A Spatial Dynamic Panel Data Model with Both Time and Individual Fixed Effects

Lung-fei Lee* 
Department of Economics 
The Ohio State University

Jihai Yu 
Department of Economics 
The Ohio State University

July 18, 2007

Abstract

This paper establishes asymptotic properties of quasi-maximum likelihood estimators for spatial dynamic panel data with both time and individual fixed effects when both the number of individuals $n$ and the number of time periods $T$ can be large. Instead of using the direct approach where we estimate both individual effects and time effects directly, we propose a data transformation approach to eliminate the time effects so that the bias of the order $O(n^{-1})$ is avoided. When $T$ is relatively larger than $n$, the estimators are $\sqrt{nT}$ consistent and asymptotically centered normal; when $n$ is asymptotically proportional to $T$, the estimators are $\sqrt{nT}$ consistent and asymptotically normal, but the limit distribution is not centered around 0; when $T$ is relatively smaller than $n$, the estimators are consistent with rate $T$ and have a degenerate limit distribution. We also propose a bias correction for our estimators. We show that when $T$ grows faster than $n^{1/3}$, the correction will asymptotically eliminate the bias and yield a centered confidence interval. This transformation approach has advantage over the direct approach especially when $n$ is relatively small as the direct approach has the bias of order $O(n^{-1})$ remained.

JEL classification: C13; C23

Keywords: Spatial autoregression, Dynamic panels, Fixed effects, Time effects, Quasi-maximum likelihood estimation, Bias correction

*We acknowledge financial support for the research from NSF under Grant No. SES-0519204.
1 Introduction

This paper investigates the properties of maximum likelihood (ML) estimator and quasi-maximum likelihood (QML) estimator for spatial dynamic panel data models with both individual effects and time effects when both the number of individuals $n$ and the number of time periods $T$ can be large.

Recently, there is a growing literature on the estimation of dynamic panel data models when both $n$ and $T$ are large (see Phillips and Moon (1999), Hahn and Kuersteiner (2002), Hahn and Newey (2004), etc). For the panel data with spatial interaction, we have Baltagi, Song and Koh (2003), Kapoor, Kelejian and Prucha (2004) and Yu, de Jong and Lee (2006), etc. Those papers focus on models with only individual effects. While in the panel data literature, we also have the two way error component regression model where we have not only unobservable individual effects but also unobserved time effects (See Wallace and Hussain (1969), Nerlove (1971) and Amemiya (1971), etc). Hence, it is natural to study the spatial dynamic panel data models when both $n$ and $T$ are large with both individual effects and time effects.

In Yu, de Jong and Lee (2006), the consistency and asymptotic distribution of the QML estimator are established when individual effects are included. Also, a bias correction procedure for the estimator is proposed. It is shown that as long as $T$ grows faster than $n^{1/3}$, the correction will asymptotically eliminate the bias of the order $O(T^{-1})$ and yield a centered confidence interval. This approach can be directly extended to the estimation of models with both individual and time fixed effects, where ‘direct’ means we estimate also the fixed time effects jointly with other parameters. When there are also time effects included in the model, we might have additional bias of the order $O(n^{-1})$ in the estimation due to the presence of time effects if we estimate the time effects directly (see Theorem 4.2). In this paper, we propose a transformation procedure to eliminate the time effects that can avoid the additional $O(n^{-1})$ order bias with the same asymptotic efficiency as the direct QML estimates when $n$ is not relatively smaller than $T$. Our transformation procedure is particularly useful when $n$ is relatively smaller than $T$. For the latter, the estimates of the transformed approach has faster rates of convergence than that of the direct estimates. The direct estimates have a degenerate limit distribution but the transformed estimates are properly centered and are asymptotically normal.

This paper is organized as follows. In Section 2, the model is introduced and the data transformation procedure is proposed. We then explain our method of estimation, which is a concentrated quasi-maximum likelihood estimation. In Section 3, we establish the consistency and asymptotic distribution of the QML estimator of the transformation approach. A bias correction procedure is proposed and the simulation result is reported. Section 4 presents the asymptotic properties of the direct approach estimator and compares the two approaches. Section 5 concludes the paper. Some useful lemmas and proofs are collected in Appendix.

2 The Model

2.1 Data Generating Process and Data Transformation

The model considered in this paper is
\[ Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{nt-1} + \rho_0 W_n Y_{nt-1} + X_{nt} \beta_0 + c_{n0} + \alpha_0 l_n + V_{nt}, \quad t = 1, 2, \ldots, T, \]  

(2.1)

where \( Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})' \) and \( V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})' \) are \( n \times 1 \) column vectors and \( v_{it} \) is i.i.d. across \( i \) and \( t \) with zero mean and variance \( \sigma^2_{n0} \). Also, \( W_n \) is an \( n \times n \) spatial weights matrix which is nonstochastic and generates the spatial dependence between cross sectional units \( y_{it} \). \( X_{nt} \) is an \( n \times k \) matrix of nonstochastic regressors, \( c_{n0} \) is \( n \times 1 \) column vector of individual fixed effects, \( \alpha_0 \) is a scalar of time effect and \( l_n \) is \( n \times 1 \) column vector of ones\(^1\). Therefore, the total number of parameters in this model is equal to the sum of the number of individuals \( n \) and the number of time periods \( T \), plus the dimension of the common parameters \((\gamma, \rho, \lambda', \sigma^2)\)'s, which is \( k_x + 4 \).

\( W_n \) is usually row normalized from a symmetric matrix such that its \( i \)th row is

\[ [c_{n,i1}, c_{n,i2}, \ldots, c_{n,in}] / \sum_{j=1}^{n} c_{n,ij}, \]

(2.2)

where \( c_{n,ij} \) represents a function of the spatial distance of different units in some space. As a normalization, \( c_{n,ii} = 0 \). It is a common practice in empirical work that \( W_n \) is row normalized, which ensures that all the weights are between 0 and 1 and weighting operations can be interpreted as an average of the neighboring values. Also, a spatial weights matrix row normalized has the property \( W_n l_n = l_n \).

Define \( S_n(\lambda) = I_n - \lambda W_n \) and \( S_n(\lambda_0) = I_n - \lambda_0 W_n \). Then, presuming \( S_n \) is invertible and denoting \( A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n) \), (2.1) can be rewritten as

\[ Y_{nt} = A_n Y_{nt-1} + S_n^{-1} X_{nt} \beta_0 + S_n^{-1} c_{n0} + \alpha_0 S_n^{-1} l_n + S_n^{-1} V_{nt}. \]

(2.3)

Assuming that the infinite sums are well-defined, by continuous substitution of (2.3),

\[ Y_{nt} = \sum_{h=0}^{\infty} A_n^h S_{n}^{-1}(c_{n0} + X_{n,t-h} \beta_0 + \alpha_0 l_n + V_{n,t-h}) = \mu_n + X_{nt} \beta_0 + \alpha_0 l_n \sum_{h=0}^{\infty} \left( \frac{\gamma_0 + \rho_0}{1 - \lambda_0} \right)^h + U_{nt}, \]

(2.4)

where \( \mu_n = \sum_{h=0}^{\infty} A_n^h S_n^{-1} c_{n0}, \quad X_{nt} = \sum_{h=0}^{\infty} A_n^h S_n^{-1} X_{n,t-h} \) and \( U_{nt} = \sum_{h=0}^{\infty} A_n^h S_n^{-1} V_{n,t-h} \).

One way to estimate (2.1) is to estimate all the parameters including the time effects and individual effects in the model, which will yield a bias of the order \( O(\max(n^{-1}, T^{-1})) \) for the common parameters (see Theorem 4.2). In this paper, we will introduce a data transformation approach (to eliminate the time effects) such that we can avoid the bias of the order \( O(n^{-1}) \), where the estimator has the same asymptotic efficiency as the direct QML estimator when \( n \) is not relatively smaller than \( T \). This transformation procedure is particularly useful when \( n \) is relatively smaller than \( T \) where the estimates of the transformed approach has faster rates of convergence than that of the direct estimates. Also, when \( n \) is relatively smaller than \( T \), the direct estimates have a degenerate limit distribution but the transformed estimates are properly centered and are asymptotically normal.

Let \( J_n = I_n - \frac{1}{n} l_n l_n' \) be the deviation from the group mean transformation. As \( I_n = J_n + \frac{1}{n} l_n l_n' \) and \( W_n l_n = l_n \), we have

---

\(^1\)Due to the presence of fixed individual and time effects, the \( X_{nt} \) will not include time invariant or individual invariant regressors.
\[ J_n W_n = J_n W_n (J_n + \frac{1}{n} l_n l_n') = J_n W_n J_n, \]

because \( J_n W_n l_n = J_n l_n = 0 \). Hence, for \( t = 1, 2, ..., T \), we have

\[
(J_n Y_{nt}) = \lambda_0 (J_n W_n) (J_n Y_{nt}) + \gamma_0 (J_n Y_{n,t-1}) + \rho_0 (J_n W_n) (J_n Y_{n,t-1}) + (J_n X_{nt}) \beta_0 + (J_n c_{nt}) + (J_n V_{nt}),
\]

which does not involve the time effects and \( J_n c_{nt} \) can be regarded as the transformed individual effects. Thus, we can estimate \( \theta_0 \) and \( J_n c_{nt} \) basing on the transformed equation (2.5) where relevant variables are premultiplied by \( J_n \).

A special feature of the transformed equation (2.5) is that the variance matrix of \( J_n V_{nt} \) is equal to \( \sigma_0^2 J_n \) so that the elements of \( J_n V_{nt} \) are correlated. Also, \( J_n \) is singular with rank \( (n-1) \) as \( J_n \) is an orthogonal projector with trace \( (n-1) \). Hence, there is a linear dependence among the elements of \( J_n V_{nt} \). An effective estimation method shall eliminate the linear dependence in sample observations. This can be done with the eigenvalues and eigenvectors decomposition as in the theory of generalized inverses for the estimation of linear regression models (see, e.g., Theil (1971), Ch. 6).

The eigenvalues of \( J_n \) are a single zero and \((n-1)\) ones. An eigenvector corresponding to the zero eigenvalue is proportional to \( l_n \). Let \( (F_{n,n-1}, l_n/\sqrt{n}) \) be the orthonormal matrix of eigenvectors of \( J_n \) where \( F_{n,n-1} \) corresponds to the eigenvalues of ones and \( l_n/\sqrt{n} \) corresponds to the eigenvalue zero. As is derived in Appendix A.1, the transformation of \( J_n Y_{nt} \) to \( Y_{nt}^* \), where \( Y_{nt}^* = F_{n,n-1}' J_n Y_{nt} \) is of dimension \((n-1)\), gives

\[
Y_{nt}^* = \lambda_0 W_n^* Y_{nt}^* + \gamma_0 Y_{n,t-1}^* + \rho_0 W_n^* Y_{n,t-1}^* + X_{nt}^* \beta_0 + c_{nt}^* + V_{nt}^*,
\]

where \( W_n^* = F_{n,n-1}' W_n F_{n,n-1}, Y_{nt}^* = F_{n,n-1}' J_n Y_{nt}, X_{nt}^* = F_{n,n-1}' J_n X_{nt}, c_{nt}^* = F_{n,n-1}' J_n c_{nt}, V_{nt}^* = F_{n,n-1}' J_n V_{nt}, \) and \( V_{nt}^* \) is an \( n-1 \) dimensional disturbance vector with zero mean and variance matrix \( \sigma_0^2 J_{n-1} \).

2.2 The Method of Maximum Likelihood Estimation

Suppose that \( V_{nt} \) is normally distributed \( N(0, \sigma_0^2 I_n) \), the transformed \( V_{nt}^* \) will be \( N(0, \sigma_0^2 I_{n-1}) \) and (2.6) can be estimated by the ML. The log likelihood function of (2.6) for \( Y_{nt}^* \) is

\[
\ln L_{n,T}(\theta, c_n^*) = \frac{(n-1)T}{2} \ln 2\pi - \frac{(n-1)T}{2} \ln \sigma^2 + T \ln |I_{n-1} - \lambda W_n^*| - \frac{1}{2\sigma^2} \sum_{t=1}^T V_{nt}^*(\theta) V_{nt}^*(\theta),
\]

where \( V_{nt}^*(\theta) = (I_{n-1} - \lambda W_n^*) Y_{nt}^* - Z_{nt}^* \delta - c_{nt}^* X_{nt} = (Y_{nt-1}, W_n Y_{nt-1}, X_{nt}) \) and \( \delta = (\gamma, \rho, \beta' \beta)' \). In order to use Equation (2.6) for effective estimation, the determinant and inverse of \((I_{n-1} - \lambda W_n^*)\) are needed. We shall show that their computations are not more complicated than those for the original matrix \( I_n - \lambda W_n \).

As is derived in Appendix A.2, \((I_{n-1} - \lambda W_n^*) = F_{n-1}' (I_n - \lambda W_n) F_{n-1} \) and we have

\[
|I_{n-1} - \lambda W_n^*| = \frac{1}{1 - \lambda} |I_n - \lambda W_n|,
\]

(2.8)
Furthermore, it is not clear whether W analyze the asymptotic distribution of the estimator just from Equation (2.10). with the disturbances asymptotic analysis in Yu, de Jong and Lee (2006) does not directly carry over to the transformed model one may maximize Equation (2.10) instead. However, although the components of panel model with only individual effects with 2.3 FOC and SOC of MLE where

\[ V_{nt}(\theta) = (I_{n-1} - \lambda W_{n})^{-1} F_{n,n-1} (I_{n-1} - \lambda W_{n})^{-1} F_{n,n-1}. \]  

(2.9)

Also, as we have

\[
(I_{n-1} - \lambda W_{n}^{*} - Z_{nt}\delta - c_{n}^{*})
= F'_{n,n-1}(I_{n} - \lambda W_{n}) F_{n,n-1} F'_{n,n-1} Y_{nt} - F'_{n,n-1} Z_{nt} \delta - F'_{n,n-1} c_{n}
= F'_{n,n-1}(I_{n} - \lambda W_{n})(I_{n} - \frac{1}{n} I_{n} l_{n}') Y_{nt} - F'_{n,n-1} Z_{nt} \delta - F'_{n,n-1} c_{n}
= F'_{n,n-1}(I_{n} - \lambda W_{n}) Y_{nt} - Z_{nt} \delta - c_{n},
\]

because $F'_{n,n-1} W_{n} l_{n} = F'_{n,n-1} l_{n} = 0$, it follows that

\[
V'_{nt}(\theta)V_{nt}(\theta)
= [(I_{n-1} - \lambda W_{n}) Y_{nt} - Z_{nt} \delta - c_{n}^{*}][[(I_{n-1} - \lambda W_{n}) Y_{nt} - Z_{nt} \delta - c_{n}^{*}]
= [(I_{n} - \lambda W_{n}) Y_{nt} - Z_{nt} \delta - c_{n}][F_{n,n-1} F'_{n,n-1} (I_{n} - \lambda W_{n}) Y_{nt} - Z_{nt} \delta - c_{n}]
= [(I_{n} - \lambda W_{n}) Y_{nt} - Z_{nt} \delta - c_{n}][J_{n}(I_{n} - \lambda W_{n}) Y_{nt} - Z_{nt} \delta - c_{n}],
\]

by the property $F_{n,n-1} F'_{n,n-1} = J_{n}$. Hence, the log likelihood function (2.7) for $Y_{nt}^{*}$ can be expressed in terms of $Y_{nt}$ as

\[
\ln L_{n,T}(\theta, c_{n}) = -\frac{(n - 1)T}{2} \ln 2\pi - \frac{(n - 1)T}{2} \ln \sigma^{2} - T \ln (1 - \lambda) + T \ln |I_{n} - \lambda W_{n}|
= \frac{-1}{2\sigma^{2}} \sum_{t=1}^{T} V'_{nt}(\theta) J_{n} V_{nt}(\theta),
\]

where $V_{nt}(\theta) = (I_{n} - \lambda W_{n}) Y_{nt} - Z_{nt} \delta - c_{n}$ and $J_{n}$ shall be read as the generalized inverse of $\sigma_{0}^{-2} Var(J_{n} V_{nt})$.

### 2.3 FOC and SOC of MLE

Therefore, we will first transform the data from $Y_{nt}$ to $Y_{nt}^{*} = F'_{n,n-1} Y_{nt}$, then we will maximize Equation (2.7) by searching over the parameter space. This is equivalent to the estimation of the spatial dynamic panel model with only individual effects with $(n - 1)$ cross-section units and $T$ time periods. Alternatively, one may maximize Equation (2.10) instead. However, although the components of $V_{nt}$ are i.i.d. in the model, the elements of $V_{nt}^{*}$ might not be independent in general, even though they are uncorrelated. The original asymptotic analysis in Yu, de Jong and Lee (2006) does not directly carry over to the transformed model with the disturbances $V_{nt}^{*}$.

As Equation (2.7) is equivalent to Equation (2.10), it turns out that we can analyze the asymptotic distribution of the estimator just from Equation (2.10).

---

2One could not treat the components of $V_{nt}^{*}$ as if they were independent when the disturbances are not normally distributed. Furthermore, it is not clear whether $W_{n}^{*}$ and $A_{n}^{*} = (I_{n-1} - \lambda_{0} W_{n}^{*})^{-1}(\gamma_{0} I_{n-1} + \rho_{0} W_{n}^{*})$ would be uniformly bounded in both row and column sums even $W_{n}$ and $A_{n}$ are.
Using first order conditions, we concentrate out \( c_n \) in (2.10) and get
\[
\ln L_{n,T}(\theta) = -\frac{(n-1)T}{2} \ln 2\pi - \frac{(n-1)T}{2} \ln \sigma^2 - T \ln (1-\lambda) + T \ln |I_n - \lambda W_n| \quad (2.11)
\]
\[
-\frac{1}{2\sigma^2} \sum_{t=1}^{T} \tilde{V}_n(t) J_n \tilde{V}_n(t),
\]
where \( \tilde{V}_n(\theta) = (I_n - \lambda W_n) \tilde{Y}_n - \tilde{Z}_n \delta_0 \) and \( J_n \tilde{V}_n(\theta) = J_n[(I_n - \lambda W_n) \tilde{Y}_n - \tilde{Z}_n \delta_0 - \tilde{\alpha}_t l_n] \) because \( J_n l_n = 0 \). The concentrated likelihood function is different from (4.2) of the direct approach\(^4\) in that we have an attachment of \( \frac{T}{2} \ln 2\pi \sigma^2 - T \ln (1-\lambda) \). For the concentrated likelihood function (2.11), the first order derivatives and the second order derivatives are Equation (C.3) and (C.4) in Appendix C.2. Denote
\[
J_n \tilde{V}_n = J_n[S_n \tilde{Y}_n - \tilde{Z}_n \delta_0 - \tilde{\alpha}_t l_n].
\]
Hence, for the first order derivative evaluated at \( \theta_0 \), denote \( G_n = W_n S_n^{-1} \), we have
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} = \begin{pmatrix}
\frac{1}{\sigma_0^2} \sqrt{(n-1)T} \sum_{t=1}^{T} \tilde{Z}_n J_n \tilde{V}_n \\
\frac{1}{\sigma_0^2} \sqrt{(n-1)T} \sum_{t=1}^{T} (G_n \tilde{Z}_n \delta_0) J_n \tilde{V}_n + \frac{1}{\sigma_0^2} \sqrt{(n-1)T} \sum_{t=1}^{T} (\tilde{V}_n G_n J_n \tilde{V}_n - \sigma_0^2 \tr J_n G_n) \\
\frac{1}{2\sigma_0^2} \sqrt{(n-1)T} \sum_{t=1}^{T} (\tilde{V}_n J_n \tilde{V}_n - (n-1) \sigma_0^2)
\end{pmatrix},
\]
which is a linear and quadratic form of \( \tilde{V}_n \). For the information matrix, denote \( \Sigma_{\theta_0,nT} = -E \left( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial \theta \partial \theta} \right) \)
and \( \mathcal{H}_nT = \frac{1}{(n-1)T} \sum_{t=1}^{T} (\tilde{Z}_n, G_n \tilde{Z}_n \delta_0) J_n(\tilde{Z}_n, G_n \tilde{Z}_n \delta_0) \), we have
\[
\Sigma_{\theta_0,nT} = \frac{1}{\sigma_0^2} \begin{pmatrix}
EH_{nT} & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix} + \begin{pmatrix}
0 & 0 & 0 \\
0 & \frac{1}{\sigma_0^2} \tr (G_n J_n G_n) + \frac{1}{\sigma_0^2 (n-1) \tr J_n G_n} & 0 \\
0 & 0 & \frac{1}{\sigma_0^2} (n-1) \tr J_n G_n
\end{pmatrix} \begin{pmatrix}
0 & 0 & 0 \\
0 & \frac{1}{\sigma_0^2} \tr J_n G_n & 0 \\
0 & 0 & \frac{1}{\sigma_0^2} \tr J_n G_n
\end{pmatrix}
\]
\[
-\begin{pmatrix}
0 & 0 & 0 \\
0 & \frac{2}{\sigma_0^2 (n-1) \tr (G_n \tilde{Z}_n \delta_0)} J_n G_n \tilde{V}_n \tilde{V}_n & 0 \\
0 & \frac{2}{\sigma_0^2 (n-1) \tr (G_n \tilde{Z}_n \delta_0)} J_n G_n \tilde{V}_n \tilde{V}_n & 0
\end{pmatrix}.
\]

3 \textbf{QML Estimators}

For our analysis of the asymptotic properties of estimator, we need the following assumptions:

\(^3\)For notational purpose, we define for any \( n \times 1 \) vector at period \( t, \ Y_{nt}, \) we have \( \tilde{Y}_{nt} = Y_{nt} - \tilde{\gamma}_{nt} \) and \( \gamma_{nt} = \gamma_{nt} - \tilde{\gamma}_{nt} \) for \( t = 1, 2, \cdots, T \) where \( \gamma_{nt} = \frac{1}{T} \sum_{t=1}^{T} Y_{nt} \) and \( \gamma_{nt} = \frac{1}{T} \sum_{t=1}^{T} \gamma_{nt} \).

\(^4\)The difference term can be regarded as an adjustment by the transformation approach to eliminate the bias of the order \( O(n^{-1}) \) which will appear in the direct approach.
Assumption 1. $W_n$ is a row normalized nonstochastic spatial weights matrix.

Assumption 2. The disturbances $\{v_{it}\}$, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, are i.i.d. across $i$ and $t$ with zero mean, variance $\sigma_i^2$ and $E|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

Assumption 3. $S_n(\lambda)$ is invertible for all $\lambda \in \Lambda$. Furthermore, $\Lambda$ is compact and the true parameter $\lambda_0$ is in the interior of $\Lambda$.

Assumption 4. The elements of $X_{nt}$ are nonstochastic and bounded, uniformly in $n$ and $t$, and $\lim_{T \to \infty} \frac{1}{nT} \sum_{t=1}^{T} X_{nt}' J_n X_{nt}$ exists and is nonsingular.

Assumption 5. The row and column sums of $W_n$ and $S_n^{-1}(\lambda)$ are bounded uniformly in $n$, also uniformly in $\lambda \in \Lambda$ for $S_n^{-1}(\lambda)$.

Assumption 6. The row and column sums of $\sum_{h=1}^{\infty} \text{abs}(A_n^h)$ are bounded uniformly in $n$, where $[\text{abs}(A_n)]_{ij} = |A_{n,ij}|$.

Assumption 7. $n$ is a nondecreasing function of $T$.

Assumption 1 is a standard normalization assumption in spatial econometrics. In many empirical applications, the rows of $W_n$ sum to 1, which ensures that all the weights are between 0 and 1. Assumption 2 provides regularity assumptions for $v_{it}$. Assumption 3 guarantees that Equation (2.3) is valid. When exogenous variables $X_{nt}$ are included in the model, it is convenient to assume that the exogenous regressors are uniformly bounded as in Assumption 4. Assumption 5 is originated by Kelejian and Prucha (1998, 2001). The uniform boundedness of $W_n$ and $S_n^{-1}(\lambda)$ is a condition that limits the spatial correlation to a manageable degree. Assumption 6 is the absolute summability condition and row/column sum boundedness condition, which will play an important role to derive asymptotic properties of QML estimator. This assumption is essential for the paper because it limits the dependence between time series and between cross sectional units. In order to justify the absolute summability of $A_n$ in Equation (2.4) and Assumption 6, a sufficient condition is $\|A_n\| < 1$ for any matrix norm (see Horn and Johnson (1985), Corollary 5.6.16) that satisfies $\|A_n\| = \|\text{abs}(A_n)\|$. When $\|A_n\| < 1$, $\sum_{h=0}^{\infty} A_n^h$ exists and can be defined as $(I_n - A_n)^{-1}$. Assumption 7 allows two cases: (i) $n \to \infty$ as $T \to \infty$; (ii) $n$ is fixed as $T \to \infty$. Because (ii) is similar to a vector autoregressive (VAR) model, our main interest is in (i). If Assumption 7 holds, then we say that $n, T \to \infty$ simultaneously.

3.1 Consistency

For the log likelihood function (2.11) divided by the effective sample size $(n-1)T$, we have corresponding

$$Q_{n,T}(\theta) = E \max_{c_n} \frac{1}{(n-1)T} \ln L_{n,T}(\theta, c_n).$$

Hence,

$$Q_{n,T}(\theta) = \frac{1}{(n-1)T} E \ln L_{n,T}(\theta)$$

$$= -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{n - 1} \ln(1 - \lambda) + \frac{1}{n - 1} \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} \frac{1}{(n-1)T} E \left( \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{\nu}_{nt}(\theta) \right).$$

(3.1)

To get the consistency proof, we need the following uniform convergence result.

---

[5] We say the row and column sums of a (sequence of $n \times n$) matrix $P_n$ are bounded uniformly in $n$ if $\sup_{1 \leq i \leq n, n \geq 1} \sum_{j=1}^{n} |p_{ij,n}| < \infty$ and $\sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^{n} |p_{ij,n}| < \infty$. 

---
Claim 3.1 Let $\Theta$ be any compact parameter space. Then under Assumptions 1-7, $\frac{1}{(n-1)T} \ln L_n,T(\theta) \overset{p}{\to} 0$ uniformly in $\theta \in \Theta$ and $Q_{n,T}(\theta)$ is uniformly equicontinuous for $\theta \in \Theta$.

**Proof.** See Appendix D.1. ■

For identification, if the information matrix $-E \left( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_n,T(\theta)}{\partial \theta \partial \theta'} \right)$ is nonsingular and $-E \left( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_n,T(\theta)}{\partial \theta \partial \theta'} \right)$ has full rank for any $\theta$ in some neighborhood $N(\theta_0)$, the parameters are locally identified (see Rothenberg (1971)). For the information matrix, using Lemma B.2 in Yu, de Jong and Lee (2006), we have

\[
\Sigma_{\theta_0,nT} = \frac{1}{\sigma_0^2} \begin{pmatrix}
E H_{nT} & 0 \\
0 & 0
\end{pmatrix} + \begin{pmatrix}
0 & -\frac{1}{\sigma_0^2} \left[ \text{tr}(G_n'J_nG_n) + \text{tr}((J_nG_n)^2) \right] \\
0 & \frac{1}{\sigma_0^2} \text{tr}(J_nG_n)
\end{pmatrix} + O\left( \frac{1}{T} \right),
\]

which is nonsingular if $E H_{nT}$ is nonsingular or $\frac{1}{n-1} \left[ \text{tr}(G_n'J_nG_n) + \text{tr}((J_nG_n)^2) - \frac{2\text{tr}^2(J_nG_n)}{n-1} \right]$ is nonzero$^6$. Also, its rank does not change in a small neighborhood of $\theta_0$ (see Equation (C.8)).

When $\lim_{T \to \infty} E H_{nT}$ is nonsingular, we can get the global identification of the parameters.

**Theorem 3.2** Under Assumptions 1-7, if $\lim_{T \to \infty} E H_{nT}$ is nonsingular, $\theta_0$ is globally identified and $\hat{\theta}_{nT} \overset{p}{\to} \theta_0$.

**Proof.** See Appendix D.3. ■

When $\lim_{T \to \infty} E H_{nT}$ is singular, global identification can still be obtained from the following theorem. Denote $\sigma_n^2(\lambda) = \frac{\sigma_n^2}{n-1} \text{tr}(S_n^{-1}S_n^{-1}(\lambda)J_nS_n(\lambda)S_n^{-1})$.

**Theorem 3.3** Under Assumptions 1-7, $\theta_0$ is globally identified if

\[
\lim_{n \to \infty} \left( \frac{1}{n} \ln \left| \sigma_n^2 S_n^{-1} S_n^{-1} \right| - \frac{1}{n} \ln \left| \sigma_n^2(\lambda) S_n^{-1}(\lambda) S_n^{-1}(\lambda) \right| \right) \neq 0 \text{ for } \lambda \neq \lambda_0^7. And if this condition holds, $\hat{\theta}_{nT} \overset{p}{\to} \theta_0$.

**Proof.** See Appendix D.4. ■

### 3.2 Asymptotic Distribution

As $Z_{nt} = (Y_{n,t-1}, W_nY_{n,t-1}, X_{nt})$ where $Y_{nt}$ is specified in Equation (2.4), we can decompose $J_n\tilde{Z}_{nt}$ such that

\[
J_n\tilde{Z}_{nt} = J_n\tilde{Z}_{nt}^{(u)} - (J_n\bar{U}_{nT,-1}, J_nW_n\bar{U}_{nT,-1}, 0),
\]

where $\tilde{Z}_{nt}^{(u)} = ((\bar{X}_{n,t-1} + U_{n,t-1}), (W_n\bar{X}_{n,t-1} + W_nU_{n,t-1}), \bar{X}_{nt})$ with $\bar{X}_{n,t-1} = X_{n,t-1} - \bar{x}_{nT,-1}$. Hence, $J_n\tilde{Z}_{nt}$ has two components: one is $J_n\tilde{Z}_{nt}^{(u)}$, which is uncorrelated with $V_{nt}$; the other is $-\hat{J}_n\bar{U}_{nT,-1}, J_nW_n\bar{U}_{nT,-1}, 0$.

---

$^6$See Appendix D.2 for proof.

$^7$When $n$ is finite, the condition is $\frac{1}{n} \ln \left| \sigma_n^2 S_n^{-1} S_n^{-1} \right| - \frac{1}{n} \ln \left| \sigma_n^2(\lambda) S_n^{-1}(\lambda) S_n^{-1}(\lambda) \right| + \left( \frac{1}{n} \ln \frac{\sigma_n^2}{(1-\lambda_0)^2} - \frac{1}{n} \ln \frac{\sigma_n^2(\lambda)}{(1-\lambda)^2} \right) \neq 0$ for $\lambda \neq \lambda_0$. See Appendix D.4.
which is correlated with \( V_{nt} \) when \( t \leq T - 1 \). Therefore, from Equation (2.12), the score can be decomposed into two parts such that

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} - \Delta_{nT},
\]

(3.4)

where

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} = \left( \begin{array}{c}
\frac{1}{\sigma_0} \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} \bar{Z}_{nt}^{(u)} J_n V_{nt} \\
\frac{1}{\sigma_0} \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} (G_n \bar{Z}_{nt}^{(u)} \delta_0)' J_n V_{nt} + \frac{1}{\sigma_0} \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} (V_n' G_n' J_n V_{nt} - \sigma^2_0 tr J_n G_n) \\
\frac{1}{2\sigma^2_0} \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} (V_n' J_n V_{nt} - (n-1)\sigma^2_0)
\end{array} \right)
\]

(3.5)

\[
\Delta_{nT} = \sqrt{\frac{n-1}{T}} \left( \begin{array}{c}
\frac{1}{\sigma_0} \frac{1}{T} \left( J_n \bar{U}_{nt}-1, J_n W_n \bar{U}_{nt}-1, 0 \right)' \bar{V}_{nt} \\
\frac{1}{\sigma_0} \frac{1}{T} \left( J_n G_n (U_n T-1, W_n U_n T-1, 0) \delta_0)' \bar{V}_{nt} + \frac{1}{\sigma_0} \frac{1}{T} \bar{V}_n' G_n' J_n \bar{V}_{nt} \right)
\end{array} \right).
\]

(3.6)

As is derived in Appendix C.3, the variance matrix of \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \) is equal to

\[
E \left( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta'} \right) = \Sigma_{\theta_0,nT} + \Omega_{\theta_0,n} + O \left( T^{-1} \right),
\]

(3.7)

where \( \Sigma_{\theta_0,nT} \) is in (3.2) and \( \Omega_{\theta_0,n} = \frac{\mu_4 - 3\sigma^4}{\sigma^4} \left( \begin{array}{ccc}
0 & 0 & 0 \\
0 & \frac{1}{n-1} \sum_{i=1}^{n} [(J_n G_n)^2]_i & 0 \\
0 & \frac{1}{2\sigma^2_0(n-1)} \text{tr} (J_n G_n) & \frac{1}{4\sigma^4_0}
\end{array} \right) \) is a symmetric matrix with \( \mu_4 \) being the fourth moment of \( v_{nt} \). When \( V_{nt} \) are normally distributed, \( \Omega_{\theta_0,n} = 0 \) because \( \mu_4 - 3\sigma^4 = 0 \) for a normal distribution. Denote \( \Sigma_{\theta_0} = \lim_{T \to \infty} \Sigma_{\theta_0,nT} \) and \( \Omega_{\theta_0} = \lim_{T \to \infty} \Omega_{\theta_0,n} \), then

\[
\lim_{T \to \infty} E \left( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta'} \right) = \Sigma_{\theta_0} + \Omega_{\theta_0}.
\]

(3.8)

The asymptotic distribution of \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \) can be derived from the central limit theorem for martingale difference arrays (Theorem B.3). For the term \( \Delta_{nT} \), from Equation (B.3) in Lemma B.1 and Equation (B.5) in Theorem B.2, \( \Delta_{nT} = \sqrt{\frac{n-1}{T}} \Theta_{\theta_0,n} + O \left( \sqrt{\frac{n-1}{T}} \right) + O_P \left( \frac{1}{\sqrt{T}} \right) \) where

\[
\Theta_{\theta_0,n} = \left( \begin{array}{c}
\frac{1}{n-1} \text{tr} \left( (J_n \sum_{h=0}^{\infty} A^h_n) S^{-1}_n \right) \\
\frac{1}{n-1} \text{tr} \left( W_n (J_n \sum_{h=0}^{\infty} A^h_n) S^{-1}_n \right) \\
\frac{1}{n-1} \gamma_0 \text{tr}(G_n (J_n \sum_{h=0}^{\infty} A^h_n) S^{-1}_n) + \frac{1}{n-1} \rho_0 \text{tr}(G_n W_n (J_n \sum_{h=0}^{\infty} A^h_n) S^{-1}_n) + \frac{1}{n-1} \text{tr} J_n G_n \\
\frac{1}{2\sigma^2_0}
\end{array} \right)
\]

(3.9)
is $O(1)$. To get the asymptotic distribution of the score, we need the following additional assumption.

**Assumption 8.** $\lim_{T \to \infty} \mathbb{E}H_{nT}$ is nonsingular or $\lim_{n \to \infty} \frac{1}{n-1} \left[ tr(G'_n J_n G_n) + tr((J_n G_n)^2) - \frac{2tr^2(J_n G_n)}{n-1} \right] \neq 0$.

Assumption 8 is a condition for the nonsingularity of the information matrix $\Sigma_{\theta_0}$. When $\lim_{T \to \infty} \mathbb{E}H_{nT}$ is singular, as long as we have $\lim_{n \to \infty} \frac{1}{n-1} tr \left[ (G'_n J_n G_n) + tr((J_n G_n)^2) - \frac{2tr^2(J_n G_n)}{n-1} \right] \neq 0$, the information matrix $\Sigma_{\theta_0}$ is still nonsingular (see Appendix D.2).

**Claim 3.4** Under Assumptions 1-8,

$$
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} + \Delta_n T \xrightarrow{d} N(0, \Sigma_{\theta_0} + \Omega_{\theta_0}).
$$
(3.10)

When $\{v_{it}\}, i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, are normal, $\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} + \Delta_n T \xrightarrow{d} N(0, \Sigma_{\theta_0}).$

**Proof.** See Appendix D.5. ■

**Claim 3.5** Under Assumptions 1-8,

$$
\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\theta)}{\partial \theta \partial \theta'} - \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial \theta \partial \theta'} = \|\theta - \theta_0\| \cdot O_p(1),
$$
(3.11)

and

$$
\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_{n,T}(\theta_0)}{\partial \theta \partial \theta'} = O_p \left( \frac{1}{\sqrt{(n-1)T}} \right).
$$
(3.12)

**Proof.** See Appendix D.6. ■

Using Claim 3.4 and Claim 3.5, we have the following theorem for the distribution of $\hat{\theta}_{nT}$.

**Theorem 3.6** Under Assumptions 1-8,

$$
\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n-1}{T}} b_{\theta_0,nT} + O_p \left( \max \left( \sqrt{\frac{n-1}{T^3}}, \sqrt{\frac{1}{T}} \right) \right) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}),
$$
(3.13)

where $b_{\theta_0,nT} = \Sigma_{\theta_0,nT}^{-1}a_{\theta_0,n}$ is $O(1)$.

When $\frac{n}{T} \to 0$,

$$
\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}).
$$
(3.14)

When $\frac{n}{T} \to k < \infty$,

$$
\sqrt{kT}(\hat{\theta}_{nT} - \theta_0) + \sqrt{k} b_{\theta_0,nT} \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}).
$$
(3.15)

When $\frac{n}{T} \to \infty$,

$$
T(\hat{\theta}_{nT} - \theta_0) + b_{\theta_0,nT} \xrightarrow{p} 0.
$$
(3.16)

**Proof.** See Appendix D.7. ■

---

Footnote: The $\lim_{T \to \infty} \mathbb{E}H_{nT}$ can be singular if, for example, $\delta_0 = 0$. 

---
From Equation (3.13), the QML estimator has the bias $-\frac{1}{T}b_{\theta_0,nT}$ where

$$b_{\theta_0,nT} = \sum_{\theta_0,nT}^{-1} a_{\theta_0,n}$$

and the confidence interval is not centered when $\frac{n}{T} \to k$ where $0 < k < \infty$. Furthermore, when $T$ is relatively smaller than $n$, the presence of $b_{\theta_0,nT}$ causes $\theta_{nT}$ to have a degenerated distribution. An analytical bias reduction procedure is to correct the bias of the estimate. Define the bias corrected estimator as

$$\hat{\theta}_{nT}^1 = \hat{\theta}_{nT} - \frac{\hat{B}_{nT}}{T},$$

where, from Theorem 3.6,

$$\hat{B}_{nT} = \left[ -\sum_{\theta_0,nT}^{-1} a_{\theta_0,n} \right] \bigg|_{\theta = \theta_{nT}}.$$  

We will show that when $n/T \to 0$, $\hat{\theta}_{nT}^1$ is $\sqrt{nT}$ consistent and asymptotically centered normal even when $n/T \to \infty$.

To show our result for the bias corrected estimator, we need the following additional assumption.

**Assumption 9.** $\sum_{h=0}^{\infty} A_n^h(\theta)$ and $\sum_{h=1}^{\infty} hA_n^{h-1}(\theta)$ are uniformly bounded in either row sum or column sums, uniformly in a neighborhood of $\theta_0$.

Assumption 9 can be verified through the following lemma.

**Lemma 3.7** If $\|A_n(\theta_0)\|_\infty < 1$ (resp: $\|A_n(\theta_0)\|_1 < 1$), then the row sum (resp: column sum) of $\sum_{h=0}^{\infty} A_n^h(\theta)$ and $\sum_{h=1}^{\infty} hA_n^{h-1}(\theta)$ are bounded uniformly in $n$ and in a neighborhood of $\theta_0$.

**Proof.** This is Lemma 3.9 in Yu, de Jong and Lee (2006). □

Our result for the bias corrected estimator is as follows.

**Theorem 3.8** Under Assumptions 1-9, if $\frac{n}{T} \to 0$,

$$\sqrt{nT}(\hat{\theta}_{nT}^1 - \theta_0) \overset{d}{\to} N(0, \Sigma_{\theta_0}^{-1} + \Sigma_{\theta_0}^{-1} \Omega_{\theta_0} \Sigma_{\theta_0}^{-1}).$$

**Proof.** See Appendix D.8. □

### 3.3 Monte Carlo Results

We conduct a small Monte Carlo experiment to evaluate the performance of our ML estimator and the bias corrected estimator for the transformation approach. We generate samples from Equation (2.1) using $\theta_0^a = (0.2, 0.2, 1, 0.2, 1)'$ and $\theta_0^b = (0.3, 0.3, 1, 0.3, 1)'$ where $\theta_0 = (\gamma_0, \rho_0, \beta_0, \lambda_0, \sigma_0^2)'$, and $X_{nt}$, $c_{nt}$, $\alpha_T = (\alpha_1, \alpha_2, \cdots, \alpha_T)$ and $V_{nt}$ are generated from independent normal distributions and the spatial weights matrix we use is a rook matrix. We use $T = 10, 50$, and $n = 16, 49$. For each set of generated sample observations, we calculate the ML estimator $\hat{\theta}_{nT}$ and evaluate the bias $\hat{\theta}_{nT} - \theta_0$; we then construct the bias corrected estimator $\hat{\theta}_{nT}^1$ and evaluate the bias $\hat{\theta}_{nT}^1 - \theta_0$. We do this for 1000 times to see if the bias is...
reduced on average by using the analytical bias correction procedure, i.e., to compare \( \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_{nT} - \theta_0)_i \) with \( \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_{nT}^1 - \theta_0)_i \). We also compare not only the empirical standard error of estimator but also the average of estimated standard error of the estimator\(^{10}\). With two different values of \( \theta_0 \) for each \( n \) and \( T \), finite sample properties of both estimators are summarized in Table 1 and Table 3, where Table 1 is for the magnitude of biases and Table 3 is for the standard errors of the estimator.

We can see that both estimators have some biases, but the bias corrected estimators reduce those biases. This is consistent with our asymptotic analysis, because the bias corrected estimator will eliminate the bias of order \( O(T^{-1}) \). Also, the bias reduction is achieved while there is no significant increase in the variance of the estimator, as can be seen from Table 3. Also, the estimated variances and the empirical variances are close, so the estimated variances constructed from the Hessian matrix may be reliable.

For different cases of \( n \) and \( T \), we can see that for each given \( n \), when \( T \) is larger, the biases of the two sets of estimators will be smaller and the variance will be smaller; for each given \( T \), when \( n \) is larger, the biases of the two sets of estimators will not be smaller but the variance will be smaller. This is consistent with our theoretical prediction, because the bias is of the order \( O(T^{-1}) \) and the variance of the estimator is of the order \( O((nT)^{-1}) \).

## 4 An Alternative Approach: Direct Estimation

An alternative approach is to estimate both the individual effects and the time effects directly\(^{11}\), which will yields bias of the order \( O(\max(n^{-1}, T^{-1})) \).

### 4.1 Model, Likelihood Function, FOC and SOC

Denote \( \alpha_T = (\alpha_1, \alpha_2, \ldots, \alpha_T) \) as the time effects, the likelihood function of (2.1) is

\[
\ln L_{n,T}^d(\theta, c_n, \alpha_T) = -\frac{nT}{2} \ln 2\pi - \frac{nT}{2} \ln \sigma^2 + T \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} \sum_{t=1}^{T} V_{nt}(\theta, c_n, \alpha_T) V_{nt}(\theta, c_n, \alpha_T), \tag{4.1}
\]

where \( V_{nt}(\theta, c_n, \alpha_T) = S_n(\lambda) Y_{nt} - Z_{nt} \delta - c_n - \alpha_t l_n \). We will use the concentrating approach to estimate the common parameters. As is derived in Appendix E.1, the likelihood function with both \( c_n \) and \( \alpha_T \) concentrated out is

\[
\ln L_{n,T}^{d,c}(\theta) = -\frac{nT}{2} \ln 2\pi - \frac{nT}{2} \ln \sigma^2 + T \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt}(\theta). \tag{4.2}
\]

\(^{10}\)The empirical variance is the diagonal elements of \( \frac{1}{1000} \sum_{i=1}^{1000} ((\hat{\theta}_{nT} - \text{mean}(\hat{\theta}_{nT}))(\hat{\theta}_{nT} - \text{mean}(\hat{\theta}_{nT}))')_i \) and the estimated variance is the diagonal elements of \( \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_{nT} - \theta_0)(\hat{\theta}_{nT} - \theta_0)'_i \).

\(^{11}\)Remark: There is an underidentification problem on \( c_n \) and \( \alpha_T \). Because components of \( c_n \) and \( \alpha_T \) appear additively, there is a normalization issue in terms of location. By adding a constant to all components of \( c_n \) and subtracting the same constant from all components of \( \alpha_T \), the different sets of fixed effects can not be identified. So a proper normalization is needed in order to identify fixed effects. A convenient location normalization is to set the sum of the components of \( c_n \) to be zero, i.e., the elements of \( c_n \) have a zero (empirical) mean.
Proof. See Appendix E.3. 

For the concentrated likelihood function (4.2), the first order derivatives are in (E.3) and (E.4) in Appendix E.1. Hence, for the first order derivative evaluated at \( \theta_0 \), we have

\[
\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}'(\theta_0)}{\partial \theta} = \left( \begin{array}{c} \frac{1}{\sigma_0^2} \sum_{t=1}^{T} \tilde{Z}_{nt}' J_n \tilde{V}_{nt} \\ \frac{1}{\sigma_0^2} \sum_{t=1}^{T} (G_n \tilde{Z}_{nt} \delta_0)' J_n \tilde{V}_{nt} + \frac{1}{\sigma_0^2} \sum_{t=1}^{T} (\tilde{V}_{nt} G_n' J_n \tilde{V}_{nt} - \sigma_0^2 \text{tr} G_n) \\ \frac{1}{2\sigma_0^2} \sum_{t=1}^{T} (\tilde{V}_{nt}' J_n \tilde{V}_{nt} - n\sigma_0^2) \end{array} \right),
\]

which is a linear and quadratic form of \( \tilde{V}_{nt} \). For the information matrix of (4.2), denote \( \Sigma_{\theta_0,nT}^d = -E \left( \frac{1}{nT} \frac{\partial^2 \ln L_{nT}'(\theta_0)}{\partial \theta \partial \theta'} \right) \) and \( \mathcal{H}_{nT}^d = \frac{1}{nT} \sum_{t=1}^{T} (\tilde{Z}_{nt} G_n \tilde{Z}_{nt} \delta_0)' J_n (\tilde{Z}_{nt} G_n \tilde{Z}_{nt} \delta_0) \), we have\(^{12}\)

\[
\Sigma_{\theta_0,nT}^d = \frac{1}{\sigma_0^2} \begin{pmatrix} E \mathcal{H}_{nT}^d & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{n} \text{tr} (G_n' J_n G_n) + \text{tr} (G_n^2) & 0 \\ 0 & \frac{1}{n} \text{tr} (J_n G_n) \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \frac{2}{n} \text{tr} (J_n G_n) \end{pmatrix}
\]

\[
- \left( \begin{array}{ccc} \frac{1}{\sigma_0^2} E (G_n \tilde{V}_{nt})' J_n \tilde{Z}_{nt} & \frac{2}{\sigma_0^2} E \left[ (G_n \tilde{Z}_{nt} \delta_0)' J_n G_n \tilde{V}_{nt} \right] + \frac{1}{n} \text{tr} (G_n' J_n G_n) & 0 \\ \frac{1}{\sigma_0^2} E (\tilde{Z}_{nt}' J_n \tilde{V}_{nt})' & \frac{2}{\sigma_0^2} E \left[ (G_n \tilde{Z}_{nt} \delta_0)' J_n \tilde{V}_{nt} \right] + \frac{1}{n} \text{tr} (G_n' J_n G_n) & 1 \end{array} \right).
\]

Here, \( \mathcal{H}_{nT}^d \) is the covariance matrix of regressors of the reduced form (C.1) after demeaning from both the time dimension and the cross sectional dimension (see Appendix C.1 for the reduced form).

### 4.2 Asymptotic Properties

Similarly as the transformation approach, we can get the asymptotic properties of the QML estimator. To do that, we need to strengthen Assumption 7 such that both \( n \) and \( T \) are large. This is so in order that the time dummies can be consistently estimated. But, compared to the transformation approach, we don’t need the row normalization of the spatial weights matrix.

**Assumption 1’.** \( W_n \) is a nonstochastic spatial weights matrix.

**Assumption 7’.** \( n \) is an increasing function of \( T \).

**Theorem 4.1** Under Assumptions 1’; 2–6, 7’, if \( \lim_{T \to \infty} E \mathcal{H}_{nT}^d \) is nonsingular or

\[
\lim_{n \to \infty} \frac{1}{n} \ln \left| \sigma_n^2 S_n^{-1} S_n^{-1} \right| - \frac{1}{n} \ln \left| \sigma_n^2 (\lambda) S_n^{-1} (\lambda) S_n^{-1} (\lambda) \right| \neq 0 \quad \text{for} \quad \lambda \neq \lambda_0,
\]

\( \theta_0 \) is globally identified and \( \hat{\theta}_{nT}^d \overset{p}{\to} \theta_0 \).

**Proof.** See Appendix E.3. 

\(^{12}\)The \( \mathcal{H}_{nT}^d \) differs from \( \mathcal{H}_{nT} \) in the transformation approach only in the division by \( n \) in \( \mathcal{H}_{nT}^d \) instead of \( (n - 1) \) in \( \mathcal{H}_{nT} \).
Theorem 4.2

From (4.3) and (3.3), the score can be decomposed into 3 parts such that

\[
\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d}(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial \theta} - \Delta_{1,nT} - \Delta_{2,nT},
\]

where

\[
\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial \theta} = \begin{pmatrix}
\frac{1}{\sigma_0} \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \tilde{Z}_{nt}^{(u)}' \tilde{J}_n V_{nt} \\
\frac{1}{\sigma_0} \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} (G_n \tilde{Z}_{nt}^{(u)} \xi_t)' \tilde{J}_n V_{nt} + \frac{1}{\sigma_0} \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} (V_n' G_n' \tilde{J}_n V_{nt} - \sigma_0^2 tr J_n G_n) \\
\frac{1}{\sigma_0} \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} (V_n' \tilde{J}_n V_{nt} - (n-1)\sigma_0^2)
\end{pmatrix},
\]

\[
\Delta_{1,nT} = \sqrt{\frac{n}{T}} \begin{pmatrix}
\frac{1}{\sigma_0} \frac{T}{T} (J_n \tilde{U}_{nt,-1}, J_n \tilde{W}_{nt,-1}, 0)' \tilde{V}_{nt} \\
\frac{1}{\sigma_0} \frac{T}{T} (J_n G_n (\tilde{U}_{nt,-1}, \tilde{W}_{nt,-1}, 0) \xi_0)' \tilde{V}_{nt} + \frac{1}{\sigma_0} \frac{T}{T} \tilde{V}_{nt}' G_n' \tilde{J}_n \tilde{V}_{nt}
\end{pmatrix},
\]

\[
\Delta_{2,nT} = \begin{pmatrix} 0 \\ \sqrt{\frac{T}{n}} \left( tr G_n - tr J_n G_n \right) \end{pmatrix} = \sqrt{\frac{T}{n}} \begin{pmatrix} 0 \\ \frac{1}{1 - \lambda_0} \frac{1}{2\sigma_0^2} \end{pmatrix}.
\]

The asymptotic distribution of \( \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial \theta} \) can be derived from the central limit theorem for martingale difference arrays (Theorem B.3). For the term \( \Delta_{1,nT} \), from Theorem B.2, \( \Delta_{1,nT} = \sqrt{\frac{T}{n}} a_{\theta_0,n,1} + O(\sqrt{\frac{T}{n}}) + O_p(\frac{1}{\sqrt{T}}) \) where

\[
a_{\theta_0,n,1} = \begin{pmatrix}
\frac{1}{n} tr \left( \left( J_n \sum_{h=0}^{\infty} A_n^h \right) S_n^{-1} \right) \\
\frac{1}{n} tr \left( \left( W_n \sum_{h=0}^{\infty} A_n^h \right) S_n^{-1} \right) \\
\frac{1}{n} \gamma_0 tr (G_n (J_n \sum_{h=0}^{\infty} A_n^h) S_n^{-1}) + \frac{1}{n} \rho_0 tr (G_n W_n (J_n \sum_{h=0}^{\infty} A_n^h) S_n^{-1}) + \frac{1}{n} tr J_n G_n
\end{pmatrix}
\]

is \( O(1) \) and

\[
a_{\theta_0,n,2} = \left( 0, \frac{1}{1 - \lambda_0} \frac{1}{2\sigma_0^2} \right)'.
\]

The distribution of the QML estimator is as follows.

Theorem 4.2 Under Assumption 1', 2-6, 7', 8,

\[
\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n}{T}} b_{\theta_0,n,T,1} + \sqrt{\frac{T}{n}} b_{\theta_0,n,T,2} + O_p \left( \max \left( \sqrt{\frac{n}{T}}, \sqrt{\frac{T}{n^3}}, \sqrt{\frac{1}{T}} \right) \right) \overset{d}{\rightarrow} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_\theta + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}),
\]

where \( b_{\theta_0,n,T,1} = (\Sigma_{\theta_0,n,T})^{-1} a_{\theta_0,n,1} \) and \( b_{\theta_0,n,T,2} = (\Sigma_{\theta_0,n,T})^{-1} a_{\theta_0,n,2} \) are \( O(1) \) and \( a_{\theta_0,n,1}, a_{\theta_0,n,2} \) are defined in (4.8) and (4.9).
When $\frac{n}{T} \to 0$,
\[ n(\hat{\theta}_{nT}^d - \theta_0) + b_{\theta_0,nT;2} \xrightarrow{p} 0. \]  
(4.11)

When $\frac{n}{T} \to k < \infty$,
\[ \sqrt{nT}(\hat{\theta}_{nT}^d - \theta_0) + \sqrt{\kappa}b_{\theta_0,nT,1} + \sqrt{k^{-1}}b_{\theta_0,nT;2} \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}). \]  
(4.12)

When $\frac{n}{T} \to \infty$,
\[ T(\hat{\theta}_{nT}^d - \theta_0) + b_{\theta_0,nT,1} \xrightarrow{p} 0. \]  
(4.13)

**Proof.** See Appendix E.4. ■

From Equation (4.10), the QML estimator has the bias $-\frac{1}{n}b_{\theta_0,nT,1}$ and $-\frac{1}{n}b_{\theta_0,nT;2}$ where
\[ b_{\theta_0,nT,1} = (\Sigma_{\theta_0,nT})^{-1} \cdot a_{\theta_0,n,1}, \]  
(4.14)
\[ b_{\theta_0,nT;2} = (\Sigma_{\theta_0,nT})^{-1} \cdot a_{\theta_0,2}. \]  
(4.15)

and the confidence interval is not centered when $\frac{n}{T} \to k$ where $0 < k < \infty$. Furthermore, when $T$ and $n$ do not have the same order, the presence of $b_{\theta_0,nT,1}$ and $b_{\theta_0,nT;2}$ causes $\hat{\theta}_{nT}$ to have a degenerated distribution. An analytical bias reduction procedure is to correct the bias $B_{1,nT} = -b_{\theta_0,nT,1}$ and $B_{2,nT} = -b_{\theta_0,nT;2}$ by constructing estimator $\hat{B}_{1,nT}, \hat{B}_{2,nT}$ and defining the bias corrected estimator as
\[ \hat{\theta}_{nT}^{d1} = \hat{\theta}_{nT}^d - \hat{B}_{1,nT} - \frac{\hat{B}_{2,nT}}{n}. \]  
(4.16)

From Equation 4.10, we can choose
\[ \hat{B}_{1,nT} = \left[-(\Sigma_{\theta_0,nT})^{-1} \cdot a_{\theta_0,n,1}\right]_{\theta = \hat{\theta}_{nT}^d}, \]  
(4.17)
\[ \hat{B}_{2,nT} = \left[-(\Sigma_{\theta_0,nT})^{-1} \cdot a_{\theta_0,2}\right]_{\theta = \hat{\theta}_{nT}^d}. \]  
(4.18)

We show that when $n/T^3 \to 0$ and $T/n^3 \to 0$, $\hat{\theta}_{nT}^{d1}$ is $\sqrt{nT}$ consistent and asymptotically centered normal. Our result for the bias corrected estimator is as follows.

**Theorem 4.3** Under Assumption 1',2-6,7',8,9, if $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$,
\[ \sqrt{nT}(\hat{\theta}_{nT}^{d1} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1} + \Sigma_{\theta_0}^{-1} \Omega_{\theta_0} \Sigma_{\theta_0}^{-1}). \]  
(4.19)

**Proof.** See Appendix E.5. ■

By comparing Theorem 4.2, Theorem 4.3 with Theorem 3.6 and Theorem 3.8, we can see that both estimators are consistent and have the same limiting distribution. For the transformed approach, the estimator has the bias of the order $O(T^{-1})$ and the bias correction requires $n/T^3 \to 0$. For the direct approach, the estimator has the bias of the order $O(\max(n^{-1}, T^{-1}))$ and the bias correction requires not only $n/T^3 \to 0$ but also $T/n^3 \to 0$. Hence, the transformed approach has an advantage over the direct approach especially when $n$ is relatively small. But the direct approach also has its own merit in that: we don’t need the weights matrix to be row normalized. We conduct a small Monte Carlo to compare the two estimators where we use the same simulated data. Table 2 is the counterpart of Table 1 and Table 4 is the counterpart of Table 3. From the tables, we can see that for the biases of estimates, the transformation approach is smaller than the direct approach when $n$ is relatively small.
Table 1: Biases of Transformation Approach Estimators

<table>
<thead>
<tr>
<th>Case</th>
<th>$T$</th>
<th>$n$</th>
<th>$\theta_0^a$</th>
<th>$\gamma$</th>
<th>$\rho$</th>
<th>$\beta$</th>
<th>$\lambda$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>10</td>
<td>49</td>
<td>$\theta_0^a$</td>
<td>-0.0627</td>
<td>-0.0045</td>
<td>-0.0101</td>
<td>-0.0013</td>
<td>-0.1159</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0038</td>
<td>-0.0023</td>
<td>-0.0026</td>
<td>0.0004</td>
<td>-0.0276</td>
</tr>
<tr>
<td>(2)</td>
<td>10</td>
<td>49</td>
<td>$\theta_0^b$</td>
<td>-0.0697</td>
<td>-0.0093</td>
<td>-0.0132</td>
<td>-0.0086</td>
<td>-0.1183</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0037</td>
<td>-0.0001</td>
<td>-0.0023</td>
<td>-0.0029</td>
<td>-0.0307</td>
</tr>
<tr>
<td>(3)</td>
<td>50</td>
<td>49</td>
<td>$\theta_0^a$</td>
<td>-0.0119</td>
<td>-0.0013</td>
<td>0.0001</td>
<td>0.0000</td>
<td>-0.0217</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
<td>-0.0013</td>
<td>0.0004</td>
<td>0.0001</td>
<td>-0.0022</td>
</tr>
<tr>
<td>(4)</td>
<td>50</td>
<td>49</td>
<td>$\theta_0^b$</td>
<td>-0.0131</td>
<td>-0.0017</td>
<td>-0.0001</td>
<td>-0.0006</td>
<td>-0.0218</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0000</td>
<td>-0.0013</td>
<td>0.0004</td>
<td>-0.0004</td>
<td>-0.0023</td>
</tr>
<tr>
<td>(5)</td>
<td>50</td>
<td>16</td>
<td>$\theta_0^a$</td>
<td>-0.0121</td>
<td>-0.0032</td>
<td>-0.0013</td>
<td>-0.0018</td>
<td>-0.0247</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0002</td>
<td>-0.0033</td>
<td>-0.0009</td>
<td>-0.0018</td>
<td>-0.0052</td>
</tr>
<tr>
<td>(6)</td>
<td>50</td>
<td>16</td>
<td>$\theta_0^b$</td>
<td>-0.0134</td>
<td>-0.0043</td>
<td>-0.0013</td>
<td>-0.0017</td>
<td>-0.0247</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0004</td>
<td>-0.0042</td>
<td>-0.0008</td>
<td>-0.0015</td>
<td>-0.0052</td>
</tr>
</tbody>
</table>

Note: $\theta_0^a = (0.2, 0.2, 1, 0.2, 1)$ and $\theta_0^b = (0.3, 0.3, 1, 0.3, 1)$.

Table 2: Biases of Direct Approach Estimators

<table>
<thead>
<tr>
<th>Case</th>
<th>$T$</th>
<th>$n$</th>
<th>$\theta_0^a$</th>
<th>$\gamma$</th>
<th>$\rho$</th>
<th>$\beta$</th>
<th>$\lambda$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>10</td>
<td>49</td>
<td>$\theta_0^a$</td>
<td>-0.0610</td>
<td>-0.0014</td>
<td>-0.0092</td>
<td>-0.0312</td>
<td>-0.1318</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0035</td>
<td>0.0026</td>
<td>-0.0022</td>
<td>-0.0009</td>
<td>-0.0298</td>
</tr>
<tr>
<td>(2)</td>
<td>10</td>
<td>49</td>
<td>$\theta_0^b$</td>
<td>-0.0669</td>
<td>-0.0014</td>
<td>-0.0117</td>
<td>-0.0391</td>
<td>-0.1327</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0032</td>
<td>0.0115</td>
<td>-0.0015</td>
<td>-0.0008</td>
<td>-0.0333</td>
</tr>
<tr>
<td>(3)</td>
<td>50</td>
<td>49</td>
<td>$\theta_0^a$</td>
<td>-0.0100</td>
<td>0.0051</td>
<td>0.0010</td>
<td>-0.0296</td>
<td>-0.0394</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0007</td>
<td>-0.0019</td>
<td>-0.0031</td>
</tr>
<tr>
<td>(4)</td>
<td>50</td>
<td>49</td>
<td>$\theta_0^b$</td>
<td>-0.0099</td>
<td>0.0108</td>
<td>0.0016</td>
<td>-0.0297</td>
<td>-0.0379</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0002</td>
<td>0.0014</td>
<td>0.0007</td>
<td>-0.0007</td>
<td>-0.0033</td>
</tr>
<tr>
<td>(5)</td>
<td>50</td>
<td>16</td>
<td>$\theta_0^a$</td>
<td>-0.0067</td>
<td>0.0115</td>
<td>-0.0030</td>
<td>-0.0973</td>
<td>-0.0842</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0012</td>
<td>0.0013</td>
<td>0.0006</td>
<td>-0.0199</td>
<td>-0.0130</td>
</tr>
<tr>
<td>(6)</td>
<td>50</td>
<td>16</td>
<td>$\theta_0^b$</td>
<td>-0.0032</td>
<td>0.0272</td>
<td>-0.0008</td>
<td>-0.1038</td>
<td>-0.0785</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0026</td>
<td>0.0082</td>
<td>0.0014</td>
<td>-0.0213</td>
<td>-0.0117</td>
</tr>
</tbody>
</table>

Note: $\theta_0^a = (0.2, 0.2, 1, 0.2, 1)$ and $\theta_0^b = (0.3, 0.3, 1, 0.3, 1)$. 
<table>
<thead>
<tr>
<th>Case</th>
<th>Empirical S.E. of $\hat{\theta}<em>{nT}$ (1st line) and $\hat{\theta}</em>{nT}^1$ (2nd line)</th>
<th>Estimated S.E. of $\hat{\theta}<em>{nT}$ (1st line) and $\hat{\theta}</em>{nT}^1$ (2nd line)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T$</td>
<td>$n$</td>
</tr>
<tr>
<td>(1)</td>
<td>10</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>49</td>
</tr>
<tr>
<td>(3)</td>
<td>50</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>49</td>
</tr>
<tr>
<td>(5)</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>16</td>
</tr>
</tbody>
</table>

17
<table>
<thead>
<tr>
<th>Case</th>
<th>Empirical S.E. of $\hat{\theta}<em>{nT}^d$ (1st line) and $\hat{\theta}</em>{nT}^{dl}$ (2nd line)</th>
<th>Estimated S.E. of $\hat{\theta}<em>{nT}^d$ (1st line) and $\hat{\theta}</em>{nT}^{dl}$ (2nd line)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T$</td>
<td>$n$</td>
</tr>
<tr>
<td>(1)</td>
<td>10</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>50</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>50</td>
<td>16</td>
</tr>
</tbody>
</table>
5 Conclusion

This paper establishes asymptotic properties of quasi-maximum likelihood estimator for spatial dynamic panel data with both individual effects and time effects when both the number of individuals $n$ and the number of time periods $T$ can be large. A possible approach is to estimate both effects directly. But instead of the direct estimation, we propose a data transformation approach so that the magnitude of the bias is only of the order $O(T^{-1})$ rather than $O(\max(T^{-1},n^{-1}))$ in the direct estimation approach. When $n$ is asymptotically proportional to $T$, the estimator is $\sqrt{nT}$ consistent and asymptotically normal, but the limit distribution is not centered around 0; when $T$ is relatively larger than $n$, the estimator is $\sqrt{nT}$ consistent and asymptotically centered normal; when $T$ is relatively smaller than $n$, the estimator is consistent with rate $T$ and has a degenerate limit distribution. We also propose a bias correction for our estimator. We show that when $T$ grows faster than $n^{1/3}$, the correction will asymptotically eliminate the bias and yield a centered confidence interval. This transformation approach has advantage over the direct approach especially when $n$ is relatively small. When $n$ is relatively small, the estimates of the transformed approach has faster rates of convergence than that of the direct estimates. The direct estimates have a degenerate limit distribution but the transformed estimates are properly centered and are asymptotically normal.
Appendices

A Some Notes

A.1 Data Transformation

Let \( J_n = I_n - \frac{1}{n} l_n l_n' \) be the "deviation from the group mean" transformation. As \( I_n = J_n + \frac{1}{n} l_n l_n' \) and \( W_n l_n = l_n \), we have

\[
J_n W_n = J_n W_n (J_n + \frac{1}{n} l_n l_n') = J_n W_n J_n,
\]

because \( J_n W_n l_n = J_n l_n = 0 \). Then, from (2.1), for \( t = 1, 2, ..., T \), we have

\[
(J_n Y_{nt}) = \lambda_0 (J_n W_n) (J_n Y_{nt}) + \gamma_0 (J_n Y_{n,t-1}) + \rho_0 (J_n W_n) (J_n Y_{n,t-1}) + (J_n X_{nt}) \beta_0 + (J_n c_{n0}) + (J_n V_{nt}),
\]

(A.1)

which does not involve the time effects and \( J_n c_{n0} \) can be regarded as the individual effects. A special feature of the transformed equation (A.1) is that the variance matrix of \( J_n V_{nt} \) is equal to \( \sigma_0^2 J_n \), which is of the rank \( n - 1 \).

The variance of \( J_n V_{nt} \) is \( \sigma_0^2 J_n' J_n' = \sigma_0^2 J_n \) and the elements of \( J_n V_{nt} \) are correlated. Here, \( J_n \) is singular because \( J_n \) has rank \( n - 1 \) as \( J_n \) is an orthogonal projector with trace \( n - 1 \). Hence, there is a linear dependence among the elements of \( J_n V_{nt} \). An effective estimation method shall eliminate the linear dependence in sample observations. This can be done with the eigenvalues and eigenvectors decomposition as in the theory of generalized inverse for the estimation of linear regression models (see, e.g., Theil (1971), Ch. 6).

The eigenvalues of \( J_n \) are a single zero and \( n - 1 \) ones. An eigenvector corresponding to the zero eigenvalue is proportional to \( l_n \). Let \( (F_{n,n-1}, l_n/\sqrt{n}) \) be the orthonormal matrix of \( J_n \) where \( F_{n,n-1} \) corresponds to the eigenvalues of ones and \( l_n/\sqrt{n} \) corresponds to the eigenvalue zero. Thus,

\[
J_n F_{n,n-1} = F_{n,n-1}, \quad F_{n,n-1}' F_{n,n-1} = I_{n-1}, \quad J_n l_n = 0, \quad F_{n,n-1}' l_n = 0, \quad F_{n,n-1}' l_n' = I_n, \quad F_{n,n-1}' F_{n,n-1} = J_n.
\]

(A.2)

To eliminate the linear dependence in \( J_n V_{nt} \), the transformation of \( J_n Y_{nt} \) to \( Y_{nt}^* \), where \( Y_{nt}^* = F_{n,n-1}' J_n Y_{nt} \) is a vector with dimension \( n - 1 \), gives

\[
Y_{nt}^* = \lambda_0 F_{n,n-1}' (J_n W_n)(J_n Y_{nt}) + \gamma_0 F_{n,n-1}' (J_n Y_{n,t-1}) + \rho_0 F_{n,n-1}' (J_n W_n)(J_n Y_{n,t-1}) + F_{n,n-1}' J_n X_{nt} \beta_0 + F_{n,n-1}' J_n c_{n0} + F_{n,n-1}' J_n V_{nt}
\]

\[
= \lambda_0 F_{n,n-1}' J_n W_n J_n Y_{nt} + \gamma_0 Y_{n,t-1}^* + \rho_0 F_{n,n-1}' J_n W_n J_n Y_{n,t-1} + X_{nt}^* \beta_0 + c_{n0}^* + V_{nt}^*,
\]

(A.3)

where \( Y_{nt}^* = F_{n,n-1}' J_n Y_{nt}, X_{nt}^* = F_{n,n-1}' J_n X_{nt}, c_{n0}^* = F_{n,n-1}' J_n c_{n0} \) and \( V_{nt}^* = F_{n,n-1}' J_n V_{nt} \). Observe that \( J_n W_n = J_n W_n (F_{n,n-1}' F_{n,n-1} + \frac{1}{n} l_n l_n') = J_n W_n F_{n,n-1}' F_{n,n-1} \) because \( J_n l_n = 0 \). Denote \( W_n' = F_{n,n-1}' J_n W_n F_{n,n-1} = F_{n,n-1}' W_n F_{n,n-1} \) (as \( F_{n,n-1}' l_n = 0 \)). One arrives at
\[ Y_{nt}^* = \lambda_0 W_n^* Y_{nt}^* + \gamma_0 Y_{n,t-1}^* + \rho_0 W_n^* X_{n,t-1}^* + X_{nt}^* \beta_0 + c_{nt}^* + V_{nt}^* , \]  

(A.4)

where \( V_{nt}^* \) is a \((n - 1)\) dimensional disturbance vector with zero mean and variance matrix \( \sigma_0^2 I_{n-1} \). Furthermore, as \( l_n' J_n Y_{nt} = 0 \), \( l_n' J_n W_n Y_{nt} = 0 \), \( l_n' J_n X_{nt} = 0 \), the singularity of \( J_n V_{nt} \) does not impose deterministic constraints on the parameters (Theil (1971)). This is expected as \((A.4)\) is derived by eliminating the time effects and there is no deterministic constraints imposed on the parameters in \((2.1)\) to begin with.

Equation \((A.4)\) shall provide the estimation of the structural parameters in the model. This equation is in the format of a typical spatial autoregressive model in panel data, where the number of observations is \( T(n - 1) \), reduced from the original sample observations by one for each period. Equation \((A.4)\) is useful as it motivates the derivation of the likelihood function for \( Y_{nt}^* \) in our approach.

**A.2 Determinant and Inverse of \( I_{n-1} - \lambda W_n^* \)**

We note that \((I_{n-1} - \lambda W_n^* ) = F_{n,n-1}' (I_n - \lambda W_n) F_{n,n-1} \). For the determinant, we have

\[
[F_{n,n-1}, l_n'/\sqrt{n}](I_n - \lambda W_n)[F_{n,n-1}, l_n'/\sqrt{n}]
\]

\[
= \begin{pmatrix}
F_{n,n-1}' (I_n - \lambda W_n) F_{n,n-1} & F_{n,n-1}' (I_n - \lambda W_n)(l_n'/\sqrt{n}) \\
(l_n'/\sqrt{n})(I_n - \lambda W_n) F_{n,n-1} & (l_n'/\sqrt{n})(I_n - \lambda W_n)(l_n'/\sqrt{n})
\end{pmatrix}
\]

\[
= \begin{pmatrix}
F_{n,n-1}' (I_n - \lambda W_n) F_{n,n-1} & 0 \\
-\lambda l_n' W_n F_{n,n-1}/\sqrt{n} & 1 - \lambda
\end{pmatrix},
\]

because \( F_{n,n-1}' W_n l_n = F_{n,n-1}' l_n = 0 \) and \( \frac{1}{n} l_n' W_n l_n = 1 \). Hence,

\[
|I_{n-1} - \lambda W_n^* | = |F_{n,n-1}' (I_n - \lambda W_n) F_{n,n-1} | = \frac{1}{1 - \lambda} |I_n - \lambda W_n| .
\]

Thus, the tractability in computing the determinant of \( I_{n-1} - \lambda W_n^* \) is exactly that of \( I_n - \lambda W_n \). When \( W_n \) is constructed as a weights matrix row normalized from an original symmetric matrix, Ord (1975) has suggested a computationally tractable method for the evaluation of \(|I_n - \lambda W_n| \) at various \( \lambda \) for the ML method. Thus, this will also be useful for evaluating the determinant of \((I_{n-1} - \lambda W_n^* )\) even though the row sums of the transformed spatial weights matrix \( W_n^* \) may not even be unity.

Furthermore, a spatial autoregressive model is an equilibrium model in the sense that the observed outcomes are determined by the equation. That is, the matrix \( I_{n-1} - \lambda W_n^* \) shall be invertible. For the transformed equation \((2.6)\), \( I_{n-1} - \lambda W_n^* \) is invertible as long as the original matrices \( I_n - \lambda W_n \) in \((2.1)\) is invertible. This is so as it shall be shown below.

The inverse of \( I_{n-1} - \lambda W_n^* \) is

\[
(I_{n-1} - \lambda W_n^*)^{-1} = F_{n,n-1}' (I_n - \lambda W_n)^{-1} F_{n,n-1} ,
\]

(A.6)
because we have

\[(I_n - \lambda W_n) \cdot F_{n,n-1}'(I_n - \lambda W_n)^{-1} F_{n,n-1} = F_{n,n-1}'(I_n - \lambda W_n)F_{n,n-1} - F_{n,n-1}'(I_n - \lambda W_n)^{-1} F_{n,n-1} = F_{n,n-1}'(I_n - \lambda W_n)J_n(I_n - \lambda W_n)^{-1} F_{n,n-1} = F_{n,n-1}' F_{n,n-1} - F_{n,n-1}'(I_n - \lambda W_n)^{\frac{l}{n}} F_{n,n-1} = I_n - 1,
\]

as \( F_{n,n-1}' l_n = 0 \) and \( F_{n,n-1}' W_n l_n = 0 \). We have also

\[W_n'(I_n - \lambda W_n)^{-1} = F_{n,n-1}'W_n F_{n,n-1} F_{n,n-1}'(I_n - \lambda W_n)^{-1} F_{n,n-1} = F_{n,n-1}'W_n'(I_n - \lambda W_n)^{-1} F_{n,n-1},
\]

and

\[(I_n - \lambda W_n)^{-1} W_n'(I_n - \lambda W_n)^{-1} = F_{n,n-1}'W_n(I_n - \lambda W_n)^{-1} F_{n,n-1}.
\]

A.3 About \( tr[(J_n G_n(\lambda))^m] \)

As is shown in (A.6) that \( S_n^{-1}(\lambda) = (I_n - \lambda W_n)^{-1} = F_{n,n-1}' S_n^{-1}(\lambda) F_{n,n-1} \), it follows that \( G_n^m(\lambda) = W_n^m S_n^{-1}(\lambda) = F_{n,n-1}' W_n F_{n,n-1} \cdot F_{n,n-1}' S_n^{-1}(\lambda) F_{n,n-1} = F_{n,n-1}' W_n J_n S_n^{-1}(\lambda) F_{n,n-1} = F_{n,n-1}' G_n(\lambda) F_{n,n-1} \) because \( F_{n,n-1}' W_n J_n = F_{n,n-1}' W_n \). Furthermore, as \( G_n(\lambda) l_n = \frac{1}{(1-\lambda)^m} l_n \) and \( F_{n,n-1}' G_n(\lambda) l_n = 0 \), by induction, one has

\[G_n^m(\lambda) = F_{n,n-1}' G_n^m(\lambda) F_{n,n-1}, \quad m = 0, 1, 2, \cdots .
\]

From this, \( tr(G_n^m(\lambda)) = tr(J_n G_n^m(\lambda)) \) because \( F_{n,n-1}' F_{n,n-1} = J_n \). Furthermore, because \( G_n^m(\lambda) l_n = \frac{1}{(1-\lambda)^m} l_n \),

\[tr(G_n^m(\lambda)) - tr(G_n^m(\lambda)) = \frac{-1}{n} tr(G_n^m(\lambda) l_n l_n') = \frac{-1}{n(1-\lambda^m)} tr(l_n l_n') = \frac{-1}{(1-\lambda)^m}, \quad m = 1, 2, \cdots .
\]

B Lemmas for Some Statistics in the Model

The following lemmas and theorems can be found in Yu, de Jong and Lee (2006).

Let \( V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})' \) be \( n \times 1 \) column vector. We assume that \( \{v_{it}\}, i = 1, 2, \cdots, n \) and \( t = 1, 2, \cdots, T \), are i.i.d. across \( i \) and \( t \) with zero mean, variance \( \sigma_i^2 \) and \( E |v_{it}|^{4+\eta} < \infty \) for some \( \eta > 0 \). Also, let \( D_{nt} \) be \( n \times 1 \) vector of uniformly bounded constants for all \( n \) and \( t \). Denote

\[U_{nt} = \sum_{h=1}^{\infty} P_{nh} V_{n,t+1-h}, \quad (B.1)
\]

where \( \{P_{nh}\}_{h=1}^{\infty} \) is a sequence of \( n \times n \) nonstochastic square matrices.

**Assumption A1.** The disturbances \( \{v_{it}\}, i = 1, 2, \ldots, n \) and \( t = 1, 2, \ldots, T \), are i.i.d across \( i \) and \( t \) with zero mean, variance \( \sigma_i^2 \) and \( E |v_{it}|^{4+\eta} < \infty \) for some \( \eta > 0 \).
**Assumption A2.** The row and column sums of $\sum_{n=1}^{\infty} \text{abs}(P_{hn})$ are bounded uniformly in $n$.

**Assumption A3.** The elements of $n \times 1$ vector $D_{nt}$ are nonstochastic and bounded, uniformly in $n$ and $t$.

**Assumption A4.** $n$ is a nondecreasing function of $T$.

**Lemma B.1** Under Assumptions A1 and A4, for an $n \times n$ nonstochastic matrix $B_n$, uniformly bounded in row and column sums,

$$\frac{1}{nT} \sum_{t=1}^{T} V'_{nt} B_n V_{nt} - E(\frac{1}{nT} \sum_{t=1}^{T} V'_{nt} B_n V_{nt}) = O_p \left( \frac{1}{\sqrt{nT}} \right), \quad (B.2)$$

$$\frac{1}{n} V'_{nt} B_n V_{nt} - E(\frac{1}{n} V'_{nt} B_n V_{nt}) = O_p \left( \frac{1}{nT^2} \right), \quad (B.3)$$

and

$$\frac{1}{nT} \sum_{t=1}^{T} V'_{nt} B_n V_{nt} - E(\frac{1}{nT} \sum_{t=1}^{T} V'_{nt} B_n V_{nt}) = O_p \left( \frac{1}{nT} \right), \quad (B.4)$$

where $E(\frac{1}{nT} \sum_{t=1}^{T} V'_{nt} B_n V_{nt}) = \frac{1}{n} \sigma_0^2 \text{tr}(B_n) = O(1)$ and $E(\frac{1}{n} V'_{nt} B_n V_{nt}) = \frac{1}{nT} \sigma_0^2 \text{tr}(B_n) = O \left( \frac{1}{T} \right)$.

**Theorem B.2** Under Assumptions A1, A2 and A4,

$$\sqrt{\frac{T}{n}} \left( \bar{U}^T_{nT,-1} \bar{V}_{nT} - E \left( \bar{U}^T_{nT,-1} \bar{V}_{nT} \right) \right) = O_p \left( \frac{1}{\sqrt{T}} \right), \quad (B.5)$$

where $\sqrt{\frac{T}{n}} E \left( \bar{U}^T_{nT,-1} \bar{V}_{nT} \right) = \sqrt{\frac{T}{n}} \sigma_0^2 \text{tr} \left( \sum_{k=1}^{\infty} P_{nh} \right) + O \left( \frac{\sqrt{T}}{n} \right)$ when $T \to \infty$.

For the theorem that follows, we will consider the following form:

$$Q_{nT} = \sum_{t=1}^{T} (V'_{nt} V_{nt} + D'_{nt} V_{nt} + V'_{nt} B_n V_{nt} - \sigma_0^2 \text{tr} B_n) = \sum_{t=1}^{T} \sum_{i=1}^{n} z_{nt,i},$$

where $B_n$ is a $n \times n$ nonstochastic symmetric matrix which is uniformly bounded in both row and column sums, and $z_{nt,i} = (u_{i,t-1} + d_{nti}) v_{it} + b_{n,ii}(v_{it}^2 - \sigma_0^2) + 2(\sum_{j=1}^{i-1} b_{n,ij} v_{jt}) v_{it}$, where $b_{n,ij}$ is the $(i, j)$ element of $B_n$ and $d_{nti}$ is the $i$th element of $D_{nt}$. Then, for the mean and variance of $Q_{nT}$, $\mu_{Q_{nT}} = 0$ and

$$\sigma_{Q_{nT}}^2 = T \sigma_0^4 \text{tr} \left( \sum_{h=1}^{\infty} P_{nh} P_{nh} \right) + \sigma_0^2 \sum_{t=1}^{T} D'_{nt} D_{nt} + T \left( (\mu_4 - 3\sigma_0^4) \sum_{i=1}^{n} b_{n,ii}^2 + 2\sigma_0^4 \text{tr}(B_n^2) \right) + 2\mu_3 \sum_{t=1}^{T} \sum_{i=1}^{n} d_{nti} b_{n,ii},$$

where $\mu_s = E v_{it}^s$ for $s = 3, 4$.

**Theorem B.3** Under Assumptions A1, A2, A3, A4 and that row and column sums of $B_n$ are bounded uniformly in $n$, if the sequence $\frac{1}{nT} \sigma_{Q_{nT}}^2$ is bounded away from zero, then,

$$\frac{Q_{nT}}{\sigma_{Q_{nT}}} \overset{d}{\sim} N(0, 1). \quad (B.6)$$
Denote $Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt})$, we are going to provide some lemmas related to $J_n \hat{Z}_{nt}$, $J_n \tilde{Z}_{nT}$ and $\tilde{V}_{nt}, \tilde{V}_{nT}$ of the model (2.1).

**Lemma B.4** Under Assumptions 1-7, for an $n \times n$ nonstochastic matrix $B_n$, uniformly bounded in row and column sums,

$$
\frac{1}{nT} \sum_{t=1}^{T} (J_n \hat{Z}_{nt})' B_n (J_n \hat{Z}_{nt}) - E \frac{1}{nT} \sum_{t=1}^{T} (J_n \hat{Z}_{nt})' B_n (J_n \hat{Z}_{nt}) = O_p \left( \frac{1}{\sqrt{nT}} \right),
$$

(7)

where $E \frac{1}{nT} \sum_{t=1}^{T} (J_n \hat{Z}_{nt})' B_n (J_n \hat{Z}_{nt})$ is $O(1)$;

$$
\frac{1}{nT} \sum_{t=1}^{T} (J_n \tilde{Z}_{nt})' B_n \tilde{V}_{nt} - E \frac{1}{nT} \sum_{t=1}^{T} (J_n \tilde{Z}_{nt})' B_n \tilde{V}_{nt} = O_p \left( \frac{1}{\sqrt{nT}} \right),
$$

(8)

where $E \frac{1}{nT} \sum_{t=1}^{T} (J_n \hat{Z}_{nt})' B_n \tilde{V}_{nt}$ is $O \left( \frac{1}{n} \right)$.

**Lemma B.5** Under Assumptions 1-7, for an $n \times n$ nonstochastic matrix $B_n$, uniformly bounded in row and column sums,

$$
\frac{1}{n} (J_n \hat{Z}_{nT})' B_n (J_n \hat{Z}_{nT}) - E \frac{1}{n} (J_n \hat{Z}_{nT})' B_n (J_n \hat{Z}_{nT}) = O_p \left( \frac{1}{\sqrt{nT}} \right),
$$

(9)

where $E \frac{1}{n} (J_n \hat{Z}_{nT})' B_n (J_n \hat{Z}_{nT})$ is $O(1)$;

$$
\frac{1}{n} (J_n \tilde{Z}_{nT})' B_n \tilde{V}_{nT} - E \frac{1}{n} (J_n \tilde{Z}_{nT})' B_n \tilde{V}_{nT} = O_p \left( \frac{1}{\sqrt{nT}} \right),
$$

(10)

where $E \frac{1}{n} (J_n \hat{Z}_{nT})' B_n \tilde{V}_{nT}$ is $O \left( \frac{1}{n} \right)$.

Also, from Equation (3.3),

$$
J_n \hat{Z}_{nt} = J_n \hat{Z}_{nt}^{(u)} - (J_n \tilde{U}_{nT,-1}, J_n W_n \tilde{U}_{nT,-1}, 0),
$$

where $\hat{Z}_{nt}^{(u)} = ((\tilde{x}_{n,t-1} + U_{n,t-1}), (W_n \tilde{x}_{n,t-1} + W_n U_{n,t-1}), X_{nt})$ with $X_{n,t-1} = \hat{x}_{n,t-1}$ and $\tilde{x}_{nT,-1}$. Hence $J_n \hat{Z}_{nt}$ has two components: one is $J_n \hat{Z}_{nt}^{(u)}$, which is uncorrelated with $V_{nt}$; the other is $- (J_n \tilde{U}_{nT,-1}, J_n W_n \tilde{U}_{nT,-1}, 0)$, but is correlated with $V_{nt}$ when $t \leq T - 1$. Following is a lemma related to $\hat{Z}_{nt}^{(u)}$ and $\tilde{Z}_{nt}$.

**Lemma B.6** Under Assumptions 1-7, for an $n \times n$ nonstochastic matrix $B_n$, uniformly bounded in row and column sums,

$$
E \frac{1}{nT} \sum_{t=1}^{T} \hat{Z}_{nt}' J_n B_n J_n \hat{Z}_{nt} - E \frac{1}{nT} \sum_{t=1}^{T} \hat{Z}_{nt}^{(u)'} J_n B_n J_n \hat{Z}_{nt}^{(u)} = O \left( \frac{1}{T} \right),
$$

(11)

where $E \frac{1}{nT} \sum_{t=1}^{T} \hat{Z}_{nt}^{(u)'} J_n B_n J_n \hat{Z}_{nt}^{(u)}$ is $O(1)$. 

24
C Concentrated QML of Transformation Approach

C.1 Reduced Form of (2.1)

From Equation (2.1), we have $Y_{nt} = S^{-1}_n(Z_{nt}\delta_0 + c_{n0} + \alpha t l_n + V_{nt})$ and $W_n Y_{nt} = G_n Z_{nt}\delta_0 + G_n c_{n0} + \alpha t G_n l_n + G_n V_{nt}$. Hence, using $S^{-1}_n = I_n + \lambda_0 G_n$, we have $Y_{nt} = Z_{nt}\delta_0 + \lambda_0 G_n Z_{nt}\delta_0 + S^{-1}_n c_{n0} + \alpha t S^{-1}_n l_n + S^{-1}_n V_{nt}$. Using $S^{-1}_n l_n = \frac{1}{1-\lambda_0} l_n$ and $J_n l_n = 0$, we have

$$
\tilde{Y}_{nt} = \tilde{Z}_{nt}\delta_0 + \lambda_0 G_n \tilde{Z}_{nt}\delta_0 + \frac{\tilde{\alpha}_t}{1-\lambda_0} l_n + \tilde{S}^{-1}_n \tilde{V}_{nt},
$$

and

$$
J_n \tilde{Y}_{nt} = J_n \tilde{Z}_{nt}\delta_0 + \lambda_0 J_n G_n \tilde{Z}_{nt}\delta_0 + J_n \tilde{S}^{-1}_n \tilde{V}_{nt}.
$$

(C.1)

Similarly, as we have $W_n \tilde{Y}_{nt} = G_n \tilde{Z}_{nt}\delta_0 + \tilde{\alpha}_t G_n l_n + G_n \tilde{V}_{nt}$, we can get

$$
J_n W_n \tilde{Y}_{nt} = J_n G_n \tilde{Z}_{nt}\delta_0 + J_n G_n \tilde{V}_{nt},
$$

(C.2)

because $J_n G_n l_n = J_n W_n S^{-1}_n l_n = J_n W_n J_n S^{-1}_n l_n = 0$.

C.2 FOC and SOC of MLE

For the concentrated likelihood function (2.11), the first order derivatives are

$$
\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta} = \begin{pmatrix}
\frac{\partial \ln L_{n,T}(\theta)}{\partial \lambda_0} \\
\frac{\partial \ln L_{n,T}(\theta)}{\partial \alpha_t} \\
\frac{\partial \ln L_{n,T}(\theta)}{\partial \sigma^2}
\end{pmatrix} = \begin{pmatrix}
\frac{1}{\sigma^2} \sum_{t=1}^{T} (J_n \tilde{Z}_{nt})' \tilde{V}_{nt}(\theta) \\
\frac{1}{\sigma^2} \sum_{t=1}^{T} (J_n W_n \tilde{Y}_{nt})' \tilde{V}_{nt}(\theta) - T \text{tr}(J_n G_n(\lambda)) \\
\frac{1}{2\sigma^4} \sum_{t=1}^{T} (\tilde{V}_{nt}(\theta)') J_n \tilde{V}_{nt}(\theta) - (n-1)\sigma^2)
\end{pmatrix},
$$

and the second order derivatives are

$$
\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial \theta \partial \theta'} = -\begin{pmatrix}
\frac{1}{\sigma^2} \sum_{t=1}^{T} \tilde{Z}_{nt}' J_n \tilde{Z}_{nt} & \frac{1}{\sigma^2} \sum_{t=1}^{T} \tilde{Z}_{nt}' J_n W_n \tilde{Y}_{nt} & \frac{1}{\sigma^2} \sum_{t=1}^{T} \tilde{Z}_{nt}' J_n \tilde{V}_{nt}(\theta) \\
* & \frac{1}{\sigma^2} \sum_{t=1}^{T} (W_n \tilde{Y}_{nt})' J_n W_n \tilde{Y}_{nt} + T \text{tr}(J_n G_n(\lambda))^2 & \frac{1}{\sigma^2} \sum_{t=1}^{T} (W_n \tilde{Y}_{nt})' J_n \tilde{V}_{nt}(\theta) \\
* & * & -\frac{(n-1)\sigma^4}{2\sigma^4} + \frac{1}{\sigma^2} \sum_{t=1}^{T} \tilde{V}_{nt}(\theta)' J_n \tilde{V}_{nt}(\theta)
\end{pmatrix}.
$$

Using $\text{tr}(G_n(\lambda)) - \text{tr}(J_n G_n(\lambda)) = \frac{1}{1-\lambda}$ and $\text{tr}(G^2_n(\lambda)) - \text{tr}((J_n G_n(\lambda))^2) = \frac{1}{(1-\lambda)^2}$ (see Appendix A.3), we can get
\[
\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta} = \left( \begin{array}{c}
\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta} \\
\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta} \\
\frac{\partial \ln L_{n,T}(\theta)}{\partial \sigma^2}
\end{array} \right) = \left( \begin{array}{c}
\frac{1}{\sigma} \sum_{t=1}^{T} (J_n \hat{\theta}_{nt})' \hat{\theta}_{nt} \\
\frac{1}{\sigma} \sum_{t=1}^{T} (J_n W_n \hat{\theta}_{nt})' \hat{\theta}_{nt} + T \text{tr} G_n (\lambda) + \frac{T}{1-\lambda} \\
\frac{1}{2\sigma^2} \sum_{t=1}^{T} (\hat{\theta}_{nt})' J_n \hat{\theta}_{nt} - (n-1)\sigma^2
\end{array} \right),
\]

and the second order derivatives are

\[
\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial \theta \partial \theta'} = \left( \begin{array}{ccc}
\frac{1}{\sigma^2} \sum_{t=1}^{T} \hat{\theta}_{nt}' J_n \hat{\theta}_{nt} & \frac{1}{\sigma^2} \sum_{t=1}^{T} \hat{\theta}_{nt}' J_n W_n \hat{\theta}_{nt} & \frac{1}{\sigma^2} \sum_{t=1}^{T} \hat{\theta}_{nt}' J_n \hat{\theta}_{nt} \\
* & \frac{1}{\sigma^2} \sum_{t=1}^{T} (W_n \hat{\theta}_{nt})' J_n W_n \hat{\theta}_{nt} + T \text{tr} (G_n^2 (\lambda)) - \frac{T}{(1-\lambda)^2} & \frac{1}{\sigma^2} \sum_{t=1}^{T} (W_n \hat{\theta}_{nt})' J_n \hat{\theta}_{nt} \\
* & * & -(n-1)\sigma^2 + \frac{T}{\sigma^2} \sum_{t=1}^{T} \hat{\theta}_{nt}' J_n \hat{\theta}_{nt}
\end{array} \right). \tag{C.4}
\]

**C.3 The Variance of the Gradient**

We are going to show that for \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(n)}(\theta_0)}{\partial \theta} \), \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(n)}(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(n)}(\theta_0)}{\partial \sigma^2} \) = \( \Sigma_{\theta_0,T} + \Omega_{\theta_0,n} + O(T^{-1}) \) where \( \Omega_{\theta_0,n} = \frac{\mu_T - 3\sigma^2}{\sigma^2} \left( \begin{array}{ccc}
0 & 0 & 0 \\
\frac{1}{n-1} \sum_{i=1}^{n} [(J_n G_n)^2]_{ii} & \frac{1}{(n-1)\sigma^2} \text{tr} J_n G_n & 0 \\
0 & \frac{2\sigma^2}{(n-1)\sigma^2} \text{tr} J_n G_n & \frac{4\sigma^2}{(n-1)\sigma^2}
\end{array} \right) \). First, we can write \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(n)}(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(n)}(\theta_0)}{\partial \sigma^2} = \)

\[
E \left[ \frac{1}{(n-1)T} \left( \begin{array}{c}
\frac{1}{\sigma^2} \sum_{t=1}^{T} (G_n \hat{\theta}_{nt} (\delta_0))' J_n \hat{\theta}_{nt} \\
\frac{1}{\sigma^2} \sum_{t=1}^{T} (V_n G_n' J_n \hat{\theta}_{nt} - \sigma^2_0 \text{tr} J_n G_n) (\sum_{t=1}^{T} \hat{\theta}_{nt}' J_n \hat{\theta}_{nt}) \\
\frac{1}{2\sigma^2} \sum_{t=1}^{T} (V_n J_n V_n - (n-1)\sigma^2_0) \left( \sum_{t=1}^{T} \hat{\theta}_{nt}' J_n \hat{\theta}_{nt} \right)^2 \\
0 & 0 & 0
\end{array} \right) \right]
\]

As \( \hat{\theta}_{nt} \) is un-
correlated with \( V_{nt} \), we have

\[
E \left( \frac{1}{\sqrt{(n-1)T}} \frac{\partial}{\partial \theta} L_{nT}^{(u)}(\theta_0) \right) \cdot \frac{1}{\sqrt{(n-1)T}} \frac{\partial}{\partial \theta} L_{nT}^{(u)}(\theta_0) = \left( \begin{array}{ccc} \frac{1}{\sigma_0^2(n-1)T} E \sum_{t=1}^{T} \tilde{Z}_{nt}^{(u)T} J_n \tilde{Z}_{nt}^{(u)} & \ast & \ast \\ \ast & \frac{1}{\sigma_0^2(n-1)T} E \sum_{t=1}^{T} (G_n \tilde{Z}_{nt}^{(u)})' J_n G_n \tilde{Z}_{nt}^{(u)} & \ast \\ 0 & 0 & 1 \end{array} \right) \cdot \left( \begin{array}{ccc} \ast & \ast & \ast \\ \ast & \frac{1}{\sigma_0^2(n-1)T} tr(J_n G_n) & \frac{1}{2\sigma_0^2} \\ 0 & 0 & \ast \end{array} \right).
\]

The first matrix is equal to \( \Sigma_{\theta_0,nt} + O(T^{-1}) \) using Lemma B.6. The second matrix is equal to \( \Omega_{\theta_0,n} \) using \( E \sum_{t=1}^{T} \tilde{Z}_{nt}^{(u)} = 0 \) and \( E \sum_{t=1}^{T} G_n \tilde{Z}_{nt}^{(u)} \delta_0 = 0 \). Hence, \( E \left( \frac{1}{\sqrt{(n-1)T}} \frac{\partial}{\partial \theta} L_{nT}^{(u)}(\theta_0) \right) \cdot \frac{1}{\sqrt{(n-1)T}} \frac{\partial}{\partial \theta} L_{nT}^{(u)}(\theta_0) = \Sigma_{\theta_0,nt} + \Omega_{\theta_0,n} + O(T^{-1}) \). When \( V_{nt} \) are normally distributed, \( \Omega_{\theta_0,n} = 0 \) because \( \mu_4 - 3\sigma_0^4 = 0 \) for a normal distribution.

**C.4 About** \(- \frac{1}{(n-1)T} \frac{\partial^2}{\partial \theta \partial \theta'} L_{nT}(\theta)\)

Denote \( \| \theta - \theta_0 \| \) as the Euclidean norm of \( \theta - \theta_0 \), and \( \Theta_1 \) as a neighborhood of \( \theta_0 \), then, we have

\[
\frac{1}{(n-1)T} \frac{\partial^2}{\partial \theta \partial \theta'} L_{nT}(\theta) - \frac{1}{(n-1)T} \frac{\partial^2}{\partial \theta \partial \theta'} L_{nT}(\theta_0) = \| \theta - \theta_0 \| \cdot O_p(1),
\]

(C.5)

\[
\frac{1}{(n-1)T} \frac{\partial^2}{\partial \theta \partial \theta'} L_{nT}(\theta_0) + \Sigma_{\theta_0,nt} = O_p \left( \frac{1}{\sqrt{nT}} \right),
\]

(C.6)

\[
\sup_{\theta \in \Theta_1} \left| \frac{1}{(n-1)T} \frac{\partial^2}{\partial \theta \partial \theta'} L_{nT}(\theta) - \frac{1}{(n-1)T} E \frac{\partial^2}{\partial \theta \partial \theta'} \ln L_{nT}(\theta) \right|_{ij} = O_p \left( \frac{1}{\sqrt{nT}} \right),
\]

(C.7)

and

\[
\sup_{\theta \in \Theta_1} \left| \frac{1}{(n-1)T} E \frac{\partial^2}{\partial \theta \partial \theta'} \ln L_{nT}(\theta) + \Sigma_{\theta_0,nt} \right|_{ij} = \sup_{\theta \in \Theta_1} \| \theta - \theta_0 \| \cdot O(1)
\]

(C.8)

for all \( i,j = 1,2,\cdots, k_x + 4 \).

These are Equations (C.7) to (C.10) in Yu, de Jong and Lee (2006).

**D Proofs for Theorems**

**D.1 Proof of Claim 3.1**

To prove \( \frac{1}{(n-1)T} \ln L_{nT}(\theta) - Q_{nT}(\theta) \overset{P}{\rightarrow} 0 \) uniformly in \( \theta \) in any compact parameter space \( \Theta \):

From \( \hat{V}_{nt}(\theta) = S_n(\lambda) \hat{Y}_{nt} - \hat{Z}_{nt} \delta \) and \( \hat{V}_{nt} = S_n \hat{Y}_{nt} - \hat{Z}_{nt} \delta_0 \), using \( J_n l_n = 0 \), we have
\begin{equation}
J_n \tilde{V}_{nt}(\theta) = J_n \tilde{V}_{nt} - (\lambda - \lambda_0) J_n W_n \tilde{V}_{nt} - J_n \tilde{Z}_{nt}(\delta - \delta_0). \tag{D.1}
\end{equation}

Then,
\begin{equation}
\tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt}(\theta) = \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt} + (\lambda - \lambda_0)^2 (W_n \tilde{V}_{nt})' J_n W_n \tilde{V}_{nt} + (\delta - \delta_0)' \tilde{Z}_{nt} J_n \tilde{Z}_{nt}(\delta - \delta_0) \tag{D.2}
+ 2(\lambda - \lambda_0)(W_n \tilde{V}_{nt})' J_n \tilde{Z}_{nt}(\delta - \delta_0) - 2(\lambda - \lambda_0)(W_n \tilde{V}_{nt})' J_n \tilde{V}_{nt} - 2(\delta - \delta_0)' \tilde{Z}_{nt} J_n \tilde{V}_{nt}
\end{equation}

where, using (C.2), we have

\[ (W_n \tilde{V}_{nt})' J_n W_n \tilde{V}_{nt} = (G_n \tilde{Z}_{nt} \delta_0)' J_n (G_n \tilde{Z}_{nt} \delta_0) + 2(G_n \tilde{Z}_{nt} \delta_0)' J_n G_n \tilde{V}_{nt} + (G_n \tilde{V}_{nt})' J_n G_n \tilde{V}_{nt}. \]

Using Lemma B.1 and Lemma B.4,
\begin{align*}
\frac{1}{(n-1)T} \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt} - E \frac{1}{(n-1)} \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt} & \xrightarrow{p} 0, \\
\frac{1}{(n-1)T} \sum_{t=1}^{T} (W_n \tilde{V}_{nt})' J_n W_n \tilde{V}_{nt} - E \frac{1}{(n-1)} \sum_{t=1}^{T} (W_n \tilde{V}_{nt})' J_n W_n \tilde{V}_{nt} & \xrightarrow{p} 0, \\
\frac{1}{(n-1)T} \sum_{t=1}^{T} \tilde{Z}_{nt} J_n \tilde{Z}_{nt} - E \frac{1}{(n-1)} \sum_{t=1}^{T} \tilde{Z}_{nt} J_n \tilde{Z}_{nt} & \xrightarrow{p} 0, \\
\frac{1}{(n-1)T} \sum_{t=1}^{T} \tilde{Z}_{nt} J_n \tilde{V}_{nt} - E \frac{1}{(n-1)} \sum_{t=1}^{T} \tilde{Z}_{nt} J_n \tilde{V}_{nt} & \xrightarrow{p} 0 \text{ and}
\end{align*}

As \( \Theta \) is compact so that \( \lambda, \delta \) are bounded in \( \Theta \), we have
\[ \frac{1}{(n-1)T} \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt}(\theta) - E \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt}(\theta) \xrightarrow{p} 0 \text{ uniformly in } \theta \text{ in } \Theta. \]

Also, from (3.1) and by using the fact that \( \sigma^2 \) is bounded away from zero in \( \Theta \),
\[ \frac{1}{(n-1)T} \ln L_n, \tau(\theta) - Q_n, \tau(\theta) = -\frac{1}{2}\sigma^2 \left( \frac{1}{(n-1)T} \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt}(\theta) - E \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt}(\theta) \right) \xrightarrow{p} 0 \]

uniformly in \( \theta \) in \( \Theta \).

To prove \( Q_n, \tau(\theta) \) is uniformly equicontinuous in \( \theta \) in any compact parameter space \( \Theta \):

We have
\[ Q_n, \tau(\theta) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{n-1} \ln (1-\lambda) \]
\[ + \frac{1}{n-1} \ln |S_n(\lambda)| - \frac{1}{2\sigma^2(n-1)T} E \sum_{t=1}^{T} \tilde{V}_{nt}(\theta) J_n \tilde{V}_{nt}(\theta). \]

As \( J_n \tilde{V}_{nt}(\theta) = J_n \left[ S_n(\lambda) \tilde{Y}_{nt} - \tilde{Z}_{nt} \delta \right] \) and \( \tilde{Y}_{nt} = S_n^{-1} \tilde{Z}_{nt} \delta_0 + S_n^{-1} \tilde{V}_{nt} + \frac{\delta_0}{1 - \lambda_0} t_n \),
\[ J_n \tilde{V}_{nt}(\theta) = J_n [S_n(\lambda) S_n^{-1} \tilde{Z}_{nt} \delta_0 - \tilde{Z}_{nt} \delta + S_n(\lambda) S_n^{-1} \tilde{V}_{nt}] \]
because $J_n l_n = 0$. Hence,

$$
E \left( \frac{1}{n-1} \sum_{i=1}^{T} \tilde{V}_{ni}(\theta) J_n \tilde{V}_{ni}(\theta) \right) = \frac{1}{(n-1)T} \sum_{i=1}^{T} (S_n(\lambda) S_n^{-1} \tilde{Z}_{ni} \delta_0 - \tilde{Z}_{ni} \delta) J_n(S_n(\lambda) S_n^{-1} \tilde{Z}_{ni} \delta_0 - \tilde{Z}_{ni} \delta) + \frac{1}{n-1} T \sigma_n^2 \operatorname{tr}(S_n^{-1} S_n') J_n S_n(\lambda) S_n^{-1}) + \frac{2}{(n-1)T} \sum_{i=1}^{T} (S_n(\lambda) S_n^{-1} \tilde{Z}_{ni} \delta_0 - \tilde{Z}_{ni} \delta) J_n S_n(\lambda) S_n^{-1} \tilde{V}_{ni}.
$$

(D.3)

The third term $\frac{2}{(n-1)T} E \sum_{i=1}^{T} (S_n(\lambda) S_n^{-1} \tilde{Z}_{ni} \delta_0 - \tilde{Z}_{ni} \delta) J_n S_n(\lambda) S_n^{-1} \tilde{V}_{ni}$ is $O(T^{-1})$ according to expected values in Lemma B.4 and the order $O(T^{-1})$ is uniformly in $\theta$ in $\Theta$ because it is a polynomial function in $\theta$ and $\Theta$ is a bounded set. The first term is equal to $(\delta' - \delta_0, \lambda - \lambda_0) E H_n(T(\delta' - \delta_0, \lambda - \lambda_0)'$ using $S_n(\lambda) S_n^{-1} = I_n - (\lambda - \lambda_0) G_n$; the second term is equal to $\frac{T-1}{T} \sigma_n^2(\lambda)$ where $\sigma_n^2(\lambda) = \frac{\sigma_n^2}{n-1} \operatorname{tr}(S_n^{-1} S_n' \lambda) J_n S_n(\lambda) S_n^{-1})$, which are all polynomial functions of $\theta$. To prove $Q_{n,T}(\theta)$ is uniformly equicontinuous in $\theta$ in the compact parameter space $\Theta$, the followings are sufficient: (1) $\ln \sigma^2$ is uniformly continuous; (2) $\frac{1}{n} \ln |S_n(\lambda)|$ is uniformly equicontinuous; (3) $(\delta' - \delta_0, \lambda - \lambda_0) H_n(T(\delta' - \delta_0, \lambda - \lambda_0)' is uniformly equicontinuous; (4) $\sigma_n^2(\lambda)$ is uniformly equicontinuous.

(1) is obvious because $\sigma^2$ is bounded away from zero in $\Theta$. For (2), $\frac{1}{n} \ln |S_n(\lambda_2)| - \frac{1}{n} \ln |S_n(\lambda_1)| = \frac{1}{n} \operatorname{tr} (W_n S_n^{-1} (\tilde{\lambda})) (\lambda_2 - \lambda_1)$ where $\lambda$ lies between $\lambda_2$ and $\lambda_1$. As $S_n^{-1}(\lambda)$ is uniformly bounded in row and column sums, uniformly in $\theta \in \Theta$, $\frac{1}{n} \operatorname{tr} (W_n S_n^{-1} (\lambda))$ is bounded, we have $\frac{1}{n} \ln |S_n(\lambda)|$ is uniformly equicontinuous.

(3) because $\delta$ and $\lambda$ are bounded and because $E H_n(T)$ is $O(1)$ according to Lemma B.4, the result follows. For (4), $\sigma_n^2(\lambda_2) - \sigma_n^2(\lambda_1) = \frac{\sigma_n^2}{n-1} \operatorname{tr}(S_n^{-1} S_n' \lambda_2) J_n S_n(\lambda_2) S_n^{-1}) - \frac{\sigma_n^2}{n-1} \operatorname{tr}(S_n^{-1} S_n' \lambda_1) J_n S_n(\lambda_1) S_n^{-1})$. Using $S_n(\lambda) S_n^{-1} = I_n - (\lambda - \lambda_0) G_n$,

$$
\sigma_n^2(\lambda_2) - \sigma_n^2(\lambda_1) = \sigma_0^2 \left[ (\lambda_2 - \lambda_1) (\lambda_2 + \lambda_1 - 2\lambda_0) \frac{\operatorname{tr} G_n' J_n G_n}{n} - (\lambda_2 - \lambda_1) \frac{\operatorname{tr} (G_n' J_n + J_n G_n)}{n} \right].
$$

As $G_n' J_n G_n$ and $J_n G_n$ are uniformly bounded in row and column sums, $\sigma_n^2(\lambda)$ is uniformly equicontinuous.

\[ \Box \]

### D.2 Proof of nonsingularity of the information matrix

We can prove the result by using an argument by contradiction. For $\Sigma_{\theta_0} \equiv \lim_{T \to -\infty} \Sigma_{\theta_0,nT}$, where $\Sigma_{\theta_0,nT}$ is (3.2), we need to prove that $\Sigma_{\theta_0} \alpha = 0$ implies $\alpha = 0$ where $\alpha = (\alpha_1', \alpha_2, \alpha_3)'$ and $\alpha_2, \alpha_3$ are scalars and $\alpha_1$ is $(k_x + 2) \times 1$ vector. If this is true, then, columns of $\Sigma_{\theta_0}$ would be linear independent so that $\Sigma_{\theta_0}$ would be nonsingular. Denote $\mathcal{H}_{\delta} = \lim_{T \to -\infty} \frac{1}{(n-1)T} \sum_{i=1}^{T} \tilde{Z}_{ni} J_n \tilde{Z}_{ni}$, $\mathcal{H}_{\delta \lambda} = \lim_{T \to -\infty} \frac{1}{(n-1)T} \sum_{i=1}^{T} \tilde{Z}_{ni} J_n G_n \tilde{Z}_{ni} \delta_0$. 

29
\( \mathcal{H}_\delta = \mathcal{H}_\delta' \) and \( \mathcal{H}_\lambda = \lim_{t \to -\infty} \frac{1}{(n-1)t} \sum_{i=1}^{T} (G_n \tilde{Z}_{ni} \delta_0)' J_n G_n \tilde{Z}_{ni} \delta_0 \). Then

\[
\Sigma_{\theta_0} = \frac{1}{\sigma_0^2} \begin{pmatrix}
\mathcal{H}_\delta & 0 \\
\mathcal{H}_\delta' & 0 \\
0 & \mathcal{H}_\lambda \\
\end{pmatrix} \begin{pmatrix}
\mathcal{H}_\delta \\
\mathcal{H}_\delta' \\
0 \\
\end{pmatrix} E \mathcal{H}_\lambda + \lim_{n \to \infty} \frac{\sigma_0^2}{n-1} \left[ tr(G_n' J_n G_n) + tr((J_n G_n)^2) \right] \lim_{n \to \infty} \frac{1}{n-1} tr(J_n G_n) \begin{pmatrix}
0 \\
\frac{1}{2\sigma_0^2} \\
\end{pmatrix}.
\]

Hence, \( \Sigma_{\theta_0} \alpha = 0 \) implies

\[
\mathcal{H}_\delta \times \alpha_1 + \mathcal{H}_\delta \times \alpha_2 = 0,
\]

\[
\frac{1}{\sigma_0^2} \mathcal{H}_\lambda \times \alpha_1 + \left( \frac{1}{\sigma_0^2} \mathcal{H}_\lambda + \lim_{n \to \infty} \frac{1}{n-1} \left[ tr(G_n' J_n G_n) + tr((J_n G_n)^2) \right] \right) \times \alpha_2 + \lim_{n \to \infty} \frac{1}{n-1} tr(J_n G_n) \times \alpha_3 = 0,
\]

\[
\lim_{n \to \infty} \frac{1}{n-1} tr(J_n G_n) \times \alpha_2 + \frac{1}{2\sigma_0^2} \times \alpha_3 = 0.
\]

From the first equation, \( \alpha_1 = -(\mathcal{H}_\delta)^{-1} \mathcal{H}_\delta \times \alpha_2 \); from the second equation, \( \alpha_3 = -2 \lim_{n \to \infty} \frac{\sigma_0^2}{n-1} tr(J_n G_n) \times \alpha_2 \). By eliminating \( \alpha_1 \) and \( \alpha_3 \), the remaining equation becomes

\[
\left\{ \left( \frac{1}{\sigma_0^2} (\mathcal{H}_\lambda - \mathcal{H}_\delta \mathcal{H}_\delta'^{-1} \mathcal{H}_\delta) \right) + \lim_{n \to \infty} \frac{1}{n-1} \left[ tr(G_n' J_n G_n) + tr((J_n G_n)^2) - \frac{2tr^2(J_n G_n)}{n-1} \right] \right\} \times \alpha_2 = 0.
\]

Denote \( C_n = J_n G_n - \frac{tr(J_n G_n)}{n-1} J_n \), then,

\[
tr(G_n' J_n G_n) + tr((J_n G_n)^2) - \frac{2tr^2(J_n G_n)}{n-1}
\]

\[
= \frac{1}{2} tr(C_n' + C_n)(C_n' + C_n),
\]

which is nonnegative. Hence, if the limit of \( E \mathcal{H}_n \mathcal{L} \) is nonsingular or the limit of \( \frac{1}{n-1} tr(G_n' J_n G_n) + tr((J_n G_n)^2) - \frac{2tr^2(J_n G_n)}{n-1} \) is nonzero, we have \( \alpha_2 = 0 \) and hence \( \alpha = 0 \). This proves the nonsingularity of \( \Sigma_{\theta_0} \). ■

### D.3 Proof of Theorem 3.2

As \( E \sum_{t=1}^{T} \tilde{V}_{nt}' J_n \tilde{V}_{nt} = (n-1)(T-1)\sigma_0^2 \) according to Lemma B.1, at \( \theta_0 \), the expected log likelihood from (3.1) implies

\[
E \ln L_n, T(\theta_0) = -\frac{(n-1)T}{2} \ln 2\pi - \frac{(n-1)T}{2} \ln \sigma_0^2 - T \ln (1 - \lambda_0) + T \ln |S_n| - \frac{(n-1)(T-1)}{2}.
\]

Denote \( \sigma_n^2(\lambda) = \frac{\sigma_0^2}{n-1} tr(S_n^{-1}S_n' \lambda)J_n S_n(\lambda)S_n^{-1} \). By using \( S_n(\lambda)S_n^{-1} = I_n + (\lambda_0 - \lambda)G_n \) for (D.3), it follows that

\[
\frac{1}{(n-1)T} E \ln L_n, T(\theta) - \frac{1}{(n-1)T} E \ln L_n, T(\theta_0)
\]

30
\[ = -\frac{1}{2}(\ln \sigma^2 - \ln \sigma_0^2) + \frac{1}{n-1} \ln |S_n(\lambda)| - \frac{1}{n-1} \ln |S_n| - \left( \frac{1}{2\sigma^2} \frac{1}{n-1} \frac{1}{T} \sum_{t=1}^{T} E \hat{V}_{nt}(\theta)J_n \hat{V}_{nt}(\theta) - \frac{T-1}{2}\right) \]

\[ - \frac{1}{n-1} (\ln(1 - \lambda) - \ln(1 - \lambda_0)) \]

\[ = T_{1,n}(\lambda, \sigma^2) - \frac{1}{2\sigma^2} T_{2,n,T}(\delta, \lambda) + O(T^{-1}), \]

where

\[ T_{1,n}(\lambda, \sigma^2) = \frac{1}{2}(\ln \sigma^2 - \ln \sigma_0^2) + \frac{1}{n-1} \ln |S_n(\lambda)| - \frac{1}{n-1} \ln |S_n| - \frac{1}{2\sigma^2} (\sigma_n^2(\lambda) - \sigma^2) - \frac{1}{n-1} (\ln(1 - \lambda) - \ln(1 - \lambda_0)), \]

and

\[ T_{2,n,T}(\delta, \lambda) = \frac{1}{(n-1)T} \sum_{t=1}^{T} E \left\{ \left[ \hat{Z}_{nt}(\delta_0 - \delta) + (\lambda_0 - \lambda)G_n \hat{Z}_{nt}\delta_0 \right] J_n \left[ \hat{Z}_{nt}(\delta_0 - \delta) + (\lambda_0 - \lambda)G_n \hat{Z}_{nt}\delta_0 \right] \right\} \]

\[ = ((\delta_0 - \delta)'(\lambda_0 - \lambda)) \cdot H_{n,T} \cdot ((\delta_0 - \delta)'(\lambda_0 - \lambda)). \]

Consider the pure spatial process \( \tilde{Y}_{nt} = \lambda_0 W_n Y_{nt} + \alpha_t I_n + V_{nt} \) for a period \( t \). Similarly as Equation (2.10) after data transformation, we can get the log likelihood function of this process as

\[ \ln L_{p,n}(\lambda, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n-1}{2} \ln \sigma^2 - \ln(1 - \lambda) + \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} \left( V_{nt}(\lambda) J_n V_{nt}(\lambda) \right), \quad \text{(D.4)} \]

where \( V_{nt}(\lambda) = S_n(\lambda)Y_{nt} \). Let \( E_p(\cdot) \) be the expectation operator for \( Y_{nt} \) based on this pure spatial autoregressive process. It follows that

\[ E_p(\frac{1}{n-1} \ln L_{p,n}(\lambda, \sigma^2)) - E_p(\frac{1}{n-1} \ln L_{p,n}(\lambda_0, \sigma_0^2)) \]

\[ = -\frac{1}{2}(\ln \sigma^2 - \ln \sigma_0^2) + \frac{1}{n-1} \ln |S_n(\lambda)| - \frac{1}{n-1} \ln |S_n(\lambda_0)| \]

\[ - \frac{1}{2\sigma^2} (\sigma_n^2(\lambda) - \sigma^2) - \frac{1}{n-1} (\ln(1 - \lambda) - \ln(1 - \lambda_0)), \]

which equals to \( T_{1,n}(\lambda, \sigma^2) \). By the information inequality, \( \ln L_{p,n}(\lambda, \sigma^2) - \ln L_{p,n}(\lambda_0, \sigma_0^2) \leq 0 \). Thus, \( T_{1,n}(\lambda, \sigma^2) \leq 0 \) for any \( (\lambda, \sigma^2) \).

For \( T_{2,n,T}(\delta, \lambda) \), it is a quadratic function of \( \delta \) and \( \lambda \). Under the assumed condition that \( \lim_{T \to \infty} E \mathcal{H}_{n,T} \) is nonsingular, \( \lim_{T \to \infty} T_{2,n,T}(\delta, \lambda) > 0 \) whenever \( (\delta, \lambda) \neq (\delta_0, \lambda_0) \). So, \( (\delta, \lambda) \) is globally identified. Given \( \lambda_0, \sigma_0^2 \) is the unique maximizer of \( T_{1,n}(\lambda_0, \sigma_2) \) for any given \( n \). In the event that \( n \to \infty \), \( \sigma_0^2 \) is the unique maximizer of \( \lim_{T \to \infty} T_{1,n}(\lambda_0, \sigma_2) \). Hence, \( (\delta, \lambda, \sigma^2) \) is globally identified.

Combined with uniform convergence and equicontinuity in Claim 3.1, the consistency follows.

**D.4 Proof of Theorem 3.3**

From the Proof of Theorem 3.2 (see Appendix D.3),

\[ \frac{1}{(n-1)T} E \ln L_{n,T}(\theta) - \frac{1}{(n-1)T} E \ln L_{n,T}(\theta_0) = T_{1,n}(\lambda, \sigma^2) - \frac{1}{2\sigma^2} T_{2,n,T}(\delta, \lambda) + O(T^{-1}). \]
When the limit of \( E \mathcal{H}_{nT} \) is singular, \( \delta_0 \) and \( \lambda_0 \) cannot be identified from \( T_{2,n,T}(\delta, \lambda) \). Global identification requires that the limit of \( T_{1,n}(\lambda, \sigma^2) \) is strictly less than zero. Thus, we require the global identification just from the pure spatial model with the likelihood function (D.4). Using the concentrating approach by concentrating out \( \sigma^2 \) in (D.4), we have the concentrated likelihood function

\[
\ln L_{p,n}(\lambda) = -\frac{n-1}{2} \ln(2\pi) + \frac{n-1}{2} \ln \sigma^2_n(\lambda) - \ln(1 - \lambda) + \ln |S_n(\lambda)|,
\]

where \( \sigma^2_n(\lambda) = \frac{1}{n-1} V^*_{xT}(\lambda) J_n V_n(\lambda) \). Also, we have corresponding \( Q_n(\lambda) = \max_{\veps \geq 2} E(\ln L_{p,n}(\lambda, \sigma^2)) = -\frac{n-1}{2} \ln(2\pi) - \frac{n-1}{2} \ln \sigma^2_n(\lambda) - \ln(1 - \lambda) + \ln |S_n(\lambda)| \). Global identification of \( \lambda_0 \) requires that \( \lim_{n \to \infty} \frac{1}{n} [Q_n(\lambda) - Q_n(\lambda_0)] \neq 0 \) whenever \( \lambda \neq \lambda_0 \), which is equivalent to \( \lim_{n \to \infty} \frac{1}{n} |\ln |S_n(\lambda)| - \ln |S_n| - \frac{n-1}{2} \ln \sigma^2_n(\lambda) - \ln \sigma^2_n(\lambda_0) - \ln(1 - \lambda) - \ln(1 - \lambda_0)| \neq 0 \) whenever \( \lambda \neq \lambda_0 \). By rearranging the terms, \( Q_n(\lambda) - Q_n(\lambda_0) \neq 0 \) whenever \( \lambda \neq \lambda_0 \) is just equivalent to

\[
\frac{1}{n} \ln |\sigma^2_n S_n^{-1} S_n^{-1}| - \frac{1}{n} \ln |\sigma^2_n(\lambda) S_n^{-1} S_n^{-1}(\lambda)| + \left( \frac{1}{n} \ln \frac{\sigma^2_n}{(1 - \lambda_0)^2} - \frac{1}{n} \ln \frac{\sigma^2_n}{(1 - \lambda)^2} \right) \neq 0
\]

for \( \lambda \neq \lambda_0 \). When \( n \to \infty \), it becomes

\[
\lim_{n \to \infty} \left( \frac{1}{n} \ln |\sigma^2_n S_n^{-1} S_n^{-1}| - \frac{1}{n} \ln |\sigma^2_n(\lambda) S_n^{-1} S_n^{-1}(\lambda)| \right) \neq 0
\]

for \( \lambda \neq \lambda_0 \). After \( \lambda_0 \) is identified, \( \sigma^2_0 \) is then identified. Also, given \( \lambda_0, \delta_0 \) can be identified from \( \lim_{T \to \infty} T_{2,n,T}(\delta, \lambda) \). Combined with uniform convergence and equicontinuity in Claim 3.1, the consistency follows.

### D.5 Proof of Claim 3.4

As \( Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt}) \) where \( Y_{nt} \) is specified in Equation (2.4), we have two components of \( J_n \vec{Z}_{nt} \) such that

\[
J_n \vec{Z}_{nt} = J_n \vec{Z}_{nt}^{(u)} - (J_n \vec{U}_{nT-1}, J_n W_n \vec{U}_{nT-1}, 0), \tag{D.6}
\]

where \( \vec{Z}_{nt}^{(u)} = (\vec{X}_{n,t-1} + U_{n,t-1}, W_n \vec{X}_{n,t-1} + W_n U_{n,t-1}, \vec{X}_{nt}) \), with \( \vec{X}_{n,t-1} = X_{n,t-1} - \vec{X}_{n,T-1} \). Hence, \( J_n \vec{Z}_{nt} \) has two components: one is \( J_n \vec{Z}_{nt}^{(u)} \), which is uncorrelated with \( V_{nt} \); the other is \( -(J_n \vec{U}_{nT-1}, J_n W_n \vec{U}_{nT-1}, 0) \), which is correlated with \( V_{nt} \) when \( t \leq T - 1 \). Then, the score can be decomposed into 2 parts such that

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} - \Delta_{nT}, \tag{D.7}
\]

where \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \) is defined in (3.5) and \( \Delta_{nT} \) is defined in (3.6).

For the first part, it’s a linear and quadratic form of \( V_{nt} \) and the asymptotic distribution can be derived from the central limit theorem for martingale difference arrays (Theorem B.3). For the term \( \Delta_{nT} \) from Equation (B.3) in Lemma B.1 and Equation (B.5) in Theorem B.2, we have \( \Delta_{nT} = \sqrt{\frac{n-1}{T} \theta_{0,n} + O(\sqrt{\frac{n-1}{T^2}})} + O_p(\frac{1}{\sqrt{T}}) \) where \( a_{\theta_{0,n}} \) is \( O(1) \) and specified in Equation (3.9). Combined together, we get the result that
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} + \Delta_{nT} \xrightarrow{d} N(0, \Sigma_{\theta_0} + \Omega_{\theta_0}). \tag{D.8}
\]

D.6 Proof of Claim 3.5

This is Equation (C.5) and (C.6).

D.7 Proof of Theorem 3.6

The Taylor expansion gives

\[
\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) = \left( -\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} + \Delta_{nT}, \tag{D.9}
\]

where \(\hat{\theta}_{nT}\) lies between \(\theta_0\) and \(\hat{\theta}_{nT}\) and

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} - \Delta_{nT}.
\]

As

\[
-\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial \theta \partial \theta'} = \left( -\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} \left( -\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial \theta \partial \theta'} \right) + \Sigma_{\theta_0,nT}
\]

where the first term is \(\|\hat{\theta}_{nT} - \theta_0\| \cdot O_p(1)\) (Equation (C.5)) and the second term is \(O_p \left( \frac{1}{(n-1)T} \right)\). Because (1) \(\hat{\theta}_{nT} - \theta_0\) is consistent and (2) \(\Sigma_{\theta_0,nT}\) is the nonsingular in the limit according to Assumption 8 and Appendix D.2, we have\(\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\) is invertible for large \(T\) and \(\left( -\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1}\) is \(O_p(1)\).

Also, we have

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Sigma_{\theta_0} + \Omega_{\theta_0})\]

and \(\Delta_{nT} = \sqrt{n-1}n\theta_{0,n} + O\left( \frac{n-1}{T} \right) + O_p\left( \frac{1}{\sqrt{T}} \right)\) with \(n\theta_{0,n} = O(1)\). Then, from (D.9), 

\[
\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) = O_p(1) \cdot \left( O_p(1) + O\left( \frac{1}{\sqrt{T}} \right) \right),
\]

which implies that

\[
\hat{\theta}_{nT} - \theta_0 = O_p \left( \max \left( \frac{1}{\sqrt{nT}}, \frac{1}{T} \right) \right). \tag{D.10}
\]

Using the fact that\(^\dag\)

\[
\left( -\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} = \Sigma_{\theta_0,nT}^{-1} + O_p \left( \max \left( \frac{1}{\sqrt{nT}}, \frac{1}{T} \right) \right), \tag{D.11}
\]

given that \(\Sigma_{\theta_0,nT}\) is nonsingular and its inverse is of order \(O(1)\), we have

\(^\dag\)For two matrices \(C_k\) and \(D_k\) which are nonsingular and \(C_k - D_k = O_p(T^{-\eta})\) for \(\eta > 0\), we have \((C_k - D_k)^{-1} = C_k^{-1}(D_k - C_k)D_k^{-1} = O_p(T^{-\eta})\) when \(C_k^{-1}\) and \(D_k^{-1}\) are of order \(O(1)\).
\[
\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) = \left( -\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial \theta^2} \right) \cdot \left( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} \right) - \Delta_{nT} = \Sigma_{\theta_0,nT}^{-1} \cdot \frac{1}{\sqrt{(n-1)T}} \left( \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} + O_p \left( \max \left( \frac{1}{\sqrt{nT}}, \frac{1}{T} \right) \right) \right) - \Sigma_{\theta_0,nT}^{-1} \cdot \Delta_{nT} - O_p \left( \max \left( \frac{1}{\sqrt{nT}}, \frac{1}{T} \right) \right) \cdot \Delta_{nT},
\]

which implies that

\[
\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) + \Sigma_{\theta_0,nT}^{-1} \cdot \Delta_{nT} + O_p \left( \max \left( \frac{1}{\sqrt{nT}}, \frac{1}{T} \right) \right) \cdot \Delta_{nT} = (\Sigma_{\theta_0,nT}^{-1} + o_p(1)) \cdot \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta}.
\]

As \(\Sigma_{\theta_0} = \lim_{T \to \infty} \Sigma_{\theta_0,nT}\) exists, then using Claim 3.4 and that \(\Delta_{nT} = \sqrt{\frac{n-1}{T}} a_{\theta_{0,n}} + O(\sqrt{\frac{n-1}{T}}) + O_p(\frac{1}{\sqrt{T}})\) with \(a_{\theta_{0,n}} = O(1),\)

\[
\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n-1}{T}} \Sigma_{\theta_0,nT}^{-1} a_{\theta_{0,n}} + O_p \left( \max \left( \sqrt{nT}, \sqrt{\frac{1}{T}} \right) \right) \overset{d}{\to} N(0, \Sigma_{\theta_0}^{-1} (\Sigma_{\theta_0} + \Omega_{\theta_0}) \Sigma_{\theta_0}^{-1}). \quad \blacksquare
\]

(D.12)

**D.8 Proof for Theorem 3.8**

Theorem 3.6 states that \(\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n-1}{T}} b_{\theta_0,nT} + O_p \left( \max \left( \sqrt{\frac{n-1}{T}}, \frac{1}{T} \right) \right) \overset{d}{\to} N(0, \Sigma_{\theta_0}^{-1} (\Sigma_{\theta_0} + \Omega_{\theta_0}) \Sigma_{\theta_0}^{-1})\). As the bias corrected estimator is

\[
\hat{\theta}_{nT}^1 = \hat{\theta}_{nT} + \frac{1}{T} \left( -\frac{1}{(n-1)T} E \frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} a_n(\hat{\theta}_{nT}),
\]

where \(a_n(\theta) = a_{\theta,n}\), we will have \(\sqrt{(n-1)T}(\hat{\theta}_{nT}^1 - \theta_0) \overset{d}{\to} N(0, \Sigma_{\theta_0}^{-1} (\Sigma_{\theta_0} + \Omega_{\theta_0}) \Sigma_{\theta_0}^{-1})\) if

\[
\sqrt{\frac{n-1}{T}} \left( -\frac{1}{(n-1)T} E \frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} a_n(\hat{\theta}_{nT}) - \Sigma_{\theta_0,nT}^{-1} a_n(\theta_0) \overset{p}{\to} 0,
\]

(D.14)

and \(\frac{n}{T} \to 0\). Assuming that \(\frac{n}{T^2} \to 0\), we are going to prove Equation (D.14).

From Equation (D.11) and that \(\hat{\theta}_{nT} - \theta_0 = O_p \left( \max \left( \frac{1}{\sqrt{nT}}, \frac{1}{T} \right) \right) \) from Equation (D.10), we have

\[
-\frac{1}{(n-1)T} E \frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} = \Sigma_{\theta_0,nT}^{-1} + O_p \left( \max \left( \frac{1}{\sqrt{nT}}, \frac{1}{T} \right) \right).
\]
Hence,

\[
\sqrt{\frac{n-1}{T}} \left\{ \left( \frac{1}{(n-1)T} \right)^{\frac{1}{2}} \right\}^{-1} a_n(\hat{\theta}_n) - \Sigma_{\theta_0} a_n(\theta_0) \right) \}
\]

\[
= \sqrt{\frac{n-1}{T}} \left\{ \left( \frac{1}{(n-1)T} \right)^{\frac{1}{2}} \right\} a_n(\hat{\theta}_n) - \Sigma_{\theta_0} a_n(\theta_0) \right) \}
\]

As \( \hat{\theta}_n - \theta_0 = O_p \left( \frac{1}{\sqrt{nT}} \right) \) and \( a_n(\theta_0) \) is \( O(1) \), according to the Taylor expansion of \( a_n(\hat{\theta}_n) \) around \( a_n(\theta_0) \), to prove Equation (D.14) is reduced to prove that the sequence of \( \frac{\partial a_n(\theta)}{\partial \theta} \) is bounded, uniformly in a small neighborhood of \( \theta_0 \) where, from (3.9),

\[
a_{\theta} = \left[ \begin{array}{c}
\frac{1}{n-1} tr \left( (J_n \sum_{h=0}^{\infty} A_n^h(\theta)) S_n^{-1}(\lambda) \right) \\
\frac{1}{n-1} tr \left( W_n (\sum_{h=0}^{\infty} A_n^h(\theta)) S_n^{-1}(\lambda) \right) \\
\left( \frac{1}{n-1} tr(G_n(\lambda) \left( J_n \sum_{h=0}^{\infty} A_n^h(\theta) S_n^{-1}(\lambda) \right) \\
+ \frac{1}{n-1} tr(G_n(\lambda) W_n \left( J_n \sum_{h=0}^{\infty} A_n^h(\theta) S_n^{-1}(\lambda) \right) + \frac{1}{n-1} tr(G_n(\lambda) W_n \left( J_n \sum_{h=0}^{\infty} A_n^h(\theta) S_n^{-1}(\lambda) \right) + \frac{1}{n-1} tr \left( W_n (\sum_{h=0}^{\infty} A_n^h(\theta)) S_n^{-1}(\lambda) \right) \\
\end{array} \right) \right].
\]

As \( A_n(\theta) = S_n^{-1}(\lambda)(\gamma I_n + \rho W_n) \) and \( G_n(\lambda) = W_n S_n^{-1}(\lambda) \), we have \( \frac{\partial A_n(\theta)}{\partial \gamma} = S_n^{-1}(\lambda), \frac{\partial A_n(\theta)}{\partial \rho} = S_n^{-1}(\lambda) W_n, \)
\[
= 0 \text{ for } i = 1, 2, \ldots, kx \text{ and } \frac{\partial \gamma}{\partial \lambda} = S_n^{-1}(\lambda) W_n S_n^{-1}(\lambda) (\gamma I_n + \rho W_n). \]

Because\(^{14}\) for each element of \( \theta, \frac{\partial A_n^h(\theta)}{\partial \theta_i} = h A_n^h(\theta) \frac{\partial A_n^h(\theta)}{\partial \theta_i} \) for \( h \geq 1 \), \( \sum_{h=0}^{\infty} \frac{\partial A_n^h(\theta)}{\partial \theta_i} = \sum_{h=1}^{\infty} h A_n^{h-1}(\theta) \frac{\partial A_n^h(\theta)}{\partial \theta_i} \). As (1) \( \sum_{h=0}^{\infty} A_n^h(\theta) \) and \( \sum_{h=1}^{\infty} h A_n^{h-1}(\theta) \) are uniformly bounded in either row sum or column sum, uniformly in a neighborhood of \( \theta_0 \), (2) \( S_n^{-1}(\lambda) \) is uniformly bounded in both row and column sums, also uniformly in \( \lambda \) in a neighborhood of \( \lambda_0 \) and (3) \( W_n \) is uniformly bounded in both row and column sums, we have the result that the sequence of \( \frac{\partial a_n(\theta)}{\partial \theta} \) will be uniformly bounded in a neighborhood of \( \theta_0 \). As \( \hat{\theta}_n \) converges in probability to \( \theta_0 \), we conclude that elements of \( \frac{\partial a_n(\theta)}{\partial \theta} \) are \( O_p(1) \). \( \blacksquare \)

E  Direct Approach

E.1  Concentrated Likelihood Function and its FOC and SOC

The likelihood function for all the parameters including both individual and time dummies of the model is (4.1) and we will concentrate out \( \alpha_t \) and \( c_n \). The first order condition for \( \alpha_t \) is

\[
\frac{\partial L^d(\theta, c_n, \alpha_t)}{\partial \alpha_t} = \frac{1}{\sigma^2} L^d V_{nt}(\theta, c_n, \alpha_t),
\]

and it implies

\(^{14}\)This can be proved by mathematical induction. See footnote 9 in Yu, de Jong and Lee (2006).
where $J_n = I_n - \frac{1}{n} l_n l_n'$, which is idempotent with rank $n - 1$ and $J_n l_n = 0$. Hence, the likelihood function with $\alpha_T$ being concentrated out is

$$
\ln L_{n,T}^d(\theta, c_n) = -\frac{nT}{2} \ln 2\pi - \frac{nT}{2} \ln \sigma^2 + T \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} \sum_{t=1}^T \left[ V_n(\theta, c_n, \alpha_T(\theta, c_n)) \right]' \left[ V_n(\theta, c_n, \alpha_T(\theta, c_n)) \right].
$$

(E.1)

Compared to (2.10), (E.1) is the concentrated likelihood function for the true data generating process where the time effects are concentrated out; while (2.10) is the likelihood function for the transformed data.

For (E.1), the first order derivative of $c_n$ is $\frac{\partial \ln L_{n,T}^d(\theta, c_n)}{\partial c_n} = \frac{1}{\sigma^2} \sum_{t=1}^T V_n(\theta, c_n, \alpha_T(\theta, c_n))$ and it implies

$$
J_n \check{c}_n(\theta) = J_n \left[ S_n(\lambda)^T Y_{nT} - \check{Z}_{nT} \delta \right]
$$

where $Z_{nt} = (Y_{n,t-1}, \ldots, Y_{n,t-1}, X_{nt})$ and $\delta = (\gamma, \rho, \beta')'$. Therefore, the likelihood function with both $c_n$ and $\alpha_T$ concentrated out is

$$
\ln L_{n,T}^d(\theta) = -\frac{nT}{2} \ln 2\pi - \frac{nT}{2} \ln \sigma^2 + T \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} \sum_{t=1}^T \check{V}_n(\theta) J_n \check{V}_n(\theta).
$$

(E.2)

Our approach here is the same as the "difference in difference" one. Denote $\tilde{Q} = I_n T - I_T \otimes \frac{1}{n} l_n l_n' - \frac{1}{n^2} l_T l_T' \otimes I_n + \frac{1}{n} J = (I_T - \frac{1}{T} I_T l_T') \otimes (J_n - \frac{1}{n} l_n l_n')$ where $J$ is an $nT \times nT$ matrix of ones, and denote $Z = (Z'_{n1}, Z'_{n2}, \ldots, Z'_{nT})'$, then, $\sum_{t=1}^T \check{Z}_n J_n \check{Z}_{nt} = Z' \tilde{Q} Z$. This is so because

$$
\sum_{t=1}^T \check{Z}_n J_n \check{Z}_{nt} = \left[ (J_n \check{Z}_{n1}), (J_n \check{Z}_{n2}), \ldots, (J_n \check{Z}_{nT})' \right] \cdot \left[ (J_n \check{Z}_{n1})', (J_n \check{Z}_{n2})', \ldots, (J_n \check{Z}_{nT})' \right]'.
$$

where

$$
J_n \check{Z}_{nt} = (I_n - \frac{1}{n^2} l_n l_n') \times \begin{bmatrix} z_{1t} \\ z_{2t} \\ \vdots \\ z_{nt} \end{bmatrix} - \frac{1}{T} \begin{bmatrix} z_{1t} + z_{12} + z_{1T} \\ z_{2t} + z_{22} + z_{2T} \\ \vdots \\ z_{nt} + z_{n2} + z_{nT} \end{bmatrix} = \begin{bmatrix} z_{1t} \\ z_{2t} \\ \vdots \\ z_{nt} \end{bmatrix} - \frac{1}{n^2} l_n l_n' \begin{bmatrix} z_{1t} \\ z_{2t} \\ \vdots \\ z_{nt} \end{bmatrix} - \frac{1}{T} \begin{bmatrix} z_{1t} + z_{12} + z_{1T} \\ z_{2t} + z_{22} + z_{2T} \\ \vdots \\ z_{nt} + z_{n2} + z_{nT} \end{bmatrix} + \frac{1}{n} \begin{bmatrix} \sum_{i=1}^t (z_{i1} + z_{i2} + z_{iT}) \\ \sum_{i=1}^t (z_{i1} + z_{i2} + z_{iT}) \\ \vdots \\ \sum_{i=1}^t (z_{i1} + z_{i2} + z_{iT}) \end{bmatrix}.
$$

Hence,

$$
\tilde{Q} = I_{nT} T - I_T \otimes \frac{1}{n^2} l_n l_n' - \frac{1}{n^2} l_T l_T' \otimes I_n + \frac{1}{n^2} J \text{ where } J \text{ is an } nT \times nT \text{ matrix of ones, and denote } Z = (Z'_{n1}, Z'_{n2}, \ldots, Z'_{nT})'.
$$

For the concentrated likelihood function (E.2), the first order derivatives are
\[
\frac{1}{\sqrt{nT}} \frac{\partial \ln L^d_{n,T}(\theta)}{\partial \theta} = \frac{1}{\sqrt{nT}} \left( \begin{array}{c}
\frac{\partial \ln L^d_{n,T}(\theta)}{\partial \theta} \\
\frac{\partial \ln L^d_{n,T}(\theta)}{\partial \sigma^2}
\end{array} \right) = \left( \begin{array}{c}
\frac{1}{\sigma^2} \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} (J_n \delta^2_n)^T \bar{V}_{nt}(\theta) \\
\frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \left((J_n W_n \bar{Y}_{nt}) \bar{V}_{nt}(\theta) - \sigma^2 tr G_n(\lambda)\right) \\
\frac{1}{2\sigma^2} \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} (\bar{V}_{nt}(\theta) J_n \bar{V}_{nt}(\theta) - n\sigma^2)
\end{array} \right), \quad (E.3)
\]

and the second order derivatives are

\[
\frac{1}{nT} \frac{\partial^2 \ln L^d_{n,T}(\theta)}{\partial \theta \partial \theta'} = \frac{1}{nT} \left( \begin{array}{ccc}
\frac{\partial^2 \ln L^d_{n,T}(\theta)}{\partial \theta^2} & \frac{\partial^2 \ln L^d_{n,T}(\theta)}{\partial \theta \partial \theta'} \\
\frac{\partial^2 \ln L^d_{n,T}(\theta)}{\partial \theta \partial \sigma^2} & \frac{\partial^2 \ln L^d_{n,T}(\theta)}{\partial \theta \partial \sigma^2}
\end{array} \right) = \left( \begin{array}{ccc}
\frac{1}{\sigma^2} \frac{1}{nT} \sum_{t=1}^{T} \left( J_n \delta^2_n \delta^2_n^T \right) \left( W_n \bar{Y}_{nt} \right) \bar{V}_{nt}(\theta) \\
\frac{1}{\sigma^2} \frac{1}{nT} \sum_{t=1}^{T} \left( J_n \delta^2_n \delta^2_n^T \right) \bar{V}_{nt}(\theta) \bar{V}_{nt}(\theta) \\
\frac{1}{\sigma^2} \frac{1}{nT} \sum_{t=1}^{T} \left( J_n \delta^2_n \delta^2_n^T \right) \left( W_n \bar{Y}_{nt} \right) \bar{V}_{nt}(\theta)
\end{array} \right).
\]

Similarly to that of \( \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \), the variance matrix of \( \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \) of (4.5) is equal to

\[
E \left( \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta'} \right) = \Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT}^d + O \left( T^{-1} \right), \quad (E.5)
\]

and \( \Omega_{\theta_0,nT}^d = \frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4} \left( \begin{array}{ccc}
0 & 0 & 0 \\
0 & \frac{1}{nT} \sum_{i=1}^{n} G_{n,ii} & \frac{1}{2\sigma_0^2} tr G_n \\
0 & \frac{1}{2\sigma_0^2} tr G_n & \frac{1}{4\sigma_0^4}
\end{array} \right) \) is a symmetric matrix with \( \mu_4 \) being the fourth moment of \( v_{it} \), where \( G_{n,ii} \) is the \((i, i)\) entry of \( G_n \). When \( V_{nt} \) are normally distributed, \( \Omega_{\theta_0,nT} = 0 \) because \( \mu_4 - 3\sigma_0^4 = 0 \) for a normal distribution. Denote \( \Sigma_{\theta_0} = \lim_{T \to \infty} \Sigma_{\theta_0,nT} \) and \( \Omega_{\theta_0} = \lim_{T \to \infty} \Omega_{\theta_0,nT} \) (note that \( \lim_{T \to \infty} \Sigma_{\theta_0,nT} = \lim_{T \to \infty} \Sigma_{\theta_0,nT} \) and \( \lim_{T \to \infty} \Omega_{\theta_0,nT} = \lim_{T \to \infty} \Omega_{\theta_0,nT} \)), then,

\[
\lim_{T \to \infty} E \left( \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial \theta'} \right) = \Sigma_{\theta_0} + \Omega_{\theta_0}. \quad (E.6)
\]

**Claim E.1** Under Assumption 1, 2, 6, 7, 8,

\[
\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d}(\theta_0)}{\partial \theta} + \Delta_{1,nT} + \Delta_{2,nT} \overset{d}{\to} N(0, \Sigma_{\theta_0} + \Omega_{\theta_0}). \quad (E.7)
\]

When \( \{v_{it}\}, i = 1, 2, ..., n \) and \( t = 1, 2, ..., T \), are normal, \( \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d}(\theta_0)}{\partial \theta} + \Delta_{1,nT} + \Delta_{2,nT} \overset{d}{\to} N(0, \Sigma_{\theta_0}). \)

**Proof.** See Appendix E.2. ■

**Claim E.2** Under Assumption 1, 2, 6, 7, 8, \( \frac{1}{nT} \frac{\partial^2 \ln L_{n,T}^{d}(\theta)}{\partial \theta \partial \theta'} - \frac{1}{nT} \frac{\partial^2 \ln L_{n,T}^{d}(\theta)}{\partial \theta \partial \theta'} = \| \theta - \theta_0 \| \cdot O_p(1). \)

**Proof.** Its proof is similar to that of Equation (C.5). ■
Claim E.3 Under Assumption 1',2-6,7',8, \( \frac{1}{nT} \partial^2 \ln L_{n,T}^d(\theta_0) - \frac{1}{nT} E \frac{\partial^2 \ln L_{n,T}^d(\theta_0)}{\partial \theta \partial \theta^T} = O_p \left( \frac{1}{\sqrt{nT}} \right) \).

**Proof.** Its proof is similar to that of Equation (C.6). □

Using Claim E.1, Claim E.2 and Claim E.3, we have Theorem 4.2.

### E.2 Proof of Claim E.1

From (4.4)-(4.7), the score is decomposed into 3 parts such that

\[
\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^d(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial \theta} - \Delta_{1,nT} - \Delta_{2,nT}. \tag{E.8}
\]

For the first part, it’s a linear and quadratic form of \( V_{nt} \) and the asymptotic distribution can be derived from the central limit theorem for martingale difference arrays (Theorem B.3). For the term \( \Delta_{1,nT} \), from Equation (B.3) in Lemma B.1 and Equation (B.5) in Theorem B.2, we have \( \Delta_{1,nT} = \sqrt{\frac{n-1}{T}} a_{\theta_0,n,1} + O(\sqrt{\frac{n-1}{T}}) + O_p(\frac{1}{\sqrt{nT}}) \) where \( a_{\theta_0,n,1} \) is \( O(1) \) and is specified in Equation (4.8). Also, \( \Delta_{2,nT} \) is specified in (4.9). Combining above together, we get the result. □

### E.3 Proof of Theorem 4.1

For the log likelihood function (4.1) divided by the sample size \( nT \), we have corresponding \( Q_{n,T}^d(\theta) = E_{\text{max}_{c_n,\alpha_T}} \frac{1}{nT} \ln L_{n,T}(\theta, c_n, \alpha_T) \). Hence,

\[
Q_{n,T}^d(\theta) = \frac{1}{nT} E \ln L_{n,T}^d(\theta) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 + \frac{1}{n} \ln |S_n(\lambda)| - \frac{1}{2 \sigma^2} E \frac{1}{nT} \sum_{t=1}^{T} \hat{v}'_{nt}(\theta) \hat{v}_{nt}(\theta). \tag{E.9}
\]

Similarly as the proof of Claim 3.1 in Appendix D.1, we can show the uniform convergence of \( \frac{1}{nT} E \ln L_{n,T}^d(\theta) - Q_{n,T}^d(\theta) \) and that \( Q_{n,T}^d(\theta) \) is uniformly equicontinuous in \( \theta \) in any compact parameter space \( \Theta \). When \( \lim_{T \to \infty} \mathcal{H}_{n,T}^d \) is nonsingular, we can show that \( \hat{\theta}_{nT}^d \) is globally identified and consistent similarly as proof of Theorem 3.2 in Appendix D.3. When \( \lim_{T \to \infty} \mathcal{H}_{n,T}^d \) is singular, as long as

\[
\lim_{n \to \infty} \left( \frac{1}{n} \ln |\sigma_0^2 \sigma_0^{-1} S_0^{-1}| - \frac{1}{n} \ln |\sigma_0^2(\lambda) S_0^{-1}(\lambda) S_0^{-1}(\lambda)| \right) \neq 0 \text{ for } \lambda \neq \lambda_0,
\]

\( \theta_0 \) is still globally identified and \( \hat{\theta}_{nT}^d \overset{p}{\to} \theta_0 \). The proof can be done similarly as Appendix D.4.

### E.4 Proof of Theorem 4.2

The Taylor expansion gives

\[
\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_0) = \left( -\frac{1}{nT} \frac{\partial^2 \ln L_{n,T}^d(\theta_{nT})}{\partial \theta \partial \theta^T} \right)^{-1} \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^d(\theta_0)}{\partial \theta}
\]

where \( \tilde{\theta}_{nT} \) lies between \( \theta_0 \) and \( \hat{\theta}_{nT}^d \) and \( \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^d(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial \theta} - \Delta_{1,nT} - \Delta_{2,nT}. \)
As
\[-\frac{1}{n^T} \frac{\partial^2 \ln L_{n,T}^d(\tilde{\theta}_{nT})}{\partial \theta \partial \theta'} = \left( -\frac{1}{n^T} \frac{\partial^2 \ln L_{n,T}^d(\tilde{\theta}_{nT})}{\partial \theta \partial \theta'} - \left( -\frac{1}{n^T} \frac{\partial^2 \ln L_{n,T}^d(\theta_0)}{\partial \theta \partial \theta'} \right) \right) + \left( -\frac{1}{n^T} \frac{\partial^2 \ln L_{n,T}^d(\theta_0)}{\partial \theta \partial \theta'} - \Sigma_{\theta_0,nT}^d \right) = \Sigma_{\theta_0,nT}^d\]

where the first term is $\|\tilde{\theta}_{nT} - \theta_0\| \cdot O_p(1)$ (Claim E.2) and the second term is $O_p\left( \frac{1}{\sqrt{n^T}} \right)$ (Claim E.3), we have
\[-\frac{1}{n^T} \frac{\partial^2 \ln L_{n,T}^d(\tilde{\theta}_{nT})}{\partial \theta \partial \theta'} = \|\tilde{\theta}_{nT} - \theta_0\| \cdot O_p(1) + O_p\left( \frac{1}{\sqrt{n^T}} \right) + \Sigma_{\theta_0,nT}^d.\]

Because (1) $\|\tilde{\theta}_{nT} - \theta_0\| = o_p(1)$ as $\tilde{\theta}_{nT}$ is consistent and (2) $\Sigma_{\theta_0,nT}^d$ is the nonsingular in the limit according to Assumption 8, we have
-1
\[-\frac{1}{n} \frac{\partial^2 \ln L_{n,T}^d(\tilde{\theta}_{nT})}{\partial \theta \partial \theta'} \]
is invertible for large $n$ and $T$ and $\left( -\frac{1}{n^T} \frac{\partial^2 \ln L_{n,T}^d(\tilde{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1}$ is $O_p(1)$.

Also, we have $\frac{1}{n^T} \frac{\partial \ln L_{n,T}^d(\tilde{\theta}_{nT})}{\partial \theta} = N(0, \Sigma_{\theta_0} + \Omega_{\theta_0})$ and $\Delta_{nT}^d = \Delta_{nT}^1 + \Delta_{nT}^2$ where $\Delta_{nT}^1$ and $\Delta_{nT}^2$ are specified in (4.6) and (4.7). Then, $\sqrt{n^T}(\tilde{\theta}_{nT} - \theta_0) = O_p(1) \cdot \left( \left( \frac{1}{n^T} \right) + O_p\left( \frac{1}{T^2} \right) \right)$, which implies that
\[
\hat{\theta}_{nT}^d - \theta_0 = O_p\left( \max\left( \frac{1}{n^T}, \frac{1}{T} \right) \right)
\]

Given that $\Sigma_{\theta_0,nT}^d$ is nonsingular and its inverse is $O(1)$, we have
\[
\left( -\frac{1}{n} \frac{\partial^2 \ln L_{n,T}^d(\tilde{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} = \Sigma_{\theta_0,nT}^{-1} = O_p\left( \max\left( \frac{1}{n}, \frac{1}{T} \right) \right)
\]

Hence,
\[
\sqrt{n^T}(\tilde{\theta}_{nT}^d - \theta_0) = \left( -\frac{1}{n^T} \frac{\partial^2 \ln L_{n,T}^d(\tilde{\theta}_{nT})}{\partial \theta \partial \theta'} \right)^{-1} - \Delta_{nT}^1 - \Delta_{nT}^2
\]
\[
= \Sigma_{\theta_0,nT}^{-1} \cdot \frac{1}{n^T} \frac{\partial \ln L_{n,T}^d(\theta_0)}{\partial \theta} + O_p\left( \max\left( \frac{1}{n^T}, \frac{1}{T} \right) \right) \cdot \frac{1}{n^T} \frac{\partial \ln L_{n,T}^d(\theta_0)}{\partial \theta} - (\Sigma_{\theta_0,nT}^{-1} \cdot \Delta_{nT}^1 + \Delta_{nT}^2) - O_p\left( \max\left( \frac{1}{n^T}, \frac{1}{T} \right) \right) \cdot (\Delta_{nT}^1 + \Delta_{nT}^2),
\]

which implies that
\[
\sqrt{n^T}(\tilde{\theta}_{nT}^d - \theta_0) + (\Sigma_{\theta_0,nT}^{-1} \cdot \Delta_{nT}^1 + \Delta_{nT}^2) + O_p\left( \max\left( \frac{1}{n^T}, \frac{1}{T} \right) \right) (\Delta_{nT}^1 + \Delta_{nT}^2)
\]
\[
= \left( (\Sigma_{\theta_0,nT}^{-1} + o_p(1)) \right) \cdot \frac{1}{n^T} \frac{\partial \ln L_{n,T}^d(\theta_0)}{\partial \theta}.
\]

As $\Sigma_{\theta_0} = \lim_{T \to \infty} \Sigma_{\theta_0,nT}^d$ exists, then using Claim E.1 and that $\Delta_{nT} = \sqrt{T}n a_n \theta_{0,1} + O(\sqrt{T^2}) + O_p\left( \frac{1}{\sqrt{T}} \right)$, we have
\[
\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n}{T}}(\Sigma_{\theta_0,nT})^{-1}a_{n,\theta_0,1} + \sqrt{\frac{T}{n}}(\Sigma_{\theta_0,nT})^{-1}a_{\theta_0,2} \\
+ O_p\left(\max\left(\sqrt{\frac{n}{T}T}, \sqrt{\frac{T}{n}}, \sqrt{\frac{1}{T}}\right)\right)
\]  
(E.13)

\[d \to N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}).\]  
(E.14)

**E.5 Proof for Theorem 4.3**

Theorem 4.2 states that \(\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_0) + \sqrt{\frac{T}{n}}b_{\theta_0,nT,1} + \sqrt{\frac{T}{n}}b_{\theta_0,nT,2} + O_p\left(\max\left(\sqrt{\frac{T}{n}T}, \sqrt{\frac{T}{n}}, \sqrt{\frac{1}{T}}\right)\right) \rightarrow N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})\). As the bias corrected estimator is

\[
\hat{\theta}_{nT}^d = \hat{\theta}_{nT} + \frac{1}{T}\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\right)^{-1}a_{1,n}(\hat{\theta}_{nT}) + \frac{1}{n}\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\right)^{-1}a_{2}(\hat{\theta}_{nT})
\]

where \(a_{1,n}(\theta) = a_{\theta,1,1}\) and \(a_{2}(\theta) = a_{\theta,2,2}\), we will have \(\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_0) \rightarrow N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})\) if

\[
\sqrt{\frac{n}{T}}\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\right)^{-1}a_{1,n}(\hat{\theta}_{nT}) + \frac{1}{n}\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\right)^{-1}a_{2}(\hat{\theta}_{nT}) \rightarrow 0,
\]  
(E.15)

\[
\sqrt{\frac{T}{n}}\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\right)^{-1}a_{2}(\hat{\theta}_{nT}) \rightarrow 0,
\]  
(E.16)

when \(\frac{n}{T} \to 0\), \(\frac{T}{n} \to 0\). Assuming that \(\frac{n}{T} \to 0\) and \(\frac{T}{n} \to 0\), we are going to prove Equation (E.15) and (E.16).

From Equation (E.11) and that \(\hat{\theta}_{nT} - \theta_0 = O_p\left(\max\left(\frac{1}{T}, \frac{1}{n}\right)\right)\) from Equation (E.10),

\[-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} = (\Sigma_{\theta_0,nT})^{-1} + O_p\left(\max\left(\frac{1}{T}, \frac{1}{n}\right)\right)\]

Hence,

\[
\sqrt{nT}\left\{\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\right)^{-1}a_{1,n}(\hat{\theta}_{nT}) + \frac{1}{n}\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\right)^{-1}a_{2}(\hat{\theta}_{nT})\right\}
\]

\[\rightarrow\]

\[\sqrt{nT}\left\{(\Sigma_{\theta_0,nT})^{-1} + O_p\left(\max\left(\frac{1}{T}, \frac{1}{n}\right)\right)\right\}a_{1,n}(\hat{\theta}_{nT}) - (\Sigma_{\theta_0,nT})^{-1}a_{1,n}(\theta_0)\}

\[=\]

\[\sqrt{nT}\left\{(\Sigma_{\theta_0,nT})^{-1}a_{1,n}(\hat{\theta}_{nT}) - a_{1,n}(\theta_0)\right\} + \frac{T}{n}a_{1,n}(\hat{\theta}_{nT}) \times O_p\left(\max\left(\frac{1}{T}, \frac{1}{n}\right)\right)\]

As \(\hat{\theta}_{nT} - \theta_0 = O_p\left(\max\left(\frac{1}{T}, \frac{1}{n}\right)\right)\) and \(a_{1,n}(\theta_0)\) is \(O(1)\), according to Taylor expansion of \(a_{1,n}(\hat{\theta}_{nT})\) around \(a_{1,n}(\theta_0)\), to prove Equation (E.15) is reduced to prove that the sequence of \(\frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial \theta \partial \theta'}\) is bounded, uniformly in a small neighborhood of \(\theta_0\), which can be proved by similar argument at the end of Appendix D.8.
For Equation (E.16), we have

\[
\sqrt{\frac{T}{n}} \left\{ -\frac{1}{nT} E \frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial \theta \partial \theta'} \right\}^{-1} a_2(\hat{\theta}_{nT}) - (\Sigma_{\theta_0,nT})^{-1} a_2(\theta_0) \right\}
\]

\[
= \sqrt{\frac{T}{n}} \left\{ \left( \Sigma_{\theta_0,nT} \right)^{-1} + O_p \left( \max \left( \frac{1}{T}, \frac{1}{n} \right) \right) \right\} a_2(\hat{\theta}_{nT}) - (\Sigma_{\theta_0,nT})^{-1} a_2(\theta_0) \right\}
\]

\[
= \sqrt{\frac{T}{n}} \left\{ \left( \Sigma_{\theta_0,nT} \right)^{-1} a_2(\hat{\theta}_{nT}) - a_2(\theta_0) \right\} + \sqrt{\frac{T}{n}} a_2(\hat{\theta}_{nT}) \times O_p \left( \max \left( \frac{1}{T}, \frac{1}{n} \right) \right).
\]

As \( \hat{\theta}_{nT} - \theta_0 = O_p \left( \max \left( \frac{1}{T}, \frac{1}{n} \right) \right) \) and \( a_2(\theta_0) \) is \( O(1) \), according to Taylor expansion of \( a_2(\hat{\theta}_{nT}) \) around \( a_2(\theta_0) \), to prove Equation (E.16) is reduced to prove that elements of \( \frac{\partial a_2(\hat{\theta}_{nT})}{\partial \theta} \) implied by (4.7), it is proved.

\[
\text{References}
\]


