_2,\sigma^2,0)}{\partial \delta^3}$. It follows that by some calculation,

$$\frac{\partial \ln L_{4n}(\lambda, \beta_1, \beta_2, \sigma^2, 0)}{\partial \tau} = \frac{1}{6\sigma^3}\big(1 - \frac{4}{\pi}\big)\sqrt{\frac{2}{\pi}}\sum_{i=1}^n \epsilon_{ni}^3(\lambda, \beta) + \frac{2}{\pi\sigma}\sqrt{\frac{2}{\pi}}\sum_{i=1}^n \epsilon_{ni}(\lambda, \beta). \tag{2.16}$$

In addition,

$$\frac{\partial \ln L_{4n}(\lambda_0, \beta_{10}, \beta_{20}, \sigma_0^2, 0)}{\partial \eta^\ddagger} = \begin{pmatrix} \frac{1}{\sigma_0^2}\epsilon_n' G_n \epsilon_n - \operatorname{tr}(G_n) + \frac{1}{\sigma_0^2}(G_n X_n \beta_0)'\epsilon_n \\ \frac{1}{\sigma_0^2} X_n' \epsilon_n \\ \frac{1}{2\sigma_0^4}(\epsilon_n'\epsilon_n - n\sigma_0^2) \end{pmatrix},$$

where $G_n = G_n(\lambda_0)$ and $\epsilon_n = [\epsilon_{n1}, \ldots, \epsilon_{nn}]'$. For the log likelihood function $\ln L_{4n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \tau)$, the information matrix is

$$\Delta_n = \begin{pmatrix} \frac{1}{\sigma_0^2}\operatorname{E}[(G_n X_n \beta_0)'(G_n X_n \beta_0)] + \operatorname{tr}(G_n G_n^{(s)}) & * & * & * \\ \frac{1}{\sigma_0^2}\operatorname{E}(X_n' G_n X_n \beta_0) & \frac{1}{\sigma_0^2}\operatorname{E}(X_n' X_n) & * & * \\ \frac{1}{\sigma_0^2}\operatorname{tr}(G_n) & 0 & \frac{n}{2\sigma_0^4} & * \\ \frac{1}{\sqrt{2\pi\sigma_0^2}}\operatorname{E}(l_n' G_n X_n \beta_0) & \frac{1}{\sqrt{2\pi\sigma_0^2}}\operatorname{E}(l_n' X_n) & 0 & \frac{n}{6\pi}(5 - \frac{16}{\pi} + \frac{32}{\pi^2}) \end{pmatrix}, \tag{2.17}$$

where $A^{(s)} = A + A'$ for any square matrix A, and l_n is an $n \times 1$ vector of ones. Under the following assumption, $\frac{1}{n}\Delta_n$ is positive definite for a large enough n.

Assumption 11. *Either (a)* $\lim_{n\to\infty} \frac{1}{n}\operatorname{E}[(G_n X_n \beta_0, X_n)' T_n (G_n X_n \beta_0, X_n)]$ *is positive definite, where* $T_n = I_n - \frac{3}{n(5-\frac{16}{\pi}+\frac{32}{\pi^2})} l_n l_n'$*, or (b)* $\lim_{n\to\infty} \frac{1}{n}\operatorname{E}(X_n' T_n X_n)$ *is positive definite and* $\lim_{n\to\infty}[\frac{1}{n}\operatorname{tr}(G_n^{(s)} G_n^{(s)}) - \frac{1}{n^2}\operatorname{tr}^2(G_n^{(s)})] > 0$.

The above assumption is similar to one for the SAR model in Lee (2004), except for the presence of the matrix T_n, which is due to the inclusion of u_{ni} in the SARSF model. Note that $T_n = (I_n - \frac{1}{n} l_n l_n') + \frac{2-16/\pi+32/\pi^2}{n(5-16/\pi+32/\pi^2)} l_n l_n'$ is positive definite because $2 - 16/\pi + 32/\pi^2$ is positive. In addition, $\frac{1}{n}\operatorname{tr}(G_n^{(s)} G_n^{(s)}) \geq \frac{1}{n^2}\operatorname{tr}^2(G_n^{(s)})$ by the Cauchy-Schwarz inequality.

As $\delta \geq 0$, $\tau \geq 0$. The MLE $[\hat\lambda, \hat\beta_1^\dagger, \hat\beta_2, \hat\sigma^{\dagger 2}, \hat\tau]$ of $[\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \tau]$ maximizes $\ln L_{4n}(\lambda, \beta_1^\dagger, \beta_2, \sigma^{\dagger 2}, \tau)$ on the transformed parameter space with $\tau \geq 0$. It is possible that the MLE occurs at the boundary with $\hat\tau = 0$. Let $\breve\eta$ be the MLE of η_0 for the SAR model, i.e., model (2.1) with $\epsilon_{ni} = v_{ni}$. Then the

MLE $[\hat{\eta}^{\ddagger\prime}, \hat{\tau}]$ is equal to $[\check{\eta}', 0]$ if and only if $\frac{\partial \ln L_{4n}(\check{\eta},0)}{\partial \tau} \leq 0$ as in Waldman (1982). As $\frac{\partial \ln L_{4n}(\check{\eta},0)}{\partial \tau} = \frac{1}{6}\frac{\partial^3 \ln L_{3n}(\check{\eta},0)}{\partial \delta^3}$, $\frac{\partial \ln L_{4n}(\check{\eta},0)}{\partial \tau} \leq 0$ if and only if $\sum_{i=1}^n \check{\epsilon}_{ni}^3 \geq 0$ by (2.16), where $\check{\epsilon}_{ni} = y_{ni} - \check{\lambda} w_{n,i\cdot} Y_n - x_{ni}'\check{\beta}$.[8] The $\check{\eta}$ satisfies $\frac{\partial \ln L_{4n}(\check{\eta},0)}{\partial \eta^\ddagger} = 0$. Then under regularity conditions, by the CLT for NED processes in Jenish and Prucha (2012),

$$\sqrt{n}(\check{\eta} - \eta_0) = \big(\frac{1}{n}\Delta_{n,11}\big)^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln L_{4n}(\lambda_0, \beta_{10}, \beta_{20}, \sigma_0^2, 0)}{\partial \eta^\ddagger} + o_p(1) \xrightarrow{d} N\Big(0, \lim_{n\to\infty}\big(\frac{1}{n}\Delta_{n,11}\big)^{-1}\Big), \quad (2.18)$$

where

$$\Delta_{n,11} = \begin{pmatrix} \frac{1}{\sigma_0^2}\,\mathrm{E}[(G_nX_n\beta_0)'(G_nX_n\beta_0)] + \mathrm{tr}(G_nG_n^{(s)}) & * & * \\ \frac{1}{\sigma_0^2}\,\mathrm{E}(X_n'G_nX_n\beta_0) & \frac{1}{\sigma_0^2}\,\mathrm{E}(X_n'X_n) & * \\ \frac{1}{\sigma_0^2}\,\mathrm{tr}(G_n) & 0 & \frac{n}{2\sigma_0^4} \end{pmatrix}.$$

Let $J = [J_1, J_2, J_3', J_4]'$ be the multivariate normal vector $N(0, \lim_{n\to\infty}(\frac{1}{n}\Delta_{n,11})^{-1})$, where J_1, J_2 and J_4 are univariate normal random variables. Under regularity conditions, by a Taylor expansion and (2.18),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \check{\epsilon}_{ni}^3 = \frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_{ni}^3 - \frac{3\sigma_0^2}{n}[\mathrm{E}(l_n'G_nX_n\beta_0, l_n'X_n, 0)]\sqrt{n}(\check{\eta} - \eta_0) + o_p(1) = \Gamma_n + o_p(1), \quad (2.19)$$

where $\Gamma_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_{ni}^3 - \frac{3\sigma_0^2}{n}[\mathrm{E}(l_n'G_nX_n\beta_0, l_n'X_n, 0)]\big(\frac{1}{n}\Delta_{n,11}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\lambda_0,\beta_{10},\beta_{20},\sigma_0^2,0)}{\partial \eta^\ddagger}$. Since

$$\mathrm{E}\Big[\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_{ni}^3\Big)\Big(\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\lambda_0, \beta_{10}, \beta_{20}, \sigma_0^2, 0)}{\partial \eta^{\ddagger\prime}}\Big)\Big] = \frac{3\sigma_0^2}{n}\,\mathrm{E}[l_n'G_nX_n\beta_0, l_n'X_n, 0],$$

Γ_n is uncorrelated with the leading order term $\big(\frac{1}{n}\Delta_{n,11}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\lambda_0,\beta_{10},\beta_{20},\sigma_0^2,0)}{\partial \eta^\ddagger}$ of $\sqrt{n}(\check{\eta} - \eta_0)$. Then $\frac{1}{\sqrt{n}}\sum_{i=1}^n \check{\epsilon}_{ni}^3$ is asymptotically uncorrelated with $\sqrt{n}(\check{\eta} - \eta_0)$. By the CLT in Jenish and Prucha (2012), $[\sqrt{n}(\check{\eta} - \eta_0)', \frac{1}{\sqrt{n}}\sum_{i=1}^n \check{\epsilon}_{ni}^3]'$ converges in distribution to the normal vector $[J', K]'$, where $K = N(0, 6\sigma_0^6)$ is independent of J; therefore, the event $\frac{\partial \ln L_{4n}(\check{\eta},0)}{\partial \tau} \leq 0$ is asymptotically independent of J.

When $\hat{\tau} > 0$, the MLE $[\hat{\eta}^{\ddagger\prime}, \hat{\tau}]'$ satisfies the first order conditions $\frac{\partial \ln L_{4n}(\hat{\eta}^\ddagger,\hat{\tau})}{\partial \eta^\ddagger} = 0$ and $\frac{\partial \ln L_{4n}(\hat{\eta}^\ddagger,\hat{\tau})}{\partial \tau} = 0$. Let $F = [F_1, F_2, F_3', F_4, F_5]'$ be the normal vector distributed as $N(0, \lim_{n\to\infty}(\frac{1}{n}\Delta_n)^{-1})$, where F_1, F_2, F_4 and F_5 are univariate random variables, and the distribution of $N(0, \lim_{n\to\infty}(\frac{1}{n}\Delta_n)^{-1})$ is the asymptotic distribution of $(\frac{1}{n}\Delta_n)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\lambda_0,\beta_{10},\beta_{20},\sigma_0^2,0)}{\partial \theta^\ddagger}$ for $\theta^\ddagger = [\eta^{\ddagger\prime}, \tau]'$. We may show that conditional on $\hat{\tau} > 0$, $\sqrt{n}[\hat{\eta}^{\ddagger\prime} - \eta_0', \hat{\tau}]'$ converges in distribution to the random vector $[F_1, F_2, F_3', F_4, |F_5|]'$, where $|F_5|$ represents the truncated normal of F_5 on the nonnegative axis.

[8]When $\delta_0 > 0$, $\frac{1}{n}\sum_{i=1}^n \check{\epsilon}_{ni}^3$ has a negative probability limit, by a proof similar to that for $\frac{1}{n}\sum_{i=1}^n \check{\epsilon}_{ni}^3$ in the proof of Proposition 2.5. However, for a finite sample size in practice, it can be the case that $\sum_{i=1}^n \check{\epsilon}_{ni}^3 \geq 0$, which implies that $[\check{\eta}', 0]'$ is a stationary point of the log likelihood function. This is the so-called "wrong skew" problem (see, e.g., Olson et al., 1980; Waldman, 1982; Simar and Wilson, 2010). Horrace and Wright (2019) study conditions for the existence of stationary points in parametric stochastic frontier models.

Denote $\hat{K} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\check{\epsilon}_{ni}^{3}$. Since $\delta = \tau^{1/3}$, conditional on $\hat{K} \geq 0$, $\hat{\delta} = 0$; conditional on $\hat{K} < 0$, $(n^{1/6}\hat{\delta})^3 = n^{1/2}\hat{\tau} \xrightarrow{d} |F_5|$, which implies that $n^{1/6}\hat{\delta} = O_p(1)$. With the asymptotic distribution of $[\hat{\eta}^{\ddagger\prime}, \hat{\tau}]'$, and the relations $\beta_1 = \beta_1^{\dagger} + \tau^{1/3}\sigma^{\dagger}(\frac{2}{\pi} + \frac{4}{\pi^2}\tau^{2/3})^{1/2}$ and $\sigma^2 = (1 + \frac{2}{\pi}\tau^{2/3})\sigma^{\dagger 2}$, conditional on $\hat{K} \geq 0$, the asymptotic distributions of $\hat{\beta}_1$ and $\hat{\sigma}^2$ are the same as those of $\hat{\beta}_1^{\dagger}$ and $\hat{\sigma}^{\dagger 2}$; conditional on $\hat{K} < 0$,

$$\begin{aligned} n^{1/6}(\hat{\beta}_1 - \beta_{10}) &= n^{1/6}(\hat{\beta}_1^{\dagger} - \beta_{10}) + n^{1/6}\hat{\tau}^{1/3}\hat{\sigma}^{\dagger}\big(\frac{2}{\pi} + \frac{4}{\pi^2}\hat{\tau}^{2/3}\big)^{1/2} \\ &= O_p(n^{1/6-1/2}) + n^{1/6}\hat{\delta}\sigma_0\sqrt{2/\pi} + o_p(1) \\ &= (n^{1/6}\hat{\delta})\sigma_0\sqrt{2/\pi} + o_p(1) \end{aligned}$$

and

$$\begin{aligned} n^{1/3}(\hat{\sigma}^2 - \sigma_0^2) &= n^{1/3}(\hat{\sigma}^{\dagger 2} - \sigma_0^2) + \frac{2}{\pi}n^{1/3}\hat{\tau}^{2/3}\hat{\sigma}^{\dagger 2} \\ &= O_p(n^{1/3-1/2}) + \frac{2}{\pi}(n^{1/6}\hat{\delta})^2\sigma_0^2 + o_p(1) \\ &= \frac{2}{\pi}(n^{1/6}\hat{\delta})^2\sigma_0^2 + o_p(1). \end{aligned}$$

The analysis requires the following assumption.

Assumption 12. *(a)* $\sup_{1\leq k\leq k_x, i,n} \mathrm{E}[|x_{ni,k}|^{14}] < \infty$; *(b)* $\alpha > \frac{17}{5}d$; *(c) for the* α*-mixing coefficients of* $\{x_{ni}\}_{i=1}^{n}$ *in Assumption 5,* $\check{\alpha}(s)$ *satisfies* $\sum_{s=1}^{\infty} s^{d[1+c_3\iota^*/(2+\iota^*)]-1}[\hat{\alpha}(s)]^{\iota^*/(4+2\iota^*)} < \infty$ *for some* $0 < \iota^* < 5$.

Since $\frac{\partial \ln L_{3n}(\eta^{\ddagger},0)}{\partial \delta} = \frac{\partial^2 \ln L_{3n}(\eta^{\ddagger},0)}{\partial \delta^2} = 0$, the analysis on the asymptotic distribution of the MLE $\hat{\theta}$ essentially relies on higher order Taylor expansions of $\ln L_{3n}(\eta^{\ddagger}, \delta)$, thus Assumption 12$(a)$ is needed so that the orders of terms in a proper higher order Taylor expansion can be derived. With $\sup_{1\leq k\leq k_x, i,n} \mathrm{E}[|x_{ni,k}|^{14}] < \infty$, Assumption 12$(b)$–$(c)$ are conditions for the applicability of the CLT in Jenish and Prucha (2012).

Proposition 2.3. *Under Assumptions 1–7, 9, 11–12 and* $\delta_0 = 0$,

- (i) *conditional on* $\hat{K} \geq 0$, $\hat{\delta} = 0$ *and* $\sqrt{n}(\hat{\eta} - \eta_0) \xrightarrow{d} J$, *where* J *is independent of* K, *the limit of* $\hat{K}$*; and*
- (ii) *conditional on* $\hat{K} < 0$, $(n^{1/6}\hat{\delta})^3 \xrightarrow{d} |F_5|$, $n^{1/6}(\hat{\beta}_1 - \beta_{10}) = \sqrt{\frac{2}{\pi}}\sigma_0(n^{1/6}\hat{\delta}) + o_p(1)$, $n^{1/3}(\hat{\sigma}^2 - \sigma_0^2) = \frac{2}{\pi}\sigma_0^2(n^{1/6}\hat{\delta})^2 + o_p(1)$, *and* $\sqrt{n}[\hat{\lambda} - \lambda_0, \hat{\beta}_2' - \beta_{20}']' \xrightarrow{d} [F_1, F_3']'$.

2.3 Tests for H_0: $\delta_0 = 0$

As the asymptotic distribution of the MLE depends on whether $\delta_0 = 0$ or not, we consider LR and score tests of $\delta_0 = 0$. For the LR test, using the relation $\ln L_{4n}(\lambda, \beta_1^{\dagger}, \beta_2, \sigma^{\dagger 2}, \tau) = \ln L_n\big(\lambda, \beta_1^{\dagger} +$

$\tau^{1/3}\sigma^{\dagger}(\frac{2}{\pi}+\frac{4}{\pi^2}\tau^{2/3})^{1/2}, \beta_2, (1+\frac{2}{\pi}\tau^{2/3})\sigma^{\dagger 2}, \tau^{1/3})$ in (2.15), we have $2[\ln L_n(\hat{\eta},\hat{\delta}) - \ln L_n(\check{\eta},0)] = 2[\ln L_{4n}(\hat{\eta}^{\ddagger},\hat{\tau}) - \ln L_{4n}(\check{\eta},0)] \cdot I(\sum_{i=1}^n \check{\epsilon}_{ni}^3 < 0)$. Then the asymptotic distribution of $2[\ln L_n(\hat{\eta},\hat{\delta}) - \ln L_n(\check{\eta},0)]$ is $\chi^2(0)\cdot I(K \geq 0) + \chi^2(1) \cdot I(K < 0)$, which is derived by a Taylor expansion, where $\chi^2(0)$ is degenerate with a unit mass at zero. Due to the irregular feature of $L_n(\theta)$, a corresponding score test should be constructed with a higher order derivative of $\ln L_{3n}(\eta^{\ddagger},\delta)$ at $[\check{\eta},0]$ (Lee and Chesher, 1986). Equivalently, we may construct a score test with $\frac{\partial \ln L_{4n}(\check{\eta},0)}{\partial \tau}$. By (2.16) and (2.19), the score test statistic, which turns out to be $\frac{n\sum_{i=1}^n \check{\epsilon}_{ni}^3}{\sqrt{6}(\sum_{i=1}^n \check{\epsilon}_{ni}^2)^{3/2}}$, is asymptotically standard normal, and the test is left sided since $\delta \geq 0$.

Proposition 2.4. *Under Assumptions 1–7, 9 and 11–12, when $\delta_0 = 0$, we have*

(a) $2[\ln L_n(\hat{\eta},\hat{\delta}) - \ln L_n(\check{\eta},0)] \xrightarrow{d} \chi^2(0)\cdot I(K\geq 0) + \chi^2(1)\cdot I(K<0)$; *and*

(b) *the score test statistic* $\frac{n\sum_{i=1}^n \check{\epsilon}_{ni}^3}{\sqrt{6}(\sum_{i=1}^n \check{\epsilon}_{ni}^2)^{3/2}} \xrightarrow{d} N(0,1)$, *and H_0: $\delta_0 = 0$ is rejected if* $\frac{n\sum_{i=1}^n \check{\epsilon}_{ni}^3}{\sqrt{6}(\sum_{i=1}^n \check{\epsilon}_{ni}^2)^{3/2}} < c_\varsigma$, *where c_ς satisfies $\Phi(c_\varsigma) = \varsigma$ for a chosen level of significance ς.*

The LR test involves both unrestricted and restricted MLEs, while the score test only involves the restricted MLE. If the LR test is used, the MLEs will be computed; if the score test is used, in the case that the null hypothesis of $\delta_0 = 0$ is not rejected, a researcher might further consider the choice of an appropriate model before computing the MLE of the SARSF model. By performing a test and then constructing an estimation based on the result of a test, the final estimator would be subject to the pretesting problem if the level of significance is fixed but does not depend on sample size. However, one might argue that it is reasonable to have the level of significance decrease as the sample size increases, then asymptotically the pretesting problem would not be an issue any more. Indeed, in practice, we can suggest such a testing procedure and then execute the proper estimation.

Also, a consistent estimator of $[\lambda_0, \beta_{20}']'$ can be derived by a two stage least squares (2SLS). While the 2SLS estimate of the intercept term might not be consistent, it can be adjusted to achieve consistency. Overall, a corrected 2SLS estimator (C2SLSE) of θ_0 can be derived similarly to the corrected ordinary least squares estimator for the SF model with no spatial dependence (Aigner et al., 1977). The details are in the next subsection. The C2SLSE is consistent but might not be asymptotically efficient under regularity conditions. It is also computationally simple for large sample sizes, since it avoids the computation of the determinant $|I_n - \lambda W_n|$.[9]

2.4 Corrected 2SLS estimation

Let Q_n be an IV matrix for $Z_n = [W_n Y_n, X_n]$, which can consist of, e.g., linearly independent columns of $[X_n, W_n X_n, W_n^2 X_n]$. Then the 2SLS estimate $\tilde{\kappa}$ of $\kappa_0 = [\lambda_0, \beta_0']'$ is $\tilde{\kappa} = (Z_n' P_n Z_n)^{-1} Z_n' P_n Y_n$, where $P_n = Q_n(Q_n'Q_n)^{-1}Q_n'$. Let $\tilde{\epsilon}_{ni} = y_{ni} - \tilde{\lambda} w_{n,i\cdot} Y_n - x_{ni}'\tilde{\beta}$. We can estimate σ_{u0}^2 by $\tilde{\sigma}_u^2 =$

[9] It can also be used as the starting point in optimization subroutines for the search of the MLE.

$\left[\frac{\pi}{\pi-4}\sqrt{\frac{\pi}{2}}\left(\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3\right)\right]^{2/3}$ if $\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3 < 0$, and $\tilde{\sigma}_u^2 = 0$ otherwise. The σ_{v0}^2 can be estimated by $\tilde{\sigma}_v^2 = \frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^2 - \frac{\pi-2}{\pi}\tilde{\sigma}_u^2$. Then estimates of σ_0^2 and δ_0 are, respectively, $\tilde{\sigma}^2 = \tilde{\sigma}_u^2 + \tilde{\sigma}_v^2$ and $\tilde{\delta} = \tilde{\sigma}_u/\tilde{\sigma}_v$. To derive a consistent estimate of β_{10}, adjust the 2SLS estimate $\tilde{\beta}_1$ to be $\tilde{\beta}_{1c} = \tilde{\beta}_1 + \sqrt{\frac{2}{\pi}}\tilde{\sigma}_u$. The C2SLSE of θ_0 is $\tilde{\theta}_c = [\tilde{\lambda}, \tilde{\beta}_{1c}, \tilde{\beta}_2', \tilde{\sigma}^2, \tilde{\delta}]'$. We maintain the following assumption for the C2SLSE, which would be satisfied by proper selection of IVs.

Assumption 13. *(a)* $\frac{1}{n}Q_n'[\epsilon_n - \mathrm{E}(\epsilon_n)] = O_p(n^{-1/2})$; *(b)* $\operatorname{plim}_{n\to\infty}\frac{1}{n}Q_n'G_n[\epsilon_n - \mathrm{E}(\epsilon_n)] = 0$; *(c)* $\operatorname{plim}_{n\to\infty}\frac{1}{n}Q_n'[G_n l_n(\beta_{10} - \sigma_{u0}\sqrt{2/\pi}) + G_n X_{2n}\beta_{20}, X_n]$ *has full column rank; (d)* $\operatorname{plim}_{n\to\infty}\frac{1}{n}Q_n'Q_n$ *is positive definite.*

Assumption 13(a) imposes a relatively stronger condition than the exogeneity condition that $\operatorname{plim}_{n\to\infty}\frac{1}{n}Q_n'[\epsilon_n - \mathrm{E}(\epsilon_n)] = 0$ of the IV matrix Q_n in terms of its rate of convergence. This is because we would like to investigate the convergence rate of the C2SLSE below. Assumption 13(b) is needed due to the presence of the spatial lag W_nY_n. These two conditions would be satisfied by a proper selection of Q_n, e.g., Q_n consists of linearly independent columns of $[X_n, W_nX_n, W_n^2X_n]$, so that the LLN can be applied and the convergence rate of such averages would be of order $\sqrt{n}$. Assumption 13(c) requires that the instruments are relevant. It has taken into account the nonzero mean of v_{ni} if $\sigma_{u0} \neq 0$. When $\beta_{10} = \sigma_{u0}\sqrt{2/\pi}$ and $\beta_{20} = 0$, Assumption 13(c) would not hold as there is no valid IV for W_nY_n. In the case that W_n is normalized to have row sums equal to one, as G_nl_n is proportional to l_n and X_n also contains l_n, Assumption 13(c) requires β_{20} to be nonzero in order to avoid this possible multicolinearity in the regressors (of the reduced form equation). Assumption 13(d) is standard.

Proposition 2.5. *Under Assumptions 1–5 and 13,*

(i) *if* $\delta_0 > 0$, $\tilde{\theta}_c = \theta_0 + O_p(n^{-1/2})$;

(ii) *if* $\delta_0 = 0$, $\tilde{\lambda} = \lambda_0 + O_p(n^{-1/2})$, $\tilde{\beta}_2 = \beta_{20} + O_p(n^{-1/2})$, $\tilde{\delta} = O_p(n^{-1/6})$, $\tilde{\beta}_{1c} = \beta_{10} + O_p(n^{-1/6})$ *and* $\tilde{\sigma}^2 = \sigma_0^2 + O_p(n^{-1/3})$.

When $\delta_0 > 0$, the C2SLSE has the $\sqrt{n}$-rate of convergence; when $\delta_0 = 0$, only $\tilde{\lambda}$ and $\tilde{\beta}_2$ have the $\sqrt{n}$-rate of convergence, and other parameter estimators have slower rates of convergence, with the rates equal to the corresponding ones of the MLE.

3 Monte Carlo

In this section, we report some Monte Carlo results on the estimates and tests considered in this paper.

We generate data from model (2.1). The spatial weights matrix W_n is based on the queen criterion and normalized to have row sums equal to one.[10] There are three variables in x_{ni}: a constant term and two variables randomly drawn from the standard normal distribution. The true value of $\beta = [\beta_1, \beta_2, \beta_3]'$ is $[0.5, 0.5, 0.5]'$, λ_0 is either 0.2 or 0.6, σ_0^2 is either 1 or 2, and δ_0 is either 0, 0.5, 1, 1.5, 2 or 2.5. The sample size n is either 144 or 400. The number of Monte Carlo repetitions is $5,000$. Due to a small percentage of outliers, we report the following robust measures of central tendency and dispersions of the MLEs and C2SLSEs: the median bias (MB), the median absolute deviation (MAD), and the interdecile range (IDR).[11] For the estimates of δ, we also report the percentages of estimates equal to zero, and the MBs, MADs and IDRs with zero estimates excluded.

The estimation results when $\delta_0 = 0$ are reported in Table 1. The MLE-r is a restricted MLE with $\delta = 0$ imposed, i.e., the MLE of a standard SAR model. The MLE-r is $\sqrt{n}$-consistent under regularity conditions (Lee, 2004). Recall that the information matrix of model (2.1) is singular when $\delta_0 = 0$. As a result, only the MLEs and C2SLSEs of λ, β_2 and β_3 have the $\sqrt{n}$-rate of convergence, and those of β_1, σ^2 and δ have slower rates of convergence. Table 1 shows that MLE-r performs the best, MLE has similar performance as that of MLE-r for λ, β_2 and β_3, but MLE performs worse than MLE-r for β_1 and σ^2. The MLEs and C2SLSEs of λ, β_2 and β_3 have relatively small MBs in all cases, while those of β_1, σ^2 and δ can have larger MBs, especially those of β_1. For λ, the C2SLSEs have much larger MBs, MADs and IDRs than those of the MLEs; for β_1, the C2SLSEs also have larger MBs, MADs and IDRs than those of the MLEs except when $n = 144$ and $\lambda_0 = 0.6$; for β_2 and β_3, the C2SLSEs have similar MBs, MADs and IDRs as those of the MLEs; for σ^2, the C2SLSEs have larger MBs and MADs than those of the MLEs, but they have slightly smaller IDRs; for δ, the C2SLSEs have slightly smaller IDRs than those of the MLEs, but neither the MLE nor the C2SLSE has a dominating performance in terms of the MB and MAD. Note that some MBs and MADs of the estimates of δ are 0.000. This is because more than 50% estimates are estimated as zero.

Table 2 reports the estimation results when $\delta_0 \neq 0$ and $n = 144$. In addition to MBs, MADs and IDRs, coverage probabilities (CP) of 95% confidence intervals are also reported for MLE-r and MLE. As MLE-r restricts the wrong restriction $\delta = 0$, it has large biases and extremely low CPs for β_1 and σ^2. The MLEs and C2SLSEs have the $\sqrt{n}$-rate of convergence when $\delta_0 \neq 0$. The MBs of the MLEs are relatively small in all cases, and they are generally smaller than those of the C2SLSEs. For λ, β_1, β_2 and β_3, the MLEs have smaller MADs and IDRs than those of the C2SLSEs in most cases; for σ^2 and δ, the C2SLSEs have smaller MADs and IDRs in some cases. When $\delta_0 = 1$, the CPs of the MLEs for λ, β_2 and β_3 are close to the nominal 95%, while the CPs

[10] Connectivity for the queen criterion is based on a grid of cells. Each cell corresponds to a reference location, so the sample size is k^2 for a $k \times k$ grid. The spatial weight $w_{n,ij}$ is 1 if cell i and cell j share a common side or vertex, and $w_{n,ij} = 0$ otherwise. See Kelejian and Robinson (1995) for more details and definitions of various types of spatial weights matrices.

[11] The IDR is the difference between the 90% quantile and 10% quantile in the empirical distribution.

of β_1, σ^2 and δ are significantly lower than 95%; when $\delta_0 = 2$, all CPs of the MLEs are close to 95%. Note that in all cases, more than a quarter of MLEs and C2SLSEs of δ are estimated as zero when $\delta_0 = 1$, but less than 2.5% of MLEs and C2SLSEs of δ are estimated as zero when $\delta_0 = 2$. The large percentages of zero estimates when $\delta_0 = 1$ explain why the CPs for MLEs and C2SLSEs of some parameters are much lower than the nominal level, while the small percentages of zero estimates when $\delta_0 = 2$ explain why the CPs are close to the nominal level. Thus, with the sample size $n = 144$, we observe a relatively severe wrong skew problem for $\delta_0 = 1$, which is mentioned in footnote 8.

Table 3 reports the estimation results when $\delta_0 \neq 0$ for a larger sample size $n = 400$. We observe that the MBs, MADs and IDRs are smaller than those in Table 2. Compared to the results with $n = 144$, the C2SLSEs of σ^2 and δ have smaller MADs and IDRs than those of the MLEs in much fewer cases, and other patterns are similar. Note that, with the sample size $n = 400$, in all cases, less than 15% of MLEs and C2SLSEs of δ are estimated as zero for $\delta_0 = 1$, and almost all MLEs and C2SLSEs of δ are positive for $\delta_0 = 2$. For $\delta_0 = 1$, with zero estimates excluded, while the MBs are larger, the MADs and IDRs are much smaller.

Empirical sizes of the score and LR tests are reported in Table 4. With $n = 144$, at the 5% level of significance, the size distortions of the score and LR tests are within, respectively, 0.9 and 0.8 percentage points; for the 10% level of significance, they are within, respectively, 1.6 and 1.3 percentage points. Size distortion generally decreases as n increases (from 144 to 400).

Table 5 reports empirical powers of the tests. The score and LR tests have similar powers. Powers increase as δ_0 or the sample size increases. For $n = 144$ and $\delta_0 = 0.5$, the powers are similar to the significance level and are small; but for $n = 400$ and $\delta_0 = 2.5$, the powers are all close to 1.

4 Conclusion

We study asymptotic properties of the MLE and a corrected 2SLSE for the SARSF model in this paper. When inefficiency exists, all model parameter estimators are $\sqrt{n}$ consistent and asymptotically normal; when there is no inefficiency, only some parameter estimators are $\sqrt{n}$ consistent and the rest of parameters have slower rates of convergence. We also derive the asymptotic distributions of the score and likelihood ratio test statistics that test for the existence of inefficiency.

For the SARSF model with exponential distribution under efficiency, some very preliminary investigation has indicated that its information matrix might still be non-singular and the rate of convergence of its ML estimator would still be regular. So it is likely that the irregularity of estimates for an SF model and an SARSF model, when all firms are efficiently operated, depends on a parametric form of a one-sided distribution of the possibly inefficient disturbance—a particular feature of a stochastic frontier model. Our analysis does not allow for distributional misspecification of inefficiency and disturbance terms. It is of interest to extend the analysis to the case with

Table 1: MBs, MADs and IDRs of parameter estimates when $\delta_0 = 0$

		$\lambda_0 = 0.2$, $\sigma_0^2 = 1$	$\lambda_0 = 0.2$, $\sigma_0^2 = 2$	$\lambda_0 = 0.6$, $\sigma_0^2 = 1$	$\lambda_0 = 0.6$, $\sigma_0^2 = 2$
$n = 144$					
λ	MLE-r	-0.027[0.084]0.336	-0.027[0.091]0.352	-0.023[0.061]0.242	-0.029[0.063]0.249
	MLE	-0.027[0.085]0.337	-0.027[0.092]0.353	-0.024[0.061]0.243	-0.029[0.063]0.250
	C2SLSE	0.068[0.174]0.706	0.135[0.229]0.917	0.063[0.112]0.465	0.114[0.140]0.596
β_1	MLE-r	0.013[0.079]0.303	0.016[0.100]0.377	0.025[0.097]0.385	0.025[0.115]0.445
	MLE	0.233[0.287]0.987	0.319[0.401]1.374	0.282[0.316]1.038	0.353[0.422]1.434
	C2SLSE	0.251[0.329]1.039	0.339[0.468]1.443	0.236[0.344]1.120	0.293[0.483]1.519
β_2	MLE-r	-0.002[0.058]0.216	0.001[0.079]0.309	0.001[0.058]0.219	0.002[0.082]0.312
	MLE	-0.001[0.058]0.218	0.000[0.079]0.310	0.001[0.058]0.221	0.002[0.083]0.313
	C2SLSE	-0.007[0.058]0.220	-0.007[0.080]0.312	-0.005[0.058]0.220	-0.009[0.084]0.315
β_3	MLE-r	0.001[0.057]0.215	-0.004[0.082]0.306	0.002[0.058]0.214	0.001[0.079]0.309
	MLE	0.000[0.057]0.216	-0.004[0.082]0.306	0.002[0.058]0.217	-0.000[0.079]0.310
	C2SLSE	-0.005[0.057]0.217	-0.012[0.082]0.310	-0.005[0.059]0.219	-0.012[0.081]0.311
σ^2	MLE-r	-0.036[0.077]0.295	-0.066[0.151]0.590	-0.027[0.080]0.298	-0.045[0.158]0.610
	MLE	0.096[0.190]0.885	0.200[0.374]1.740	0.106[0.192]0.902	0.210[0.377]1.838
	C2SLSE	0.114[0.192]0.829	0.261[0.394]1.677	0.105[0.193]0.850	0.245[0.392]1.727
δ	MLE-r	0.000[0.000]0.000	0.000[0.000]0.000	0.000[0.000]0.000	0.000[0.000]0.000
	MLE	0.000[0.000]1.494	0.124[0.124]1.472	0.064[0.064]1.487	0.000[0.000]1.497
		1.026[0.339]1.306{0.501}	1.027[0.329]1.280{0.494}	1.037[0.334]1.347{0.498}	1.038[0.342]1.307{0.504}
	C2SLSE	0.000[0.000]1.354	0.217[0.217]1.337	0.000[0.000]1.373	0.162[0.162]1.367
		0.981[0.287]1.090{0.501}	0.979[0.282]1.052{0.496}	0.988[0.285]1.118{0.504}	0.982[0.289]1.094{0.499}
$n = 400$					
λ	MLE-r	-0.011[0.052]0.198	-0.012[0.055]0.208	-0.009[0.036]0.141	-0.011[0.039]0.148
	MLE	-0.010[0.052]0.198	-0.012[0.055]0.210	-0.009[0.036]0.141	-0.011[0.039]0.148
	C2SLSE	0.023[0.110]0.428	0.049[0.153]0.589	0.025[0.075]0.298	0.050[0.099]0.401
β_1	MLE-r	0.005[0.048]0.184	0.006[0.061]0.229	0.008[0.059]0.219	0.010[0.068]0.261
	MLE	0.157[0.219]0.794	0.211[0.297]1.114	0.185[0.238]0.811	0.232[0.313]1.134
	C2SLSE	0.209[0.267]0.845	0.305[0.392]1.184	0.220[0.284]0.883	0.290[0.403]1.256
β_2	MLE-r	-0.001[0.035]0.128	-0.001[0.047]0.179	0.001[0.034]0.129	-0.001[0.047]0.183
	MLE	-0.001[0.035]0.128	-0.000[0.047]0.180	0.000[0.035]0.129	-0.001[0.047]0.183
	C2SLSE	-0.004[0.035]0.131	-0.006[0.047]0.179	-0.003[0.035]0.129	-0.006[0.048]0.183
β_3	MLE-r	-0.002[0.033]0.129	-0.001[0.048]0.184	0.000[0.035]0.129	0.001[0.048]0.184
	MLE	-0.002[0.033]0.129	-0.000[0.048]0.183	0.000[0.035]0.128	0.001[0.048]0.183
	C2SLSE	-0.004[0.034]0.130	-0.003[0.049]0.185	-0.002[0.035]0.128	-0.005[0.049]0.184
σ^2	MLE-r	-0.014[0.049]0.181	-0.024[0.095]0.361	-0.007[0.047]0.184	-0.016[0.100]0.378
	MLE	0.075[0.126]0.601	0.160[0.258]1.194	0.081[0.129]0.592	0.163[0.262]1.217
	C2SLSE	0.082[0.128]0.590	0.188[0.271]1.180	0.089[0.134]0.582	0.182[0.282]1.206
δ	MLE-r	0.000[0.000]0.000	0.000[0.000]0.000	0.000[0.000]0.000	0.000[0.000]0.000
	MLE	0.075[0.075]1.077	0.106[0.106]1.086	0.081[0.081]1.087	0.000[0.000]1.086
		0.810[0.220]0.829{0.499}	0.809[0.226]0.862{0.495}	0.801[0.223]0.828{0.498}	0.804[0.224]0.833{0.502}
	C2SLSE	0.000[0.000]1.062	0.169[0.169]1.077	0.000[0.000]1.060	0.000[0.000]1.064
		0.806[0.201]0.772{0.501}	0.805[0.208]0.774{0.497}	0.799[0.205]0.777{0.502}	0.806[0.208]0.770{0.503}

[a] For parameters except δ, the three numbers in each cell are MB[MAD]IDR; for δ, the additional rows for MLE and C2SLSE have four numbers in each cell, where the first three numbers are MB[MAD]IDR with zero estimates excluded and the last number in curly parentheses is the percentage of zero estimates. MB: median bias; MAD: median absolute deviation; IDR: interdecile range.

[b] $\beta_0 = [0.5, 0.5, 0.5]'$, and the number of Monte Carlo repetitions is $5,000$.

Table 2: MBs, MADs and IDRs of parameter estimates when $\delta_0 \neq 0$ and $n = 144$

		$\lambda_0 = 0.2$, $\sigma_0^2 = 1$	$\lambda_0 = 0.2$, $\sigma_0^2 = 2$	$\lambda_0 = 0.6$, $\sigma_0^2 = 1$	$\lambda_0 = 0.6$, $\sigma_0^2 = 2$
$\delta_0 = 1$					
λ	MLE-r	-0.025[0.085]0.325(0.939)	-0.026[0.087]0.338(0.946)	-0.023[0.058]0.228(0.943)	-0.027[0.063]0.241(0.943)
	MLE	-0.026[0.085]0.324(0.938)	-0.027[0.088]0.340(0.944)	-0.022[0.058]0.229(0.939)	-0.027[0.063]0.241(0.944)
	C2SLSE	0.055[0.150]0.592(***)	0.101[0.191]0.772(***)	0.046[0.098]0.401(***)	0.087[0.126]0.535(***)
β_1	MLE-r	-0.567[0.048]0.187(0.000)	-0.804[0.075]0.288(0.000)	-0.568[0.051]0.200(0.000)	-0.812[0.081]0.326(0.000)
	MLE	-0.036[0.245]0.871(0.654)	-0.056[0.339]1.211(0.667)	-0.029[0.232]0.866(0.666)	-0.056[0.337]1.227(0.676)
	C2SLSE	-0.049[0.228]0.839(***)	-0.059[0.344]1.182(***)	-0.039[0.219]0.829(***)	-0.019[0.347]1.216(***)
β_2	MLE-r	-0.002[0.048]0.183(0.942)	-0.003[0.068]0.254(0.945)	-0.000[0.048]0.178(0.950)	-0.003[0.069]0.258(0.940)
	MLE	-0.002[0.048]0.183(0.936)	-0.002[0.068]0.256(0.944)	-0.000[0.048]0.179(0.948)	-0.003[0.068]0.260(0.940)
	C2SLSE	-0.005[0.048]0.183(***)	-0.008[0.068]0.258(***)	-0.006[0.048]0.177(***)	-0.011[0.069]0.261(***)
β_3	MLE-r	-0.000[0.047]0.178(0.941)	-0.000[0.066]0.251(0.949)	0.001[0.046]0.175(0.947)	0.000[0.067]0.257(0.947)
	MLE	0.001[0.047]0.178(0.939)	-0.000[0.066]0.252(0.947)	0.000[0.046]0.177(0.944)	0.001[0.068]0.258(0.945)
	C2SLSE	-0.003[0.047]0.179(***)	-0.007[0.066]0.252(***)	-0.005[0.046]0.176(***)	-0.009[0.068]0.256(***)
σ^2	MLE-r	-0.341[0.053]0.206(0.040)	-0.685[0.109]0.406(0.039)	-0.341[0.056]0.207(0.044)	-0.672[0.108]0.411(0.046)
	MLE	-0.065[0.237]0.785(0.718)	-0.120[0.477]1.565(0.716)	-0.052[0.241]0.788(0.720)	-0.077[0.488]1.566(0.731)
	C2SLSE	-0.079[0.215]0.737(***)	-0.148[0.434]1.456(***)	-0.077[0.220]0.743(***)	-0.134[0.442]1.503(***)
δ	MLE-r	-1.000[0.000]0.000(0.000)	-1.000[0.000]0.000(0.000)	-1.000[0.000]0.000(0.000)	-1.000[0.000]0.000(0.000)
	MLE	-0.026[0.617]1.971(0.714)	-0.014[0.604]1.898(0.714)	-0.014[0.604]1.959(0.715)	0.005[0.589]1.944(0.724)
		0.236[0.394]1.573{0.268}	0.238[0.367]1.474{0.272}	0.249[0.378]1.510{0.268}	0.245[0.376]1.455{0.261}
	C2SLSE	-0.080[0.548]1.776(***)	-0.080[0.537]1.736(***)	-0.059[0.530]1.783(***)	-0.043[0.519]1.786(***)
		0.171[0.343]1.295{0.275}	0.162[0.319]1.246{0.285}	0.177[0.328]1.321{0.274}	0.174[0.327]1.281{0.272}
$\delta_0 = 2$					
λ	MLE-r	-0.021[0.077]0.294(0.950)	-0.028[0.085]0.328(0.943)	-0.022[0.055]0.212(0.945)	-0.025[0.060]0.232(0.948)
	MLE	-0.018[0.076]0.289(0.944)	-0.026[0.082]0.325(0.937)	-0.019[0.055]0.208(0.938)	-0.025[0.058]0.225(0.941)
	C2SLSE	0.038[0.129]0.504(***)	0.077[0.172]0.676(***)	0.033[0.086]0.338(***)	0.061[0.112]0.449(***)
β_1	MLE-r	-0.716[0.046]0.176(0.000)	-1.020[0.081]0.306(0.000)	-0.720[0.052]0.200(0.000)	-1.035[0.095]0.376(0.000)
	MLE	-0.011[0.096]0.386(0.928)	-0.025[0.141]0.576(0.936)	-0.015[0.096]0.403(0.927)	-0.047[0.150]0.609(0.936)
	C2SLSE	-0.031[0.106]0.418(***)	-0.019[0.176]0.712(***)	-0.024[0.109]0.435(***)	-0.001[0.195]0.795(***)
β_2	MLE-r	-0.002[0.040]0.151(0.942)	0.002[0.056]0.216(0.943)	0.000[0.040]0.151(0.948)	-0.001[0.056]0.216(0.939)
	MLE	-0.001[0.039]0.147(0.941)	0.002[0.054]0.209(0.939)	0.001[0.038]0.146(0.935)	-0.002[0.054]0.211(0.931)
	C2SLSE	-0.004[0.041]0.153(***)	-0.002[0.057]0.217(***)	-0.004[0.040]0.153(***)	-0.008[0.057]0.219(***)
β_3	MLE-r	-0.001[0.040]0.154(0.949)	-0.002[0.058]0.219(0.944)	0.001[0.040]0.152(0.942)	0.003[0.055]0.216(0.947)
	MLE	-0.001[0.039]0.148(0.942)	-0.001[0.056]0.212(0.930)	0.001[0.037]0.148(0.934)	0.004[0.055]0.208(0.942)
	C2SLSE	-0.003[0.041]0.154(***)	-0.006[0.058]0.220(***)	-0.003[0.041]0.151(***)	-0.004[0.056]0.216(***)
σ^2	MLE-r	-0.526[0.041]0.156(0.000)	-1.056[0.082]0.309(0.000)	-0.523[0.041]0.162(0.000)	-1.050[0.084]0.320(0.000)
	MLE	-0.026[0.156]0.592(0.944)	-0.062[0.309]1.181(0.943)	-0.022[0.152]0.600(0.939)	-0.059[0.301]1.186(0.940)
	C2SLSE	-0.068[0.154]0.596(***)	-0.140[0.308]1.206(***)	-0.068[0.154]0.600(***)	-0.160[0.311]1.180(***)
δ	MLE-r	-2.000[0.000]0.000(0.000)	-2.000[0.000]0.000(0.000)	-2.000[0.000]0.000(0.000)	-2.000[0.000]0.000(0.000)
	MLE	0.077[0.529]2.253(0.977)	0.087[0.537]2.276(0.976)	0.072[0.530]2.203(0.971)	0.075[0.512]2.134(0.980)
		0.093[0.521]2.188{0.018}	0.110[0.528]2.235{0.017}	0.091[0.520]2.150{0.021}	0.094[0.505]2.075{0.016}
	C2SLSE	-0.152[0.479]2.145(***)	-0.158[0.490]2.214(***)	-0.147[0.505]2.176(***)	-0.176[0.467]2.113(***)
		-0.135[0.468]2.105{0.021}	-0.145[0.478]2.180{0.018}	-0.126[0.494]2.138{0.022}	-0.159[0.457]2.078{0.019}

[a] For parameters except δ, the four numbers in each cell are MB[MAD]IDR(CP); for δ, the additional rows for MLE and C2SLSE have four numbers in each cell, where the first three numbers are MB[MAD]IDR with zero estimates excluded and the last number in curly parentheses is the percentage of zero estimates. MB: median bias; MAD: median absolute deviation; IDR: interdecile range; CP: coverage probability of a 95% confidence interval.

[b] $\beta_0 = [0.5, 0.5, 0.5]'$, and the number of Monte Carlo repetitions is $5,000$.

Table 3: MBs, MADs and IDRs of parameter estimates when $\delta_0 \neq 0$ and $n = 400$

		$\lambda_0 = 0.2$, $\sigma_0^2 = 1$	$\lambda_0 = 0.2$, $\sigma_0^2 = 2$	$\lambda_0 = 0.6$, $\sigma_0^2 = 1$	$\lambda_0 = 0.6$, $\sigma_0^2 = 2$
$\delta_0 = 1$					
λ	MLE-r	-0.009[0.049]0.192(0.944)	-0.009[0.055]0.204(0.950)	-0.006[0.034]0.133(0.945)	-0.011[0.037]0.142(0.948)
	MLE	-0.009[0.049]0.191(0.943)	-0.009[0.054]0.204(0.949)	-0.006[0.034]0.133(0.945)	-0.011[0.037]0.143(0.946)
	C2SLSE	0.020[0.094]0.361(***)	0.045[0.129]0.503(***)	0.016[0.062]0.255(***)	0.032[0.088]0.346(***)
β_1	MLE-r	-0.565[0.029]0.108(0.000)	-0.801[0.044]0.171(0.000)	-0.565[0.029]0.113(0.000)	-0.805[0.050]0.191(0.000)
	MLE	-0.008[0.125]0.719(0.808)	-0.015[0.183]1.015(0.813)	-0.006[0.127]0.724(0.797)	-0.014[0.180]1.009(0.802)
	C2SLSE	-0.012[0.121]0.708(***)	-0.011[0.186]1.003(***)	-0.012[0.125]0.715(***)	-0.007[0.190]0.994(***)
β_2	MLE-r	-0.000[0.028]0.105(0.945)	0.000[0.040]0.150(0.948)	0.000[0.028]0.104(0.955)	0.001[0.039]0.146(0.956)
	MLE	0.000[0.028]0.105(0.947)	-0.000[0.040]0.151(0.947)	-0.000[0.028]0.105(0.955)	0.000[0.039]0.145(0.954)
	C2SLSE	-0.001[0.028]0.105(***)	-0.003[0.041]0.152(***)	-0.002[0.028]0.105(***)	-0.004[0.039]0.150(***)
β_3	MLE-r	-0.002[0.028]0.106(0.954)	-0.001[0.040]0.149(0.954)	0.000[0.028]0.105(0.949)	-0.000[0.040]0.153(0.945)
	MLE	-0.002[0.028]0.105(0.955)	-0.001[0.039]0.149(0.955)	0.000[0.028]0.105(0.949)	0.001[0.040]0.152(0.944)
	C2SLSE	-0.003[0.028]0.105(***)	-0.004[0.040]0.150(***)	-0.002[0.028]0.107(***)	-0.003[0.041]0.153(***)
σ^2	MLE-r	-0.327[0.033]0.126(0.000)	-0.652[0.066]0.251(0.000)	-0.326[0.034]0.128(0.000)	-0.649[0.068]0.255(0.000)
	MLE	-0.019[0.149]0.563(0.842)	-0.033[0.307]1.132(0.841)	-0.015[0.155]0.571(0.835)	-0.018[0.308]1.143(0.838)
	C2SLSE	-0.026[0.145]0.547(***)	-0.040[0.296]1.109(***)	-0.024[0.153]0.560(***)	-0.044[0.299]1.120(***)
δ	MLE-r	-1.000[0.000]0.000(0.000)	-1.000[0.000]0.000(0.000)	-1.000[0.000]0.000(0.000)	-1.000[0.000]0.000(0.000)
	MLE	-0.003[0.293]1.516(0.836)	-0.004[0.305]1.499(0.837)	0.007[0.307]1.514(0.829)	0.015[0.303]1.543(0.830)
		0.061[0.240]0.947{0.132}	0.069[0.251]0.929{0.138}	0.084[0.245]0.957{0.143}	0.090[0.248]0.962{0.138}
	C2SLSE	-0.016[0.283]1.475(***)	-0.022[0.294]1.472(***)	-0.014[0.300]1.470(***)	-0.003[0.291]1.501(***)
		0.049[0.235]0.886{0.135}	0.048[0.239]0.898{0.141}	0.065[0.238]0.901{0.147}	0.067[0.235]0.921{0.142}
$\delta_0 = 2$					
λ	MLE-r	-0.007[0.048]0.181(0.954)	-0.008[0.051]0.197(0.950)	-0.007[0.033]0.125(0.947)	-0.008[0.036]0.140(0.944)
	MLE	-0.007[0.046]0.176(0.954)	-0.007[0.050]0.195(0.952)	-0.007[0.032]0.121(0.946)	-0.008[0.034]0.136(0.946)
	C2SLSE	0.012[0.079]0.314(***)	0.030[0.112]0.432(***)	0.011[0.055]0.213(***)	0.022[0.078]0.302(***)
β_1	MLE-r	-0.716[0.027]0.102(0.000)	-1.012[0.047]0.178(0.000)	-0.717[0.029]0.114(0.000)	-1.019[0.059]0.221(0.000)
	MLE	-0.003[0.052]0.204(0.961)	-0.011[0.080]0.307(0.956)	-0.007[0.054]0.211(0.957)	-0.018[0.086]0.325(0.954)
	C2SLSE	-0.011[0.062]0.238(***)	-0.005[0.105]0.423(***)	-0.010[0.064]0.243(***)	0.002[0.124]0.486(***)
β_2	MLE-r	-0.001[0.024]0.090(0.954)	-0.000[0.033]0.129(0.942)	0.000[0.023]0.090(0.953)	-0.001[0.033]0.128(0.953)
	MLE	-0.000[0.023]0.086(0.946)	-0.000[0.032]0.123(0.940)	0.000[0.022]0.085(0.954)	-0.002[0.032]0.121(0.951)
	C2SLSE	-0.002[0.024]0.091(***)	-0.003[0.034]0.131(***)	-0.001[0.024]0.090(***)	-0.003[0.033]0.129(***)
β_3	MLE-r	0.001[0.025]0.092(0.942)	0.001[0.033]0.127(0.956)	-0.000[0.023]0.090(0.948)	0.002[0.034]0.129(0.945)
	MLE	0.001[0.023]0.087(0.943)	0.000[0.032]0.121(0.951)	-0.000[0.022]0.087(0.948)	0.001[0.032]0.123(0.946)
	C2SLSE	-0.000[0.025]0.091(***)	-0.002[0.033]0.127(***)	-0.002[0.024]0.090(***)	-0.002[0.035]0.129(***)
σ^2	MLE-r	-0.515[0.025]0.094(0.000)	-1.031[0.049]0.190(0.000)	-0.514[0.026]0.099(0.000)	-1.028[0.050]0.190(0.000)
	MLE	-0.006[0.087]0.334(0.950)	-0.024[0.177]0.681(0.950)	-0.009[0.087]0.338(0.955)	-0.019[0.179]0.670(0.951)
	C2SLSE	-0.024[0.094]0.358(***)	-0.053[0.191]0.728(***)	-0.025[0.095]0.361(***)	-0.055[0.192]0.737(***)
δ	MLE-r	-2.000[0.000]0.000(0.000)	-2.000[0.000]0.000(0.000)	-2.000[0.000]0.000(0.000)	-2.000[0.000]0.000(0.000)
	MLE	0.035[0.284]1.116(0.963)	0.024[0.289]1.111(0.964)	0.033[0.285]1.113(0.963)	0.027[0.287]1.104(0.962)
		0.035[0.284]1.116{0.000}	0.024[0.289]1.111{0.000}	0.033[0.285]1.112{0.000}	0.027[0.287]1.104{0.000}
	C2SLSE	-0.056[0.307]1.248(***)	-0.070[0.316]1.265(***)	-0.064[0.310]1.287(***)	-0.064[0.317]1.250(***)
		-0.056[0.307]1.248{0.000}	-0.070[0.316]1.265{0.000}	-0.064[0.310]1.286{0.000}	-0.064[0.317]1.250{0.000}

[a] For parameters except δ, the four numbers in each cell are MB[MAD]IDR(CP); for δ, the additional rows for MLE and C2SLSE have four numbers in each cell, where the first three numbers are MB[MAD]IDR with zero estimates excluded and the last number in curly parentheses is the percentage of zero estimates. MB: median bias; MAD: median absolute deviation; IDR: interdecile range; CP: coverage probability of a 95% confidence interval.

[b] $\beta_0 = [0.5, 0.5, 0.5]'$, and the number of Monte Carlo repetitions is $5,000$.

Table 4: Empirical sizes of the score and LR tests

	$\varsigma = 5\%$		$\varsigma = 10\%$	
	Score	LR	Score	LR
$n = 144,\ \lambda_0 = 0.2,\ \sigma_0^2 = 1$	0.043	0.053	0.084	0.098
$n = 144,\ \lambda_0 = 0.2,\ \sigma_0^2 = 2$	0.041	0.051	0.090	0.102
$n = 144,\ \lambda_0 = 0.6,\ \sigma_0^2 = 1$	0.044	0.055	0.096	0.113
$n = 144,\ \lambda_0 = 0.6,\ \sigma_0^2 = 2$	0.046	0.058	0.093	0.103
$n = 400,\ \lambda_0 = 0.2,\ \sigma_0^2 = 1$	0.048	0.051	0.099	0.100
$n = 400,\ \lambda_0 = 0.2,\ \sigma_0^2 = 2$	0.051	0.054	0.102	0.104
$n = 400,\ \lambda_0 = 0.6,\ \sigma_0^2 = 1$	0.050	0.051	0.094	0.097
$n = 400,\ \lambda_0 = 0.6,\ \sigma_0^2 = 2$	0.048	0.049	0.098	0.101

[a] $\beta_0 = [0.5, 0.5, 0.5]'$, ς is the level of siginificance, and the number of Monte Carlo repetitions is $5,000$.

Table 5: Empirical powers of the score and LR tests

		Score					LR				
		$\delta_0 = 0.5$	$\delta_0 = 1$	$\delta_0 = 1.5$	$\delta_0 = 2$	$\delta_0 = 2.5$	$\delta_0 = 0.5$	$\delta_0 = 1$	$\delta_0 = 1.5$	$\delta_0 = 2$	$\delta_0 = 2.5$
$\varsigma = 5\%$	$n = 144,\ \lambda_0 = 0.2,\ \sigma_0^2 = 1$	0.054	0.152	0.393	0.645	0.824	0.067	0.174	0.434	0.700	0.870
	$n = 144,\ \lambda_0 = 0.2,\ \sigma_0^2 = 2$	0.055	0.152	0.386	0.653	0.824	0.066	0.174	0.433	0.713	0.870
	$n = 144,\ \lambda_0 = 0.6,\ \sigma_0^2 = 1$	0.058	0.147	0.379	0.640	0.832	0.069	0.165	0.425	0.699	0.881
	$n = 144,\ \lambda_0 = 0.6,\ \sigma_0^2 = 2$	0.061	0.147	0.382	0.644	0.835	0.075	0.167	0.423	0.696	0.878
	$n = 400,\ \lambda_0 = 0.2,\ \sigma_0^2 = 1$	0.066	0.287	0.750	0.974	0.998	0.069	0.298	0.764	0.979	0.998
	$n = 400,\ \lambda_0 = 0.2,\ \sigma_0^2 = 2$	0.078	0.290	0.756	0.975	0.998	0.082	0.297	0.769	0.978	0.998
	$n = 400,\ \lambda_0 = 0.6,\ \sigma_0^2 = 1$	0.076	0.285	0.756	0.972	0.998	0.080	0.289	0.770	0.975	0.999
	$n = 400,\ \lambda_0 = 0.6,\ \sigma_0^2 = 2$	0.065	0.303	0.756	0.972	0.997	0.067	0.308	0.770	0.974	0.998
$\varsigma = 10\%$	$n = 144,\ \lambda_0 = 0.2,\ \sigma_0^2 = 1$	0.109	0.253	0.526	0.772	0.904	0.124	0.277	0.559	0.806	0.925
	$n = 144,\ \lambda_0 = 0.2,\ \sigma_0^2 = 2$	0.105	0.246	0.527	0.784	0.903	0.120	0.270	0.562	0.814	0.925
	$n = 144,\ \lambda_0 = 0.6,\ \sigma_0^2 = 1$	0.116	0.242	0.517	0.762	0.914	0.133	0.267	0.552	0.794	0.932
	$n = 144,\ \lambda_0 = 0.6,\ \sigma_0^2 = 2$	0.126	0.244	0.524	0.764	0.911	0.138	0.269	0.553	0.800	0.932
	$n = 400,\ \lambda_0 = 0.2,\ \sigma_0^2 = 1$	0.126	0.421	0.851	0.990	0.999	0.130	0.426	0.856	0.991	0.999
	$n = 400,\ \lambda_0 = 0.2,\ \sigma_0^2 = 2$	0.147	0.418	0.851	0.990	1.000	0.151	0.422	0.858	0.992	1.000
	$n = 400,\ \lambda_0 = 0.6,\ \sigma_0^2 = 1$	0.141	0.411	0.856	0.987	0.999	0.144	0.416	0.863	0.987	0.999
	$n = 400,\ \lambda_0 = 0.6,\ \sigma_0^2 = 2$	0.131	0.429	0.852	0.988	0.999	0.134	0.433	0.859	0.989	0.999

[a] $\beta_0 = [0.5, 0.5, 0.5]'$, ς is the level of siginificance, and the number of Monte Carlo repetitions is $5,000$.

unknown distribution of inefficiency and disturbance terms in future research.

Acknowledgement. We are grateful to the editor James LeSage and three anonymous referees for helpful comments that lead to improvements of this paper. Fei Jin gratefully acknowledges the financial support from the National Natural Science Foundation of China (No. 71973030 and No. 71833004).

Appendix A Second order derivatives of $\ln L_n(\theta)$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \lambda^2} = -\operatorname{tr}[G_n^2(\lambda)] - \frac{1}{\sigma^2}\sum_{i=1}^n (w_{n,i\cdot}Y_n)^2 + \frac{\delta^2}{\sigma^2}\sum_{i=1}^n (w_{n,i\cdot}Y_n)^2 f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{A.1}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \lambda \partial \beta} = -\frac{1}{\sigma^2}\sum_{i=1}^n w_{n,i\cdot}Y_n x_{ni} + \frac{\delta^2}{\sigma^2}\sum_{i=1}^n w_{n,i\cdot}Y_n x_{ni} f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{A.2}$$

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial \lambda \partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^n w_{n,i\cdot}Y_n \epsilon_{ni}(\lambda,\beta) - \frac{\delta}{2\sigma^3}\sum_{i=1}^n w_{n,i\cdot}Y_n f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big) \\ &\quad + \frac{\delta^2}{2\sigma^4}\sum_{i=1}^n w_{n,i\cdot}Y_n \epsilon_{ni}(\lambda,\beta) f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big),\end{aligned} \tag{A.3}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \lambda \partial \delta} = \frac{1}{\sigma}\sum_{i=1}^n w_{n,i\cdot}Y_n f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big) - \frac{\delta}{\sigma^2}\sum_{i=1}^n w_{n,i\cdot}Y_n \epsilon_{ni}(\lambda,\beta) f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{A.4}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \beta \partial \beta'} = -\frac{1}{\sigma^2}\sum_{i=1}^n x_{ni}x_{ni}' + \frac{\delta^2}{\sigma^2}\sum_{i=1}^n x_{ni}x_{ni}' f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{A.5}$$

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial \beta \partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^n x_{ni}\epsilon_{ni}(\lambda,\beta) - \frac{\delta}{2\sigma^3}\sum_{i=1}^n x_{ni} f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big) \\ &\quad + \frac{\delta^2}{2\sigma^4}\sum_{i=1}^n x_{ni}\epsilon_{ni}(\lambda,\beta) f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big),\end{aligned} \tag{A.6}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \beta \partial \delta} = \frac{1}{\sigma}\sum_{i=1}^n x_{ni} f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big) - \frac{\delta}{\sigma^2}\sum_{i=1}^n x_{ni}\epsilon_{ni}(\lambda,\beta) f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{A.7}$$

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial^2 \sigma^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n \epsilon_{ni}^2(\lambda,\beta) - \frac{3\delta}{4\sigma^5}\sum_{i=1}^n \epsilon_{ni}(\lambda,\beta) f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big) \\ &\quad + \frac{\delta^2}{4\sigma^6}\sum_{i=1}^n \epsilon_{ni}^2(\lambda,\beta) f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big),\end{aligned} \tag{A.8}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \sigma^2 \partial \delta} = \frac{1}{2\sigma^3}\sum_{i=1}^n \epsilon_{ni}(\lambda,\beta) f\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big) - \frac{\delta}{2\sigma^4}\sum_{i=1}^n \epsilon_{ni}^2(\lambda,\beta) f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{A.9}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \delta^2} = \frac{1}{\sigma^2}\sum_{i=1}^n \epsilon_{ni}^2(\lambda,\beta) f^{(1)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big), \tag{A.10}$$

where $f^{(1)}(t) = \frac{\partial f(t)}{\partial t} = -tf(t) - f^2(t)$.

Appendix B Proofs

For any random variable t with a finite pth absolute moment, where $p \geq 1$, denote its L_p-norm by $\|t\|_p = \mathrm{E}[|t|^p]^{1/p}$. Let $g = \{g_{ni}, i \in D_n, n \geq 1\}$ and $\nu = \{\nu_{ni}, i \in D_n, n \geq 1\}$ be two random fields, where D_n satisfies Assumption 1. Assume that g is uniformly L_p bounded for some $p \geq 1$, i.e., $\sup_{i,n} \|g_{ni}\|_p < \infty$. Let $\mathcal{F}_{ni}(s)$ be the σ-field generated by the random variables ν_{nj}'s with units j's located within the ball $B_i(s)$, where $B_i(s)$ is centered at i with radius s. The random field g is said to be L_p-NED on ν if $\|g_{ni} - \mathrm{E}(g_{ni}|\mathcal{F}_{ni}(s))\|_p \leq d_{ni}\psi(s)$ for some arrays of finite positive constants $\{d_{ni}, i \in D_n, n \geq 1\}$ and for some sequence $\psi(s) \geq 0$ such that $\lim_{s\to\infty} \psi(s) = 0$. The $\psi(s)$ is called the NED coefficient. If we also have $\sup_n \sup_{i\in D_n} d_{ni} < \infty$, g is said to be uniformly L_p-NED on ν.

The results in the following lemmas are frequently used in subsequent proofs, so we collect them in lemmas. For $j \geq 0$, $f^{(j)}(t)$ denotes the jth derivative of $f(t)$. In the following proofs, c will denote a generic positive constant that may be different in different cases.

Lemma B.1. *Suppose that Assumptions 1–4 and 7 hold. Let Λ and $\mathcal{B}$ be, respectively, the parameter spaces of λ and β.*

(a) If $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p \geq 1$, then $\sup_{i,n} \|y_{ni}\|_p < \infty$, $\sup_{i,n} \|w_{n,i\cdot}Y_n\|_p < \infty$, and $\sup_{i,n,\lambda\in\Lambda,\beta\in\mathcal{B}} \|\epsilon_{ni}(\lambda, \beta)\|_p < \infty$.

(b) If $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p \geq 2$, $\sup_{i,n,\lambda\in\Lambda,\beta\in\mathcal{B}} \|\ln \Phi(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda, \beta))\|_{p/2} < \infty$; if $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p \geq j + 1$ with $j \geq 0$,

$$\sup_{i,n,\lambda\in\Lambda,\beta\in\mathcal{B}} \big\|f^{(j)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda, \beta)\big)\big\|_{p/(j+1)} < \infty.$$

Proof. (a) The reduced form of y_{ni} is $y_{ni} = \sum_{j=1}^n t_{n,ij}(x'_{nj}\beta_0 + v_{nj} - u_{nj})$, where $t_{n,ij}$ is the (i, j)th element of $(I_n - \lambda_0 W_n)^{-1}$. For any matrix $A = [a_{ij}]$, let $abs(A) = [|a_{ij}|]$. Note that $abs((I_n - \lambda_0 W_n)^{-1}) \leq^* (I_n - abs(\lambda_0 W_n))^{-1}$, where $B \leq^* C$ for conformable matrices $B = [b_{ij}]$ and $C = [c_{ij}]$ means that $b_{ij} \leq c_{ij}$ for all i, j. Denote $M_n = (I_n - abs(\lambda_0 W_n))^{-1} = [m_{n,ij}]$. Then $\sup_{i,n} \|y_{ni}\|_p \leq \sup_{i,n} \sum_{j=1}^n m_{n,ij}(\sum_{k=1}^{k_x} \|x_{nj,k}\|_p|\beta_{0k}| + \|v_{nj}\|_p + \|u_{nj}\|_p)$ by the Minkowski inequality. As $\lambda_m \sup_n \|W_n\|_\infty < \infty$, $\sup_{1\leq k\leq k_x,j,n} \|x_{nj,k}\|_p < \infty$, $\sup_{j,n} \|v_{nj}\|_p < \infty$ and $\sup_{j,n} \|u_{nj}\|_p < \infty$, we have $\sup_{i,n} \|y_{ni}\|_p < \infty$. So is $\{w_{n,i\cdot}Y_n\}_{i=1}^n$. As $\epsilon_{ni}(\lambda, \beta) = y_{ni} - \lambda w_{n,i\cdot}Y_n - x'_{ni}\beta$, by the Minkowski inequality, $\sup_{i,n,\lambda\in\Lambda,\beta\in\mathcal{B}} \|\epsilon_{ni}(\lambda, \beta)\|_p < \infty$.

(b) By the proof of Lemma A.9 in Xu and Lee (2015), $|\ln \Phi(t)| \leq c(t^2 + |t| + 1)$ and $|f(t)| \leq 2|t| + c$. Then the first result follows by the Minkowski inequality. Since $f^{(1)}(t) = -tf(t) - f^2(t)$, $f^{(j)}(t)$ for $j > 1$ can be derived recursively and it can be regarded as a $(j + 1)$-order polynomial function of $[t, f(t)]$. Thus, $|f^{(j)}(t)| \leq c(|t|^{j+1} + \cdots + 1)$. With $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p \geq j+1$, the second result follows by the Minkowski inequality and (a). $\square$

Lemma B.2. *Suppose that Assumptions 1–4 and 7 hold. Let Λ and $\mathcal{B}$ be, respectively, the parameter spaces of λ and β.*

(a) If $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p \geq 2$, $\{y_{ni}\}_{i=1}^n$ is uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$ with NED coefficient $\psi(s) = c_1^{s/d_0}$ under Assumption 3(a), where c_1 is defined in Assumption 2; and $\psi(s) = s^{-(\alpha-d)}$ under Assumption 3(b). The same holds for $\{w_{n,i\cdot}Y_n\}_{i=1}^n$ and $\{\epsilon_{ni}(\lambda, \beta)\}_{i=1}^n$.

(b) If $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p > 4$, then $\{w_{n,i\cdot}Y_n\epsilon_{ni}\}_{i=1}^n$, $\{w_{n,i\cdot}Y_n x_{ni,j}\}_{i=1}^n$ and $\{w_{n,i\cdot}Y_n f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})\}_{i=1}^n$ are uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$ with NED coefficient $\psi(s) = c_1^{s(p-4)/[d_0(2p-4)]}$ under Assumption 3(a), and $\psi(s) = s^{-(\alpha-d)(p-4)/(2p-4)}$ under Assumption 3(b).

(c) If $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p > 4$, $\{\epsilon_{ni}^2(\lambda, \beta)\}_{i=1}^n$ and $\{\ln \Phi(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda, \beta))\}_{i=1}^n$ are uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$; if $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p > 2j$, $\{(w_{n,i\cdot}Y_n)^j\}_{i=1}^n$ is uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$; if $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$ for some $p > 6$, $\{f(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda, \beta))\}_{i=1}^n$ is uniformly L_2-NED.

Proof. (a) By Proposition 1 in Jenish and Prucha (2012), $\{y_{ni}\}_{i=1}^n$ is uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$ if $\lim_{s\to\infty} \sup_{i,n} \sum_{j:d(i,j)>s} m_{n,ij} = 0$ and $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_2 < \infty$. By the proof of Proposition 1 in Xu and Lee (2015), $\sup_{i,n} \sum_{j:d(i,j)>s} m_{n,ij} \leq c c_1^{s/d_0}$ under Assumption 3(a), and $\sup_{i,n} \sum_{j:d(i,j)>s} m_{n,ij} \leq cs^{-(\alpha-d)}$ under Assumption 3(b). Thus $\{y_{ni}\}_{i=1}^n$ is uniformlyL_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$ with NED coefficient $\psi(s) = c_1^{s/d_0}$ under Assumption 3(a) and $\psi(s) = s^{-(\alpha-d)}$ under Assumption 3(b). With the NED property of $\{y_{ni}\}_{i=1}^n$, by the proof of Proposition 1 in Xu and Lee (2015), $\{w_{n,i\cdot}Y_n\}_{i=1}^n$ is also uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$ with the same NED coefficient as that of $\{y_{ni}\}_{i=1}^n$. The same holds for $\epsilon_{ni}(\lambda, \beta)$ as $\epsilon_{ni}(\lambda, \beta) = y_{ni} - \lambda w_{n,i\cdot}Y_n - x_{ni}'\beta$ is linear in y_{ni}, $w_{n,i\cdot}Y_n$ and x_{ni}.

(b) With the NED property of $\{w_{n,i\cdot}Y_n\}_{i=1}^n$ in (a), the results directly follow by Lemma A.2 in Xu and Lee (2015).

(c) By the mean value theorem, $|t_1^j - t_2^j| \leq j|\bar{t}|^{j-1} \cdot |t_1 - t_2| \leq j(|t_1|^{j-1} + |t_2|^{j-1}) \cdot |t_1 - t_2|$, $|\ln \Phi(t_1) - \ln \Phi(t_2)| \leq |f(\dot{t})| \cdot |t_1 - t_2| \leq (2|t_1| + 2|t_2| + c)|t_1 - t_2|$ and $|f(t_1) - f(t_2)| \leq |f^{(1)}(\ddot{t})| \cdot |t_1 - t_2| \leq c(t_1^2 + t_2^2 + 1) \cdot |t_1 - t_2|$, where $\bar{t}$, $\dot{t}$ and $\ddot{t}$ are between t_1 and t_2. By Lemma B.1(a), if $\sup_{1\leq k\leq k_x,i,n} \|x_{ni,k}\|_p < \infty$, then $\sup_{i,n} \|\epsilon_{ni}(\lambda, \beta)\|_p \leq \infty$ and $\sup_{i,n} \|w_{n,i\cdot}Y_n\|_p \leq \infty$. The NED results in the lemma on functions of $\epsilon_{ni}(\lambda, \beta)$ and $w_{n,i\cdot}Y_n$ then follow by Lemma A.4 in Xu and Lee (2015). $\square$

Proof of Proposition 2.1. We first prove the uniform convergence of $\ln L_n(\theta)$ that $\sup_{\theta\in\Theta} \frac{1}{n}| \ln L_n(\theta)-$

$\mathrm{E}[\ln L_n(\theta)]| = o_p(1)$. Note that

$$\frac{1}{n}[\ln L_n(\theta) - \mathrm{E}[\ln L_n(\theta)]] = a_{1n}(\theta) + a_{2n}(\theta),$$

where $a_{1n}(\theta) = -\frac{1}{2n\sigma^2}\sum_{i=1}^n\{\epsilon_{ni}^2(\lambda,\beta) - \mathrm{E}[\epsilon_{ni}^2(\lambda,\beta)]\}$ and $a_{2n}(\theta) = \frac{1}{n}\sum_{i=1}^n\big\{\ln\Phi\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big) - \mathrm{E}\big[\ln\Phi\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big)\big]\big\}$. We shall prove that $\sup_{\theta\in\Theta}|a_{1n}(\theta)| = o_p(1)$ and $\sup_{\theta\in\Theta}|a_{2n}(\theta)| = o_p(1)$. Under Assumptions 1–5, by Lemma B.1, $\{\epsilon_{ni}^2(\lambda,\beta)\}_{i=1}^n$ and $\{\ln\Phi(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta))\}_{i=1}^n$ are uniformly $L_{2+\iota/2}$ bounded; by Lemma B.2(c), they are uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$. Then $a_{1n}(\theta) = o_p(1)$ and $a_{2n}(\theta) = o_p(1)$ by Theorem 1 in Jenish and Prucha (2012). Since $a_{1n}(\theta)$ is quadratic in $[\lambda,\beta']$ and the parameter space of σ^2 is compact, $\sup_{\theta\in\Theta}|a_{1n}(\theta)| = o_p(1)$. For the proof of $\sup_{\theta\in\Theta}|a_{2n}(\theta)| = o_p(1)$, with $a_{2n}(\theta) = o_p(1)$, by Theorem 1 in Andrews (1992), we only need to prove that $a_{2n}(\theta)$ is stochastically equicontinuous (SE). Denote $\dot\lambda = -\frac{\delta}{\sigma}\lambda$, $\dot\beta = -\frac{\delta}{\sigma}\beta$, $\dot\delta = -\frac{\delta}{\sigma}$ and $\dot\theta = [\dot\lambda, \dot\beta', \sigma^2, \dot\delta]'$. Since the parameter space of θ is compact, the elements of $\frac{\partial\dot\lambda}{\partial\theta}$, $\frac{\partial\dot\beta}{\partial\theta'}$ and $\frac{\partial\dot\delta}{\partial\theta}$ are bounded. Then by Lemma A.5 in Xu and Lee (2015), we only need to prove that $\frac{1}{n}\sum_{i=1}^n\{\ln\Phi(\dot\epsilon_{ni}(\dot\theta)) - \mathrm{E}[\ln\Phi(\dot\epsilon_{ni}(\dot\theta)]\}$ is SE, where $\dot\epsilon_{ni}(\dot\theta) = \dot\delta y_{ni} - \dot\lambda w_{n,i\cdot}Y_n - x_{ni}'\dot\beta$. By the proof of Lemma B.2, $|\ln\Phi(t_1) - \ln\Phi(t_2)| \le (2|t_1| + 2|t_2| + c)|t_1 - t_2|$. Then

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^n \ln\Phi(\dot\epsilon_{ni}(\dot\theta_1)) - \frac{1}{n}\sum_{i=1}^n \ln\Phi(\dot\epsilon_{ni}(\dot\theta_2))\Big| \\
&\le \frac{1}{n}\sum_{i=1}^n [2|\dot\epsilon_{ni}(\dot\theta_1)| + 2|\dot\epsilon_{ni}(\dot\theta_2)| + c]\cdot|\dot\epsilon_{ni}(\dot\theta_1) - \dot\epsilon_{ni}(\dot\theta_2)| \\
&\le \frac{1}{n}\sum_{i=1}^n [2|\dot\epsilon_{ni}(\dot\theta_1)| + 2|\dot\epsilon_{ni}(\dot\theta_2)| + c]\Big[|y_{ni}| + |w_{n,i\cdot}Y_n| + \sum_{k=1}^{k_x}|x_{ni,k}|\Big] \\
&\quad \times\Big(|\dot\delta_1 - \dot\delta_2| + |\dot\lambda_1 - \dot\lambda_2| + \sum_{k=1}^{k_x}|\dot\beta_{1k} - \dot\beta_{2k}|\Big).
\end{aligned}$$

Since $\{y_{ni}\}_{i=1}^n$, $\{w_{n,i\cdot}Y_n\}_{i=1}^n$ and $\{x_{ni,k}\}_{i=1}^n$ are uniformly L_4 bounded, each term of $[2|\dot\epsilon_{ni}(\dot\theta_1)| + 2|\dot\epsilon_{ni}(\dot\theta_2)| + c][|y_{ni}| + |w_{n,i\cdot}Y_n| + \sum_{k=1}^{k_x}|x_{ni,k}|]$ is uniformly L_2 bounded by the Cauchy-Schwarz inequality. It follows that $\frac{1}{n}\sum_{i=1}^n \mathrm{E}[\ln\Phi(\dot\epsilon_{ni}(\dot\theta)]$ is equicontinuous. Thus, $\frac{1}{n}\sum_{i=1}^n\{\ln\Phi(\dot\epsilon_{ni}(\dot\theta)) - \mathrm{E}[\ln\Phi(\dot\epsilon_{ni}(\dot\theta)]\}$ is SE by Lemma 1(a) in Andrews (1992). As the parameter space of σ^2 is compact and $\epsilon_{ni}(\lambda,\beta)$ is linear in θ, $-\frac{1}{2n\sigma^2}\sum_{i=1}^n \mathrm{E}[\epsilon_{ni}^2(\lambda,\beta)]$ is equicontinuous. It follows that $\frac{1}{n}\mathrm{E}[\ln L_n(\theta)]$ is equicontinuous.

With the identification condition in Assumption 6, the uniform convergence $\sup_{\theta\in\Theta}\frac{1}{n}|\ln L_n(\theta) - \mathrm{E}[\ln L_n(\theta)]| = o_p(1)$ and the equicontinuity of $\frac{1}{n}\mathrm{E}[\ln L_n(\theta)]$, the consistency of the MLE follows. □

Proof of Proposition 2.2. By the mean value theorem, $0 = \frac{\partial\ln L_n(\hat\theta)}{\partial\theta} = \frac{\partial\ln L_n(\theta_0)}{\partial\theta} + \frac{\partial^2\ln L_n(\bar\theta)}{\partial\theta\partial\theta'}(\hat\theta - \theta_0)$, where $\bar{\theta}$ lies between $\hat{\theta}$ and θ_0. Then

$$\sqrt{n}(\hat{\theta}-\theta_0) = \Big(-\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta})}{\partial\theta\partial\theta'}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}.$$

We first prove the asymptotic normality of $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}$. The elements of $\frac{\partial \ln L_n(\theta_0)}{\partial\theta}$ have the following forms: $c\sum_{i=1}^n w_{n,i\cdot}Y_n\epsilon_{ni}$, $c\sum_{i=1}^n w_{n,i\cdot}Y_n f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})$, $c\sum_{i=1}^n x_{ni}\epsilon_{ni}$, $c\sum_{i=1}^n x_{ni}f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})$, $c\sum_{i=1}^n \epsilon_{ni}^2$, $c\sum_{i=1}^n \epsilon_{ni}f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})$ and c. By Lemma B.1(a) and the Cauchy-Schwarz inequality, $\{w_{n,i\cdot}Y_n\epsilon_{ni}\}_{i=1}^n$ is uniformly $L_{p/2}$ bounded; by Lemma B.2(b), $\{w_{n,i\cdot}Y_n\epsilon_{ni}\}_{i=1}^n$ is uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$ with NED coefficient $\psi(s) = c_1^{s(p-4)/[d_0(2p-4)]}$ under Assumption 3(a) and $\psi(s) = s^{-(\alpha-d)(p-4)/(2p-4)}$ under Assumption 3(b), where $p = 6$ under Assumption 10(a). To apply the CLT in Theorem 2 of Jenish and Prucha (2012) to $\{w_{n,i\cdot}Y_n\epsilon_{ni}\}_{i=1}^n$, $\psi(s)$ should satisfy $\sum_{s=1}^\infty s^{d-1}\psi(s) < \infty$. If $\psi(s) = c_1^{s(p-4)/[d_0(2p-4)]}$, then $\sum_{s=1}^\infty s^{d-1}\psi(s) < \infty$ as $0 < c_1^{(p-4)/[d_0(2p-4)]} < 1$; if $\psi(s) = s^{-(\alpha-d)(p-4)/(2p-4)}$, $\sum_{s=1}^\infty s^{d-1}\psi(s) < \infty$ requires $d-(\alpha-d)(p-4)/(2p-4) < 0$, i.e. $\alpha > (3+\frac{4}{p-4})d$. With $p = 6$, $\alpha > 5d$ is maintained in Assumption 10(b). In addition, for the α-mixing coefficient of $\{x_{ni}\}_{i=1}^n$, $\hat{\alpha}(s)$ should satisfy Assumption 3 in Jenish and Prucha (2012): $\sum_{s=1}^\infty s^{d[1+c_3\iota^*/(2+\iota^*)]-1}[\hat{\alpha}(s)]^{\iota^*/(4+2\iota^*)} < \infty$, where ι^* is some positive number smaller than $p/2-2$ so that Assumption 4(a) in Jenish and Prucha (2012) is satisfied. This condition is maintained in Assumption 10(c). By Lemma B.2, $\{w_{n,i\cdot}Y_n f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})\}_{i=1}^n$ has the same NED property as $\{w_{n,i\cdot}Y_n\epsilon_{ni}\}_{i=1}^n$. The rest of random fields $\{cx_{ni}\epsilon_{ni}\}_{i=1}^n$, $\{cx_{ni}f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})\}_{i=1}^n$, $\{c\epsilon_{ni}^2\}_{i=1}^n$ and $\{c\epsilon_{ni}f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})\}_{i=1}^n$ involved in $\frac{\partial \ln L_n(\theta_0)}{\partial\theta}$ are trivially L_2-NED with $\psi(s) = 0$. Then by the CLT in Theorem 2 of Jenish and Prucha (2012), $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, \lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\frac{\partial \ln L_n(\theta_0)}{\partial\theta'}))$.

As $\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) = -\frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\frac{\partial \ln L_n(\theta_0)}{\partial\theta'})$, it remains to prove that $\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta})}{\partial\theta\partial\theta'} = \mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})+o_p(1)$. We shall prove that (i) $\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta})}{\partial\theta\partial\theta'} = \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}+o_p(1)$ and (ii) $\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} = \mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})+o_p(1)$. To prove (i), we first prove a general result on the order of the jth derivative of $\frac{1}{n}\ln L_n(\theta)$. Except $-\operatorname{tr}(G_n^k(\lambda)) = O(n)$ for $1 \le k \le j$, other terms in the jth derivative of $\ln L_n(\theta)$ have the form $c(\theta)\sum_{i=1}^n h_{ni,1}\cdots h_{ni,j}$ or $c(\theta)\sum_{i=1}^n h_{ni,1}\cdots h_{ni,j}f^{(k)}(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta))$, where $c(\theta)$ is a function of θ, $0 \le k \le j-1$, and $h_{ni,r}$ for $1 \le r \le j$ is either 1, $\epsilon_{ni}(\lambda,\beta)$, $w_{n,i\cdot}Y_n$, or an element of x_{ni}. By Lemma B.1(b), $\sup_{i,n,\lambda\in\Lambda,\beta\in\mathcal{B}}\|f^{(k)}(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta))\|_{p/(k+1)} < \infty$ if $\sup_{i,k,n}\|x_{ni,k}\|_p < \infty$ for $p \ge k+1$. If $\sup_{i,r,n}\mathrm{E}(|h_{ni,r}|^{2j}) < \infty$, then $\sup_{i,n}\mathrm{E}[|h_{ni,1}\cdots h_{ni,j}|] \le \sup_{i,n}\|h_{ni,1}\|_j\cdots\|h_{ni,j}\|_j < \infty$ and

$$\begin{aligned}&\sup_{i,n,\lambda\in\Lambda,\beta\in\mathcal{B}}\mathrm{E}\big[\big|h_{ni,1}\cdots h_{ni,j}f^{(k)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big)\big|\big]\\ &\le \sup_{i,n,\lambda\in\Lambda,\beta\in\mathcal{B}}\|h_{ni,1}\|_{2j}\cdots\|h_{ni,j}\|_{2j}\big\|f^{(k)}\big(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\big)\big\|_2 < \infty\end{aligned}$$

for $1 \le k \le j-1$ by the generalized Hölder's inequality and Lemma B.1(a). Thus, if

$$\sup_{1\le k\le k_x,i,n}\mathrm{E}(|x_{ni,k}|^{2j}) < \infty,$$

the jth derivative of $\frac{1}{n}\ln L_n(\theta)$ is $O_p(1)$. In particular, if $\sup_{1\le k\le k_x,i,n}\mathrm{E}(|x_{ni,k}|^6)<\infty$, the third derivative of $\frac{1}{n}\ln L_n(\theta)$ is $O_p(1)$. Hence, (i) holds by the mean value theorem under Assumption 10(a).

We next prove (ii). By the above argument, except $-\operatorname{tr}(G_n^2)$, elements of $\frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'}$ have the form $c\sum_{i=1}^n h_{ni,1}h_{ni,2}$, $c\sum_{i=1}^n h_{ni,1}h_{ni,2}f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})$ or $c\sum_{i=1}^n h_{ni,1}h_{ni,2}f^{(1)}(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})$, where $h_{ni,r}$ for $r=1,2$ is either 1, ϵ_{ni}, $w_{n,i\cdot}Y_n$, or an element of x_{ni}. By Lemma B.1 and the Cauchy-Schwarz inequality, each $\{h_{ni,1}h_{ni,2}\}_{i=1}^n$ is uniformly L_3 bounded; by Lemma B.2, each $\{h_{ni,1}h_{ni,2}\}_{i=1}^n$ is uniformly L_2-NED. By the generalized Hölder's inequality,

$$\big\|h_{ni,1}h_{ni,2}f^{(1)}\big(-\frac{\delta_0}{\sigma_0}\epsilon_{ni}\big)\big\|_{3/2}\le\|h_{ni,1}\|_6\|h_{ni,2}\|_6\big\|f^{(1)}\big(-\frac{\delta_0}{\sigma_0}\epsilon_{ni}\big)\big\|_3.$$

Thus, each $\{h_{ni,1}h_{ni,2}f^{(1)}(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})\}_{i=1}^n$ is uniformly $L_{3/2}$ bounded. Since

$$\begin{aligned}&\Big\|h_{ni,1}h_{ni,2}f^{(1)}\big(-\frac{\delta_0}{\sigma_0}\epsilon_{ni}\big)-\mathrm{E}\big[h_{ni,1}h_{ni,2}f^{(1)}\big(-\frac{\delta_0}{\sigma_0}\epsilon_{ni}\big)\big|\mathcal{F}_{ni}(s)\big]\Big\|\\&=\Big\|[h_{ni,1}h_{ni,2}-\mathrm{E}(h_{ni,1}h_{ni,2}|\mathcal{F}_{ni}(s))]f^{(1)}\big(-\frac{\delta_0}{\sigma_0}\epsilon_{ni}\big)\Big\|\\&\le\|h_{ni,1}h_{ni,2}-\mathrm{E}[h_{ni,1}h_{ni,2}|\mathcal{F}_{ni}(s)]\|_2\Big\|f^{(1)}\big(-\frac{\delta_0}{\sigma_0}\epsilon_{ni}\big)\Big\|_2,\end{aligned}$$

each $\{h_{ni,1}h_{ni,2}f^{(1)}(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})\}_{i=1}^n$ is uniformly L_1-NED. Similarly, each $\{h_{ni,1}h_{ni,2}f(-\frac{\delta_0}{\sigma_0}\epsilon_{ni})\}_{i=1}^n$ is uniformly $L_{3/2}$ bounded and uniformly L_1-NED. Thus, by the LLN in Theorem 1 of Jenish and Prucha (2012), (ii) holds. In consequence, the asymptotic distribution of $\hat\theta$ in the proposition follows. □

Proof of Proposition 2.3. As the proof is sketched in the main text, here we only verify some claims that have not been proved.

We first prove that $\lim_{n\to\infty}\frac{1}{n}\Delta_n$ is positive definite (PD) under Assumption 11. For a block matrix $E=\left(\begin{smallmatrix}A&B\\C&D\end{smallmatrix}\right)$, where A and D are square matrices and D is invertible,

$$\begin{pmatrix}I&-BD^{-1}\\0&I\end{pmatrix}\begin{pmatrix}A&B\\C&D\end{pmatrix}\begin{pmatrix}I&0\\-D^{-1}C&I\end{pmatrix}=\begin{pmatrix}A-BD^{-1}C&0\\0&D\end{pmatrix}.\tag{B.1}$$

If E is symmetric and D is PD, then E is PD when $A-BD^{-1}C$ is PD. Partition Δ_n in (2.17) into a 2×2 block matrix so that the $(2,2)$th block is the scalar $\frac{n}{6\pi}(5-\frac{16}{\pi}+\frac{32}{\pi^2})$, which corresponds to the D block in (B.1). By (B.1), Δ_n is PD if

$$\begin{pmatrix} \frac{1}{\sigma_0^2}\,\mathrm{E}[(G_nX_n\beta_0)'T_n(G_nX_n\beta_0)] + \mathrm{tr}(G_nG_n^{(s)}) & * & * \\ \frac{1}{\sigma_0^2}\,\mathrm{E}(X_n'T_nG_nX_n\beta_0) & \frac{1}{\sigma_0^2}\,\mathrm{E}(X_n'T_nX_n) & * \\ \frac{1}{\sigma_0^2}\,\mathrm{tr}(G_n) & 0 & \frac{n}{2\sigma_0^4} \end{pmatrix} + \frac{3}{n\sigma_0^2(5-\frac{16}{\pi}+\frac{32}{\pi^2})}[\mathrm{E}(\Xi_{1n}'\Xi_{1n}) - \mathrm{E}(\Xi_{1n}')\,\mathrm{E}(\Xi_{1n})] \tag{B.2}$$

is PD, where $T_n = I_n - \frac{3}{n(5-\frac{16}{\pi}+\frac{32}{\pi^2})}l_nl_n'$ is PD and $\Xi_{1n} = [l_n'G_nX_n\beta_0, l_n'X_n, 0]$. As $\mathrm{E}(\Xi_{1n}'\Xi_{1n}) - \mathrm{E}(\Xi_{1n}')\,\mathrm{E}(\Xi_{1n})$ is positive semidefinite, applying (B.1) to the first matrix in (B.2), Δ_n is PD if

$$\begin{pmatrix} \frac{1}{\sigma_0^2}\,\mathrm{E}[(G_nX_n\beta_0)'T_n(G_nX_n\beta_0)] + \mathrm{tr}(G_nG_n^{(s)}) - \frac{2}{n}\,\mathrm{tr}^2(G_n) & * \\ \frac{1}{\sigma_0^2}\,\mathrm{E}(X_n'T_nG_nX_n\beta_0) & \frac{1}{\sigma_0^2}\,\mathrm{E}(X_n'T_nX_n) \end{pmatrix} \tag{B.3}$$

is PD. If $\lim_{n\to\infty}\frac{1}{n}\,\mathrm{E}[(G_nX_n\beta_0, X_n)'T_n(G_nX_n\beta_0, X_n)]$ is PD, since

$$\mathrm{tr}(G_nG_n^{(s)}) - \frac{2}{n}\,\mathrm{tr}^2(G_n) = \frac{1}{2}\,\mathrm{tr}(G_n^{(s)}G_n^{(s)}) - \frac{1}{2n}\,\mathrm{tr}^2(G_n^{(s)}) \geq 0,$$

$\lim_{n\to\infty}\frac{1}{n}\Delta_n$ is PD. Alternatively, if $\mathrm{E}(X_n'T_nX_n)$ is PD, applying (B.1) to (B.3), Δ_n is PD when $\frac{1}{\sigma_0^2}\Xi_{2n} + [\mathrm{tr}(G_nG_n^{(s)}) - \frac{2}{n}\,\mathrm{tr}^2(G_n)] > 0$, where

$$\Xi_{2n} = \mathrm{E}[(G_nX_n\beta_0)'T_n(G_nX_n\beta_0)] - \mathrm{E}[(G_nX_n\beta_0)'T_nX_n][\mathrm{E}(X_n'T_nX_n)]^{-1}\,\mathrm{E}(X_n'T_nG_nX_n\beta_0) \geq 0.$$

Thus, $\lim\limits_{n\to\infty}\frac{1}{n}\Delta_n$ is PD under Assumption 11.

We next prove (2.18). By the mean value theorem, $0 = \frac{\partial \ln L_{4n}(\check{\eta},0)}{\partial \eta^\ddagger} = \frac{\partial \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger} + \frac{\partial^2 \ln L_{4n}(\bar{\eta},0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}}(\check{\eta}-\eta_0)$, where $\bar{\eta}$ lies between $\check{\eta}$ and η_0. Then $\sqrt{n}(\check{\eta}-\eta_0) = -(\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\bar{\eta},0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}})^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger}$. We first prove the asymptotic normality of $\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger}$. As $\frac{\partial \ln L_{4n}(\eta,0)}{\partial \eta^\ddagger} = \frac{\partial \ln L_n(\eta,0)}{\partial \eta}$ is a subvector of $\frac{\partial \ln L_n(\theta_0)}{\partial \theta}$ with $\delta_0 = 0$, by the proof of Proposition 2.2, $\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger} \xrightarrow{d} N(0, \lim_{n\to\infty}(\frac{1}{n}\Delta_{n,11})^{-1})$, and under Assumption 12(a) with $p = 14$, $\alpha > (3+\frac{4}{p-4})d = \frac{17}{5}d$ and $0 < \iota^* < p/2 - 2 = 5$, which are maintained in Assumption 12(b)–(c).

As $\mathrm{E}(\frac{\partial^2 \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}}) = -\Delta_{n,11}$, it remains to prove that $\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\bar{\eta},0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}} = \mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}}) + o_p(1)$. We shall prove that $\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}} = \mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}}) + o_p(1)$ and $\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\check{\eta},0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}} = \frac{1}{n}\frac{\partial \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}} + o_p(1)$. By (A.1)–(A.10), except $-\mathrm{tr}(G_n^2)$, elements of $\frac{\partial^2 \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}}$ have the form $c\sum_{i=1}^n \zeta_{ni}$, where ζ_{ni} is either $(w_{n,i\cdot}Y_n)^2$, $w_{n,i\cdot}Y_nx_{ni,j}$, $w_{n,i\cdot}Y_n\epsilon_{ni}$, $w_{n,i\cdot}Y_n$, $x_{ni,j}x_{ni,k}$, $x_{ni,j}\epsilon_{ni}$ ϵ_{ni}^2, ϵ_{ni} or $x_{ni,j}$. By Lemma B.2 and the Cauchy-Schwarz inequality, each $\{\zeta_{ni}\}_{i=1}^n$ is uniformly L_2 bounded. Furthermore, each $\{\zeta_{ni}\}_{i=1}^n$ is uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$. Thus, $\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}} = \mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}}) + o_p(1)$. As $\frac{\partial^2 \ln L_{4n}(\check{\eta},0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}} = \frac{\partial^2 \ln L_n(\check{\eta},0)}{\partial \eta \partial \eta'}$ and each term for the third derivative of $\ln L_n(\theta)$ is shown to be $O_p(1)$ in the proof of Proposition 2.2, by the mean value theorem, $\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\check{\eta},0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}} = \frac{1}{n}\frac{\partial^2 \ln L_{4n}(\eta_0,0)}{\partial \eta^\ddagger \partial \eta^{\ddagger\prime}} + o_p(1)$.

We further prove that (2.19) holds. Let $z_{ni} = [w_{n,i\cdot}Y_n, x'_{ni}]'$, $\kappa_0 = [\lambda_0, \beta'_0]'$ and $\check{\kappa} = [\check{\lambda}, \check{\beta}']'$. Then $\check{\epsilon}_{ni} = \epsilon_{ni} + z'_{ni}(\kappa_0 - \check{\kappa})$. It follows that

$$\frac{1}{n}\sum_{i=1}^{n}\check{\epsilon}_{ni}^3 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_{ni}^3 + \frac{3}{n}\sum_{i=1}^{n}\epsilon_{ni}^2 z'_{ni}(\kappa_0 - \check{\kappa}) + \frac{3}{n}\sum_{i=1}^{n}\epsilon_{ni}(\kappa_0 - \check{\kappa})' z_{ni} z'_{ni}(\kappa_0 - \check{\kappa})$$
$$+ \frac{1}{n}\sum_{i=1}^{n}(\kappa_0 - \check{\kappa})' z_{ni} z'_{ni}(\kappa_0 - \check{\kappa}) z'_{ni}(\kappa_0 - \check{\kappa}).$$

By the Lindeberg-Lévy CLT, $\frac{1}{n}\sum_{i=1}^{n}\epsilon_{ni}^3 = \mathrm{E}(\epsilon_{ni}^3) + O_p(n^{-1/2}) = O_p(n^{-1/2})$. By Theorem 1 in Jenish and Prucha (2012), $\frac{1}{n}\sum_{i=1}^{n}\epsilon_{ni}^2 z_{ni} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\epsilon_{ni}^2 z_{ni}) + o_p(1) = \frac{\sigma_0^2}{n}\mathrm{E}[(G_n X_n \beta_0, X_n)' l_n] + o_p(1) = O_p(1)$. Our proof above shows that $\frac{1}{n}\sum_{i=1}^{n} z_{ni} z'_{ni} z_{nij} = O_p(1)$ and $\frac{1}{n}\sum_{i=1}^{n} z_{ni} z'_{ni}\epsilon_{ni} = O_p(1)$, where z_{nij} is the jth element of z_{ni}. Hence, (2.19) holds.

We continue to prove that the asymptotic distribution of $\sqrt{n}[\hat{\eta}^{\ddagger\prime}, \hat{\tau}]'$ conditional on $\sum_{i=1}^{n}\check{\epsilon}_{ni}^3 \leq 0$ is $[F_1, F_2, F'_3, F_4, |F_5|]'$. Since $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\check{\epsilon}_{ni}^3$ is asymptotically normal with mean zero, we only need to prove that the MLE $\dot{\omega}$ of $\omega = [\eta^{\ddagger\prime}, \tau]'$ from the log likelihood function $\ln L_{4n}(\omega)$ with no nonnegativity restriction on τ has the asymptotic distribution $N(0, \lim_{n\to\infty}(\frac{1}{n}\Delta_n)^{-1})$. By the mean value theorem, $0 = \frac{\partial \ln L_{4n}(\dot{\omega})}{\partial \omega} = \frac{\partial \ln L_{4n}(\omega_0)}{\partial \omega} + \frac{\partial^2 \ln L_{4n}(\bar{\omega})}{\partial \omega \partial \omega'}(\hat{\omega} - \omega_0)$, where $\bar{\omega}$ lies between $\dot{\omega}$ and ω_0. Then $\sqrt{n}(\dot{\omega} - \omega_0) = (-\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\bar{\omega})}{\partial \omega \partial \omega'})^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\omega_0)}{\partial \omega}$. Compared to $\frac{\partial \ln L_{4n}(\eta_0, 0)}{\partial \eta^{\ddagger}}$, by (2.16), $\frac{\partial \ln L_{4n}(\omega_0)}{\partial \omega}$ has the additional element $\sum_{i=1}^{n}[\frac{1}{6\sigma_0^3}(1 - \frac{4}{\pi})\sqrt{\frac{2}{\pi}}\epsilon_{ni}^3 + \frac{2}{\pi\sigma_0}\sqrt{\frac{2}{\pi}}\epsilon_{ni}]$, which is a sum of i.i.d. elements. Thus, as shown above for $\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\eta_0, 0)}{\partial \eta^{\ddagger}}$, by Theorem 2 in Jenish and Prucha (2012), $\frac{1}{\sqrt{n}}\frac{\partial \ln L_{4n}(\omega_0)}{\partial \omega} \xrightarrow{d} N(0, \lim_{n\to\infty}(\frac{1}{n}\Delta_n)^{-1})$.

For the asymptotic distribution of $\sqrt{n}[\hat{\eta}^{\ddagger\prime}, \hat{\tau}]'$ conditional on $\sum_{i=1}^{n}\check{\epsilon}_{ni}^3 \leq 0$, it remains to prove that $\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\bar{\omega})}{\partial \omega \partial \omega'} = -\mathrm{E}(\frac{1}{n}\frac{\partial \ln L_{4n}(\omega_0)}{\partial \omega}\frac{\partial \ln L_{4n}(\omega_0)}{\partial \omega'}) + o_p(1)$. For that purpose, we shall prove the following:

$$\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\bar{\omega})}{\partial \omega \partial \omega'} = \frac{1}{n}\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial \omega \partial \omega'} + o_p(1), \tag{B.4}$$

$$\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial \omega \partial \omega'} = \mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial \omega \partial \omega'}) + o_p(1), \tag{B.5}$$

$$\mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial \omega \partial \omega'}) = -\mathrm{E}(\frac{1}{n}\frac{\partial \ln L_{4n}(\omega_0)}{\partial \omega}\frac{\partial \ln L_{4n}(\omega_0)}{\partial \omega'}). \tag{B.6}$$

We first prove (B.6). As $\frac{\partial \ln L_{4n}(\omega)}{\partial \tau} = \frac{1}{3}\tau^{-2/3}\frac{\partial \ln L_{3n}(\eta^{\ddagger}, \tau^{1/3})}{\partial \delta}$,

$$\frac{\partial^2 \ln L_{4n}(\omega)}{\partial \tau^2} = \frac{1}{9}\tau^{-4/3}\frac{\partial^2 \ln L_{3n}(\eta^{\ddagger}, \tau^{1/3})}{\partial \delta^2} - \frac{2}{9}\tau^{-5/3}\frac{\partial \ln L_{3n}(\eta^{\ddagger}, \tau^{1/3})}{\partial \delta}.$$

Thus, by L'Hôpital's rule,

$$\frac{\partial^2 \ln L_{4n}(\eta^{\ddagger}, 0)}{\partial \tau^2} = \lim_{\delta\to 0}(\frac{1}{9\delta^4}\frac{\partial^2 \ln L_{3n}(\eta^{\ddagger}, \delta)}{\partial \delta^2} - \frac{2}{9\delta^5}\frac{\partial \ln L_{3n}(\eta^{\ddagger}, \delta)}{\partial \delta})$$

$$= \lim_{\delta\to 0} \frac{1}{9\delta^5}\big(\delta\frac{\partial^2 \ln L_{3n}(\eta^\ddagger,\delta)}{\partial\delta^2} - 2\frac{\partial \ln L_{3n}(\eta^\ddagger,\delta)}{\partial\delta}\big)$$
$$= \lim_{\delta\to 0} \frac{1}{45\delta^4}\big(\delta\frac{\partial^3 \ln L_{3n}(\eta^\ddagger,\delta)}{\partial\delta^3} - \frac{\partial^2 \ln L_{3n}(\eta^\ddagger,\delta)}{\partial\delta^2}\big)$$
$$= \lim_{\delta\to 0} \frac{1}{180\delta^2}\frac{\partial^4 \ln L_{3n}(\eta^\ddagger,\delta)}{\partial\delta^4}$$
$$= \frac{1}{360}\frac{\partial^6 \ln L_{3n}(\eta^\ddagger,0)}{\partial\delta^6},$$

and

$$\frac{\partial^2 \ln L_{4n}(\eta^\ddagger,\delta)}{\partial\tau\partial\eta^\ddagger} = \lim_{\delta\to 0}\frac{1}{3\delta^2}\frac{\partial^2 \ln L_{3n}(\eta^\ddagger,\delta)}{\partial\delta\partial\eta^\ddagger} = \frac{1}{6}\frac{\partial^4 \ln L_{3n}(\eta^\ddagger,0)}{\partial\delta^3\partial\eta^\ddagger}.$$

As $\frac{\partial \ln L_{3n}(\eta^\ddagger,0)}{\partial\delta} = \frac{\partial^2 \ln L_{3n}(\eta^\ddagger,0)}{\partial\delta^2} = 0$, by Lemma 1 in Rotnitzky et al. (2000),

$$\frac{\partial^6 \ln L_{3n}(\omega_0)}{\partial\delta^6} = \frac{1}{L_{3n}(\omega_0)}\frac{\partial^6 L_{3n}(\omega_0)}{\partial\delta^6} - \frac{6!}{2\times(3!)^2}\big(\frac{\partial^3 \ln L_{3n}(\omega_0)}{\partial\delta^3}\big)^2$$

and $\frac{\partial^4 \ln L_{3n}(\omega_0)}{\partial\delta^3\partial\eta^\ddagger} = \frac{1}{L_{3n}(\omega_0)}\frac{\partial^4 L_{3n}(\omega_0)}{\partial\delta^3\partial\eta^\ddagger} - \frac{\partial^3 \ln L_{3n}(\omega_0)}{\partial\delta^3}\frac{\partial \ln L_{3n}(\omega_0)}{\partial\eta^\ddagger}$. As $\frac{\partial \ln L_{4n}(\omega_0)}{\partial\tau} = \frac{1}{6}\frac{\partial^3 \ln L_{3n}(\omega_0)}{\partial\delta^3}$ and $\frac{\partial \ln L_{4n}(\omega_0)}{\partial\eta^\ddagger} = \frac{\partial \ln L_{3n}(\omega_0)}{\partial\eta^\ddagger}$ by Proposition 3 in Lee (1993), we have

$$\mathrm{E}\big(\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial\tau^2}\big) = -\,\mathrm{E}\big[\big(\frac{\partial \ln L_{4n}(\omega_0)}{\partial\tau}\big)^2\big] = -\frac{1}{36}\,\mathrm{E}\big[\big(\frac{\partial^3 \ln L_{3n}(\omega_0)}{\partial\delta^3}\big)^2\big]$$

and $\mathrm{E}\big(\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial\tau\partial\eta^\ddagger}\big) = -\,\mathrm{E}\big(\frac{\partial \ln L_{4n}(\omega_0)}{\partial\tau}\frac{\partial \ln L_{4n}(\omega_0)}{\partial\eta^\ddagger}\big) = -\frac{1}{6}\,\mathrm{E}\big(\frac{\partial^3 \ln L_{3n}(\omega_0)}{\partial\delta^3}\frac{\partial \ln L_{3n}(\omega_0)}{\partial\eta^\ddagger}\big)$. Hence, (B.6) holds.

We next prove (B.4). As shown above, $\frac{\partial^2 \ln L_{4n}(\omega)}{\partial\omega\partial\omega'}$ as $\tau\to 0$ involves the sixth order derivatives of $\ln L_n(\theta)$. Under Assumption 12, by the proof of Proposition 2.2, each term in the seventh order derivatives of $\frac{1}{n}\ln L_n(\theta)$ is $O_p(1)$. Then by the mean value theorem, $\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\bar\omega)}{\partial\omega\partial\omega'} = \frac{1}{n}\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial\omega\partial\omega'} + o_p(1)$.

For (B.5), note that except $-\operatorname{tr}(G_n^2)$, each element of $\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial\omega\partial\omega'}$ has the form

$$c\sum_{i=1}^n (w_{n,i\cdot}Y_n)^j h_{ni,1}\cdots h_{ni,k},$$

where $0\le j\le 6$, $j+k\le 6$, and $h_{ni,r}$ for $1\le r\le k$ is either ϵ_{ni} or an element of x_{ni}. We shall prove that each $\{(w_{n,i\cdot}Y_n)^j h_{ni,1}\cdots h_{ni,k}\}_{i=1}^n$ with $1\le j\le 6$ and $j+k\le 6$ is uniformly L_2-NED on $\{x_{ni},v_{ni},u_{ni}\}_{i=1}^n$ under Assumption 12. By Lemma B.2(c), since $\sup_{1\le k\le k_x,i,n}\|x_{ni,k}\|_{14}<\infty$, $\{(w_{n,i\cdot}Y_n)^j\}_{i=1}^n$ is uniformly L_2-NED on $\{x_{ni},v_{ni},u_{ni}\}_{i=1}^n$ for $1\le j\le 6$. Then for $1\le j\le 3$ and $1\le k\le 3$, $\{(w_{n,i\cdot}Y_n)^j h_{ni,1}\cdots h_{ni,k}\}_{i=1}^n$ is uniformly L_2-NED on $\{x_{ni},v_{ni},u_{ni}\}_{i=1}^n$ by Lemma A.2 in Xu and Lee (2015), since $\sup_{i,n}\|(w_{n,i\cdot}Y_n)^j\|_{14/3}<\infty$ and $\sup_{i,n}\|h_{ni,1}\cdots h_{ni,k}\|_{14/3}<\infty$. Similarly, $\{w_{n,i\cdot}Y_n h_{ni,1}\cdots h_{ni,k}\}_{i=1}^n$ for $k=4$ or 5, $\{(w_{n,i\cdot}Y_n)^2 h_{ni,1}\cdots h_{ni,4}\}_{i=1}^n$, $\{(w_{n,i\cdot}Y_n)^4 h_{ni,1}\cdots h_{ni,k}\}_{i=1}^n$ for $k=1$ or 2, and $\{(w_{n,i\cdot}Y_n)^5 h_{ni,1}\}_{i=1}^n$ are all uniformly L_2-NED on $\{x_{ni},v_{ni},u_{ni}\}_{i=1}^n$, since the random variables in these random fields can be written as, respectively, $[w_{n,i\cdot}Y_n h_{ni,1}h_{ni,2}]\cdot[h_{ni,3}\cdots h_{ni,k}]$,

$[(w_{n,i\cdot}Y_n)^2 h_{ni,1}] \cdot [h_{ni,2}h_{ni,3}h_{ni,4}]$, $[(w_{n,i\cdot}Y_n)^3] \cdot [w_{n,i\cdot}Y_n h_{ni,1} \cdots h_{ni,k}]$, and $[(w_{n,i\cdot}Y_n)^3] \cdot [(w_{n,i\cdot}Y_n)^2 h_{ni,1}]$, where each term in the square brackets is $L_{14/3}$ bounded uniformly in i and n by the generalized Hölder's inequality and uniformly L_2-NED. Hence, each $\{(w_{n,i\cdot}Y_n)^j h_{ni,1} \cdots h_{ni,k}\}_{i=1}^n$ with $1 \le j \le 6$ and $j + k \le 6$ is uniformly L_2-NED on $\{x_{ni}, v_{ni}, u_{ni}\}_{i=1}^n$. It follows that $\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial\omega\partial\omega'} = \mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln L_{4n}(\omega_0)}{\partial\omega\partial\omega'}) + o_p(1)$. □

Proof of Proposition 2.4. The asymptotic distribution of the score test statistic follows by using (2.19), so it remains only for us to prove the asymptotic distribution of the LR test statistic. Denote $\omega = [\eta^{\ddagger\prime}, \tau]'$ and $\breve{\omega} = [\breve{\eta}', 0]'$. When $\sum_{i=1}^n \breve{\epsilon}_{ni}^3 < 0$, $\sqrt{n}(\hat{\omega} - \omega_0) = (\frac{1}{n}\Delta_n)^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{4n}(\omega_0)}{\partial\omega} + o_p(1)$. As $\sqrt{n}(\breve{\eta} - \eta_0) = (\frac{1}{n}\Delta_{n,11})^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{4n}(\omega_0)}{\partial\eta^\ddagger} + o_p(1)$, $\sqrt{n}(\breve{\omega} - \hat{\omega}) = \Xi_{3n}\frac{1}{\sqrt{n}}\frac{\partial L_{4n}(\omega_0)}{\partial\omega} + o_p(1)$ when $\sum_{i=1}^n \breve{\epsilon}_{ni}^3 < 0$, where $\Xi_{3n} = (\frac{1}{n}\Delta_n)^{-1} - \begin{pmatrix} (\frac{1}{n}\Delta_{n,11})^{-1} & 0 \\ 0 & 0 \end{pmatrix}$. By a first order Taylor expansion,

$$
\begin{aligned}
2[\ln L_n(\hat{\theta}) - \ln L_n(\breve{\eta}, 0)] &= -2[\ln L_{4n}(\breve{\omega}) - \ln L_{4n}(\hat{\omega})] \\
&= -\sqrt{n}(\breve{\omega} - \hat{\omega})'\frac{1}{n}\frac{\partial^2 L_{4n}(\bar{\omega})}{\partial\omega\partial\omega'}\sqrt{n}(\breve{\omega} - \hat{\omega}) \\
&= (\frac{1}{\sqrt{n}}\frac{\partial L_{4n}(\omega_0)}{\partial\omega'})\Xi_{3n}(\frac{1}{n}\Delta_n)\Xi_{3n}(\frac{1}{\sqrt{n}}\frac{\partial L_{4n}(\omega_0)}{\partial\omega})I\Big(\sum_{i=1}^n \breve{\epsilon}_{ni}^3 < 0\Big) + o_p(1) \\
&= [(\frac{1}{n}\Delta_n)^{-1/2}\frac{1}{\sqrt{n}}\frac{\partial L_{4n}(\omega_0)}{\partial\omega}]'(\frac{1}{n}\Delta_n)^{1/2}\Xi_{3n}(\frac{1}{n}\Delta_n)\Xi_{3n}(\frac{1}{n}\Delta_n)^{1/2} \\
&\quad \cdot [(\frac{1}{n}\Delta_n)^{-1/2}\frac{1}{\sqrt{n}}\frac{\partial L_{4n}(\omega_0)}{\partial\omega}]I\Big(\sum_{i=1}^n \breve{\epsilon}_{ni}^3 < 0\Big) + o_p(1),
\end{aligned}
$$

where $\bar{\omega}$ lies between $\breve{\omega}$ and $\hat{\omega}$, and $(\frac{1}{n}\Delta_n)^{-1/2}\frac{1}{\sqrt{n}}\frac{\partial L_{4n}(\omega_0)}{\partial\omega} \xrightarrow{d} N(0, I_{k_x+3})$. Partition Δ_n into a 2×2 block matrix such that $\Delta_n = \begin{pmatrix} \Delta_{n,11} & \Delta_{n,12} \\ \Delta_{n,21} & \Delta_{n,22} \end{pmatrix}$. It can be shown by the block matrix inverse formula that

$$
\Xi_{3n}(\frac{1}{n}\Delta_n)\Xi_{3n} = [-\Delta_{n,21}\Delta_{n,11}^{-1}, I_{k_x+2}]'(\frac{1}{n}\Delta_{n,22} - \frac{1}{n}\Delta_{n,21}\Delta_{n,11}^{-1}\Delta_{n,12})^{-1}[-\Delta_{n,21}\Delta_{n,11}^{-1}, I_{k_x+2}],
$$

and $\frac{1}{n}\Delta_{n,22} - \frac{1}{n}\Delta_{n,21}\Delta_{n,11}^{-1}\Delta_{n,12} = [-\Delta_{n,21}\Delta_{n,11}^{-1}, I_{k_x+2}]\frac{1}{n}\Delta_n[-\Delta_{n,21}\Delta_{n,11}^{-1}, I_{k_x+2}]'$. Thus,

$$
(\frac{1}{n}\Delta_n)^{1/2}\Xi_{3n}(\frac{1}{n}\Delta_n)\Xi_{3n}(\frac{1}{n}\Delta_n)^{1/2}
$$

is a projection matrix with rank being 1. Hence, $2[\ln L_n(\hat{\theta}) - \ln L_n(\breve{\eta}, 0)] \xrightarrow{d} \chi^2(0) \cdot I(K \ge 0) + \chi^2(1) \cdot I(K < 0)$. □

Proof of Proposition 2.5. Note that $\mathrm{E}(\epsilon_{ni}) = -\sigma_{u0}\sqrt{\frac{2}{\pi}}$. Denote $\kappa_a = [\lambda_0, \beta_{10} - \sigma_{u0}\sqrt{\frac{2}{\pi}}, \beta_{20}']'$. The 2SLS estimator $\tilde{\kappa}$ satisfies

$$
\tilde{\kappa} = (Z_n'P_nZ_n)^{-1}Z_n'P_n(Z_n\kappa_0 + \epsilon_n) = \kappa_a + (Z_n'P_nZ_n)^{-1}Z_n'P_n[\epsilon_n - \mathrm{E}(\epsilon_n)]
$$

$$= \kappa_a + \big[\frac{1}{n}Z_n' Q_n\big(\frac{1}{n}Q_n' Q_n\big)^{-1}\frac{1}{n}Q_n' Z_n\big]^{-1}\frac{1}{n}Z_n' Q_n\big(\frac{1}{n}Q_n' Q_n\big)^{-1}\frac{1}{n}Q_n'[\epsilon_n - \mathrm{E}(\epsilon_n)],$$

where $\frac{1}{n}Q_n' Z_n = \frac{1}{n}Q_n'[G_n(X_n\beta_0 + \mathrm{E}(\epsilon_n)), X_n] + \frac{1}{n}Q_n' G_n[\epsilon_n - \mathrm{E}(\epsilon_n), 0]$. Thus, under Assumption 13, $\tilde{\kappa} = \kappa_a + O_p(n^{-1/2})$. As $\tilde{\epsilon}_{ni} = y_{ni} - z_{ni}'\kappa_a + z_{ni}'(\kappa_a - \tilde{\kappa}) = \zeta_{ni} + z_{ni}'(\kappa_a - \tilde{\kappa})$, where $z_{ni} = [w_{n,i\cdot}Y_n, x_{ni}']'$ and $\zeta_{ni} = v_{ni} - (u_{ni} - \sigma_{u0}\sqrt{\frac{2}{\pi}})$, it follows that

$$\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3 = \frac{1}{n}\sum_{i=1}^n \zeta_{ni}^3 + \frac{3}{n}\sum_{i=1}^n \zeta_{ni}^2 z_{ni}'(\kappa_a - \tilde{\kappa}) + \frac{3}{n}\sum_{i=1}^n \zeta_{ni}(\kappa_a - \tilde{\kappa})' z_{ni} z_{ni}'(\kappa_a - \tilde{\kappa})$$
$$+ \frac{1}{n}\sum_{i=1}^n (\kappa_a - \tilde{\kappa})' z_{ni} z_{ni}'(\tilde{\kappa} - \kappa_a) z_{ni}'(\kappa_a - \tilde{\kappa}).$$

By the Lindeberg-Lévy CLT, $\frac{1}{n}\sum_{i=1}^n \zeta_{ni}^3 = \mathrm{E}(\zeta_{ni}^3) + O_p(n^{-1/2}) = \frac{(\pi-4)\sigma_{u0}^3}{\pi}\sqrt{\frac{2}{\pi}} + O_p(n^{-1/2})$. For $1 \leq j \leq 3$, let h_{nij} be 1 or an element of z_{ni}. Then by the generalized Hölder's inequality, $\sup_{i,n} \mathrm{E}\,|\zeta_{ni}^k h_{ni1} h_{ni2} h_{ni3}| \leq \sup_{i,n} [\mathrm{E}(\zeta_{ni}^{4k})]^{1/4}[\mathrm{E}(h_{ni1}^4)]^{1/4}[\mathrm{E}(h_{ni2}^4)]^{1/4}[\mathrm{E}(h_{ni3}^4)]^{1/4} < \infty$, where $k = 0, 1$ or 2. Thus, $\frac{1}{n}\sum_{i=1}^n \zeta_{ni}^2 z_{ni}' = O_p(1)$, $\frac{1}{n}\sum_{i=1}^n \zeta_{ni} z_{ni} z_{ni}' = O_p(1)$, and $\frac{1}{n}\sum_{i=1}^n z_{ni} z_{ni}' z_{nij} = O_p(1)$. Hence, $\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3 = \frac{(\pi-4)\sigma_{u0}^3}{\pi}\sqrt{\frac{2}{\pi}} + O_p(n^{-1/2})$. In addition,

$$\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^2 = \frac{1}{n}\sum_{i=1}^n \zeta_{ni}^2 + \frac{2}{n}\sum_{i=1}^n \zeta_{ni} z_{ni}'(\kappa_a - \tilde{\kappa}) + \frac{1}{n}\sum_{i=1}^n (\kappa_a - \tilde{\kappa})' z_{ni} z_{ni}'(\kappa_a - \tilde{\kappa})$$
$$= \frac{\pi - 2}{\pi}\sigma_{u0}^2 + \sigma_{v0}^2 + O_p(n^{-1/2}),$$

since $\frac{1}{n}\sum_{i=1}^n \zeta_{ni}^2 = \frac{\pi-2}{\pi}\sigma_{u0}^2 + \sigma_{v0}^2 + O_p(n^{-1/2})$.

If $\sigma_{u0} \neq 0$, then $\mathrm{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3 = \frac{(\pi-4)\sigma_{u0}^3}{\pi}\sqrt{\frac{2}{\pi}} < 0$. Thus $\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_n^3 < 0$ with probability approaching one and $\tilde{\sigma}_u^2 = \sigma_{u0}^2 + o_p(1)$. For the function $g(t) = t^{2/3}$, by the mean value theorem, $\tilde{\sigma}_u^2 = g(\frac{\pi}{\pi-4}\sqrt{\frac{\pi}{2}}(\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3)) = g(\sigma_{u0}^3) + \frac{2}{3}(\bar{\sigma}_u^3)^{-1/3}[\frac{\pi}{\pi-4}\sqrt{\frac{\pi}{2}}(\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3) - \sigma_{u0}^3]$, where $\bar{\sigma}_u^3$ lies between $\frac{\pi}{\pi-4}\sqrt{\frac{\pi}{2}}(\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3)$ and σ_{u0}^3. Thus $\tilde{\sigma}_u^2 = \sigma_{u0}^2 + O_p(n^{-1/2})$. It follows that $\tilde{\sigma}_v^2 = \sigma_{v0}^2 + O_p(n^{-1/2})$. Similarly, $\tilde{\beta}_{1c} = \beta_{10} + O_p(n^{-1/2})$ by the mean value theorem.

If $\sigma_{u0} = 0$, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\epsilon}_{ni}^3 = \frac{1}{\sqrt{n}}\sum_{i=1}^n \zeta_{ni}^3 - \frac{3}{n}\sum_{i=1}^n \zeta_{ni}^2 z_{ni}'(\frac{1}{n}Z_n' P_n Z_n)^{-1}\frac{1}{\sqrt{n}}Z_n' P_n[\epsilon_n - \mathrm{E}(\epsilon_n)] + o_p(1) = O_p(1),$$

since $\mathrm{E}[(\frac{1}{\sqrt{n}}\sum_{i=1}^n \zeta_{ni}^3)^2] = \mathrm{E}(\zeta_{ni}^6) < \infty$. When $\sum_{i=1}^n \tilde{\epsilon}_n^3 \geq 0$, $\tilde{\sigma}_u^2 = 0$, $\tilde{\delta} = 0$ and $\tilde{\beta}_{1c} - \beta_{10} = O_p(n^{-1/2})$; when $\sum_{i=1}^n \tilde{\epsilon}_n^3 < 0$, $\tilde{\sigma}_u^2 = O_p(n^{-1/3})$, $\tilde{\sigma}_v^2 = \sigma_{v0}^2 + O_p(n^{-1/3})$, $\tilde{\delta} = O_p(n^{-1/6})$, and $\tilde{\beta}_{1c} - \beta_{10} = O_p(n^{-1/6})$. □

References

Adetutu, M., Glass, A.J., Kenjegalieva, K., Sickles, R.C., 2015. The effects of efficiency and TFP growth on pollution in europe: A multistage spatial analysis. Journal of Productivity Analysis 43, 307–326.

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

Andrews, D.W.K., 1992. Generic uniform convergence. Econometric Theory 8, 241–257.

Areal, F.J., Balcombe, K., Tiffin, R., 2012. Integrating spatial dependence into stochastic frontier analysis. Australian Journal of Agricultural and Resource Economics 56, 521–541.

Brehm, S., 2013. Fiscal incentives, public spending, and productivity—county-level evidence from a Chinese province. World Development 46, 92–103.

Carvalho, A., 2018. Efficiency spillovers in Bayesian stochastic frontier models: Application to electricity distribution in New Zealand. Spatial Economic Analysis 13, 171–190.

Cliff, A., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

Cliff, A., Ord, J.K., 1981. Spatial Process: Models and Applications. Pion, London.

Druska, V., Horrace, W.C., 2004. Generalized moments estimation for spatial panel data: Indonesian rice farming. American Journal of Agricultural Economics 86, 185–198.

Fusco, E., Vidoli, F., 2013. Spatial stochastic frontier models: Controlling spatial global and local heterogeneity. International Review of Applied Economics 27, 679–694.

Glass, A., Kenjegalieva, K., Paez-Farrell, J., 2013. Productivity growth decomposition using a spatial autoregressive frontier model. Economics Letters 119, 291–295.

Glass, A., Kenjegalieva, K., Sickles, R.C., 2014. Estimating efficiency spillovers with state level evidence for manufacturing in the US. Economics Letters 123, 154–159.

Glass, A.J., Kenjegalieva, K., Sickles, R.C., 2016. A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers. Journal of Econometrics 190, 289–300.

Greene, W.H., 1990. A Gamma-distributed stochastic frontier model. Journal of Econometrics 46, 141–163.

Horrace, W.C., Wright, I.A., 2019. Stationary points for parametric stochastic frontier models. Forthcoming in the Journal of Business & Economic Statsitics.

Jenish, N., Prucha, I.R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of Econometrics 150, 86–98.

Jenish, N., Prucha, I.R., 2012. On spatial processes and asymptotic inference under near-epoch dependence. Journal of Econometrics 170, 178–190.

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.

Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.

Kelejian, H.H., Robinson, D.P., 1995. Spatial correlation: A suggested alternative to the autoregressive model, in: Anselin, L., Florax, R.J. (Eds.), New directions in spatial econometrics. Berlin: Springer.

Kumbhakar, S.C., Parmeter, C.F., Tsionas, E.G., 2013. A zero inefficiency stochastic frontier model. Journal of Econometrics 172, 66–76.

Lahiri, S.N., 1996. On inconsistency of estimators based on spatial data under infill asymptotics. Sankhyā 58, 403–417.

Lee, L.F., 1993. Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix. Econometric Theory 9, 413–430.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lee, L.F., Chesher, A., 1986. Specification testing when score test statistics are identically zero. Journal of Econometrics 31, 121–149.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman & Hall/CRC.

Mastromarco, C., Serlenga, L., Shin, Y., 2013. Globalisation and technological convergence in the EU. Journal of Productivity Analysis 40, 15–29.

Mastromarco, C., Serlenga, L., Shin, Y., 2016. Modelling technical efficiency in cross sectionally dependent stochastic frontier panels. Journal of Applied Econometrics 31, 281–297.

Meeusen, W., van Den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18, 435–444.

Olson, J.A., Schmidt, P., Waldman, D.M., 1980. A Monte Carlo study of estimators of stochastic frontier production functions. Journal of Econometrics 13, 67–82.

Ord, K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120–126.

Pavlyuk, D., 2011. Application of the spatial stochastic frontier model for analysis of a regional tourism sector. Transport and Telecommunication 12, 28–38.

Pavlyuk, D., 2013. Distinguishing between spatial heterogeneity and inefficiency: Spatial stochastic frontier analysis of European airports. Transport and Telecommunication 14, 29–38.

Rotnitzky, A., Cox, D.R., Bottai, M., Robins, J., 2000. Likelihood-based inference with singular information matrix. Bernoulli 6, 243–284.

Schmidt, A.M., Moreira, A.R.B., Helfand, S.M., Fonseca, T.C.O., 2009. Spatial stochastic frontier models: Accounting for unobserved local determinants of inefficiency. Journal of Productivity Analysis 31, 101–112.

Simar, L., Wilson, P.W., 2010. Inferences from cross-sectional, stochastic frontier models. Econometric Reviews 29, 62–98.

Stevenson, R.E., 1980. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13, 57–66.

Tsionas, E.G., Michaelides, P.G., 2016. A spatial stochastic frontier model with spillovers: Evidence from Italian regions. Scottish Journal of Political Economy 63, 243–257.

Vidolia, F., Cardillob, C., Fuscoc, E., Canello, J., 2016. Spatial nonstationarity in the stochastic frontier model: An application to the Italian wine industry. Regional Science and Urban Economics 61, 153–164.

Waldman, D.M., 1982. A stationary point for the stochastic frontier likelihood. Journal of Econometrics 18, 275–279.

White, H., 1994. Estimation, Inference and Specification Analysis. New York: Cambridge University Press.

Xu, X., Lee, L.F., 2015. Maximum likelihood estimation of a spatial autoregressive Tobit model. Journal of Econometrics 188, 264–280.