Threshold Spatial Panel Regression with Fixed Effects

Xiaoyu Meng[1] and Zhenlin Yang[2]

[1]*School of Economics, Nankai University, Tianjin, China*

[2]*School of Economics, Singapore Management University, Singapore*

February 21, 2025

Abstract

We introduce general estimation and inference methods for threshold spatial panel regression with two-way fixed effects in a diminishing-threshold-effects framework. A valid objective function is obtained through a simple adjustment on the concentrated quasi log-likelihood with fixed effects being concentrated out, which leads to a consistent estimation of all common parameters. We show that the estimation of threshold parameter has a negligible effect on the asymptotic distribution of the main parameter estimators and thereby regular inference methods apply, though a bias correction may be necessary. The limiting distribution of the threshold parameter estimator is shown to be non-regular and infeasible, and for inference, a likelihood ratio test procedure is proposed. The test for the non-existence of threshold effects faces an identification issue at the null. To overcome this difficulty, we propose a sup-Wald test and a bootstrap method for its critical values. Monte Carlo results show that the proposed methods perform well in finite samples. An empirical application is presented on age-of-leader effects on political competitions across Chinese cities.

Key Words: Fixed effects; Spatial panels; Threshold effects; sup-Wald test; Bootstrap critical values; Time-varying spatial weights.

JEL classifications: C10, C13, C21, C23, C15

1. Introduction

Spatial panel regression (SPR) model has been receiving increasing attention from researchers since Anselin (1988), for its ability to model cross-sectional dependence while maintaining full control of unobservable heterogeneity. See, among others, Anselin et al. (2008), Baltagi et al. (2003), Lee and Yu (2010a), Baltagi and Yang (2013a,b), Yang et al. (2016), Kelejian and Piras (2016), and Liu and Yang (2020). See Lee and Yu (2010b, 2015) for surveys.

The threshold panel regression (TPR) model is another popular model specification, which divides observations into distinct regimes, depending on whether the value of an observable variable (threshold variable) exceeds some threshold value. See, among others, Hansen (1996, 1999, 2000), and Miao et al. (2020). Hansen (2011) presents an overview of the development of threshold models in both econometrics and economics literature.

Spatial effecs and threshold effects are both relevant in a wide range of empirical applications. In social sciences, Schelling (1971) identifies “neighborhood tipping” in residential segregation, where threshold effects occur as a new minority reaches a critical mass in a neighborhood, triggering spatial interactions between the new minority and existing residents, which influences behaviors such as moving out and resistance to integration. Strong empirical evidence for this phenomenon was found by Card et al. (2008). In public economics, Glaeser et al. (1996) analyze crime and social interactions, finding that the extent of spatial interactions in criminal behavior depends on the severity of the crime, and such interactions drive the spread of crime through social networks within communities. In finance, Pesaran and Pick (2007) argue that financial crises spread across markets due to threshold effects, where interdependence and financial contagion are triggered only after financial stress or instability across critical levels.

Apparently, empirical applications “have rushed ahead” of econometric theory in this field, as in many other fields. Therefore, there is an urgent need for solid econometric foundations for spatial panel regressions with threshold effects, which would formally join the two strands of literature, the SPR and the TPR. Some preliminary attempts have been made. Li (2018) studies an SPR model with structural change, a special case of the threshold SPR model with the threshold variable being simply the time variable. A direct quasi-maximum likelihood (QML) approach is proposed for model estimation, where the *incidental parameters problem* (Neyman and Scott, 1948) due to the estimation of fixed effects is not addressed. Wei et al. (2021) extend Hansen (1999) to allow spatial autoregressive (SAR) structure in a panel model, but the asymptotic properties of the estimators are not studied and the inference methods for the threshold parameter are not given. In addition, the estimation method they propose is a twostage least squares (2SLS) estimation (Caner and Hansen, 2004), which is inefficient compared to ML-type estimation for spatial models. Both works require spatial weight matrices to be time-invariant and do not consider spatial error dependence. Both studies do not consider the time-specific fixed effects, but these effects might be important and have empirical implications in many economic studies (e.g., Ertur and Koch 2007; Elhorst and Fréret 2009). Moreover, the additional time fixed effects further complicate the incidental parameters problem as shown in an SPR framework (Lee and Yu, 2010a). Wu and Matsuda (2021) extend the M-estimation of dynamic SPR model with short panels of Yang (2018) by adding threshold effects where the threshold parameter is assumed to be known and the threshold variable to be time-invariant.

In this paper, we carry out a formal and general study by relaxing the above mentioned restrictions, to provide general estimation, testing, and inference methods for the threshold spatial panel regression (TSPR) models with an unknown threshold parameter and a general threshold variable, both individual- and time-specific fixed effects, time-varying spatial weights, and spatial error dependence. The key challenges in this study are: (i) spatial effects introduce nonlinearity in the estimation and the least-squares methods of Hansen (1999, 2000) cannot be applied; (ii) the two-way fixed effects introduce incidental parameters problem (Neyman and Scott, 1948) and the standard transformation methods (Lee and Yu, 2010a) cannot be applied to solve this problem due to the time-varying nature of spatial-threshold effects; (iii) the threshold parameter estimate does not have a proper asymptotic distribution that hinders the statistical inference; and (iv) the threshold parameter is not identified at the null hypothesis of no threshold effects, making the test for the non-existence of threshold effects a great challenge.

The contributions of this paper are four-fold: (i) we proposed an adjusted QML method for the estimation of the key model parameters by extending Hansen's (1999, 2000) *diminishing-threshold-effect* (DTE) asymptotic framework, to deal with the challenges arising from the existence of threshold effects in both regression slopes and spatial parameter(s) and the existence of incidental two-way fixed effects (FEs). Here, the QML method addresses the issue of nonlinearity and the adjustment to the concentrated quasi-likelihood solves the incidental parameters problem when T (the time dimension) is smaller than n (the cross-section dimension). (ii) When n and T are of the same magnitude, we derive a bias correction on the adjusted QML estimates to remove the potential asymptotic biases caused by the estimation of incidental FEs. These methods (in (i) and (ii)) do away from transformation by carrying out *concentration and correction* and allow the spatial weight matrices to be time-varying. They lead to simple inference methods for the main model parameters valid irrespective of relative magnitude of n and T (see Section 2.3.2). (iii) We show that the limiting distribution of the estimator of the threshold parameter is a scalar multiple of a functional of Brownian motion, where the scalar cannot be accurately estimated without prior knowledge on the diminishing rate of the threshold effect. To overcome this difficulty, we introduce a likelihood ratio (LR) test method for inference on the threshold parameter under the DTE framework. (*iv*) We introduce a sup-Wald test to test the non-existence of threshold effects. We show that the limiting distribution of the sup-Wald test is a functional of chi-square processes and thus its asymptotic critical values cannot be tabulated. To overcome this difficulty, we introduce a bootstrap method which is shown to give a valid approximation to the asymptotic critical values. The following details may help to appreciate the methods used and the results obtained.

First, our adjusted QML estimator addresses the incidental parameters problem by concentrating out fixed effects and adjusting the concentrated loglikelihood to account for degrees of freedom loss caused by estimating the fixed effects. This adjustment ensures a consistent estimation of the threshold, spatial, and regression coefficients, as well as the error variance, provided that T does not grow faster than n. Second, we establish that the threshold estimator converges at a rate linked to DTE, and its estimation error has an asymptotically negligible impact on other estimators, thereby preserving their asymptotic normality with the standard convergence rate. When T and n grow proportionally, these estimators exhibit asymptotic bias due to the incidental parameters problem. Third, unlike Hansen (1999), the LR test statistic, proposed to facilitate inference on the threshold parameter, is based on the adjusted log-likelihood function. Its limiting distribution is unaffected by the asymptotic bias of other estimators under their standard convergence rate. Furthermore, its asymptotic distribution is not pivotal under homoskedasticity alone, as its limiting distribution depends on the third and fourth moments of the errors. However, it becomes pivotal under both normality and homoskedasticity, and a nonparametric correction can be applied otherwise to estimate the unknown parameters in its limit distribution. Finally, the sup-Wald statistic, building on Hansen (1996), is constructed to test for the non-existence of threshold effects using adjusted QML estimators, with its limiting distribution derived under local-to-null alternatives. The proposed bootstrap procedure works on an asymptotically equivalent version of the sup-Wald statistic and provides asymptotically correct critical values. Monte Carlo results show that the proposed methods perform well.

The practical relevance of allowing for threshold effects in SPR models is illustrated by studying the threshold effect of leaders' age on political competitions across 338 cities in China over the periods 2010 to 2012. Political competition among city leaders of the same level is identified by spatial effects in total investments across city level. We find that the competitions are strong for local leaders who are younger than a threshold age, but tend to vanish for those who are older than the threshold level, approaching the retirement age.

The rest of the paper goes as follows. Section 2 introduces our model and assumptions, discusses the adjusted QML estimation and its asymptotic properties, and studies the likelihood ratio test on the threshold value. Section 3 studies the hypothesis test on threshold effects. Monte Carlo simulation findings are given in Section 4. Section 5 applies our method to study the effects of the age of leaders on political competitions across Chinese cities. Section 6 discusses some extensions. Section 7 concludes. Proofs are collected in the appendices.

Notation. I_m denotes an $m \times m$ identity matrix, l_m an $m \times 1$ vector of ones, and $0_{n\times m}$ an $n \times m$ matrix of zeros. For a square matrix, $|\cdot|$ denotes its determinant, $\mathtt{tr}(\cdot)$ its trace, and $\mathtt{diagv}(\cdot)$ a column vector of its diagonal elements. $\mathtt{diag}(\cdot)$ forms a diagonal matrix using the diagonal elements of a square matrix or a vector. $\mathtt{bdiag}(\cdots)$ forms a block-diagonal matrix based on given matrices/vectors/scalars. For a real symmetric matrix, $\rho_{\min}(\cdot)$ denotes its smallest eigenvalue. For a real $n \times m$ matrix A with elements a_{ij}, $\|A\|$ denotes its Frobenius norm, $\|A\|_1 = \max_{1\leqslant j\leqslant m} \sum_{i=1}^{n} |a_{ij}|$ its maximum column sum norm, and $\|A\|_\infty = \max_{1\leqslant i\leqslant n} \sum_{j=1}^{m} |a_{ij}|$ its maximum row sum norm. The operator $\otimes$ represents a Kronecker product, The true value of a parameter is denoted by adding a subscript 0. Finally, $A^s = A + A'$.

2. Model and Adjusted QML Estimation

2.1. Threshold SPR model with fixed effects

Consider a total of n spatial units, interconnected at time t through an $n \times n$ spatial weight matrix W_t. There exists a threshold variable q_{it} such that depending on its value the spatial and regression coefficients may differ. Let $d_{it}(\gamma) = \mathbb{1}(q_{it} \leqslant \gamma)$, where $\mathbb{1}(\cdot)$ is the indicator function and γ is the **threshold parameter** taking values in a bounded set $\Gamma = [\underline{\gamma},\ \overline{\gamma}]$. Define $d_t(\gamma) = \mathtt{diag}\{d_{1t}(\gamma), \ldots, d_{nt}(\gamma)\}$. We first consider a model of the form:

$$Y_t = \lambda_{10} W_t Y_t + \lambda_{20} d_t(\gamma_0) W_t Y_t + X_t \beta_{10} + d_t(\gamma_0) X_t \beta_{20} + \mu_0 + \alpha_{t0} l_n + V_t, \tag{2.1}$$

$t = 1, \ldots, T$, where $Y_t = (y_{1t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of responses at time t, $X_t = (x_{1t}, \ldots, x_{nt})'$ is an $n \times k$ matrices containing the values of k time-varying regressors, and $V_t = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ is an $n \times 1$ vector of idiosyncratic errors. The λ_{10} and β_{10} $(k \times 1)$ characterize the baseline spatial lag effect and covariates effects, and λ_{20} and β_{20} $(k \times 1)$ are the corresponding **threshold effects**. $\mu_0 = \{\mu_{i0}\}_{i=1}^{n}$ is an $n \times 1$ vector of individual-specific effects and $\alpha_0 = \{\alpha_{t0}\}_{t=1}^{T}$ is a $T \times 1$ vector of time-specific effects, which are allowed to be correlated with X_t in an arbitrary manner. Therefore, the model is referred to, in this paper, as the *threshold spatial panel regression* (TSPR) model with *two-way fixed effects* (2FE).[1]

The neighborhood structure of the n spatial units at period t is captured by a time-varying spatial weight matrix W_t, and the magnitude of the interaction effects from its neighbors is measured by the spatial lag parameters. Thus, Model (2.1) implies that each spatial unit i in any period t receives a certain level of interaction effects from its neighbors (measured by λ_{10} or $\lambda_{10} + \lambda_{20}$), depending on the level of its threshold variable q_{it}. When γ_0 is known, our model can simply be treated as a second-order SPR model of two spatial lag terms with weight matrices being W_t and $d_t(\gamma_0)W_t$, respectively. An additional complication arises because $d_t(\gamma_0)W_t$ is no longer time-invariant, even if W_t is. This renders the transformation method (Lee and Yu, 2010a), which requires time-invariant and row-normalized weights, inapplicable for Model (2.1). Thus, an alternative method that accommodates time-varying $d_t(\gamma)$ and allows for non-row-normalized, time-varying spatial weights is needed. When γ_0 is unknown and must be estimated, the situation becomes much more challenging due to the involvement of step functions $d_{it}(\gamma)$ in the likelihood that renders standard likelihood inference methods invalid.

To overcome these difficulties, we approach the estimation and inference problems for the TSPR-2FE model using an adjusted QML method. We show that under DTE the estimation of the threshold parameter γ does not have an asymptotic impact on the joint asymptotic distribution of the other common parameters, leading to valid estimation and inference methods for these parameters. An LR test procedure is proposed for inference for γ.

2.2. Adjusted QML estimation

Denote $\lambda = (\lambda_1, \lambda_2)'$, $\beta = (\beta_1', \beta_2')'$, and $\theta = (\beta', \lambda', \sigma^2)'$. Define $\mathbf{W} = \texttt{bdiag}(W_1, \ldots, W_T)$, $\mathbf{D}(\gamma) = \texttt{bdiag}(d_1(\gamma), \ldots, d_T(\gamma))$, $\mathbf{Y} = (Y_1', \ldots, Y_T')'$, $\mathbf{X} = (X_1', \ldots, X_T')'$, $\mathbf{C}_\mu = l_T \otimes I_n$, $\mathbf{C}_\alpha = I_T \otimes l_n$, $\mathbf{V} = (V_1', \ldots, V_T')'$, and $\mathbb{X}(\gamma) = [\mathbf{X},\ \mathbf{D}(\gamma)\mathbf{X}]$. Model (2.1) can be written in matrix form: $\mathbf{Y} = \lambda_1 \mathbf{W}\mathbf{Y} + \lambda_2 \mathbf{D}(\gamma)\mathbf{W}\mathbf{Y} + \mathbb{X}(\gamma_0)\beta_0 + \mathbf{C}_\mu \mu_0 + \mathbf{C}_\alpha \alpha_0 + \mathbf{V}$. To avoid the unidentification of μ_{i0} and α_{t0} as $\mu_{i0} + \alpha_{t0} = (\mu_{i0} + c) + (\alpha_{t0} - c)$ for an arbitrary c, we impose a zero-sum constraint on α_t so that $\alpha_1 = -\sum_{t=2}^{T} \alpha_t$. Thus, the QML estimation is based on the model form:

$$\mathbf{Y} = \lambda_1 \mathbf{W}\mathbf{Y} + \lambda_2 \mathbf{D}(\gamma)\mathbf{W}\mathbf{Y} + \mathbb{X}(\gamma_0)\beta_0 + \mathbf{C}\psi_0 + \mathbf{V}, \tag{2.2}$$

where $\mathbf{C} = [\mathbf{C}_\mu,\ \mathbf{C}_\alpha^\star]$, $\mathbf{C}_\alpha^\star = [-l_n l_{T-1}';\ I_{T-1} \otimes l_n]$, $\psi = (\mu', \alpha^{*\prime})'$, and $\alpha^* = (\alpha_2 \ldots, \alpha_T)'$. Under the exogeneity of (q_{it}, X_t, W_t), the quasi Gaussian loglikelihood **as if** $\{v_{it}\}$ were iid $N(0, \sigma_0^2)$ is,

$$\ell_{nT}(\theta, \gamma, \psi) = -\frac{nT}{2} \ln(2\pi\sigma^2) + \ln |\mathbf{A}(\lambda, \gamma)| - \frac{1}{2\sigma^2} \mathbf{V}'(\beta, \lambda, \gamma, \psi)\mathbf{V}(\beta, \lambda, \gamma, \psi), \tag{2.3}$$

[1]Model (2.1) allows spatial Durbin effects to be a part of X_t. It can be readily extended to include spatial error dependence and serial correlation. Further extensions are also possible. See Sec. 6 for details.

where $\mathbf{A}(\lambda,\gamma) = I_{nT} - \lambda_1 \mathbf{W} - \lambda_2 \mathbf{D}(\gamma)\mathbf{W}$, and $\mathbf{V}(\beta,\lambda,\gamma,\psi) = \mathbf{A}(\lambda,\gamma)\mathbf{Y} - \mathbb{X}(\gamma)\beta - \mathbf{C}\psi$. Given the set of *common parameters* (θ,γ), the first-order condition for the FE parameters ψ implies:

$$\hat{\psi}_{nT}(\beta,\lambda,\gamma) = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'[\mathbf{A}(\lambda,\gamma)\mathbf{Y} - \mathbb{X}(\gamma)\beta]. \tag{2.4}$$

Upon substitution, we have the concentrated quasi Gaussian loglikelihood function for (θ,γ):

$$\ell^c_{nT}(\theta,\gamma) = -\frac{nT}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{A}(\lambda,\gamma)| - \frac{1}{2\sigma^2}\tilde{\mathbf{V}}'(\beta,\lambda,\gamma)\tilde{\mathbf{V}}(\beta,\lambda,\gamma), \tag{2.5}$$

where $\tilde{\mathbf{V}}(\beta,\lambda,\gamma) \equiv \mathbf{Q}_{nT}[\mathbf{A}(\lambda,\gamma)\mathbf{Y} - \mathbb{X}(\gamma)\beta]$ and $\mathbf{Q}_{nT} = I_{nT} - \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}' = (I_T - \frac{l_T l_T'}{T}) \otimes (I_n - \frac{l_n l_n'}{n})$. Maximizing $\ell^c_{nT}(\theta,\gamma)$ gives the *direct* QML estimators $\hat{\theta}^d_{nT}$ and $\hat{\gamma}^d_{nT}$ of θ and γ.

Note that, for the regular SPR-2FE model (i.e., $\beta_2 = 0$ and $\lambda_2 = 0$), Lee and Yu (2010a) show that when T is fixed the direct approach only gives consistent estimators for spatial and regression parameters β and λ but not for the variance parameter σ^2 - the well known *incidental parameters problem* of Neyman and Scott (1948). Furthermore, even when n and T are both large, the asymptotic distribution of the normed estimators of common parameters would not be properly centered. An intuitive interpretation is that the direct QML estimator of σ^2 failed to "recover" the effect of degrees of freedom loss due to the estimation of $(n + T - 1)$ FE parameters μ and α. The TSPR-2FE model may face the same issues.

Considering the fact that $\tilde{\mathbf{V}}'(\beta,\lambda,\gamma)\tilde{\mathbf{V}}(\beta,\lambda,\gamma)$ has $(n-1)(T-1)$ degrees of freedom, that is, $\mathrm{E}[\tilde{\mathbf{V}}'(\beta_0,\lambda_0,\gamma_0)\tilde{\mathbf{V}}(\beta_0,\lambda_0,\gamma_0)] = O((n-1)(T-1))$, we find that a simple adjustment to $\ell^c_{nT}(\theta,\gamma)$ will achieve the consistency in the joint estimation of θ and γ:

$$\ell^*_{nT}(\theta,\gamma) = -\frac{nT}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{A}(\lambda,\gamma)| - \frac{c_{nT}}{2\sigma^2}\tilde{\mathbf{V}}'(\beta,\lambda,\gamma)\tilde{\mathbf{V}}(\beta,\lambda,\gamma), \tag{2.6}$$

where $c_{nT} = \frac{nT}{(n-1)(T-1)}$ (see below for a detailed theoretical reasoning). Therefore, the adjusted QML (AQML) estimators of θ and γ are defined as follows

$$(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) = \underset{(\theta,\gamma)\in\Theta\times\Gamma}{\operatorname{argmax}}\ \ell^*_{nT}(\theta,\gamma),$$

where Θ is the parameter space for θ and Γ is the parameter space for γ defined above (2.1).

To solve the above maximization problem, we first maximize the above objective function for a given γ to obtain an estimate $\hat{\theta}_{nT}(\gamma)$ of θ. Then, we define $\ell^{*c}_{nT}(\gamma) \equiv \ell^*_{nT}(\hat{\theta}_{nT}(\gamma),\gamma)$, and search over $\Gamma_{nT} = \Gamma \cap \{q_{it}, 1 \leqslant i \leqslant n, 1 \leqslant t \leqslant T\}$ for $\hat{\gamma}_{nT}$ that maximizes $\ell^{*c}_{nT}(\gamma)$. This is because the objective function $\ell^{*c}_{nT}(\gamma)$ is a step function with at most nT steps as it depends on γ only through the indicator function $\mathbb{1}\{q_{it} \leqslant \gamma\}$. When nT is large, Hansen (1999) suggests that, to reduce the computational burden, the search can be restricted to a grid of N_0 specific quantiles for some $N_0 < nT$, $\Gamma_{N_0} = \{q_{(1)}, \ldots, q_{(N_0)}\}$, where $q_{(j)}$ is the $[\eta + \frac{j-1}{N_0-1}(1-2\eta)]$th quantile of the sample q_{it} and $\eta = 1\%$ or 5%. Then $\hat{\gamma}_{N_0} = \operatorname{argmax}_{\gamma\in\Gamma_{N_0}} \ell^{*c}_{nT}(\gamma)$ is a good approximation to

$\hat{\gamma}_{nT}$. Given $\hat{\gamma}_{nT}$, the AQMLE of θ is just $\hat{\theta}_{nT} \equiv \hat{\theta}_{nT}(\hat{\gamma}_{nT})$.

Validity of the objective function $\ell^*_{nT}(\theta,\gamma)$. We define $\mathbf{G}(\lambda,\gamma) = \mathbf{W}\mathbf{A}^{-1}(\lambda,\gamma)$. In what follows, a parametric quantity at true parameter(s) is denoted by dropping its argument(s), e.g., $\mathbf{A} = \mathbf{A}(\lambda_0,\gamma_0)$ and $\mathbf{G} = \mathbf{G}(\lambda_0,\gamma_0)$. For $\hat{\theta}_{nT}$ to be consistent, i.e., $\hat{\theta}_{nT} - \theta_0 = o_p(1)$, it is necessary that $\text{plim}_{n,T\to\infty}\frac{1}{nT}S^*_{\theta,nT}(\theta_0,\gamma_0) = 0$, where,

$$S^*_{\theta,nT}(\theta,\gamma) = \frac{\partial}{\partial\theta}\ell^*_{nT}(\theta,\gamma) = \begin{cases} \frac{c_{nT}}{\sigma^2}\mathbb{X}'(\gamma)\tilde{\mathbf{V}}(\beta,\lambda,\gamma), \\ \frac{c_{nT}}{\sigma^2}\mathbf{Y}'\mathbf{W}'\tilde{\mathbf{V}}(\beta,\lambda,\gamma) - \mathtt{tr}[\mathbf{G}(\lambda,\gamma)], \\ \frac{c_{nT}}{\sigma^2}\mathbf{Y}'\mathbf{W}'\mathbf{D}'(\gamma)\tilde{\mathbf{V}}(\beta,\lambda,\gamma) - \mathtt{tr}[\mathbf{D}(\gamma)\mathbf{G}(\lambda,\gamma)], \\ \frac{c_{nT}}{2\sigma^4}\tilde{\mathbf{V}}'(\beta,\lambda,\gamma)\tilde{\mathbf{V}}(\beta,\lambda,\gamma) - \frac{nT}{2\sigma^2}. \end{cases} \tag{2.7}$$

Under Assumption A below and using the facts: $\tilde{\mathbf{V}}(\beta_0,\lambda_0,\gamma_0) = \mathbf{Q}_{nT}\mathbf{V}$, $\mathbf{W}\mathbf{Y} = \mathbf{G}\mathbf{A}\mathbf{Y} = \mathbf{G}(\mathbb{X}\beta_0 + \mathbf{C}\psi_0 + \mathbf{V})$, $\mathbf{G}$ is block diagonal, and $\mathbf{Q}_{nT} = (I_T - \frac{l_T l_T'}{T}) \otimes (I_n - \frac{l_n l_n'}{n})$, we have,

$$\mathrm{E}[S^*_{\theta,nT}(\theta_0,\gamma_0)] = \mathrm{E}\big(0_{1\times 2k}, -\tfrac{1}{n-1}\mathtt{tr}(\bar{\mathbf{G}}\mathbf{J}), -\tfrac{1}{n-1}\mathtt{tr}[\mathbf{D}(\gamma_0)\bar{\mathbf{G}}\mathbf{J}],\ 0\big), \tag{2.8}$$

where $\bar{\mathbf{G}} = \mathbf{G} - \mathtt{diag}(\mathbf{G})$ and $\mathbf{J} = I_T \otimes l_n l_n'$. As $\mathbf{G}$ is bounded in both row and column sum norms (Lemma A.1 and Assumptions C and D), $\mathrm{E}[\mathtt{tr}(\bar{\mathbf{G}}\mathbf{J})]$ and $\mathrm{E}[\mathtt{tr}(\mathbf{D}(\gamma_0)\bar{\mathbf{G}}\mathbf{J})]$ are both $O(nT)$. Thus, $\mathrm{E}[S^*_{\theta,nT}(\theta_0,\gamma_0)] = O(T)$. The adjustment corrects the degrees of freedom loss, and the effective sample size becomes $N = (n-1)(T-1)$. It follows that $\frac{1}{N}S^*_{\theta,nT}(\theta_0,\gamma_0) = O_p(\frac{1}{n})$. Hence, a consistent estimation is possible for all common parameters based on maximizing $\ell^*_{nT}(\theta,\gamma)$, whether T is fixed or increases with n. Furthermore, $\frac{1}{\sqrt{N}}S^*_{\theta,nT}(\theta_0,\gamma_0) = O_p(\sqrt{\frac{T}{n}})$, and therefore the asymptotic distribution of $\sqrt{N}(\hat{\theta}_{nT} - \theta_0)$ would be centered as long as $\frac{T}{n} = o(1)$.[2] When $\frac{T}{n} = O(1)$, the asymptotic distribution of $\sqrt{N}(\hat{\theta}_{nT} - \theta_0)$ may not be centered, and in this case (2.8) provides a simple way for bias correction. See Sec. 2.3.2. for details.

2.3. Asymptotic properties of the adjusted QML estimators

We study the consistency and asymptotic distributions of the adjusted QML estimators $\hat{\theta}_{nT}$ and $\hat{\gamma}_{nT}$. We show that $\hat{\theta}_{nT}$ has regular asymptotic properties but $\hat{\gamma}_{nT}$ has a non-regular convergence rate and asymptotic distribution, due to the DTE assumption.

Denote $\mathbf{Z} = \mathbf{Z}(\beta_0,\lambda_0,\psi_0,\gamma_0) = \mathbf{G}(\lambda_0,\gamma_0)(\mathbb{X}(\gamma)\beta_0 + \mathbf{C}\psi_0)$, which acts as additional regressors as discussed below Assumption G. Let $\mathbf{H} = [\mathbf{X},\ \mathbf{Z}]$ with rows $\{h_{it}'\}$, and $M(\gamma) = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \mathrm{E}(h_{it}h_{it}'|q_{it} = \gamma)$. Let $f(\cdot)$ be the probability density function of q_{it} and $f_2(\cdot,\cdot)$ the joint probability density function of (q_{it}, q_{jt}). The value of $f(\cdot)$ at $q_{it} = \gamma_0$, $f(\gamma_0)$, plays an

[2] Our adjusted QML approach falls in spirit to the "*Bias-Correction of the Concentrated Likelihood function*" of Arellano and Hahn (2007). Cox and Reid's (1987) adjusted profile likelihood approach also belongs to this category but requires that the parameters of interest are orthogonal to the nuisance parameters.

important role. Note, with the new notational convention, $M = M(\gamma_0)$. Formal asymptotic analyses are based on the following assumptions.

Assumption A: *The innovations v_{it} are independent and identically distributed (i.i.d.) in i and t, having mean zero, variance σ_0^2, and $E|v_{it}|^8 < \infty$.*

Assumption B:(i) *The regressors and the threshold variable are exogenous with elements (x_{it}, q_{it}) being iid across i and t, (ii) For all $\gamma \in \Gamma$, $\mathrm{E}(\|h_{it}\|^4 | q_{it} = \gamma) < \infty$ and $f(\gamma) \leqslant c$ for some $c < \infty$, (iii) For all $\gamma_1, \gamma_2 \in \Gamma$, $\mathrm{E}(\|h_{it}\|^4 | q_{it} = \gamma_1, q_{jt} = \gamma_2) < \infty$ and $f_2(\gamma_1, \gamma_2) < \infty$, (iv) $M(\gamma)f(\gamma)$ is continuous at $\gamma = \gamma_0$, (v) $0 < Mf(\gamma_0) < \infty$, and (vi) the limit of $\frac{1}{nT}\mathrm{E}[\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]$ exists and is nonsingular.*

Assumption C: *The spatial weight matrices $\{W_t\}$ are exogenous with zero diagonals, allowed to vary in time. Both $\|\mathbf{W}\|_1$ and $\|\mathbf{W}\|_\infty$ are bounded.*

Assumption D: *The true λ_0 lies in the interior of a compact space Λ. Conditional on the threshold variable $\{q_{it}\}$ and for $(\gamma, \lambda) \in \Gamma \times \Lambda$, (i) $\mathbf{A}(\lambda, \gamma)$ is invertible; (ii) both $\|\mathbf{A}^{-1}(\lambda, \gamma)\|_1$ and $\|\mathbf{A}^{-1}(\lambda, \gamma)\|_\infty$ are bounded.*

Assumption E: *n is large, and T can be finite or large but cannot grow faster than n, i.e., $\frac{T}{n} \to c$, where $0 \leqslant c < \infty$.*

Assumption F: *Threshold effects λ_{20} and β_{20} satisfy $\lambda_{20} = (nT)^{-\tau} l_0$ and $\beta_{20} = (nT)^{-\tau} b_0$ for some $\tau \in (0, 1/2)$ with $l_0 \in \mathbf{R}$, $l_0 \neq 0$ and $b_0 \in \mathbf{R}^k$, $b_0 \neq 0$.*

The iid assumption in A is standard in the spatial econometrics literature (see, e.g., Lee and Yu, 2010a; Li, 2018), but the finite eighth moment condition on errors is more stringent than in the standard SPR models, where only a finite fourth moment condition is required. This stronger assumption is needed to establish the weak convergence in functional space of some linear-quadratic (LQ) forms that depend on γ through the indication function $\mathbb{1}\{q_{it} \leqslant \gamma\}$ (see Lemma B.1, Appendix B), which is crucial for asymptotic studies on our estimators. Assumption B(i) assumes that regressors and threshold variable are both exogenous as in Hansen (1999, 2000). As $\{h_{it}\}$ can be treated as the model regressors in a reduced form of (2.1) (see (2.15)), Assumptions B(ii)–B(v) are also common in the threshold literature (e.g., Hansen (1999); Li and Lin (2024)). Assumption B(vi) is the identification condition for β. Assumptions C and D are standard in the spatial econometrics literature. Assumption E allows T/n to be $O(1)$ or $o(1)$ (including the case of fixed T). All these scenarios encounter the incidental parameters problem of Neyman and Scott (1948) due to the estimation of the individual and time fixed effects. The assumption that T cannot grow faster than n is used to establish the consistency of $\hat{\gamma}_{nT}$. Assumption F is in the spirit of Hansen (2000) so that the asymptotic distribution of the threshold estimator is free of nuisance parameters, and thus statistical inference on γ is possible. In contrast, if threshold effects are fixed (that is, $\tau = 0$), according to Chan (1993), we can expect that the asymptotic distribution of $\hat{\gamma}_{nT}$ will involve nuisance parameters such as the marginal distribution of x_{it}.

2.3.1. Consistency and convergence rate

To study the consistency of $(\hat{\theta}_{nT}, \hat{\gamma}_{nT})$, it is crucial to first establish the consistency of $(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})$. Given λ and γ, the $\ell^*_{nT}(\theta, \gamma)$ given in (2.6) is partially maximized at

$$\hat{\beta}_{nT}(\lambda, \gamma) = [\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]^{-1}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbf{A}(\lambda, \gamma)\mathbf{Y}, \text{ and} \tag{2.9}$$

$$\hat{\sigma}^2_{nT}(\lambda, \gamma) = \tfrac{1}{N}\tilde{\mathbf{V}}'(\hat{\beta}_{nT}(\lambda, \gamma), \lambda, \gamma)\tilde{\mathbf{V}}(\hat{\beta}_{nT}(\lambda, \gamma), \lambda, \gamma). \tag{2.10}$$

Hence, the concentrated AQL (adjusted quasi loglikelihood) function of λ and γ is

$$\ell^{*c}_{nT}(\lambda, \gamma) = -\tfrac{nT}{2}(\ln 2\pi + 1) - \tfrac{nT}{2}\ln \hat{\sigma}^2_{nT}(\lambda, \gamma) + \ln|\mathbf{A}(\lambda, \gamma)|. \tag{2.11}$$

Maximizing $\ell^{*c}_{nT}(\lambda, \gamma)$ gives the AQMLEs $\hat{\lambda}_{nT}$ and $\hat{\gamma}_{nT}$ of λ and γ.[3]

The population counterpart of $\ell^{*c}_{nT}(\lambda, \gamma)$ is $\bar{\ell}^{*c}_{nT}(\lambda, \gamma) = \max_{\beta,\sigma^2} \mathrm{E}[\ell^*_{nT}(\theta, \gamma)]$. Given λ and γ, $\mathrm{E}[\ell^*_{nT}(\theta, \gamma)]$ is partially maximized at

$$\bar{\beta}_{nT}(\lambda, \gamma) = [\mathrm{E}(\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma))]^{-1}\mathrm{E}[\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbf{A}(\lambda, \gamma)\mathbf{Y}], \text{ and} \tag{2.12}$$

$$\bar{\sigma}^2_{nT}(\lambda, \gamma) = \tfrac{1}{N}\mathrm{E}[\tilde{\mathbf{V}}'(\bar{\beta}_{nT}(\lambda, \gamma), \lambda, \gamma)\tilde{\mathbf{V}}(\bar{\beta}_{nT}(\lambda, \gamma), \lambda, \gamma)]. \tag{2.13}$$

Thus, upon substituting $\bar{\beta}_{nT}(\lambda, \gamma)$ and $\bar{\sigma}^2_{nT}(\lambda, \gamma)$ back into $\mathrm{E}[\ell^*_{nT}(\theta, \gamma)]$, we obtain,

$$\bar{\ell}^{*c}_{nT}(\lambda, \gamma) = \tfrac{nT}{2}(\ln 2\pi + 1) - \tfrac{nT}{2}\ln \bar{\sigma}^2_{nT}(\lambda, \gamma) + \mathrm{E}(\ln|\mathbf{A}(\lambda, \gamma)|). \tag{2.14}$$

For standard extremum-type estimation problems, the consistency of $\hat{\lambda}_{nT}$ and $\hat{\gamma}_{nT}$ can be established if (λ_0, γ_0) uniquely maximizes $\frac{1}{N}\bar{\ell}^{*c}_{nT}(\lambda, \gamma)$ and $\frac{1}{N}[\ell^{*c}_{nT}(\lambda, \gamma) - \bar{\ell}^{*c}_{nT}(\lambda, \gamma)]$ converges to 0 uniformly in $(\lambda, \gamma) \in \Lambda \times \Gamma$. However, under the diminishing threshold assumption, the limit of either $\frac{1}{N}\ell^{*c}_{nT}(\lambda_0, \gamma)$ or $\frac{1}{N}\bar{\ell}^{*c}_{nT}(\lambda_0, \gamma)$ does not depend on the threshold parameter γ, implying that the consistency of $\hat{\lambda}_{nT}$ and $\hat{\gamma}_{nT}$ cannot be established simultaneously.

If γ_0 were known, standard asymptotic arguments (e.g., Theorem 5.9 of Van der Vaart, 1998), the uniform convergence of $\frac{1}{N}[\ell^{*c}_{nT}(\lambda, \gamma_0) - \bar{\ell}^{*c}_{nT}(\lambda, \gamma_0)]$ to 0 in $\lambda \in \Lambda$ and the global identification of λ_0, i.e., λ_0 uniquely maximizes the limit of $\frac{1}{N}\bar{\ell}^{*c}_{nT}(\lambda, \gamma_0)$ on Λ, lead to consistency of $\hat{\lambda}_{nT}(\gamma_0)$. When γ_0 is unknown, the convergence of $\sup_{\lambda\in\Lambda} \frac{1}{N}[\ell^{*c}_{nT}(\lambda, \gamma) - \bar{\ell}^{*c}_{nT}(\lambda, \gamma)]$ is still useful to establish the consistency of $\hat{\lambda}_{nT}$, uniformly in $\gamma \in \Gamma$, although it cannot provide useful

[3] It is worth noting that $\ell^{cc}_{nT}(\lambda, \gamma) = \max_{\beta,\sigma^2} \ell^{c}_{nT}(\theta, \gamma) = \ell^{*c}_{nT}(\lambda, \gamma) + \frac{nT}{2}\ln c_{nT}$. Thus, $\ell^{*c}_{nT}(\lambda, \gamma)$ and $\ell^{cc}_{nT}(\lambda, \gamma)$ yield the same maximizer, leading to identical estimates for λ, γ and β. However, as discussed above, $\ell^*_{nT}(\theta, \gamma)$ is a valid joint objective function that allows us to provide joint inference methods for all common parameters.

information to study the asymptotic behavior of $\hat{\gamma}_{nT}$ as threshold effects become zero at the limit. Therefore, we first show $\hat{\lambda}_{nT}(\gamma)$ is consistent, uniformly in $\gamma \in \Gamma$.[4] Then, based on this result, we establish the consistency of $\hat{\gamma}_{nT}$. Let $\sigma^2_{nT}(\lambda, \gamma) = \frac{\sigma_0^2}{N}\mathtt{tr}[\mathbf{A}'^{-1}\mathbf{A}'(\lambda,\gamma)\mathbf{Q}_{nT}\mathbf{A}(\lambda,\gamma)\mathbf{A}^{-1}]$ and $\mathbb{H}(\gamma) = [\mathbf{H},\ \mathbf{D}(\gamma)\mathbf{H}]$. We provide the identification conditions for λ as follows.

Assumption G: *Either* (i) *the limit of* $\frac{1}{N}[\mathbb{H}'(\gamma)\mathbf{Q}_{nT}\mathbb{H}(\gamma)]$ *exists and is nonsingular, or* (ii) *the limit of* $\frac{1}{N}(\ln|\sigma^2_{nT}(\lambda_0,\gamma)\mathbf{A}^{-1}(\lambda_0,\gamma)\mathbf{A}'^{-1}(\lambda_0,\gamma)| - \ln|\sigma^2_{nT}(\lambda,\gamma)\mathbf{A}^{-1}(\lambda,\gamma)\mathbf{A}'^{-1}(\lambda,\gamma)|)$ *is none zero for* $\lambda \neq \lambda_0$*, uniformly in* $\gamma \in \Gamma$.

Assumption G generalizes the global identification conditions for the SPR model of Lee and Yu (2010a) to our TSPR model. To gain more intuition on this assumption, noting that $\mathbf{WY} = \mathbf{WA}^{-1}(\mathbf{X}\beta_{10} + \mathbf{D}(\gamma_0)\mathbf{X}\beta_{20} + \mathbf{C}\psi_0 + \mathbf{V})$, we have from Model (2.2),

$$\mathbf{Y} = \mathbf{X}\beta_{10} + \mathbf{Z}\lambda_{10} + \mathbf{D}(\gamma_0)\mathbf{X}\beta_{20} + \mathbf{D}(\gamma_0)\mathbf{Z}\lambda_{20} + \mathbf{C}\psi_0 + \mathbf{A}^{-1}\mathbf{V}, \tag{2.15}$$

because $\mathbf{A}^{-1} = I_{nT} + \lambda_{10}\mathbf{G} + \lambda_{20}\mathbf{D}(\gamma_0)\mathbf{G}$, which comes from $I_{nT} = \mathbf{A} + \lambda_{10}\mathbf{W} + \lambda_{20}\mathbf{D}(\gamma_0)\mathbf{W}$ by right-multiplying $\mathbf{A}^{-1}$ on both sides of the equation. Clearly, the above equation can be treated as a standard panel data model with regressor matrix $\mathbb{H}(\gamma_0)$. Thus, it is standard to impose the non-singularity or full-rank condition on the limit of $\frac{1}{N}\mathrm{E}[\mathbb{H}'(\gamma_0)\mathbf{Q}_{nT}\mathbb{H}(\gamma_0)]$ to identify β and λ. As the consistency of $\hat{\gamma}_{nT}$ has not been established, a uniform (in γ) existence condition, Assumption G(i), must be imposed. In addition, λ can also be identified by the uniqueness of the conditional variance of $\mathbf{Y}$, $\sigma_0^2\mathbf{A}^{-1}\mathbf{A}'^{-1}$, given $\mathbf{X}$ and the threshold variable q_{it}. Again, without knowing the consistency of $\hat{\gamma}_{nT}$, a uniform (in γ) Assumption G(ii) must be imposed.

The AQMLEs of β and σ^2 are $\hat{\beta}_{nT} \equiv \hat{\beta}_{nT}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})$ and $\hat{\sigma}^2_{nT} \equiv \hat{\sigma}^2_{nT}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})$. It is interesting to note from (B.4) and (B.5) in Appendix B that, under Assumptions B(vi) and F, the consistencies of $\hat{\beta}_{nT}$ and $\hat{\sigma}^2_{nT}$ follow that of $\hat{\lambda}_{nT}$, whether $\hat{\gamma}_{nT}$ is consistent or not. These suggest that the consistency of $\hat{\theta}_{nT}$ does not rely on that of $\hat{\gamma}_{nT}$ under the assumption of diminishing threshold effect, and thus it can be established separately. With the above identification conditions and the convergence of $\sup_{\lambda\in\Lambda}\frac{1}{N}[\ell^{*c}_{nT}(\lambda,\gamma) - \bar{\ell}^{*c}_{nT}(\lambda,\gamma)]$ to 0 uniformly in $\gamma \in \Gamma$, we have the following theorem.

Theorem 2.1. *Suppose Assumptions A-G hold. We have* $\hat{\theta}_{nT} - \theta_0 \xrightarrow{p} 0$.

As the adjustment to the concentrated loglikelihood function in (2.6) can help to "recover" the degrees of freedom loss caused by the estimation of the incidental parameters, we see that all common estimators are consistent even when T is fixed. As discussed above, although $\hat{\lambda}_{nT}$ is shown to be consistent, the convergence of the original objective function $\frac{1}{N}\ell^{*c}_{nT}(\lambda,\gamma)$ is still

[4] One can also see from (2.17) that $\lim_{N\to\infty}\frac{1}{N}\mathrm{E}S^*_{\theta,nT}(\theta_0,\gamma) = 0$, uniformly in $\gamma \in \Gamma$, a necessary condition to have consistency of $\hat{\theta}_{nT}(\gamma)$.

too fast to be useful for studying the limiting behavior of $\hat{\gamma}_{nT}$, when the threshold effects shrink to zero at the rate $(nT)^{-\tau}$. However, we find the re-scaled objective function:

$$\ell_{nT}^{**}(\gamma) = \tfrac{(nT)^{2\tau}}{nT}[\ell_{nT}^{*c}(\hat{\lambda}_{nT}(\gamma), \gamma) - \ell_{nT}^{*c}(\lambda_0, \gamma_0)] \tag{2.16}$$

can be very useful. Specifically, multiplying $(nT)^{2\tau}$ gives us the non-diminishing threshold effects, while taking the differences removes the terms that are not asymptotically negligible, i.e., those with an order of magnitude bigger than $O_p((nT)^{1-2\tau})$.

The consistency of $\hat{\gamma}_{nT}$ follows if the maximizer of $\ell_{nT}^{**}(\gamma)$ has an asymptotically negligible distance from γ_0, i.e., the identification condition for γ. Let $\mathbf{D}(\gamma_1, \gamma_2) = \mathbf{D}(\gamma_1) - \mathbf{D}(\gamma_2)$ and $\mathcal{H}(\gamma) = [\mathbb{H}(\gamma),\ \mathbf{D}(\gamma_0, \gamma)\mathbf{H}]$. Let $\mathbb{C}(\gamma)$ be a 3×3 matrix of elements $\frac{1}{nT}\mathtt{tr}[\mathbb{C}_a^s(\gamma)\mathbb{C}_b^s(\gamma)]$, $a, b = 1, 2, 3$, where $\mathbb{C}_1 = \mathbf{G} - \frac{1}{nT}\mathtt{tr}(\mathbf{G})I_{nT}$, $\mathbb{C}_2(\gamma) = \mathbf{D}(\gamma)\mathbf{G} - \frac{1}{nT}\mathtt{tr}(\mathbf{D}(\gamma)\mathbf{G})I_{nT}$, and $\mathbb{C}_3(\gamma) = \mathbf{D}(\gamma_0, \gamma)\mathbf{G} - \frac{1}{nT}\mathtt{tr}(\mathbf{D}(\gamma_0, \gamma)\mathbf{G})I_{nT}$. We introduce the identification condition for γ.

Assumption H: $\exists\, c > 0$ *s.t.* $\rho_{\min}\big(\frac{1}{N}\mathcal{H}'(\gamma)\mathbf{Q}_{nT}\mathcal{H}(\gamma)\big) \geq c|\gamma - \gamma_0|$ **or** $\rho_{\min}\big(\mathbb{C}(\gamma)\big) \geq c|\gamma - \gamma_0|$.

As shown in Appendix B, to study the asymptotic properties of $\ell_{nT}^{**}(\hat{\gamma}_{nT})$, one has to establish a rough convergence rate for $\hat{\lambda}_{nT} - \lambda_0$. As the objective function is highly nonlinear in λ and there is no closed-form solution for its AQMLE, we have to rely on the study of the θ-component of the concentrated quasi score (CQS) function given in (2.7). We start with a Taylor expansion of $S_{\theta,nT}^*(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) = 0$ at θ_0, then justify the non-singularity of the probability limit of the negative Hessian matrix under Assumption H, and finally study the order of the component CQS function $S_{\theta,nT}^*(\theta_0, \hat{\gamma}_{nT})$. By Theorem 2.1, we show that $(nT)^\tau(\hat{\theta}_{nT} - \theta_0)$ is at most $O_p(1)$ under Assumption E, regardless of the consistency of $\hat{\gamma}_{nT}$. With this preliminary convergence rate, we show $\ell_{nT}^{**}(\hat{\gamma}_{nT}) \leq -\bar{c}|\hat{\gamma}_{nT} - \gamma_0| + o_p(1)$ for some positive constant $\bar{c} < \infty$ under Assumption H. Meanwhile, the definition of $\hat{\gamma}_{nT}$ ensures that $\ell_{nT}^{**}(\hat{\gamma}_{nT}) \geq 0$. Therefore, we conclude that $\hat{\gamma}_{nT}$ must be sufficiently close to γ_0 in probability, which establishes the consistency of $\hat{\gamma}_{nT}$.

Theorem 2.2. *Suppose Assumptions A-H hold. We have* $\hat{\gamma}_{nT} - \gamma_0 \xrightarrow{p} 0$.

To establish the convergence rate for $\hat{\gamma}_{nT}$, one needs a more precise knowledge of the convergence rate of $\hat{\theta}_{nT}$. With the consistency of $\hat{\gamma}_{nT}$ and the Taylor expansion mentioned above, we further show that $(nT)^\tau(\hat{\theta}_{nT} - \theta_0) = o_p(1)$. Thus, we have the following theorem.

Theorem 2.3. *Under Assumptions A-H,* $a_{nT}(\hat{\gamma}_{nT} - \gamma_0) = O_p(1)$, *where* $a_{nT} = (nT)^{1-2\tau}$.

Theorem 2.3 shows that the convergence rate of $\hat{\gamma}_{nT}$ is a_{nT}, in line with Hansen (1999). Intuitively, $\hat{\gamma}_{nT}$ converges to γ_0 faster when the threshold effects (λ_{20} and β_{20}) are greater (i.e., τ is smaller or the threshold diminishing rate $(nT)^{-\tau}$ is slower), as in this case more sample information is obtained about γ and hence a more precise estimate can be made.

2.3.2. Asymptotic distribution of $\hat{\theta}_{nT}$

Theorem 2.3 is crucial for establishing the asymptotic distribution of the AQMLE $\hat{\theta}_{nT}$. The generic Theorem 2.4 reveals that (i) when $T/n = o(1)$, $\sqrt{N}(\hat{\theta}_{nT} - \theta_0)$ has a centered limiting normal distribution, and inference about θ_0 can be made simply based on the AQMLE $\hat{\theta}_{nT}$; and (ii) when $T/n = O(1)$, $\sqrt{N}(\hat{\theta}_{nT} - \theta_0)$ has a non-centered limiting normal distribution, and a bias correction must be made on $\hat{\theta}_{nT}$ for valid inference about θ_0.

To derive the result, a Taylor series expansion of $S^*_{\theta,nT}(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) = 0$ at θ_0 gives

$$\sqrt{N}(\hat{\theta}_{nT} - \theta_0) = \big[\tfrac{1}{nT} H^*_{nT}(\bar{\theta}, \hat{\gamma}_{nT})\big]^{-1} \tfrac{\sqrt{N}}{nT} S^*_{\theta,nT}(\theta_0, \hat{\gamma}_{nT}),$$

where we use $H^*_{nT}(\bar{\theta}, \gamma)$ to denote $-\frac{\partial}{\partial\theta'} S^*_{\theta,nT}(\theta, \gamma)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}$ for simplicity and $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}_{nT}$ and θ_0. Thus, the asymptotic property of $\sqrt{N}(\hat{\theta}_{nT} - \theta_0)$ depends on that of $\frac{\sqrt{N}}{nT} S^*_{\theta,nT}(\theta_0, \hat{\gamma}_{nT})$ provided that $\text{plim} \frac{1}{nT} H^*_{nT}(\bar{\theta}, \hat{\gamma}_{nT})$ is nonsingular. We have

$$\tfrac{\sqrt{N}}{nT} S^*_{\theta,nT}(\theta_0, \hat{\gamma}_{nT}) = \tfrac{1}{\sqrt{N}} S^{*u}_{\theta,nT}(\theta_0, \hat{\gamma}_{nT}) + \sqrt{\tfrac{T}{n}}\, b_{\theta,nT}(\lambda_0, \hat{\gamma}_{nT}) + o_p(1), \tag{2.17}$$

where $b_{\theta,nT}(\lambda_0, \gamma) = \big\{0_{1\times 2k},\ -\frac{1}{nT}\texttt{tr}(\bar{\mathbf{G}}\mathbf{J}),\ -\frac{1}{nT}\texttt{tr}[\mathbf{D}(\gamma)\bar{\mathbf{G}}\mathbf{J}],\ 0\big\}'$, which is (2.8) rescaled, and

$$S^{*u}_{\theta,nT}(\theta_0, \gamma) = \begin{cases} \frac{1}{\sigma_0^2}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbf{V}, \\ \frac{1}{\sigma_0^2}\mathbf{Z}'\mathbf{Q}_{nT}\mathbf{V} + \frac{1}{\sigma_0^2}[\mathbf{V}'\mathbf{G}'\mathbf{Q}_{nT}\mathbf{V} - \sigma_0^2\texttt{tr}(\mathbf{Q}_{nT}\mathbf{G})], \\ \frac{1}{\sigma_0^2}\mathbf{Z}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{V} + \frac{1}{\sigma_0^2}[\mathbf{V}'\mathbf{G}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{V} - \sigma_0^2\texttt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G})], \\ \frac{1}{2\sigma_0^4}(\mathbf{V}'\mathbf{Q}_{nT}\mathbf{V} - N\sigma_0^2). \end{cases} \tag{2.18}$$

It is easy to verify that $\mathrm{E}[S^{*u}_{\theta,nT}(\theta_0, \gamma_0)] = 0$. Furthermore, with the convergence rate of $\hat{\gamma}$ proved in Theorem 2.3, we show that $\frac{1}{\sqrt{N}} S^{*u}_{\theta,nT}(\theta_0, \hat{\gamma}_{nT})$ converges to a Gaussian distribution with mean zero and variance $\frac{1}{N}\mathrm{Var}[S^{*u}_{\theta,nT}(\theta_0, \gamma_0)]$ and that $b_{\theta,nT}(\lambda_0, \hat{\gamma}_{nT}) = O_p(1)$. Then, by showing $\frac{1}{\sqrt{N}}\|S^{*u}_{\theta,nT}(\theta_0, \hat{\gamma}_{nT}) - S^{*u}_{\theta,nT}(\lambda_0, \gamma_0)\| \xrightarrow{p} 0$ and $\|b_{\theta,nT}(\lambda_0, \hat{\gamma}_{nT}) - b_{\theta,nT}(\lambda_0, \gamma_0)\| \xrightarrow{p} 0$, we show that $\hat{\theta}_{nT} - \hat{\theta}_{nT}(\gamma_0) = o_p(1)$. We have the following theorem.

Theorem 2.4. *Under Assumptions A-H, we have*

(i) $\sqrt{N}(\hat{\theta}_{nT} - \theta_0) - \sqrt{\frac{T}{n}}\Sigma_{nT}^{-1}\, b_{\theta,nT} \xrightarrow{D} N\big(0, \lim_{nT\to\infty} \Sigma_{nT}^{-1}\Omega_{nT}\Sigma_{nT}^{-1}\big)$,

(ii) $\sqrt{N}(\hat{\theta}_{nT} - \hat{\theta}_{nT}(\gamma_0)) \xrightarrow{p} 0$,

where $b_{\theta,nT} \equiv b_{\theta,nT}(\lambda_0, \gamma_0)$, $\Sigma_{nT} = -\frac{1}{nT}\mathrm{E}[\frac{\partial}{\partial\theta'} S^*_{\theta,nT}(\theta_0, \gamma_0)]$, *and* $\Omega_{nT} = \frac{1}{N}\mathrm{Var}[S^{*u}_{\theta,nT}(\theta_0, \gamma_0)]$.

Several important conclusions are drawn from Theorem 2.4. First, Theorem 2.4(i) shows that the convergence rate of $\hat{\theta}_{nT}$ is $\sqrt{N}$ in general. Second, if T is fixed or small relative to n, then the bias term $\sqrt{\frac{T}{n}}\Sigma_{nT}^{-1}\, b_{\theta,nT}$ disappears as $N \to \infty$ and hence the estimators of all common parameters are asymptotically unbiased and inference about θ_0 proceeds as normal.[5] Third, when $\frac{T}{n} \to c \neq 0$, $\sqrt{\frac{T}{n}}\Sigma_{nT}^{-1} b_{\theta,nT} = O(1)$ and thus a bias correction on $\hat{\theta}_{nT}$ must be made for valid inference. The *bias-corrected* AQMLE is:

$$\hat{\theta}_{nT}^{\rm bc} = \hat{\theta}_{nT} - \sqrt{\tfrac{T}{nN}}\widehat{\Sigma}_{nT}^{-1}\, \hat{b}_{\theta,nT}, \tag{2.19}$$

where $\hat{b}_{\theta,nT}$ and $\widehat{\Sigma}_{nT}$ are the consistent estimators of $b_{\theta,nT}$ and Σ_{nT} given below. It is easy to see that $\hat{\theta}_{nT}^{\rm bc}$ has the same asymptotic variance as $\hat{\theta}_{nT}$, that is, $\frac{1}{N}\Sigma_{nT}^{-1}\Omega_{nT}\Sigma_{nT}^{-1}$. Fourth, from Theorem 2.4(ii), we also see that given the convergence rate of $\hat{\gamma}_{nT}$ in Theorem 2.3, $\hat{\gamma}_{nT}$ can be treated as if it is the true value of γ. In other words, the estimation error associated with $\hat{\gamma}_{nT}$ has asymptotically negligible effects on the asymptotic property of the AQMLEs of the common θ. In summary, Theorem 2.1 shows that $\hat{\theta}_{nT}(\gamma)$ is consistent for θ_0 for any $\gamma \in \Gamma$, while Theorem 2.4(ii) says that $\hat{\theta}_{nT} \equiv \hat{\theta}_{nT}(\hat{\gamma}_{nT})$ and $\hat{\theta}_{nT}(\gamma_0)$ are asymptotically equivalent.

Inference for θ. In practice, to conduct statistical inference on θ, one relies on consistent estimations of Σ_{nT} and Ω_{nT}. For the former, its analytical expression is

$$\Sigma_{nT} = \begin{bmatrix} \frac{1}{N\sigma_0^2}\mathrm{E}[\mathbb{X}'(\gamma_0)\mathbf{Q}_{nT}\mathbb{X}(\gamma_0)], & \frac{1}{N\sigma_0^2}\mathrm{E}[\mathbb{X}'(\gamma_0)\mathbf{Q}_{nT}\mathbb{Z}(\gamma_0)], & 0_{2k\times 1} \\ \sim, & \Sigma_{22,nT}(\gamma_0), & \frac{1}{\sigma_0^2}\mathrm{E}[\mathcal{S}_{nT}(\gamma_0)] \\ \sim, & \sim, & \frac{1}{2\sigma_0^4} \end{bmatrix}, \tag{2.20}$$

where $\Sigma_{22,nT}(\gamma) = \frac{1}{N\sigma_0^2}\mathrm{E}[\mathbb{Z}'(\gamma)\mathbf{Q}_{nT}\mathbb{Z}(\gamma)] + \mathrm{E}[\mathbb{S}_{nT}(\gamma)]$, $\mathcal{S}_{nT}(\gamma) = \frac{1}{nT}\{\mathtt{tr}(\mathbf{G}),\ \mathtt{tr}[\mathbf{G}^{\circ}(\gamma)]\}'$, $\mathbb{S}_{nT}(\gamma) = \frac{1}{nT}\{\mathtt{tr}(\mathbf{G}\mathbf{G}^{s}),\ \mathtt{tr}[\mathbf{G}^{\circ}(\gamma)\mathbf{G}^{s}]; \mathtt{tr}[\mathbf{G}^{\circ s}(\gamma)\mathbf{G}],\ \mathtt{tr}[\mathbf{G}^{\circ}(\gamma)\mathbf{G}^{\circ s}(\gamma)]\}$, $\mathbb{Z}(\gamma) = [\mathbf{Z},\ \mathbf{D}(\gamma)\mathbf{Z}]$, and $\mathbf{G}^{\circ}(\gamma) = \mathbf{D}(\gamma)\mathbf{G}$. Thus, the sample analogue, $\widehat{\Sigma}_N^{*} = -\frac{1}{nT}\frac{\partial}{\partial\theta'}S_{\theta,nT}^{*}(\theta,\gamma)|_{\theta=\hat{\theta}_{nT},\gamma=\hat{\gamma}_{nT}}$, consistently estimates Σ_{nT}; see the proof of Theorem 2.2 given in Appendix B.

For the latter, we have $\Omega_{nT} = \Sigma_{nT} + \Gamma_{nT}$, where

$$\Gamma_{nT} = \begin{bmatrix} 0_{2k\times 2k}, & \frac{\bar{T}\kappa_3}{N\sigma_0}\mathrm{E}[\mathbb{X}'(\gamma_0)\mathbf{Q}_{nT}\mathbb{R}(\gamma_0)], & 0_{2k\times 1} \\ \sim, & \Gamma_{22,nT}(\gamma_0), & \frac{\kappa_4\bar{T}^2}{2N\sigma_0^2}\mathrm{E}[\mathbb{R}'(\gamma_0)l_{nT}] \\ \sim, & \sim, & \frac{\kappa_4\bar{T}}{4\sigma_0^4} \end{bmatrix}, \tag{2.21}$$

$\mathbb{R}(\gamma) = [\mathtt{diagv}(\mathbf{G}),\ \ \mathbf{D}(\gamma)\mathtt{diagv}(\mathbf{G})]$, $\Gamma_{22,nT}(\gamma) = \frac{2\kappa_3\bar{T}}{N\sigma_0}\mathrm{E}[\mathbb{Z}'(\gamma)\mathbf{Q}_{nT}\mathbb{R}(\gamma)] + \frac{\kappa_4\bar{T}^2}{N}\mathrm{E}[\mathbb{R}'(\gamma)\mathbb{R}(\gamma)] + \mathrm{E}[\mathbb{B}_{nT}(\gamma)]$, κ_3 and κ_4 are the skewness and excess kurtosis of the errors, and $\mathbb{B}_{nT}(\gamma)$ is 2×2 with elements: $\mathbb{B}_{11,nT} = \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[(G_t - G_k)G_t]$, $\mathbb{B}_{12,nT}(\gamma) = \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[(d_t(\gamma)G_t - d_k(\gamma)G_k)G_t] = \mathbb{B}_{21,nT}'(\gamma)$, and $\mathbb{B}_{22,nT}(\gamma) = \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[(d_t(\gamma)G_t - d_k(\gamma)G_k)d_t(\gamma)G_t]$.[6] A consistent estimator of $\Gamma_{nT}(\theta_0,\gamma_0)$ is obtained by plugging in $\hat{\theta}_{nT}$ or $\hat{\theta}_{nT}^{\rm bc}$ for θ_0, $\hat{\gamma}_{nT}$ for γ_0, and

[5] This is in contrast to the direct approach of Li (2018) where the estimator of error variance is inconsistent.

[6] The result for the standard SPR model Lee and Yu (2010a) does not involve κ_3. This is due to the time-varying feature of our model, reflected by the terms $\{G_t\}$ and $\{d_t(\gamma_0)G_t\}$ in $\mathbb{B}_{nT}(\gamma_0)$.

the consistent estimators of κ_3 and κ_4 derived based on the ideas in Meng and Yang (2024):

$$\hat{\kappa}_{3,nT} = \frac{\sum_{j=1}^{nT} \hat{v}_{j,nT}^3}{\hat{\sigma}_{nT}^3 \sum_{j=1}^{nT} \sum_{k=1}^{nT} q_{jk}^3} \text{ and } \hat{\kappa}_{4,nT} = \frac{\sum_{j=1}^{nT} \hat{v}_{j,nT}^4 - 3\hat{\sigma}_{nT}^4 \sum_{j=1}^{nT} \sum_{k=1}^{nT} \sum_{l=1}^{nT} q_{jk}^2 q_{jl}^2}{\hat{\sigma}_{nT}^4 \sum_{j=1}^{nT} \sum_{k=1}^{nT} q_{jk}^4},$$

where $\hat{v}_{j,nT}$ is the jth element of $\hat{\mathbf{V}} \equiv \tilde{\mathbf{V}}(\hat{\beta}_{nT}, \hat{\lambda}_{nT}, \hat{\gamma}_{nT}) \equiv \mathbf{Q}_{nT}[\mathbf{A}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})\mathbf{Y} - \mathbb{X}(\hat{\gamma}_{nT})\hat{\beta}_{nT}]$, and q_{jk} is the (j,k)th element of $\mathbf{Q}_{nT}$. The $\hat{\theta}_{nT}$ are to be replaced by $\hat{\theta}_{nT}^{\text{bc}}$ if $T/n = O(1)$.

2.3.3. Asymptotic distribution of $\hat{\gamma}_{nT}$

Next, we establish the asymptotic distribution of $\hat{\gamma}_{nT}$. Its convergence rate, which is a_{nT}, has already been established in Theorem 2.3. To use these results in Theorem 2.4, we define $\ell_{nT}^{\ddagger}(\gamma) = \ell_{nT}^{*}(\hat{\theta}_{nT}, \gamma)$, which must be uniquely maximized at $\gamma = \hat{\gamma}_{nT}$ because otherwise it contradicts that $(\hat{\theta}_{nT}, \hat{\gamma}_{nT})$ maximizes $\ell_{nT}^{*}(\theta, \gamma)$. With a constant term $\ell_{nT}^{\ddagger}(\gamma_0)$ subtracted, we still have $\hat{\gamma}_{nT} = \underset{\gamma\in\Gamma}{\operatorname{argmax}}[\ell_{nT}^{\ddagger}(\gamma) - \ell_{nT}^{\ddagger}(\gamma_0)]$. We then make the change-of-variable $\gamma = \gamma_0 + v/a_{nT}$ and therefore $a_{nT}(\hat{\gamma}_{nT} - \gamma_0) = \underset{v}{\operatorname{argmax}}[\ell_{nT}^{\ddagger}(\gamma_0 + v/a_{nT}) - \ell_{nT}^{\ddagger}(\gamma_0)]$. Thus, the asymptotic analysis of $a_{nT}(\hat{\gamma}_{nT} - \gamma_0)$ relies on that of the objective function $\ell_{nT}^{\ddagger}(\gamma_0 + v/a_{nT}) - \ell_{nT}^{\ddagger}(\gamma_0)$, whose limiting distribution involves following quantities:

$$\Xi_1 = \lim_{nT\to\infty} \bar{T}[\delta_0' M \delta_0 + l_0^2 \sigma_0^2 (\pi_1 + \pi_2)] \text{ and } \Xi_2 = \lim_{nT\to\infty} \bar{T}^2 (2 l_0 \sigma_0 \kappa_3 \delta_0' \pi_3 + l_0^2 \sigma_0^2 \kappa_4 \pi_2),$$

where $\pi_1(\gamma) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \mathrm{E}(\sum_{j=1}^{n} g_{ij,t}^2 | q_{it} = \gamma)$, $\pi_2(\gamma) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \mathrm{E}(g_{ii,t}^2 | q_{it} = \gamma)$, $\pi_3(\gamma) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \mathrm{E}(g_{ii,t} h_{it} | q_{it} = \gamma)$ with $g_{ij,t}$ being the (i,j)th entry of $\mathbf{G}$'s tth block G_t, and $\delta_0 = (b_0', l_0)'$. Under Assumption B(v), it is easy to see that Ξ_1 must be strictly positive. Thus, the following theorem provides the asymptotic distribution of $\hat{\gamma}_{nT}$.

Theorem 2.5. *Under Assumptions A-H, we have*

$$a_{nT}(\hat{\gamma}_{nT} - \gamma_0) \xrightarrow{D} \frac{\sigma_0^2}{f} \frac{\Xi}{\Xi_1^2} \underset{-\infty<r<\infty}{\operatorname{argmax}} \Big[-\frac{|r|}{2} + W(r) \Big],$$

where $\Xi = \Xi_1 + \Xi_2$ and $W(r)$ as a two-sided standard Brownian motion on the real line, i.e., $W(r) = W_a(-r)\mathbb{1}\{r \leqslant 0\} + W_b(r)\mathbb{1}\{r > 0\}$, and $W_a(\cdot)$ and $W_b(\cdot)$ are two independent standard Brownian motions on $[0, \infty)$ with $W_a(0) = W_b(0) = 0$.

According to Chan (1993), when the threshold effects are fixed over sample size (i.e., $\tau = 0$), it may be possible to demonstrate that $nT(\hat{\gamma}_{nT} - \gamma_0) = O_p(1)$, but the asymptotic distribution of $nT(\hat{\gamma}_{nT} - \gamma_0)$ might be a functional of a compound Poisson process that depends on the marginal distribution of x_{it}, and hence is not useful for making inference on γ. In contrast, under the shrinking threshold effects assumption, Theorem 2.5 shows that the limiting distribution of $\hat{\gamma}_{nT}$ does not involve this undue component. However, in order to conduct inference on γ directly through the above theorem, one has to find a consistent estimate for the scale component $\frac{\Xi}{\Xi_1^2 f}$. Note that both Ξ_1 and Ξ_2 involve δ_0 or l_0, neither of which can be estimated accurately without prior knowledge of the nuisance parameter τ.

2.4. Inference for $\hat{\gamma}_{nT}$ based on the likelihood ratio test

It is of practical interest to infer more precisely on the true value of γ. While Theorem 2.5 provides a theoretical base, its application is hindered by the difficulty in the estimation of δ_0 or l_0 and thus an alternative method is desired. Inspired by Hansen (1999) on regular threshold panel regression models, we introduce an inference procedure for the threshold parameter γ in the TSPR model by inverting a likelihood ratio test. Recall $\ell_{nT}^{*c}(\gamma) \equiv \ell_{nT}^{*}(\hat{\theta}_{nT}(\gamma), \gamma)$. The likelihood ratio statistic for testing the null hypothesis H_0: $\gamma = \gamma_0$ is:

$$LR_{nT}(\gamma) = \frac{2}{c_{nT}}[\ell_{nT}^{*c}(\hat{\gamma}_{nT}) - \ell_{nT}^{*c}(\gamma)]. \tag{2.22}$$

Theorem 2.6. *Under Assumptions A-H, we have*

$$LR_{nT}(\gamma_0) \xrightarrow{D} \varpi^2\, \mho,$$

where $\varpi^2 = 1+\frac{\Xi_2}{\Xi_1}$ *with* Ξ_1 *and* Ξ_2 *being given in Theorem 2.4, and* $\mho = \max_{-\infty<r<\infty}[-|r|+2W(r)]$, *of which the distribution function is characterized by* $P(\mho \leqslant z) = (1-e^{-z/2})^2$.

Note that $\varpi^2 = \frac{\Xi}{\Xi_1}$, since $\Xi = \Xi_1 + \Xi_2$. Thus, ϖ^2 must be strictly positive, because Ξ_1 is so and $\sigma_0^2\Xi$ is the variance of some LQ form (see Lemma B.3 in Appendix B). In a special case where errors are iid normal, one has $\Xi_2 = 0$ because $\kappa_3 = \kappa_4 = 0$. It follows that $\varpi^2 = 1$ and $LR_{nT}(\gamma_0)$ is an asymptotic pivot. Our result differs from Hansen's (1999, Theorem 1) for standard panels, in that his $\sigma_0^2\Xi$ term does not involve the third and fourth moments, κ_3 and κ_4, of errors as it is just the variance of some linear form.

When errors are not normally distributed, ϖ^2 must be estimated consistently. Let $\theta_{20} = (\beta_{20}', \lambda_{20})'$, the collection of the threshold effects. Then, by Assumption F, we have

$$\frac{\Xi_2}{\Xi_1} = \frac{(nT)^{-2\tau}\Xi_2}{(nT)^{-2\tau}\Xi_1} = \frac{\lim_{nT\to\infty} \bar{T}[2\lambda_{20}\sigma_0\kappa_3\theta_{20}'\pi_3 + \lambda_{20}^2\sigma_0^2\kappa_4\pi_2]}{\lim_{nT\to\infty}[\theta_{20}'M\theta_{20} + \lambda_{20}^2\sigma_0^2(\pi_1+\pi_2)]}. \tag{2.23}$$

Thus, by multiplying $(nT)^{-2\tau}$ on both the numerator and denominator, we replace δ_0 and l_0 with θ_{20} and λ_{20}, respectively. As we already have consistent estimators for θ_{20} and λ_{20}, the need for τ is eliminated. Let $\mathbf{Z} = (Z_1', \ldots, Z_T')'$. Note that $\theta_{20}'M\theta_{20} = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \mathrm{E}[(x_{it}'\beta_{20})^2 + 2\lambda_{20}Z_{it}x_{it}'\beta_{20} + Z_{it}^2\lambda_{20}^2|q_{it} = \gamma_0]$ and $\theta_{20}'\pi_3 = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \mathrm{E}[g_{ii,t}(x_{it}'\beta_{20} + Z_{it}\lambda_{20})|q_{it} = \gamma_0]$, where Z_{it} is the ith element of Z_t. As $Y_t = A_t^{-1}(\mathbb{X}_t\beta_0 + \mu_0 + \alpha_{t0}l_n + V_t)$, we have $\mathcal{Y}_t \equiv W_tY_t = Z_t + G_tV_t$, which implies $\mathrm{E}(Z_{it}x_{it}'|q_{it}) = \mathrm{E}(\mathcal{Y}_{it}x_{it}'|q_{it})$, $\mathrm{E}(g_{ii,t}Z_{it}|q_{it}) = \mathrm{E}(g_{ii,t}\mathcal{Y}_{it}|q_{it})$ and

$\mathrm{E}(Z_{it}^2|q_{it}) = \mathrm{E}(\mathcal{Y}_{it}^2|q_{it}) - \sigma_0^2\mathrm{E}(\sum_{j=1}^n g_{ij,t}^2|q_{it})$, where $\mathcal{Y}_{it}$ is the ith element of $\mathcal{Y}_t$. Thus, we have

$$\frac{\Xi_2}{\Xi_1} = \frac{\lim_{nT\to\infty} \sum_{i=1}^n \sum_{t=1}^T \mathrm{E}(\vartheta_{2,it}|q_{it} = \gamma_0)}{\lim_{nT\to\infty} \sum_{i=1}^n \sum_{t=1}^T \mathrm{E}(\vartheta_{1,it}|q_{it} = \gamma_0)},$$

where $\vartheta_{1,it} = (x_{it}'\beta_{20})^2 + 2\lambda_{20}\mathcal{Y}_{it}x_{it}'\beta_{20} + \mathcal{Y}_{it}^2\lambda_{20}^2 + \lambda_{20}^2\sigma_0^2 g_{ii,t}^2$ and $\vartheta_{2,it} = \bar{T}[2\lambda_{20}\sigma_0\kappa_3 g_{ii,t}(x_{it}'\beta_{20} + \mathcal{Y}_{it}\lambda_{20}) + \lambda_{20}^2\sigma_0^2\kappa_4 g_{ii,t}^2]$. To find their sample counterparts, we let $\hat{\kappa}_3$ and $\hat{\kappa}_4$ be the consistent estimations of κ_3 and κ_4, which are standard to find in the literature (see Li, 2018), $\hat{g}_{ij,t}$ the ith row and jth column of $G_t(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})$. Thus, their sample counterparts are just $\hat{\vartheta}_{1,it} = (x_{it}'\hat{\beta}_{2,nT})^2 + 2\hat{\lambda}_{2,nT}\mathcal{Y}_{it}x_{it}'\hat{\beta}_{2,nT} + \mathcal{Y}_{it}^2\hat{\lambda}_{2,nT}^2 + \hat{\lambda}_{2,nT}^2\hat{\sigma}_{nT}^2\hat{g}_{ii,t}^2$ and $\hat{\vartheta}_{2,it} = \bar{T}[2\hat{\lambda}_{2,nT}\hat{\sigma}_{nT}\hat{\kappa}_3\hat{g}_{ii,t}(x_{it}'\hat{\beta}_{2,nT} + \mathcal{Y}_{it}\hat{\lambda}_{2,nT}) + \hat{\lambda}_{2,nT}^2\hat{\sigma}_{nT}^2\hat{\kappa}_4\hat{g}_{ii,t}^2]$, respectively. Therefore, we finally propose to estimate ϖ^2 by

$$\hat{\varpi}^2 = 1 + \frac{\sum_{i=1}^n \sum_{t=1}^T K_h(q_{it} - \hat{\gamma}_{nT})\hat{\vartheta}_{2,it}}{\sum_{i=1}^n \sum_{t=1}^T K_h(q_{it} - \hat{\gamma}_{nT})\hat{\vartheta}_{1,it}},$$

where $K_h(u) = h^{-1}k(u/h)$ for some bandwidth $h \to 0$ and kernel function $k(\cdot)$. With this, a test of $H_0 : \gamma = \gamma_0$ rejects at the asymptotic level of α if $LR_{nT}(\gamma_0)/\hat{\varpi}^2$ exceeds $\mho_{1-\alpha}$, where $\mho_{1-\alpha} = -2\ln(1 - \sqrt{1-\alpha})$ is the $1 - \alpha$ quantile of $\mho$. From Table I of Hansen (2000), we have $\mho_{1-\alpha} = 5.94$, 7.35 and 10.59 for $\alpha = 0.1$, 0.05 and 0.01, respectively.

3. Testing for the Existence of Threshold Effects

A threshold value γ_0 may always exist but whether the **threshold effects** enter the model depends on whether $\theta_2 = (\beta_2', \lambda_2)'$ is non-zero. Therefore, a test of

$$H_0 : \theta_{20} = 0$$

is of great practical interest. However, at the null, the threshold parameter γ is not identified, which poses a great challenge in constructing such a test. In particular, the asymptotic distributions of the classical tests are nonstandard and it is impossible to tabulate their critical values.[7] To overcome the difficulties facing the classical tests, we propose a sup-Wald test and study its asymptotic property by adopting a local-to-null reparameterization similar in ideas to Hansen (1996): $\theta_{20} = \frac{c}{\sqrt{nT}}$, where $c = (c_b', c_l)'$, c_b is a $k \times 1$ vector and c_l is a scalar. In this sequence of alternatives, we see that the diminishing rate is faster than the one specified in Assumption F, which facilitates our study of the distributional theory of the test statistic. We propose a novel bootstrap procedure to bootstrap the critical values of the test statistic.

[7]This is in fact a classical problem raised by Davies (1977) and investigated by Andrews (1993) and Hansen (1996), under regular panel regression models.

3.1. The sup-Wald test and its asymptotic distribution

The construction of the sup-Wald-type test statistic is straightforward. For each $\gamma \in \Gamma$, we first obtain $\hat{\theta}_{nT}(\gamma)$. Under the null, one has (see Appendix B for details)

$$\sqrt{N}(\hat{\theta}_{nT}(\gamma) - \theta_0) = \Sigma_{nT}^{*-1}(\gamma,\gamma)\big[\tfrac{1}{\sqrt{N}} S_{\theta,nT}^{*u}(\theta_0,\gamma) + \sqrt{\tfrac{T}{n}}\, b_{\theta,nT}(\lambda_0,\gamma)\big] + o_p(1), \tag{3.1}$$

where the analytical expression for $\Sigma_{nT}^{*}(\gamma_1,\gamma_2)$ is in (B.22). It can be shown that $\frac{1}{\sqrt{N}} S_{\theta,nT}^{*u}(\theta_0,\gamma)$ converges weakly (in distribution) to $S_\theta(\gamma)$, a mean-zero Gaussian process, under the uniform metric. The covariance kernel of $S_\theta(\gamma)$ is defined by $\lim_{nT\to\infty} \Omega_{nT}^{*}(\gamma_1,\gamma_2)$, where $\Omega_{nT}^{*}(\gamma_1,\gamma_2)$ is given in (B.24). Here, the two arguments γ_1 and γ_2 are used because they are treated as distinct parameters in the subsequent derivations. A bias-corrected estimator $\hat{\theta}_{nT}^{\rm bc}(\gamma)$ is then obtained by substituting $\hat{\theta}_{nT}$ with $\hat{\theta}_{nT}(\gamma)$ in the expression for $\hat{\theta}_{nT}^{\rm bc}$ given in (2.19), while keeping the general parameter γ. Let $\widehat{\mathbb{Q}}_{nT}(\gamma,\gamma) \equiv \widehat{\Sigma}_{nT}^{*-1}(\gamma,\gamma)\widehat{\Omega}_{nT}^{*}(\gamma,\gamma)\widehat{\Sigma}_{nT}^{*-1}(\gamma,\gamma)$, where $\widehat{\Sigma}_{nT}^{*}(\gamma,\gamma) = \Sigma_{nT}^{*}(\gamma,\gamma)|_{\theta=\hat{\theta}_{nT}(\gamma)}$ and $\widehat{\Omega}_{nT}^{*}(\gamma,\gamma) = \Omega_{nT}^{*}(\gamma,\gamma)|_{\theta=\hat{\theta}_{nT}(\gamma)}$ are the respective plug-in estimators. Thus, the Wald statistic for a given γ is simply

$$W_{nT}(\gamma) = N\hat{\theta}_{nT}^{\rm bc\prime}(\gamma)\mathbf{L}[\mathbf{L}'\widehat{\mathbb{Q}}_{nT}(\gamma,\gamma)\mathbf{L}]^{-1}\mathbf{L}'\hat{\theta}_{nT}^{\rm bc}(\gamma), \tag{3.2}$$

where $\mathbf{L}$ is a selection matrix defined as $\mathbf{L} = [0_{k\times k}\ \ I_k\ \ 0_{k\times 1}\ \ 0_{k\times 1}\ \ 0_{k\times 1};\quad 0_{1\times k}\ \ 0_{1\times k}\ \ 0\ \ 1\ \ 0]'$. This statistic is, however, practically infeasible as its asymptotic distribution depends on the nuisance parameter γ. We define a sup-Wald test statistic,

$$\sup W_n = \sup_{\gamma\in\Gamma} W_{nT}(\gamma), \tag{3.3}$$

in the same spirit of Hansen (1996) for a linear threshold regression, which is shown to be an asymptotic pivot with asymptotic distribution being free from model parameters including γ. Thus, we can approximate the distribution of $\sup W_n$ via simulation!

Let $\mathbb{Q}(\gamma_1,\gamma_2) = \Sigma^{*-1}(\gamma_1,\gamma_2)\Omega^{*}(\gamma_1,\gamma_2)\Sigma^{*-1}(\gamma_1,\gamma_2)$, where $\Sigma^{*}(\gamma_1,\gamma_2) = \lim_{nT\to\infty} \Sigma_{nT}^{*}(\gamma_1,\gamma_2)$ and $\Omega^{*}(\gamma_1,\gamma_2) = \lim_{nT\to\infty} \Omega_{nT}^{*}(\gamma_1,\gamma_2)$. We have the following theorem.

Theorem 3.1. *Under Assumptions A-E, G, and the alternatives* $H_1\colon \theta_{20} = \frac{c}{\sqrt{nT}}$,

$\sup W_{nT} \xrightarrow{D} \sup_{\gamma\in\Gamma} W^c(\gamma)$, *with*

$W^c(\gamma) = [\mathbf{L}'\Sigma^{*-1}(\gamma,\gamma)S_\theta(\gamma) + \bar{\Sigma}(\gamma)c]'[\mathbf{L}'\mathbb{Q}(\gamma,\gamma)\mathbf{L}]^{-1}[\mathbf{L}'\Sigma^{*-1}(\gamma,\gamma)S_\theta(\gamma) + \bar{\Sigma}(\gamma)c]$,

where $\bar{\Sigma}(\gamma) = \sqrt{T}\mathbf{L}'\Sigma^{*-1}(\gamma,\gamma)\Sigma^{*}(\gamma,\gamma_0)\mathbf{L}$ *and* $S_\theta(\gamma)$ *is a mean-zero Gaussian process with covariance kernel* $\Omega^{*}(\gamma_1,\gamma_2)$.

Under null hypothesis, $c = 0$ and $\sup W_{nT}$ converges in distribution to $\sup_{\gamma\in\Gamma} W^0(\gamma) = \sup_{\gamma\in\Gamma} S_\theta(\gamma)'\Sigma^{*-1}(\gamma,\gamma)\mathbf{L}[\mathbf{L}'\mathbb{Q}(\gamma,\gamma)\mathbf{L}]^{-1}\mathbf{L}'\Sigma^{*-1}(\gamma,\gamma)S_\theta(\gamma)$. It is a functional of chi-square processes and thus its asymptotic critical values cannot be tabulated in general. In special cases of testing for the existence of a structure change, Andrews (1993) and Li (2018) show that the critical values depend only on the column dimension of regressors and the parameter space of γ_0 so that they can be tabulated. For threshold SPR models, an alternative approach is desired.

3.2. Bootstrap critical values for sup-Wald test

We propose a bootstrap procedure, the *estimating function bootstrap*, to approximate the asymptotic null distribution of the test statistic, extending the idea of Hansen (1996). The challenge lies in the way of simulating the asymptotic distribution with the nonlinear spatial components in the model, in contrast to Hansen (1999) for a regular panel regression.

From Theorem 3.1, $\frac{1}{\sqrt{N}} S^{*u}_{\theta,nT}(\theta_0, \gamma) \xrightarrow{D} S_\theta(\gamma)$, where $S^{*u}_{\theta,nT}(\theta_0, \gamma)$ is in (2.18). To simulate the null distribution of the sup-Wald statistic, we need the expression of $S^{*u}_{\theta,nT}(\theta_0, \gamma)$ at the null, which is of the form by (2.18),

$$S^{*u}_{\theta,nT}(\theta_{10}, \gamma) = \begin{cases} \frac{1}{\sigma_0^2} \mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbf{V}, \\ \frac{1}{\sigma_0^2} \mathbf{Y}'\mathbf{W}'\mathbf{Q}_{nT}\mathbf{V} - \mathtt{tr}[\mathbf{Q}_{nT}\mathbf{G}_1], \\ \frac{1}{\sigma_0^2} \mathbf{Y}'\mathbf{W}'\mathbf{D}'(\gamma)\mathbf{Q}_{nT}\mathbf{V} - \mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}_1], \\ \frac{1}{2\sigma_0^4} \mathbf{V}'\mathbf{Q}_{nT}\mathbf{V} - \frac{N}{2\sigma_0^2}, \end{cases} \tag{3.4}$$

where $\theta_1 = (\beta_1', \lambda_1, \sigma^2)'$, $\mathbf{G}_1 = \mathbf{W}\mathbf{A}_1^{-1}$, and $\mathbf{A}_1 = \mathbf{A}_1(\lambda_{10})$ with $\mathbf{A}_1(\lambda_1) = \mathbf{A}(\lambda, \gamma)|_{\lambda_2=0}$. The fact that $\tilde{\mathbf{V}}(\phi_0, \gamma) = \mathbf{Q}_{nT}\mathbf{V}$ at the null has been used in the above derivation, where $\phi = (\beta', \lambda')'$. Our idea is to bootstrap $S^{*u}_{\theta,nT}(\theta_{10}, \gamma)$, and thus the distribution of $W^0(\gamma)$. Note that the real world null DGP (data generating process) is

$$\mathbf{A}_1\mathbf{Y} = \mathbf{X}\beta_{10} + \mathbf{C}\psi_0 + \mathbf{V}.$$

Its sample analogue is $\mathbf{A}_1(\hat{\lambda}_1)\mathbf{Y} = \mathbf{X}\hat{\beta}_1 + \mathbf{C}\hat{\psi} + \hat{\mathbf{V}}$, which can potentially be used as a bootstrap DGP that mimics the real world DPG at the null. However, this is infeasible as due to the incidental parameters problem one cannot find a consistent estimator for $\mathbf{V}$ when T is fixed. Furthermore, the involvement of incidental parameters ψ_0 would invalidate any potential bootstrap procedure. Although a $\mathbf{Q}_{nT}$ transformation of the null DGP can remove ψ_0 and $\mathbf{Q}_{nT}\mathbf{V}$ can be 'consistently' estimated by the *unrestricted* residuals $\tilde{\mathbf{V}}(\hat{\phi}_{nT}, \hat{\gamma}_{nT})$, the elements of $\mathbf{Q}_{nT}\mathbf{V}$ are correlated and cannot be used for bootstrapping.

Very interestingly, we find that the spectral decomposition, $\mathbf{Q}_{nT} = \mathbf{S}\mathbf{S}'$, helps to do the trick that leads to an asymptotically valid bootstrap DGP, where $\mathbf{S}$ is the $nT \times N$ eigenvector matrix corresponding to eigenvalues of one. The details are as follows. First, $\mathbf{Q}_{nT}\mathbf{V} = \mathbf{S}\mathbf{S}'\mathbf{V}$.

Since $\mathbf{S}'\mathbf{S} = I_N$, the elements of $\mathbf{S}'\mathbf{V}$ are iid normal if $\{v_{it}\}$ are iid normal, and are uncorrelated if $\{v_{it}\}$ are iid. In addition, $\mathbf{S}'\mathbf{V} \equiv \mathbf{S}'\mathbf{Q}_{nT}\mathbf{V}$ can be 'consistently' estimated by $\tilde{\mathbf{V}}^*(\hat{\phi}_{nT}, \hat{\gamma}_{nT}) \equiv \mathbf{S}'\tilde{\mathbf{V}}(\hat{\phi}_{nT}, \hat{\gamma}_{nT})$, which leads to a potentially valid way of simulating $\mathbf{Q}_{nT}\mathbf{V}$ through bootstrapping $\tilde{\mathbf{V}}^*(\hat{\phi}_{nT}, \hat{\gamma}_{nT})$.

Another challenge is the generation of bootstrapped values of $\mathbf{Y}$ or $\mathbf{WY}$ as it appears in both the λ_1- and λ_2-components of $S^{*u}_{\theta,nT}(\theta_{10}, \gamma)$.[8] We show how $\mathbf{WY}$ can be related to $\mathbf{S}'\mathbf{V}$. Under H_0, $\mathbf{WY} = \mathbf{G}_1\mathbf{A}_1\mathbf{Y}$ and $\mathbf{A}_1\mathbf{Y} = \mathbf{X}\beta_{10} + \mathbf{C}\psi_0 + \mathbf{V}$. Let $\mathbf{P}_{nT} = I_{nT} - \mathbf{Q}_{nT}$ and write

$$\mathbf{A}_1\mathbf{Y} = (\mathbf{P}_{nT} + \mathbf{Q}_{nT})\mathbf{A}_1\mathbf{Y} = \mathbf{P}_{nT}\mathbf{A}_1\mathbf{Y} + \mathbf{Q}_{nT}\mathbf{X}\beta_{10} + \mathbf{Q}_{nT}\mathbf{V}.$$

The first term is uncorrelated with $\mathbf{Q}_{nT}\mathbf{V}$ and can thus be treated as constant during bootstrap draws. With $\mathbf{Q}_{nT}\mathbf{V} = \mathbf{S}\mathbf{S}'\mathbf{V}$ and the elements of $\mathbf{V}^* = \mathbf{S}'\mathbf{V}$ being iid or uncorrelated, a bootstrap value of $\mathbf{V}^*$ is therefore translated into a bootstrap value of $\mathbf{A}_1\mathbf{Y}$ or $\mathbf{Y}$ or $\mathbf{WY}$. These lead to the bootstrap DGPs for $\mathbf{Y}$ and $\mathbf{WY}$ mimicking the real world DGPs at null:

$$\mathbf{Y}^* = \mathbf{A}_1^{-1}(\hat{\lambda}_{1,nT})[\mu(\hat{\beta}_{1,nT}, \hat{\lambda}_{1,nT}) + \mathbf{S}\tilde{\mathbf{V}}^*(\hat{\phi}_{nT}, \hat{\gamma}_{nT})], \tag{3.5}$$

$$(\mathbf{WY})^* = \eta(\hat{\beta}_{1,nT}, \hat{\lambda}_{1,nT}) + \mathbf{G}_1(\hat{\lambda}_{1,nT})\mathbf{S}\tilde{\mathbf{V}}^*(\hat{\phi}_{nT}, \hat{\gamma}_{nT}), \tag{3.6}$$

where $\mu(\beta_1, \lambda_1) = \mathbf{P}_{nT}\mathbf{A}_1(\lambda_1)\mathbf{Y} + \mathbf{Q}_{nT}\mathbf{X}\beta_1$, $\eta(\beta_1, \lambda_1) = \mathbf{G}_1(\lambda_1)\mu(\beta_1, \lambda_1)$, and $\hat{\beta}_{1,nT}$ and $\hat{\lambda}_{1,nT}$ are the *unrestricted* estimates of β_1 and λ_1. In bootstrapping, one can choose (3.5) or (3.6). The following bootstrap algorithm uses the latter.

`Estimating Function Bootstrap:`

1. *Calculate the unrestricted AQML estimators $\hat{\theta}_{nT}$ and $\hat{\gamma}_{nT}$, the unrestricted transformed residuals $\tilde{\mathbf{V}}^*(\hat{\phi}_{nT}, \hat{\gamma}_{nT})$, and $\eta(\hat{\beta}_{1,nT}, \hat{\lambda}_{1,nT})$;*

2. *$\forall\gamma \in \Gamma$, calculate and save $\widetilde{\mathbb{Q}}_{nT}(\gamma, \gamma)$ and $\widetilde{\Sigma}^*_{nT}(\gamma, \gamma)$, which are the plug-in estimators of $\mathbb{Q}_{nT}(\gamma, \gamma)$ and $\Sigma^*_{nT}(\gamma, \gamma)$ with $\theta_1 = \hat{\theta}_{1,nT}$ and $\theta_2 = 0$;*

3. *Let $\mathcal{F}_{nT}$ be the empirical distribution function (EDF) defined by the centered $\tilde{\mathbf{V}}^*(\hat{\phi}_{nT}, \hat{\gamma}_{nT})$. Draw a random sample of size N from $\mathcal{F}_{nT}$ and denote it by $\tilde{\mathbf{V}}^b_N$. Compute the bootstrap value $(\mathbf{WY})^b = \eta(\hat{\beta}_{1,nT}, \hat{\lambda}_{1,nT}) + \mathbf{G}_1(\hat{\lambda}_{1,nT})\mathbf{S}\tilde{\mathbf{V}}^b_N$ through (3.6);*

4. *For each γ, calculate a bootstrap value of $S^{*u}_{\theta,nT}(\hat{\theta}_{1,nT}, \gamma)$ and denote it by $\widetilde{S}^b_{\theta,nT}(\gamma)$ based on $\tilde{\mathbf{V}}^b_N$ and $(\mathbf{WY})^b$;*

5. *Compute sup$W^b_{nT} \equiv \sup_{\gamma\in\Gamma} \frac{1}{N}\widetilde{S}^{b\prime}_{\theta,nT}(\gamma)\widetilde{\Sigma}^{*-1}_{nT}(\gamma, \gamma)\mathbf{L}[\mathbf{L}'\widetilde{\mathbb{Q}}_{nT}(\gamma, \gamma)\mathbf{L}]^{-1}\mathbf{L}'\widetilde{\Sigma}^{*-1}_{nT}(\gamma, \gamma)\widetilde{S}^b_{\theta,nT}(\gamma)$;*

6. *Repeat steps 3-5 B times;*

[8] This issue is not involved in the standard linear panel regression model in Hansen (1999) as only the linear β-component of $S^{*u}_{\theta,nT}(\theta_{10}, \gamma)$ is needed in the simulation.

7. *Calculate the bootstrap p-value of the test:* $p_W^b = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{supW_{nT}^b \geq supW_{nT}\}$, *and reject the null when* p_W^b *is less than the pre-chosen level of significance.*

The following theorem justifies the asymptotic validity of the above procedure.

Proposition 1. *Suppose Assumptions A-E, G, H and the null hypothesis hold, we have*

$$\text{sup}W_{nT}^b \xrightarrow{D^b} \sup_{\gamma\in\Gamma} W^0(\gamma),$$

where D^b *denotes the bootstrap distribution.*

For the proof of Proposition 1, it is sufficient to show that $\frac{1}{\sqrt{N}}\widetilde{S}_{\theta,nT}^b(\gamma)$ converges weakly in $\mathcal{F}_{nT}$ to a Gaussian process with covariance $\Omega^*(\gamma_1, \gamma_2)$. The details of the proof are given in Appendix B. The above theorem implies that we can approximate the asymptotic null distribution of the statistic $\text{sup}W_{nT}$ by the EDF of $\{\text{sup}W_{nT}^b, b = 1, \ldots, B\}$ for a sufficiently large B. Therefore, we can reject the null at the significance level of α when $p_W^b < \alpha$.

Remark 3.1. The above bootstrap procedure approximates the limiting distribution of $\text{sup}W_{nT}$. As $\text{sup}W_{nT}$ is an asymptotic pivot, directly bootstrapping on $\text{sup}W_{nT}$ can potentially lead to refined inference as it approximates the finite sample distribution of $\text{sup}W_{nT}$. However, this procedure, though possible in theory, may not be practical, as an estimation of the full model is required for every bootstrap sample, which can be computationally intensive.

Remark 3.2. Our proposed bootstrap procedure does not require re-estimation of the model, null or full, and thus is computationally simple. Step 3 of the bootstrap algorithm can instead be as follows: generate $\mathbf{Y}^*$ using DGP (3.5), estimate the null model to give $\hat{\theta}_{1,nT}^b$, and then compute a bootstrap value of $S_{\theta,nT}^{*u}(\theta_{10}, \gamma)$ using $\mathbf{Y}^*$ and $\hat{\theta}_{1,nT}^b$. This evidently increases the computational burden. Our simulation results show that this does not make a significant difference. Thus, we recommend the proposed bootstrap procedure for computational simplicity.

Remark 3.3. When bootstrapping the null distribution of a test statistic where only the estimation of the null model may be required, it is important to use unrestricted estimates and unrestricted residuals to set up the bootstrap DGP that mimics the real world at the null, as we never know whether the null is true or untrue, an important point made by Yang (2015b).

4. Monte Carlo Study

Monte Carlo experiments are performed to evaluate the finite sample performance of the proposed estimators and the test statistics. The following data-generating process is used:

$$Y_t = \lambda_1 W_t Y_t + \lambda_2 d_t(\gamma) W_t Y_t + X_t\beta_1 + d_t(\gamma) X_t \beta_2 + \mu + \alpha_t l_n + V_t, \;\; t = 1, \ldots, T$$

where the time-varying weight matrices W_t's are generated according to `Queen` contiguity, x_{it} are generated from $N(0, 2^2)$, the fixed effects μ are generated according to $\frac{1}{T}\Sigma_{t=1}^{T} X_t + e$, where $e \sim N(0, I_N)$, and the time fixed effects α are generated from $N(0, I_T)$. The distributions of the error term can be (i) normal, (ii) normal mixture (90% $N(0,1)$ and 10% $N(0, 4^2)$), or (iii) chi-square with 3 degrees of freedom. In both (ii) and (iii), the error distributions are standardized to have mean zero and variance $\sigma^2 = 1$. We set $\beta_1 = 1$, $\lambda_1 = 0.2$, $\beta_2 = \lambda_2 = (nT)^{-0.2}$ and $\gamma = 0$. The number of Monte Carlo runs under each parameter configuration is 1000.

Table 1 presents the Monte Carlo results for AQMLE based on (2.6) and bias-corrected AQMLE (bc-AQMLE) based on (2.19), for various combinations of $n = 50, 100, 200$ and $T = 5, 10, 20, 40$. As direct QMLE based on (2.5) shares the same estimated values as AQMLE for all parameters except σ^2 (see Footnote 3), only the results of direct QMLE of σ^2 are provided at the bottom of the table. As a comparison, the 2SLS estimator of Wei et al. (2021) is also included in our Monte Carlo experiments. Monte Carlo biases and standard deviations (sd, reported in brackets) are presented for all parameters, with empirical averages of the robust standard error estimates ($\hat{sd}$) also shown for AQMLE and bc-AQMLE. Note that bc-AQMLE of γ is the same as AQMLE of γ as the latter does not incur asymptotic bias. Only the estimation of θ is subject to asymptotic bias, as shown in Theorems 2.4 and 2.5.

The results indicate that the finite sample performance of the 2SLS estimator can be poor, exhibiting large biases and high standard deviations. The direct QML method improves the estimation of all parameters but the error variance, aligning with our theory that the direct QML estimator of σ^2 is inconsistent when T is fixed due to the incidental parameters problem. In contrast, the results demonstrate an excellent finite-sample performance of our AQMLE in terms of both consistency and efficiency of the estimation. Our bc-AQMLE performs even better, particularly when both n and T are large. All estimators improve as the sample size increases, regardless of the error distribution. The $\sqrt{N}$ convergence rate of both AQMLE and bc-AQMLE is clearly demonstrated by the Monte Carlo standard deviations. Moreover, the robust standard error estimates ($\hat{sd}$) are, on average, very close to the corresponding Monte Carlo standard deviations, with this closeness increasing as the sample size grows.

Table 2 reports the coverage probabilities of the nominal 95% confidence intervals for γ obtained by inverting the LR test introduced in Section 2.4. The results show that the coverage probabilities are close to the nominal levels under different error distributions.

Table 3 summarizes Monte Carlo results for the size and local power of the supWald test for the non-existence of threshold effects introduced in Section 3. Under the null hypothesis ($\beta_{20} = \lambda_{20} = 0$), the rejection rates align with the nominal levels under different error distributions, showing that the proposed bootstrap procedure works well. We also examine the local power of the test in the last two columns of the table. As expected, the rejection rates increase rapidly as β_{20} and λ_{20} deviate from zero. When $\beta_{20} = \lambda_{20} = 10/\sqrt{nT}$, the power of our test reaches 100% for all sample sizes.

5. An Empirical Application

In this section, we apply our method to study the age-of-leader effects on political competitions across Chinese cities. The tournament competition among Chinese city government leaders has been an important topic in China's economic growth literature (Yao and Zhang, 2015). Local government leaders compete against one another in enhancing local investment and promoting the local economy's growth so as to increase their chances for political promotion. Based on this theory, Yu et al. (2016) document a strong spatial effect for the city-level total investment. Besides, the age of a local leader is another pivotal factor determining the leader's chances of promotion. A leader's chance diminishes quickly as he or she gets older (Yao and Zhang, 2015; Yu et al., 2016). Thus, those leaders who are close to retirement age (60) should have less incentive to join this tournament competition than the young leaders, and thus a weaker spatial effect should be expected among the old leaders. That is, we would expect that the spatial correlation of city-level total investment has a threshold effect based on the leaders' age. In contrast to Yu et al. (2016) who try various cutoff ages to see the change of the spatial correlation over leader's age, our threshold SPR model can directly estimate the threshold age. First, a test of no threshold effects is carried out using the sup-Wald test developed in Section 3, and then if this test is rejected, a confidence interval for the threshold parameter is constructed by inverting the LR test given in Section 2.4.

Model and data. Following the above discussions, we consider the following model:

$$inv_{it} = \lambda_1 \sum_{j=1}^{n} w_{ij,t} inv_{jt} + \lambda_2 \sum_{j=1}^{n} w_{ij,t} inv_{jt} \mathbf{1}\{age_{it} \leqslant \gamma\} + x_{it}\beta + \mu_i + \alpha_t + v_{it},$$

where inv_{it} denotes the total investment of local government of city i in year t, age_{it} denotes the age of the local leader of city i in year t, x_{it} is a vector of time-varying regressors including fiscal revenue, fiscal expenditure, population, manufacturing ratio, GDP per capita and a set of province level variables: fiscal revenue, fiscal expenditure, and public capital investment, μ_i and α_t are the two-way fixed effects, and v_{it} is the idiosyncratic error. We follow Yu et al. (2016) and define those same-province cities whose within-province rankings of GDP per capita are either one place above or below a city as this city's spatial neighbors, because they are the main competitors in the tournament competition. Because there is no theoretical evidence to justify the threshold effects for regression coefficients, they are not included.

We analyze the annual total investments (in RMB) of 338 cities in the 27 provinces in mainland China from 2010 to 2012. Economic data is from *Fiscal Statistics of Cities and Counties in China*, *China City Statistical Yearbook* and *China Statistical Yearbook for Regional Economy* for the period 2010-2012. The ages of leaders are obtained from local government websites. The data is standardized to make all the variables have comparable scales.

Test for the presence of threshold effects. Before the estimation of the model, we conduct hypothesis testing on the presence of the age-of-leader threshold effect. In China's local official system, there are two types of leaders in the local governments, party secretaries and mayors. Party secretaries are mainly responsible for personnel work and overall decision-making while the mayors are for the formulation and implementation of specific economic and social policies so that Yao and Zhang (2013) find the weight of economic performance is lower for the party secretary than for the mayor in the assessment of local leaders. Consequently, we consider only mayors in our analysis. In addition, as Yu et al. (2016) find the age-of-leader effects on political competitions are more clear among old leaders and old leaders have different spatial responsiveness to their young and old neighbors, we also separate the leaders into young and old groups. Here, "old" leaders are defined as those whose ages exceed the median age (49 years and 8 months).

We conduct a sup-Wald test on mayors, examining four types of competitive effects: "all vs all," which examines spatial correlation among all mayors; "old vs all," which measures spatial correlation between old mayors and all their neighboring counterparts; "old vs old," which considers spatial correlation solely among old mayors; and "old vs young," which looks at the spatial correlation between old mayors and their younger neighbors. The resulting $\sup W_{nT}$ statistics and associated bootstrap p-values, based on 699 bootstrap replications, are $(3.785, 0.336)$, $(8.360, 0.060)$, $(11.543, 0.015)$, and $(1.885, 0.845)$, respectively. Consequently, we can reject the null hypothesis of no threshold effect at the 10% significance level for the "old vs all" pattern and at the 5% level for the "old vs old" pattern.

Estimation results. Table 4 reports the regression results for the two scenarios when we can reject the null hypothesis of no threshold effect. "Model 1" and "Model 2" correspond to the "old vs all" pattern and "old vs old" pattern, respectively. The estimations of threshold coefficient γ are 55.33 (55 years and 4 months)and 54.58 (54 years and 7 months) for these two models, respectively. We also report the 95% confidence intervals that are based on the likelihood ratio test. The estimations of λ_1 in these two models suggest that the spatial correlations when the ages of local mayors are beyond the threshold levels are slightly negative

(-0.033 and -0.041) but not significant. In contrast, when the ages of local mayors are below the threshold levels, the spatial correlations among local investments are the estimations of $\lambda_1 + \lambda_2$ and thus become strongly positive as λ_2 are positive with a much larger magnitude. These empirical findings are in line with our theoretical expectation, considering that mayors normally take office in their forties or fifties and the mandatory retirement age for them is 60. A more comprehensive study on this topic is of interest as future research.

6. Extensions

We have by far focused on a threshold SPR model (2.1) that contains only a spatial lag (SL) structure with additive fixed effects, for ease of exposition. The proposed estimation and inference methods are in fact quite general and can be extended to include additional features in the model such as spatial error dependence, serial correlation, time dynamics, multiple threshold effects, threshold effects on error parameters, interactive fixed effects, etc. An immediate and much-needed extension is the inclusion of spatial error (SE) effect:

$$Y_t = \lambda_{10} W_t Y_t + \lambda_{20} d_t(\gamma_0) W_t Y_t + X_t \beta_{10} + d_t(\gamma_0) X_t \beta_{20} + \mu_0 + \alpha_{t0} l_n + U_t,\ U_t = \rho_0 W_t^e U_t + V_t,$$

for $t = 1, \ldots, T$, where parameter ρ and weight matrices $\{W_t^e\}$ together characterize the SE effects, and the other parts are defined in Model (2.1). Let $\mathbf{B}(\rho) = I_{nT} - \rho \mathbf{W}^e$, where $\mathbf{W}^e = \mathtt{bdiag}(W_1^e, \ldots, W_T^e)$. We redefine $\theta = (\beta', \lambda', \rho, \sigma^2)'$ to accommodate the extra spatial error parameter. The quasi Gaussian loglikelihood function of all the parameters becomes

$$\ell_{nT}(\theta, \gamma, \psi) = -\frac{nT}{2} \ln(2\pi\sigma^2) + \ln|\mathbf{A}(\lambda, \gamma)| + \ln|\mathbf{B}(\rho)| - \frac{1}{2\sigma^2} \mathbf{V}'(\theta, \gamma, \psi) \mathbf{V}(\theta, \gamma, \psi),$$

where $\mathbf{V}(\theta, \gamma, \psi) = \mathbf{B}(\rho)[\mathbf{A}(\lambda, \gamma)\mathbf{Y} - \mathbb{X}(\gamma)\beta - \mathbf{C}\psi]$. $\ell_{nT}(\theta, \gamma, \psi)$ is partially maximized at

$$\hat{\psi}(\theta, \gamma) = [\mathbb{C}'(\rho)\mathbb{C}(\rho)]^{-1} \mathbb{C}'(\rho) \mathbf{B}(\rho)[\mathbf{A}(\lambda, \gamma)\mathbf{Y} - \mathbb{X}(\gamma)\beta],$$

where $\mathbb{C}(\rho) = \mathbf{B}(\rho)\mathbf{C}$. Thus, the adjusted concentrated quasi loglikelihood function corresponding to (2.6) becomes

$$\ell^*_{nT}(\theta, \gamma) = -\frac{nT}{2} \ln(2\pi\sigma^2) + \ln|\mathbf{A}(\lambda, \gamma)| + \ln|\mathbf{B}(\rho)| - \frac{c_{nT}}{2\sigma^2} \ddot{\mathbf{V}}'(\theta, \gamma) \ddot{\mathbf{V}}(\theta, \gamma),$$

where $\ddot{\mathbf{V}}(\theta, \gamma) = \mathbf{Q}_{\mathbb{C}}(\rho)\mathbf{B}(\rho)[\mathbf{A}(\lambda, \gamma)\mathbf{Y} - \mathbb{X}(\gamma)\beta]$ and $\mathbf{Q}_{\mathbb{C}}(\rho) = I_{nT} - \mathbb{C}(\rho)[\mathbb{C}'(\rho)\mathbb{C}(\rho)]^{-1}\mathbb{C}'(\rho)$. The adjusted QML estimators of θ, ρ and γ are simply

$$(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) = \operatorname*{argmax}_{(\theta, \gamma) \in \Theta \times \Gamma} \ell^*_{nT}(\theta, \gamma),$$

where Θ is now extended to include the the parameter space for ρ, which is denoted by Δ_ρ.

In practice, we can also first maximize the objective function conditional on γ to get $\hat{\theta}_{nT}(\gamma)$, and then apply the grid search algorithm in Section 2.2 to obtain $\hat{\gamma}_{nT}$. With some additional conditions (e.g., both $\|\mathbf{W}^{\mathbf{e}}\|_1$ and $\|\mathbf{W}^{\mathbf{e}}\|_\infty$ are bounded; both $\|\mathbf{B}^{-1}(\rho)\|_1$ and $\|\mathbf{B}^{-1}(\rho)\|_\infty$ are bounded on Δ_ρ; ρ is identifiable), we expect the estimation error of $\hat{\gamma}_{nT}$ still have asymptotically negligible effects on $\hat{\theta}_{nT}$, and thus we can establish similar results to those in Theorems 2.1 - 2.5. Moreover, to construct a confidence interval for γ, we construct the LR statistic in the same way as in Section 2.4,

$$LR_{nT}(\gamma) = \frac{2}{c_{nT}}[\ell_{nT}^*(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) - \ell_{nT}^*(\hat{\theta}_{nT}(\gamma), \gamma)].$$

When errors are normally distributed, the asymptotic distribution of $LR_{nT}(\gamma_0)$ is still pivotal, following the distribution of $\mho$. In this case, the asymptotic $1-\alpha$ confidence interval for γ is the set of values of γ satisfying $LR_{nT}(\gamma) \leqslant \mho_{1-\alpha}$. Finally, to test the presence of threshold effects, we just follow Section 3 and establish the new sup-Wald test statistic and bootstrap procedure in a similar manner.

Our estimation and inference methods can also be extended to handle models with other additional features. Firstly, an extension to allow for serial correlation in the error term (e.g., $v_{it} = \varrho v_{i,t-1} + e_{it}$ with $|\varrho| < 1$) is also straightforward like the above one with the `SE` structure. We expect the arguments and ideas behind estimation and inference methods can still be applied with minor modifications. Secondly, we can generalize our model to the dynamic SPR framework. When T is large, the direct QML approach should provide a consistent estimation for all the parameters, and thus the asymptotic properties of these AQMLEs can be derived in a standard manner. When T is fixed, the analysis will become complicated as adjustments to the concentrated QML function are required to deal with the incidental parameters problems coming from both the initial condition and the concentration. Thirdly, the extension to include multiple thresholds (Hansen, 1999) is also of theoretical and practical interest. For this extension, our AQML approach is still appropriate and the objective function with multiple thresholds corresponding (2.6) is also straightforward to construct. Thus, the adjusted QML estimators of all the parameters including multiple threshold parameters jointly maximize the new objective function. In practice, the grid search over multiple thresholds may require an excessive amount of computation. We recommend using the sequential estimation method with refinement (Bai, 1997; Hansen, 1999) to avoid this computational burden.

Fourthly, our methods can also be extended to include the threshold effects on error parameters, e.g., error variance (Miao et al., 2020). In this case, the threshold effects on error parameters need to be incorporated into the QML function. For example, when error variance has threshold effects, for each observation the variance parameter will appear in the form of $\sigma_1^2 + \sigma_2^2 \mathbb{1}(q_{it} \leqslant \gamma)$, where σ_1^2 is the baseline parameter and σ_2^2 is its threshold effect. Finally, our methods can be extended to allow the individual and time fixed effects to appear in the model interactively. According to Miao et al. (2020), we would expect the concentrated QML estimation (with common factors being concentrated out) can provide a consistent estimation for all the parameters, including threshold parameter and factor loadings, when both n and T are large. Besides, we expect that the estimation error of the threshold estimate still has no asymptotic effect on the asymptotic properties of the other estimators and that the inference methods in this paper can still be applied. However, formal studies on these extensions are still quite involved and can only be handled in future research.

7. Conclusion

In this paper, we consider estimation and inference for a threshold spatial panel data model with both individual and time fixed effects, where threshold effects are allowed for both spatial and regression parameters. The presence of the threshold effects renders the commonly used orthogonal transformation approach inapplicable to wipe out fixed effects. We propose an *adjusted quasi maximum likelihood* estimation method, where the objective function is obtained by adjusting the concentrated quasi loglikelihood function (with fixed effects being concentrated out) to "recover" the effect of degrees of freedom loss due to the estimation of these incidental parameters. We study the asymptotic properties of the adjusted QML estimators in the diminishing-threshold-effect framework and propose a likelihood ratio statistic to construct confidence intervals for the threshold parameter. We also consider the hypothesis testing on the presence of threshold effects and a sup-Wald statistic based on the bias-corrected adjusted QML estimation is proposed. Monte Carlo results show excellent performance of the proposed estimation and inference methods. We apply our model to study the age-of-leader effects on political competitions across Chinese cities and find competitions only exist among city leaders who are younger than a threshold age.

Appendix A: Some Basic Lemmas

Lemma A.1. *(Lee, 2002): Let* $\{A_n\}$ *and* $\{B_n\}$ *be two sequences of* $n \times n$ *matrices that are uniformly bounded in both row and column sums. Let* C_n *be a sequence of conformable matrices whose elements are uniformly* $O(h_n^{-1})$*. Then*

(i) *the sequence* $\{A_n B_n\}$ *are uniformly bounded in both row and column sums,*

(ii) *the elements of* A_n *are uniformly bounded and* $\texttt{tr}(A_n) = O(n)$*, and*

(iii) *the elements of* $A_n C_n$ *and* $C_n A_n$ *are uniformly* $O(h_n^{-1})$*.*

Lemma A.2. *(Lemma B.4, Yang, 2015a): Let* $\{A_n\}$ *be a sequence of* $n \times n$ *matrices that are uniformly bounded in both row and column sums. Suppose that the elements* $a_{n,ij}$ *of* A_n *are bounded uniformly in all* i *and* j*, and* $a_{n,ii} \neq 0$ *for some* i*. Let* v_n *be a random* n*-vector of iid elements with mean zero, variance* σ^2 *and finite 4th moment, and* b_n *a random* n*-vector independent of* v_n *such that* $\{\mathrm{E}(b_{ni}^2)\}$ *are bounded. Then*

(i) $\mathrm{E}(v_n' A_n v_n) = O(n)$*,* *(ii)* $\mathrm{Var}(v_n' A_n v_n) = O(n)$*,*

(iii) $\mathrm{Var}(v_n' A_n v_n + b_n' v_n) = O(n)$*,* *(iv)* $v_n' A_n v_n = O_p(n)$*,*

(v) $v_n' A_n v_n - \mathrm{E}(v_n' A_n v_n) = O_p(n^{\frac{1}{2}})$*,* *(vi)* $v_n' A_n b_n = O_p(n^{\frac{1}{2}})$*.*

Lemma A.3. *(Lemma A.5, Yang, 2018): Let* $\{\Phi_n\}$ *be a sequence of* $n \times n$ *matrices with row and column sums uniformly bounded, and elements of uniform order* $O(h_n^{-1})$*. Let* $v_n = (v_1, \dots, v_n)'$ *be a random vector of iid elements with mean zero, variance* σ^2*, and finite* $(4+2\epsilon_0)$*th moment for some* $\epsilon_0 > 0$*. Let* $b_n = \{b_{ni}\}$ *be an* $n \times 1$ *random vector, independent of* v_n*, such that (i)* $\{\mathrm{E}(b_{ni}^2)\}$ *are of uniform order* $O(h_n^{-1})$*, (ii)* $\sup_i E|b_{ni}|^{2+\epsilon_0} < \infty$*, (iii)* $\frac{h_n}{n} \sum_{i=1}^n [\phi_{n,ii}(b_{ni} - \mathrm{E} b_{ni})] = o_p(1)$ *where* $\{\phi_{n,ii}\}$ *are the diagonal elements of* Φ_n*, and (iv)* $\frac{h_n}{n} \sum_{i=1}^n [b_{ni}^2 - \mathrm{E}(b_{ni}^2)] = o_p(1)$*. Define the bilinear-quadratic form:*

$$Q_n = b_n' v_n + v_n' \Phi_n v_n - \sigma^2 \texttt{tr}(\Phi_n),$$

and let $\sigma_{Q_n}^2$ *be the variance of* Q_n*. If* $\lim_{n\to\infty} h_n^{1+2/\epsilon_0}/n = 0$ *and* $\{\frac{h_n}{n}\sigma_{Q_n}^2\}$ *are bounded away from zero, then* $Q_n/\sigma_{Q_n} \xrightarrow{D} N(0,1)$*.*

Lemma A.4. *(Adapted from Lemma 1, Hansen, 1996): If* $\{w_i\}$ *are iid,* $\mathrm{E}[\Psi(w_i)] < \infty$ *for a matrix function* Ψ *of* w_i*, and* w_i *has a continuous distribution, then*

$$\sup_{\gamma\in\Gamma} \Big\| \frac{1}{n} \sum_{i=1}^n \Psi(w_i) \mathbb{1}\{w_i \leqslant \gamma\} - \mathrm{E}[\Psi(w_i)\mathbb{1}\{w_i \leqslant \gamma\}] \Big\| \longrightarrow 0 \ \ a.s.$$

Appendix B: Proofs of the Theorems

This appendix presents proofs of the main theorems of the paper. The proofs of the main theorems are greatly facilitated by Lemmas B.1-B.4 given below. Their proofs are given in `Supplementary Material`. All quantities involved are defined in Sections 2.2 and 2.3.

Lemma B.1. *Under Assumptions A-E, we have*

$$\mathcal{J}_{1,nT}(\gamma) = \tfrac{1}{\sqrt{N}}\mathbf{H}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{V} \;\Rightarrow\; \mathcal{J}_1(\gamma),$$
$$\mathcal{J}_{2,nT}(\gamma) = \tfrac{1}{\sqrt{N}}[\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}\mathbf{V} - \sigma_0^2\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G})] \;\Rightarrow\; \mathcal{J}_2(\gamma),$$

where " $\Rightarrow$ " denotes weak convergence with respect to the uniform metric, and both $\mathcal{J}_1(\gamma)$ and $\mathcal{J}_2(\gamma)$ are mean-zero Gaussian processes with almost surely continuous sample paths and variances $\lim\limits_{nT\to\infty}\frac{\sigma_0^2}{N}\mathrm{E}[\mathbf{H}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{H}]$ *and* $\lim\limits_{nT\to\infty}\frac{\sigma_0^4}{N}\mathrm{E}[(\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G})^2+\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}\mathbf{G}'\mathbf{D}(\gamma))]+\frac{\kappa_4\sigma_0^4}{N}\mathrm{E}[\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G})'\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G})]$, *respectively.*

Lemma B.2. *Under Assumptions A-E, we have,*

$$\mathcal{F}_{nT}(v) = \tfrac{a_{nT}}{nT}\delta_0'\mathbf{H}'\mathbf{D}(\gamma_{nT},\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_{nT},\gamma_0)\mathbf{H}\delta_0 \;\Rightarrow\; \lim_{nT\to\infty}\bar{T}\delta_0'M\delta_0 f|v|,$$
$$\mathcal{K}_{nT}(v) = \tfrac{a_{nT}}{nT}l_0^2\mathbf{V}'\mathbf{G}'\mathbf{D}(\gamma_{nT},\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_{nT},\gamma_0)\mathbf{G}\mathbf{V} \;\Rightarrow\; \lim_{nT\to\infty} l_0^2\sigma_0^2\bar{T}\pi_1 f|v|,$$
$$\mathcal{L}_{nT}(v) = \tfrac{a_{nT}}{nT}l_0^2\mathtt{tr}[(\mathbf{D}(\gamma_{nT},\gamma_0)\mathbf{G})^2] \;\Rightarrow\; \lim_{nT\to\infty} l_0^2\pi_2 f|v|,$$

where $\gamma_{nT} = \gamma_0 + v/a_{nT}$, v is on a compact set $\Upsilon = [-\bar{v},\bar{v}]$, π_1 and π_2 are defined above Theorem 2.5, $\bar{T} = \frac{T-1}{T}$, and $\mathbf{D}(\gamma_1,\gamma_2) = \mathbf{D}(\gamma_1) - \mathbf{D}(\gamma_2)$.

Lemma B.3. *Under Assumptions A-E and for $\mathcal{J}_{r,nT}(\gamma)$ in Lemma B.1, $r = 1,2$, we have,*

$$\mathcal{R}_{nT}(v) = \sqrt{a_{nT}}\left[\delta_0'\mathcal{J}_{1,nT}(\gamma_{nT},\gamma_0) + l_0\mathcal{J}_{2,nT}(\gamma_{nT},\gamma_0)\right] \Rightarrow \sqrt{\sigma_0^2\Xi f}\; W(v)$$

where $\mathcal{J}_{r,nT}(\gamma_{nT},\gamma_0) = \mathcal{J}_{r,nT}(\gamma_{nT}) - \mathcal{J}_{r,nT}(\gamma_0)$, $\gamma_{nT} = \gamma_0 + v/a_{nT}$ with v being on a compact set $[-\bar{v},\bar{v}]$, and $W(v)$ is a standard Brownian motion with Ξ being given in Theorem 2.5.

Lemma B.4. *Under Assumptions A-E, there exist constants $B > 0$, $0 < k < \infty$, and $0 < l < \infty$, such that for all $\eta > 0$, and $\epsilon > 0$, there exists a $\bar{v} < \infty$ such that*

$$(a)\ P\Big(\inf_{\gamma\in\mathcal{N}_{nT}}\tfrac{D_{r,nT}(\gamma)}{|\gamma-\gamma_0|} < (1-\eta)k\Big) \leqslant \epsilon, \quad (b)\ P\Big(\sup_{\gamma\in\mathcal{N}_{nT}}\tfrac{\|F_{s,nT}(\gamma)\|}{|\gamma-\gamma_0|} > (1+\eta)l\Big) \leqslant \epsilon,$$
$$(c)\ P\Big(\sup_{\gamma\in\mathcal{N}_{nT}}\tfrac{|K_{s,nT}(\gamma)|}{|\gamma-\gamma_0|} > (1+\eta)l\Big) \leqslant \epsilon, \quad (d)\ P\Big(\sup_{\gamma\in\mathcal{N}_{nT}}\tfrac{|L_{r,nT}(\gamma)|}{|\gamma-\gamma_0|} > (1+\eta)l\Big) \leqslant \epsilon,$$
$$(e)\ P\Big(\sup_{\gamma\in\mathcal{N}_{nT}}\tfrac{\|P_{r,nT}(\gamma)\|}{|\gamma-\gamma_0|} > \eta\Big) \leqslant \epsilon, \quad (f)\ P\Big(\sup_{\gamma\in\mathcal{N}_{nT}}\tfrac{\|\mathcal{J}_{s,nT}(\gamma,\gamma_0)\|}{\sqrt{a_{nT}}|\gamma-\gamma_0|} > \eta\Big) \leqslant \epsilon,$$

for large enough (n,T) and $r = 1,2,3$ and $s = 1,2$, where $\mathcal{N}_{nT} = \big\{\gamma : \frac{\bar{v}}{a_{nT}} \leqslant |\gamma-\gamma_0| \leqslant B\big\}$,

$$D_{1,nT}(\gamma) = \delta_0'F_{1,nT}(\gamma)\delta_0, \quad D_{2,nT}(\gamma) = l_0^2K_{1,nT}(\gamma), \quad D_{3,nT}(\gamma) = l_0^2L_{2,nT}(\gamma),$$

$F_{1,nT}(\gamma) = \frac{1}{nT}\mathbf{H}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}$, $\quad F_{2,nT}(\gamma) = \frac{1}{nT}\mathbb{H}'(\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}$,
$K_{1,nT}(\gamma) = \frac{1}{nT}\mathbf{V}'\mathbf{G}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$, $\quad K_{2,nT}(\gamma) = \frac{1}{nT}\mathbb{V}'(\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$,
$L_{1,nT}(\gamma) = \frac{1}{nT}\mathtt{tr}[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}]$, $\quad L_{2,nT}(\gamma) = \frac{1}{nT}\mathtt{tr}[(\mathbf{D}(\gamma_0,\gamma)\mathbf{G})^2]$,
$L_{3,nT}(\gamma) = \frac{1}{nT}\mathtt{tr}[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{J}]$, $\quad P_{1,nT}(\gamma) = \frac{1}{nT}\mathbf{H}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$,
$P_{2,nT}(\gamma) = \frac{1}{nT}\mathbb{H}'(\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$, $\quad P_{3,nT}(\gamma) = \frac{1}{nT}\mathbb{V}'(\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}$,
$\mathcal{J}_{r,nT}(\gamma)$, $r = 1, 2$, are defined in Lemma B.1, and $\mathbf{D}(\gamma_0,\gamma)$ is defined in Lemma B.2.

Equipped with Lemmas B.1-B.4, recall the notations defined in Section 2:

- $\beta = (\beta_1', \beta_2')'$, $\lambda = (\lambda_1, \lambda_2)'$, $\phi = (\beta', \lambda')'$, $\delta_0 = (b_0', l_0)'$, $\theta_2 = (\beta_2', \lambda_2)'$;
- $\mathbf{D}(\gamma) = \mathtt{bdiag}(d_1(\gamma), \ldots, d_T(\gamma))$, $\quad \mathbf{A}(\lambda,\gamma) = I_{nT} - \lambda_1\mathbf{W} - \lambda_2\mathbf{D}(\gamma)\mathbf{W}$,
  $\mathbf{G}(\lambda,\gamma) = \mathbf{W}\mathbf{A}^{-1}(\lambda,\gamma)$, $\quad \bar{\mathbf{G}}(\lambda,\gamma) = \mathbf{G}(\lambda,\gamma) - \mathtt{diag}(\mathbf{G}(\lambda,\gamma))$;
- $\mathbf{Z} = \mathbf{G}(\mathbb{X}\beta_0 + \mathbf{C}\psi_0)$, $\quad \mathbf{H} = [\mathbf{X},\ \mathbf{Z}]$;
- $\mathbb{X}(\gamma) = [\mathbf{X},\ \mathbf{D}(\gamma)\mathbf{X}]$, $\quad \mathbb{Z}(\gamma) = [\mathbf{Z},\ \mathbf{D}(\gamma)\mathbf{Z}]$, $\quad \mathbb{H}(\gamma) = [\mathbf{H},\ \mathbf{D}(\gamma)\mathbf{H}]$,
  $\mathbb{V}(\gamma) = [\mathbf{G}\mathbf{V},\ \mathbf{D}(\gamma)\mathbf{G}\mathbf{V}]$, $\quad \mathbb{R}(\gamma) = [\mathtt{diagv}(\mathbf{G}),\ \mathbf{D}(\gamma)\mathtt{diagv}(\mathbf{G})]$,

and define the following new notations: $A^{[k]}$ denotes the submatrix of A, consisting of the first k rows and columns; and $h^{[k]}$ denotes the subvector of h, consisting of the first k elements. We are ready to prove Theorems 2.1-2.3.

Proof of Theorem 2.1: We first prove the convergence of $\hat{\beta}_{nT}(\hat{\lambda}_{nT},\gamma)$ and $\hat{\sigma}^2_{nT}(\hat{\lambda}_{nT},\gamma)$, uniformly in $\gamma \in \Gamma$. We have $\mathbf{A}(\lambda,\gamma)\mathbf{A}^{-1} = I_n + (\lambda_{10} - \lambda_1)\mathbf{G} + (\lambda_{20} - \lambda_2)\mathbf{D}(\gamma)\mathbf{G} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}$, noting that $\mathbf{A}^{-1} = I_n + \lambda_{10}\mathbf{G} + \lambda_{20}\mathbf{D}(\gamma_0)\mathbf{G}$. By $\mathbf{Y} = \mathbf{A}^{-1}(\mathbb{X}\beta_0 + \mathbf{C}\psi_0 + \mathbf{V})$ and $\mathbb{X}\beta_0 = \mathbb{X}(\gamma)\beta_0 + \mathbf{D}(\gamma_0,\gamma)\mathbf{X}\beta_{20}$, we have

$$\mathbf{A}(\lambda,\gamma)\mathbf{Y} = \mathbb{X}(\gamma)\beta_0 + [\mathbb{Z}(\gamma) + \mathbb{V}(\gamma)](\lambda_0 - \lambda) + \mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V} + \mathbf{C}\psi_0 + \mathbf{V}, \quad \text{(B.1)}$$

Combining it with (2.9) and (2.10), we have

$$\begin{aligned}\hat{\beta}_{nT}(\lambda,\gamma) &= \beta_0 + [\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]^{-1}\{\mathbb{X}'(\gamma)\mathbf{Q}_{nT}[\mathbb{Z}(\gamma) + \mathbb{V}(\gamma)](\lambda_0 - \lambda) \qquad \text{(B.2)}\\ &\quad + \mathbb{X}'(\gamma)\mathbf{Q}_{nT}[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V} + \mathbf{V}]\},\ \text{and}\\ \hat{\sigma}^2_{nT}(\lambda,\gamma) &= \tfrac{1}{N}(\lambda_0 - \lambda)'[\mathbb{Z}(\gamma) + \mathbb{V}(\gamma)]'\mathbb{M}_{nT}(\gamma)[\mathbb{Z}(\gamma) + \mathbb{V}(\gamma)](\lambda_0 - \lambda) \qquad \text{(B.3)}\\ &\quad + \tfrac{2}{N}(\lambda_0 - \lambda)'[\mathbb{Z}(\gamma) + \mathbb{V}(\gamma)]'\mathbb{M}_{nT}(\gamma)[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V} + \mathbf{V}]\\ &\quad + \tfrac{1}{N}[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V} + \mathbf{V}]\mathbb{M}_{nT}(\gamma)[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V} + \mathbf{V}],\end{aligned}$$

where $\mathbb{M}_{nT}(\gamma) = \mathbf{Q}_{nT} - \mathbf{Q}_{nT}\mathbb{X}(\gamma)[\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]^{-1}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}$.

Under Assumption B(vi), the limit of $\frac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)$ exists and is nonsingular. In addition, uniformly in $\gamma \in \Gamma$, $\frac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbf{V}$ and $\frac{1}{N}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})'\mathbf{Q}_{nT}\mathbf{V}$ are $O_p((N)^{-1/2})$ by

Lemma B.1; $\frac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}(\mathbb{V}(\gamma), \mathbf{D}(\gamma_0,\gamma)\mathbf{GV})$ and $\frac{1}{N}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})'\mathbf{Q}_{nT}(\mathbb{V}(\gamma), \mathbf{D}(\gamma_0,\gamma)\mathbf{GV})$ are $o_p(1)$ by Lemma A.4; similarly, $\frac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})$, $\frac{1}{N}\mathbf{V}'\mathbf{Q}_{nT}(\mathbb{V}(\gamma), \mathbf{D}(\gamma_0,\gamma)\mathbf{GV})$, $\frac{1}{N}(\mathbb{V}(\gamma), \mathbf{D}(\gamma_0,\gamma)\mathbf{GV})'\mathbf{Q}_{nT}(\mathbb{V}(\gamma), \mathbf{D}(\gamma_0,\gamma)\mathbf{GV})$ and $\frac{1}{N}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})'\mathbf{Q}_{nT}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})$ are all $O_p(1)$ by Lemma A.4, as their expectations are all $O(1)$. Besides, $\frac{1}{N}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{V}-\sigma_0^2 = o_p(1)$ by Lemma A.2 and $\theta_{20} = O((nT)^{-\tau})$ by Assumption F. These together lead to

$$\hat{\beta}_{nT}(\lambda,\gamma) = \beta_0 + [\tfrac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]^{-1}\tfrac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{Z}(\gamma)(\lambda_0-\lambda) + o_p(1), \quad \text{and} \tag{B.4}$$

$$\begin{aligned}\hat{\sigma}^2_{nT}(\lambda,\gamma) = \sigma_0^2 &+ \tfrac{2}{N}\mathbf{V}'\mathbf{Q}_{nT}\mathbb{V}(\gamma)(\lambda_0-\lambda) + (\lambda_0-\lambda)'\{\tfrac{1}{N}\mathbb{V}'(\gamma)\mathbf{Q}_{nT}\mathbb{V}(\gamma) + \tfrac{1}{N}\mathbb{Z}'(\gamma)\mathbf{Q}_{nT}\mathbb{Z}(\gamma)\\ &- \tfrac{1}{N}\mathbb{Z}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)[\tfrac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]^{-1}\tfrac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{Z}(\gamma)\}(\lambda_0-\lambda) + o_p(1).\end{aligned} \tag{B.5}$$

These imply that

$$\hat{\beta}_{nT}(\hat{\lambda}_{nT},\gamma) = \beta_0 + o_p(1) \text{ and } \hat{\sigma}^2_{nT}(\hat{\lambda}_{nT},\gamma) = \sigma_0^2 + o_p(1), \tag{B.6}$$

uniformly in $\gamma \in \Gamma$, as long as $\hat{\lambda}_{nT} = \lambda_0 + o_p(1)$. That is, to show the consistency of $\hat{\theta}_{nT}$, we only need to show that of $\hat{\lambda}_{nT}$. By Theorem 2.5 of Newey and McFadden (1994), the consistency of $\hat{\lambda}_{nT}$ follows if

(a) $\sup_{(\lambda,\gamma)\in\Lambda\times\Gamma} \frac{1}{N}|\ell^{*c}_{nT}(\lambda,\gamma) - \bar{\ell}^{*c}_{nT}(\lambda,\gamma)| = o_p(1)$,

(b) $\lim_{nT\to\infty} \frac{1}{N}\bar{\ell}^{*c}_{nT}(\lambda,\gamma)$ is uniformly equicontinuous in λ for any γ,

(c) λ_0 uniquely maximizes $\lim_{nT\to\infty} \frac{1}{N}\bar{\ell}^{*c}_{nT}(\lambda,\gamma)$ over $(\lambda,\gamma)\in\Lambda\times\Gamma$.

Proof of (a): For simplicity, we establish $\sup_{(\lambda,\gamma)\in\Lambda\times\Gamma} \frac{1}{nT}|\ell^{*c}_{nT}(\lambda,\gamma) - \bar{\ell}^{*c}_{nT}(\lambda,\gamma)| = o_p(1)$ instead, as $\frac{N}{nT} = O(1)$. Note from (2.11) and (2.14),

$$\tfrac{1}{nT}[\ell^{*c}_{nT}(\lambda,\gamma) - \bar{\ell}^{*c}_{nT}(\lambda,\gamma)] = -\tfrac{1}{2}[\ln\hat{\sigma}^2_{nT}(\lambda,\gamma) - \ln\bar{\sigma}^2_{nT}(\lambda,\gamma)] + \tfrac{1}{nT}[\ln|\mathbf{A}(\lambda,\gamma)| - \mathrm{E}(\ln|\mathbf{A}(\lambda,\gamma)|)].$$

For the second term, Lemma A.4 implies that $\sup_{\gamma\in\Gamma}\frac{1}{nT}[\ln|\mathbf{A}(\lambda,\gamma)| - \mathrm{E}(\ln|\mathbf{A}(\lambda,\gamma)|)] = o_p(1)$ for any given λ. Also, for a given γ, it is $o_p(1)$ for each λ and uniformly equicontinuous in λ (see the proof of (b)). Hence, $\sup_{(\lambda,\gamma)\in\Lambda\times\Gamma}\frac{1}{nT}[\ln|\mathbf{A}(\lambda,\gamma)| - \mathrm{E}(\ln|\mathbf{A}(\lambda,\gamma)|)] = o_p(1)$.

For the first term, **if**, uniformly in $(\lambda,\gamma)\in\Lambda\times\Gamma$, $\bar{\sigma}^2_{nT}(\lambda,\gamma) > c > 0$ and $\hat{\sigma}^2_{nT}(\lambda,\gamma) - \bar{\sigma}^2_{nT}(\lambda,\gamma) = o_p(1)$, then $\ln\hat{\sigma}^2_{nT}(\lambda,\gamma) - \ln\bar{\sigma}^2_{nT}(\lambda,\gamma) = \ln[1 + \bar{\sigma}^{-2}_{nT}(\lambda,\gamma)(\hat{\sigma}^2_{nT}(\lambda,\gamma) - \bar{\sigma}^2_{nT}(\lambda,\gamma))] = o_p(1)$ uniformly in $(\lambda,\gamma)\in\Lambda\times\Gamma$. From (B.1), rewrite $\mathbf{A}(\lambda,\gamma)\mathbf{Y} = \mathbb{X}(\gamma)\beta_0 + (\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})\phi^\dagger + \mathbf{C}\psi_0 + \mathbf{A}(\lambda,\gamma)\mathbf{A}^{-1}\mathbf{V}$, where $\phi^\dagger = ((\lambda_0-\lambda)', \theta_{20}')'$. Then by (2.13),

$$\begin{aligned}\bar{\sigma}^2_{nT}(\lambda,\gamma) = \mathrm{E}\sigma^2_{nT}(\lambda,\gamma) &+ \phi^{\dagger\prime}\{\mathrm{E}[\tfrac{1}{N}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})'\mathbf{Q}_{nT}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})]\\ &- \mathrm{E}[\tfrac{1}{N}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)'\mathbf{Q}_{nT}\mathbf{H})\mathbb{X}(\gamma)]\mathrm{E}[\tfrac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]^{-1}\mathrm{E}[\tfrac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}(\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H})]\}\phi^\dagger,\end{aligned} \tag{B.7}$$

where $\sigma^2_{nT}(\lambda,\gamma)$ is given above Assumption G. The remaining term in (B.7) is non-negative as the quantity in the curly brackets is a Schur complement of $\frac{1}{N}\mathrm{E}[\mathcal{H}(\gamma)\mathbf{Q}_{nT}\mathcal{H}(\gamma)]$ and thus must be positive semi-definite (p.s.d.), where $\mathcal{H}(\gamma) = [\mathbb{H}(\gamma),\ \mathbf{D}(\gamma_0,\gamma)\mathbf{H}]$.

For the first term in (B.7), we have,

$$\begin{aligned}\sigma^2_{nT}(\lambda,\gamma) &= \tfrac{\sigma_0^2}{(n-1)T}\textstyle\sum_{t=1}^T \mathtt{tr}(A_t'^{-1}A_t'(\lambda,\gamma)A_t(\lambda,\gamma)A_t^{-1}) - \tfrac{\sigma_0^2}{n(n-1)T}\textstyle\sum_{t=1}^T l_n' A_t(\lambda,\gamma)A_t^{-1}A_t'^{-1}A_t'(\lambda,\gamma)l_n \\ &= \ddot\sigma^2_{nT}(\lambda,\gamma) + O_p(\tfrac{1}{n}),\end{aligned}$$

where $\ddot\sigma^2_{nT}(\lambda,\gamma) = \frac{\sigma_0^2}{nT}\sum_{t=1}^T \mathtt{tr}(A_t'^{-1}A_t'(\lambda,\gamma)A_t(\lambda,\gamma)A_t^{-1})$. Note that

$$\begin{aligned}&-\tfrac{1}{2}[\ln\ddot\sigma^2_{nT}(\lambda,\gamma) - \ln\sigma_0^2] + \tfrac{1}{nT}[\ln|\mathbf{A}(\lambda,\gamma)| - \ln|\mathbf{A}|] \\ =\ &-\tfrac{1}{2}\big[\ln(\tfrac{1}{nT}\textstyle\sum_{t=1}^T \mathtt{tr}(A_t'^{-1}A_t'(\lambda,\gamma)A_t(\lambda,\gamma)A_t^{-1})) - \ln(\prod_{t=1}^T |A_t'^{-1}A_t'(\lambda,\gamma)A_t(\lambda,\gamma)A_t^{-1}|)^{\frac{1}{nT}}\big] \leqslant 0,\end{aligned}$$

due to the fact that arithmetic mean is no less than geometric means. As $\sigma^2_{nT}(\lambda,\gamma) = \ddot\sigma^2_{nT}(\lambda,\gamma) + O_p(\frac{1}{n})$, the above inequality implies

$$-\tfrac{1}{2}\ln\sigma^2_{nT}(\lambda,\gamma) \leqslant -\tfrac{1}{2}\ln\sigma_0^2 - \tfrac{1}{nT}[\ln|\mathbf{A}(\lambda,\gamma)| - \ln|\mathbf{A}|] + O_p(\tfrac{1}{n}) = O_p(1). \tag{B.8}$$

Hence, we conclude that $\sigma^2_{nT}(\lambda,\gamma)$ is bounded away from zero on $\Lambda\times\Gamma$, and so is $\bar\sigma^2_{nT}(\lambda,\gamma)$.

Thus, it is left to show $\hat\sigma^2_{nT}(\lambda,\gamma) - \bar\sigma^2_{nT}(\lambda,\gamma) = o_p(1)$, uniformly in $(\lambda,\gamma)\in\Lambda\times\Gamma$. Firstly, using $\mathbf{A}(\lambda,\gamma)\mathbf{A}^{-1} = I_n + (\lambda_{10}-\lambda_1)\mathbf{G} + (\lambda_{20}-\lambda_2)\mathbf{D}(\gamma)\mathbf{G} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}$, we have

$$\sigma^2_{nT}(\lambda,\gamma) = \sigma_0^2 + 2\sigma_0^2\mathcal{G}'_{1,nT}(\gamma)\lambda^\dagger + \sigma_0^2\lambda^{\dagger\prime}\mathbb{G}_{1,nT}(\gamma)\lambda^\dagger, \tag{B.9}$$

where $\lambda^\dagger = ((\lambda_0-\lambda)', \lambda_{20})'$, $\mathcal{G}_{1,nT}(\gamma) = \frac{1}{N}[\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{G}), \mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}), \mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G})]'$ and

$$\mathbb{G}_{1,nT}(\gamma) = \frac{1}{N}\begin{bmatrix}\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{G}\mathbf{G}'), & \mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}\mathbf{G}'), & \mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{G}'), \\ \sim, & \mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}\mathbf{G}'\mathbf{D}(\gamma)), & \mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{G}'\mathbf{D}(\gamma)), \\ \sim, & \sim, & \mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{G}'\mathbf{D}(\gamma_0,\gamma)),\end{bmatrix}.$$

By plugging (B.9) into (B.7) and using the fact that the elements of $\mathrm{E}[\mathcal{G}_{1,nT}(\gamma)]$ and $\mathrm{E}[\mathbb{G}_{1,nT}(\gamma)]$ are uniformly bounded on Γ by Assumption C and D and $\theta_{20} = O((nT)^{-\tau})$ by Assumption F, we have , corresponding to (B.5),

$$\begin{aligned}\bar\sigma^2_{nT}(\lambda,\gamma) &= \sigma_0^2 + 2\sigma_0^2\mathrm{E}[\mathcal{G}^{[2]\prime}_{1,nT}(\gamma)](\lambda_0-\lambda) + (\lambda_0-\lambda)'\big\{\sigma_0^2\mathrm{E}[\mathbb{G}^{[2]}_{1,nT}(\gamma)] + \mathrm{E}[\tfrac{1}{N}\mathbb{Z}'(\gamma)\mathbf{Q}_{nT}\mathbb{Z}(\gamma)] \\ &- \mathrm{E}[\tfrac{1}{N}\mathbb{Z}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]\mathrm{E}[\tfrac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]^{-1}\mathrm{E}[\tfrac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{Z}(\gamma)]\big\}(\lambda_0-\lambda) + o(1).\end{aligned} \tag{B.10}$$

We see that $\mathrm{E}[\frac{1}{N}\mathbf{V}'\mathbf{Q}_{nT}\mathbb{V}(\gamma)] = \sigma_0^2\mathrm{E}[\mathcal{G}^{[2]\prime}_{1,nT}(\gamma)]$ and $\mathrm{E}[\frac{1}{N}\mathbb{V}'(\gamma)\mathbf{Q}_{nT}\mathbb{V}(\gamma)] = \sigma_0^2\mathrm{E}[\mathbb{G}^{[2]}_{1,nT}(\gamma)]$. Thus, Lemma A.4 implies $\frac{1}{N}\mathbf{V}'\mathbf{Q}_{nT}\mathbb{V}(\gamma) - \sigma_0^2\mathrm{E}[\mathcal{G}^{[2]\prime}_{1,nT}(\gamma)] \xrightarrow{a.s.} 0$, $\frac{1}{N}\mathbb{V}'(\gamma)\mathbf{Q}_{nT}\mathbb{V}(\gamma) - \sigma_0^2\mathrm{E}[\mathbb{G}^{[2]}_{1,nT}(\gamma)] \xrightarrow{a.s.} 0$, and $\frac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)$, $\frac{1}{N}\mathbb{Z}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)$ and $\frac{1}{N}\mathbb{Z}'(\gamma)\mathbf{Q}_{nT}\mathbb{Z}(\gamma)$ converge to their expectations almost surely, uniformly in $\gamma\in\Gamma$. The convergence of $\hat\sigma^2_{nT}(\lambda,\gamma)$ to $\bar\sigma^2_{nT}(\lambda,\gamma)$ is also uniform on Λ because λ appears simply as linear or quadratic factors in these terms. Therefore, we have $\hat\sigma^2_{nT}(\lambda,\gamma) - \bar\sigma^2_{nT}(\lambda,\gamma) = o_p(1)$, uniformly in $(\lambda,\gamma)\in\Lambda\times\Gamma$.

Proof of (b): Recall from (2.14),

$$\tfrac{1}{N}\bar{\ell}^{*c}_{nT}(\lambda,\gamma) = -\tfrac{c_{nT}}{2}(\ln 2\pi + 1) - \tfrac{c_{nT}}{2}\ln\bar{\sigma}^2_{nT}(\lambda,\gamma) + \tfrac{1}{N}\mathrm{E}(\ln|\mathbf{A}(\lambda,\gamma)|).$$

From (B.10), we see that the limit of $\bar{\sigma}^2_{nT}(\lambda,\gamma)$ is uniformly equicontinuous on Λ given γ, as its terms are linear or quadratic in λ with the corresponding vector or matrices being bounded. To see the uniform equicontinuity of $\frac{1}{N}\mathrm{E}(\ln|\mathbf{A}(\lambda,\gamma)|)$ on Λ, a Taylor expansion around λ_0 gives,

$$\tfrac{1}{N}\mathrm{E}(\ln|\mathbf{A}(\lambda,\gamma)|) = \tfrac{1}{N}\mathrm{E}[(\ln|\mathbf{A}(\lambda_0,\gamma)|) + \mathtt{tr}(\mathbf{G}(\dot{\lambda},\gamma))(\lambda_1-\lambda_{10}) + \mathtt{tr}(\mathbf{D}(\gamma)\mathbf{G}(\dot{\lambda},\gamma))(\lambda_2-\lambda_{20})],$$

where $\dot{\lambda}$ lies between λ and λ_0. As $\frac{1}{N}\mathtt{tr}(\mathbf{G}(\dot{\lambda},\gamma))$ and $\frac{1}{N}\mathtt{tr}(\mathbf{D}(\gamma)\mathbf{G}(\dot{\lambda},\gamma))$ are uniformly bounded by Assumptions C and D for any $\dot{\lambda}$ and γ, we conclude $\frac{1}{N}\mathrm{E}(\ln|\mathbf{A}(\lambda,\gamma)|)$ is also uniformly equicontinuous on Λ for any γ.

Proof of (c): Again, for simplicity, we show that λ_0 uniquely maximizes $\lim_{nT\to\infty}\frac{1}{nT}\bar{\ell}^{*c}_{nT}(\lambda,\gamma)$ over $(\lambda,\gamma)\in\Lambda\times\Gamma$. Letting $\check{\sigma}^2_{nT}(\lambda,\gamma) = \mathrm{E}\sigma^2_{nT}(\lambda,\gamma)$, we have

$$\begin{aligned}
&\tfrac{1}{nT}[\bar{\ell}^{*c}_{nT}(\lambda,\gamma) - \bar{\ell}^{*c}_{nT}(\lambda_0,\gamma_0)]\\
=\ &-\tfrac{1}{2}[\ln\bar{\sigma}^2_{nT}(\lambda,\gamma) - \ln\check{\sigma}^2_{nT}(\lambda,\gamma)] - \tfrac{1}{2}[\ln\check{\sigma}^2_{nT}(\lambda,\gamma) - \ln\bar{\sigma}^2_{nT}(\lambda_0,\gamma)] - \tfrac{1}{2}[\ln\bar{\sigma}^2_{nT}(\lambda_0,\gamma)\\
&-\ln\bar{\sigma}^2_{nT}(\lambda_0,\gamma_0)] + \tfrac{1}{nT}\mathrm{E}[\ln|\mathbf{A}(\lambda,\gamma)| - \ln|\mathbf{A}(\lambda_0,\gamma)|] + \tfrac{1}{nT}\mathrm{E}[\ln|\mathbf{A}(\lambda_0,\gamma)| - \ln|\mathbf{A}|]\\
=\ &-\tfrac{1}{2}[\ln\bar{\sigma}^2_{nT}(\lambda,\gamma) - \ln\check{\sigma}^2_{nT}(\lambda,\gamma)] - \tfrac{1}{2}[\ln\check{\sigma}^2_{nT}(\lambda,\gamma) - \ln\bar{\sigma}^2_{nT}(\lambda_0,\gamma)]\\
&+\tfrac{1}{nT}\mathrm{E}[\ln|\mathbf{A}(\lambda,\gamma)| - \ln|\mathbf{A}(\lambda_0,\gamma)|] + o(1),
\end{aligned}$$

where the last equation holds as $\bar{\sigma}^2_{nT}(\lambda_0,\gamma) - \bar{\sigma}^2_{nT}(\lambda_0,\gamma_0) = o(1)$ by (B.10) and $\frac{1}{nT}\mathrm{E}[\ln|\mathbf{A}(\lambda_0,\gamma)| - \ln|\mathbf{A}|] = \frac{1}{nT}\mathrm{E}[\ln|\mathbf{A}(\lambda_0,\gamma)\mathbf{A}^{-1}|] = \frac{1}{nT}\mathrm{E}[\ln|I_{nT} + \lambda_{20}\mathbf{d}(\gamma_0,\gamma)\mathbf{G}|] = o(1)$ by $\lambda_{20} = O((nT)^{-\tau})$. Thus, it amounts to showing the last three terms are always negative for $\lambda\neq\lambda_0$ given any γ.

For the first term, noting that $\sigma^2_{nT}(\lambda,\gamma) - \check{\sigma}^2_{nT}(\lambda,\gamma) = o_p(1)$, uniformly in $(\lambda,\gamma)\in\Lambda\times\Gamma$, and using the results in (a), we have

$$\begin{aligned}
&-\tfrac{1}{2}[\ln\bar{\sigma}^2_{nT}(\lambda,\gamma) - \ln\check{\sigma}^2_{nT}(\lambda,\gamma)]\\
=&-\tfrac{1}{2}\ln[1 + \tfrac{1}{N\sigma^2_{nT}(\lambda,\gamma)}(\lambda_0-\lambda)'\mathbb{Z}'(\gamma)\mathbb{M}(\gamma)\mathbb{Z}(\gamma)(\lambda_0-\lambda)] + o_p(1).
\end{aligned}$$

We note that $\frac{1}{N}\mathbb{Z}'(\gamma)\mathbb{M}(\gamma)\mathbb{Z}(\gamma)$ is the Schur complement of $\frac{1}{N}[\mathbb{H}'(\gamma)\mathbf{Q}_{nT}\mathbb{H}(\gamma)]$ so that it is p.s.d.. Thus, the limit of the above equation is non-positive.

For the second and third terms, using $\check{\sigma}^2_{nT}(\lambda,\gamma) - \sigma^2_{nT}(\lambda,\gamma) = o_p(1)$ and $\bar{\sigma}^2_{nT}(\lambda_0,\gamma) - \sigma^2_{nT}(\lambda_0,\gamma) = o_p(1)$, shown in the proof of (a), we have

$$\begin{aligned}
&-\tfrac{1}{2}[\ln\check{\sigma}^2_{nT}(\lambda,\gamma) - \ln\bar{\sigma}^2_{nT}(\lambda_0,\gamma)] + \tfrac{1}{nT}[\mathrm{E}(\ln|\mathbf{A}(\lambda,\gamma)|) - \mathrm{E}(\ln|\mathbf{A}(\lambda_0,\gamma)|)]\\
=&-\tfrac{1}{2}[\ln\sigma^2_{nT}(\lambda,\gamma) - \ln\sigma^2_{nT}(\lambda_0,\gamma)] + \tfrac{1}{nT}[\ln|\mathbf{A}(\lambda,\gamma)| - \ln|\mathbf{A}(\lambda_0,\gamma)|] + o_p(1),
\end{aligned}$$

the limit of which is also non-positive, implied by (B.8) as $\sigma^2_{nT}(\lambda_0,\gamma) = \sigma^2_0 + o_p(1)$ by (B.9) and

$\frac{1}{nT}[\ln|\mathbf{A}(\lambda_0,\gamma)| - \ln|\mathbf{A}|] = \frac{1}{nT}[\ln|I_{nT} + \lambda_{20}\mathbf{d}(\gamma_0,\gamma)\mathbf{G}|] = o_p(1)$. Together, we have

$$\begin{aligned}
&\tfrac{1}{nT}[\bar\ell^{*c}_{nT}(\lambda,\gamma) - \bar\ell^{*c}_{nT}(\lambda_0,\gamma_0)] \\
= &-\tfrac{1}{2}\ln[1 + \tfrac{1}{N\sigma^2_{nT}(\lambda,\gamma)}(\lambda_0-\lambda)'\mathbb{Z}'(\gamma)\mathbb{M}(\gamma)\mathbb{Z}(\gamma)(\lambda_0-\lambda)] \\
&-\tfrac{1}{2}[\ln\sigma^2_{nT}(\lambda,\gamma) - \ln\sigma^2_{nT}(\lambda_0,\gamma)] + \tfrac{1}{nT}[\ln|\mathbf{A}(\lambda,\gamma)| - \ln|\mathbf{A}(\lambda_0,\gamma)|] + o_p(1).
\end{aligned}$$

As discussed above, we have $\lim_{nT\to\infty}\frac{1}{nT}[\bar\ell^{*c}_{nT}(\lambda,\gamma) - \bar\ell^{*c}_{nT}(\lambda_0,\gamma_0)] \leqslant 0$. From the matrix partition formula, $\lim_{nT\to\infty}\frac{1}{N}\mathbb{H}'(\gamma)\mathbf{Q}_{nT}\mathbb{H}(\gamma)$ is non-singular if and only if $\lim_{nT\to\infty}\frac{1}{N}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)$ and $\lim_{nT\to\infty}\frac{1}{N}\mathbb{Z}'(\gamma)\mathbb{M}(\gamma)\mathbb{Z}(\gamma)$ are non-singular. Hence, if Assumption G(i) holds, then we have $\lim_{nT\to\infty}\frac{1}{N}\mathbb{Z}'(\gamma)\mathbb{M}(\gamma)\mathbb{Z}(\gamma)$ is positive definite (p.d.) and the limit of $\frac{1}{nT}[\bar\ell^{*c}_{nT}(\lambda,\gamma) - \bar\ell^{*c}_{nT}(\lambda_0,\gamma_0)]$ is strictly less than zero unless $\lambda = \lambda_0$, i.e., λ_0 is the unique maximizer of $\frac{1}{nT}\bar\ell^{*c}_{nT}(\lambda,\gamma)$. If Assumption G($i$) fails, identification requires that the limit of $-\frac{1}{2}[\ln\sigma^2_{nT}(\lambda,\gamma) - \ln\sigma^2_{nT}(\lambda_0,\gamma)] + \frac{1}{nT}[(\ln|\mathbf{A}(\lambda,\gamma)|) - (\ln|\mathbf{A}(\lambda_0,\gamma)|)]$ is strictly less than zero for any γ and $\lambda \neq \lambda_0$, which is equivalent to Assumption G(ii). ∎

Proof of Theorem 2.2: We show the consistency of $\hat\gamma_{nT}$ in two steps:

(a) We derive a preliminary convergence rate for $\hat\theta_{nT}$, $(nT)^\tau(\hat\theta_{nT} - \theta_0) = O_p(1)$;

(b) Based on the convergence rate, we then establish the consistency of $\hat\gamma_{nT}$.

Proof of (a): Applying the mean value theorem (MVT) to each element of $S^*_{\theta,nT}(\hat\theta_{nT},\hat\gamma_{nT})$ where $S^*_{\theta,nT}(\theta,\gamma)$ is given in (2.7), we have

$$0 = S^*_{\theta,nT}(\hat\theta_{nT},\hat\gamma_{nT}) = S^*_{\theta,nT}(\theta_0,\hat\gamma_{nT}) + \big[\tfrac{\partial}{\partial\theta'}S^*_{\theta,nT}(\theta,\hat\gamma_{nT})\big|_{\theta=\bar\theta_r \text{ in } r\text{th row}}\big](\hat\theta_{nT} - \theta_0),$$

where $\{\bar\theta_r\}$ are on the line segment between $\hat\theta_{nT}$ and θ_0. In the following, we use $\bar\theta$ to denote $\{\bar\theta_r\}$ and $H^*_{nT}(\bar\theta,\gamma)$ to denote $-\frac{\partial}{\partial\theta'}S^*_{\theta,nT}(\theta,\gamma)\big|_{\theta=\bar\theta_r \text{ in } r\text{th row}}$ for simplicity. Thus, we have

$$(nT)^\tau(\hat\theta_{nT} - \theta_0) = \big[\tfrac{1}{nT}H^*_{nT}(\bar\theta,\hat\gamma_{nT})\big]^{-1}\tfrac{(nT)^\tau}{nT}S^*_{\theta,nT}(\theta_0,\hat\gamma_{nT}). \tag{B.11}$$

Therefore, the proof of the result in (a) is equivalent to showing for any given γ,

(i) $\frac{1}{nT}[H^*_{nT}(\bar\theta,\gamma) - H^*_{nT}(\theta_0,\gamma)] = o_p(1)$,

(ii) $\frac{1}{nT}[H^*_{nT}(\theta_0,\gamma) - \mathrm{E}(H^*_{nT}(\theta_0,\gamma))] = o_p(1)$,

(iii) The limit of $\frac{1}{nT}\mathrm{E}[H^*_{nT}(\theta_0,\gamma)]$ is non-singular,

(iv) $\frac{(nT)^\tau}{nT}S^*_{\theta,nT}(\theta_0,\gamma) = O_p(1)$.

The Hessian matrix $H^*_{nT}(\theta,\gamma)$ has the following components:

$H^*_{\beta\theta} = \frac{c_{nT}}{\sigma^2}[\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma),\ \mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbf{WY},\ \mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{WY},\ \frac{1}{\sigma^2}\mathbb{X}'(\gamma)\tilde{\mathbf{V}}(\phi,\gamma)]$,
$H^*_{\lambda_1\lambda_1} = \frac{c_{nT}}{\sigma^2}\mathbf{Y}'\mathbf{W}'\mathbf{Q}_{nT}\mathbf{WY} + \mathtt{tr}(\mathbf{G}^2(\lambda,\gamma))$, $H^*_{\lambda_1\lambda_2} = \frac{c_{nT}}{\sigma^2}\mathbf{Y}'\mathbf{W}'\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{WY} + \mathtt{tr}(\mathbf{D}(\gamma)\mathbf{G}^2(\lambda,\gamma))$,
$H^*_{\lambda_1\sigma^2} = \frac{c_{nT}}{\sigma^4}\mathbf{Y}'\mathbf{W}'\tilde{\mathbf{V}}(\phi,\gamma)$, $H^*_{\lambda_2\lambda_2} = \frac{c_{nT}}{\sigma^2}\mathbf{Y}'\mathbf{W}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{WY} + \mathtt{tr}[(\mathbf{D}(\gamma)\mathbf{G}(\lambda,\gamma))^2]$,

$H^*_{\lambda_2\sigma^2} = \frac{c_{nT}}{\sigma^4}\mathbf{Y}'\mathbf{W}'\mathbf{D}(\gamma)\tilde{\mathbf{V}}(\phi,\gamma)$, $H^*_{\sigma^2\sigma^2} = \frac{c_{nT}}{2\sigma^6}[2\tilde{\mathbf{V}}'(\phi,\gamma)\tilde{\mathbf{V}}(\phi,\gamma) - N\sigma^2]$.

To prove (i), we note that $\mathbf{WY} = \mathbf{Z} + \mathbf{GV}$ and $\tilde{\mathbf{V}}(\phi,\gamma) = \mathbf{Q}_{nT}[\mathbf{A}(\lambda,\gamma)\mathbf{Y} - \mathbb{X}(\gamma)\beta] = \mathbf{Q}_{nT}[\mathbb{X}(\gamma)(\beta_0 - \beta) + (\mathbb{Z}(\gamma) + \mathbb{V}(\gamma))(\lambda_0 - \lambda) + \mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV} + \mathbf{V}]$ by (B.1). Hence, for any given γ, $\frac{1}{nT}H^*_{nT}(\bar{\theta},\gamma) = O_p(1)$ by Lemma A.1 and A.2. As $\hat{\theta}_{nT} - \theta_0 \xrightarrow{p} 0$, we have $\bar{\theta} - \theta_0 = o_p(1)$. Noting that σ^{-p} appears in $H^*_{nT}(\theta)$ multiplicatively for $p = 2, 4, 6$ and $\bar{\sigma}^{-p} = \sigma_0^{-p} + o_p(1)$, we have $\frac{1}{nT}H^*_{nT}(\bar{\theta},\gamma) = \frac{1}{nT}H^*_{nT}(\bar{\phi},\sigma_0^2,\gamma) + o_p(1)$. Thus, it is equivalent to showing $\frac{1}{nT}[H^*_{nT}(\bar{\phi},\sigma_0^2,\gamma) - H^*_{nT}(\theta_0,\gamma)] \xrightarrow{p} 0$. As proof for each component in $H^*_{nT}(\bar{\phi},\sigma_0^2,\gamma)$ is similar, we only show one of them for example,

$$\frac{1}{nT}[H^*_{\lambda_1\sigma^2}(\bar{\phi},\sigma_0^2,\gamma) - H^*_{\lambda_1\sigma^2}(\theta_0,\gamma)] = \frac{1}{N\sigma_0^4}\mathbf{Y}'\mathbf{W}'[\tilde{\mathbf{V}}(\bar{\phi},\gamma) - \tilde{\mathbf{V}}(\phi_0,\gamma)]$$
$$= -\frac{1}{N\sigma_0^4}(\mathbf{Z} + \mathbf{GV})'\mathbf{Q}_{nT}[\mathbb{X}(\gamma)(\bar{\beta} - \beta_0) + (\mathbb{Z}(\gamma) + \mathbb{V}(\gamma))(\bar{\lambda} - \lambda_0)] = o_p(1),$$

by Lemmas A.1 and A.2, and $\bar{\theta} - \theta_0 = o_p(1)$.

To prove (ii), we note that $\tilde{\mathbf{V}}(\phi_0,\gamma) = \mathbf{Q}_{nT}\mathbf{V} + \mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}$ by (B.1). Hence, $\frac{1}{nT}[H^*_{nT}(\theta_0,\gamma) - \mathrm{E}(H^*_{nT}(\theta_0,\gamma))] = o_p(1)$ follows directly from Lemma A.4 and $\theta_{20} = O((nT)^{-\alpha})$.

To prove (iii), using the facts that $\lambda_{20} = O((nT)^{-\tau})$ and the elements of $\mathbf{G}(\lambda_0,\gamma)\mathbf{D}(\gamma,\gamma_0)\mathbf{G}$ are uniformly bounded, we have $\frac{1}{nT}\mathtt{tr}[\mathbf{G}(\lambda_0,\gamma) - \mathbf{G}] = \frac{1}{nT}\mathtt{tr}[\mathbf{G}(\lambda_0,\gamma)(I_{nT} - \mathbf{A}(\lambda_0,\gamma)\mathbf{A}^{-1})] = \frac{\lambda_{20}}{nT}\mathtt{tr}[\mathbf{G}(\lambda_0,\gamma)\mathbf{D}(\gamma,\gamma_0)\mathbf{G}] = O_p((nT)^{-\tau})$. Meanwhile, as $\mathbf{Q}_{nT} = (I_T - \frac{l_T l_T'}{T}) \otimes (I_n - \frac{l_n l_n'}{n})$, we have $\frac{1}{N}\mathtt{tr}(\mathbf{Q}_{nT}\Pi) - \frac{1}{nT}\mathtt{tr}(\Pi) = O_p(\frac{1}{n})$ for $\Pi = \mathbf{GG}'$, $\mathbf{D}(\gamma)\mathbf{GG}'$ or $\mathbf{D}(\gamma)\mathbf{GG}'\mathbf{D}'(\gamma)$. Thus, one shows that $\frac{1}{nT}\mathrm{E}[H^*_{nT}(\theta_0,\gamma)] - \Sigma_{nT}(\theta_0,\gamma) = o(1)$ for any γ, where $\Sigma_{nT}(\theta_0,\gamma_0)$ is in (2.20).

Therefore, it amounts to proving the limit of $\Sigma_{nT}(\theta_0,\gamma)$ is nonsingular on Γ, which follows if $\Sigma_{nT}(\theta_0,\gamma)p = 0$ implies $p = 0$, where $p = (p_1', p_2', p_3)$, p_1 is a $2k \times 1$ vector, p_2 a 2×1 vector, and p_3 a scalar. The first row block of the linear equation system $\Sigma_{nT}(\theta_0,\gamma)p = 0$ implies $p_1 = -[\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{X}(\gamma)]^{-1}[\mathbb{X}'(\gamma)\mathbf{Q}_{nT}\mathbb{Z}(\gamma)]p_2$, while the last row shows $p_3 = -2\sigma_0^2\mathcal{S}'_{nT}(\gamma)p_2$. Substituting them into the remaining equation of the linear system gives us

$$[\tfrac{1}{N\sigma_0^2}\mathbb{Z}'(\gamma)\mathbb{M}(\gamma)\mathbb{Z}(\gamma) + \tfrac{1}{2nT}\mathbb{C}^{[2]}(\gamma)]p_2 = 0,$$

where $\mathbb{C}^{[2]}(\gamma)$ is the submatrix of $\mathbb{C}(\gamma)$ by deleting its third row and column. As shown before, the first term in the square bracket is p.s.d.. Also, $\frac{1}{nT}\mathbb{C}^{[2]}(\gamma)$ is p.s.d. because

$$\tfrac{1}{nT}z'\mathbb{C}^{[2]}(\gamma)z = \tfrac{1}{2nT}\mathtt{tr}[(z_1\mathbb{C}_1^s + z_2\mathbb{C}_2^s(\gamma))'(z_1\mathbb{C}_1^s + z_2\mathbb{C}_2^s(\gamma))] \geq 0,$$

for all $z = (z_1, z_2)'$ in $\mathbf{R}^2$. Under Assumption H, either $\frac{1}{N}\mathbb{Z}'(\gamma)\mathbb{M}(\gamma)\mathbb{Z}(\gamma)$ or $\frac{1}{nT}\mathbb{C}^{[2]}(\gamma)$ is strictly p.d.. Therefore, we must have $p_2 = 0$ from the above equation, implying both $p_1 = 0$ and $p_3 = 0$ by the first and last equations of the linear system. Hence, the non-singularity of

$\lim_{nT\to\infty} \Sigma_{nT}(\theta_0,\gamma)$ follows.

To prove (iv), using $\mathbf{WY} = \mathbf{Z} + \mathbf{GV}$ and $\tilde{\mathbf{V}}(\phi_0,\gamma) = \mathbf{Q}_{nT}\mathbf{V} + \mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}$ and with $S^{*u}_{\theta,nT}(\theta_0,\gamma)$ defined in (2.18), we have

$$S^{*}_{\theta,nT}(\theta_0,\gamma) = c_{nT}[S^{*u}_{\theta,nT}(\theta_0,\gamma) + \textstyle\sum_{r=1}^{3} B_{r,nT}(\gamma)] \tag{B.12}$$

where $B_{1,nT}(\gamma) = -\big\{0_{1\times 2k},\ \frac{T-1}{nT}\mathtt{tr}[\bar{\mathbf{G}}(\lambda_0,\gamma)\mathbf{J}],\ \frac{T-1}{nT}\mathtt{tr}[\mathbf{D}(\gamma)\bar{\mathbf{G}}(\lambda_0,\gamma)\mathbf{J}], 0\big\}'$,
$B_{2,nT}(\gamma) = -\big\{0_{1\times 2k},\ \mathtt{tr}(\mathbf{Q}_{nT}(\mathbf{G}(\lambda_0,\gamma)-\mathbf{G})),\ \mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)(\mathbf{G}(\lambda_0,\gamma)-\mathbf{G})), 0\big\}'$, and

$$B_{3,nT}(\gamma) = \begin{cases} \frac{1}{\sigma_0^2}\mathbb{X}'(\gamma)\mathbf{Q}_{nT}[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}], \\ \frac{1}{\sigma_0^2}(\mathbf{Z}+\mathbf{GV})'\mathbf{Q}_{nT}[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}], \\ \frac{1}{\sigma_0^2}[\mathbf{Z}+\mathbf{GV}]'\mathbf{D}(\gamma)\mathbf{Q}_{nT}[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}] \\ \frac{1}{2\sigma_0^4}[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}]'\mathbf{Q}_{nT}[2\mathbf{V} + \mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}]. \end{cases}$$

As $c_{nT} = O(1)$, it is sufficient to show that $(nT)^{\tau-1}S^{*u}_{\theta,nT}(\theta_0,\gamma)$ and $(nT)^{\tau-1}B_{r,nT}(\gamma)$, $r = 1,2,3$ are all bounded for any γ. By Lemma A.3 and Lemma B.1, $S^{*u}_{\theta,nT}(\theta_0,\gamma) = O_p(\sqrt{nT})$, uniformly in $\gamma\in\Gamma$. Since $\tau\in(0,\frac{1}{2})$, $(nT)^{\tau-1}S^{*u}_{\theta,nT}(\theta_0,\gamma) = \frac{(nT)^{\tau-\frac{1}{2}}}{\sqrt{nT}}S^{*u}_{\theta,nT}(\theta_0,\gamma) = o_p(1)$. As for $B_{3,nT}(\gamma)$, note that $\theta_{20} = (nT)^{-\tau}\delta_0$, where $\delta_0 = (b_0', l_0)'$, by Assumption F. Thus, it is easy to see that $(nT)^{\tau-1}B_{3,nT}(\gamma) = O_p(1)$ uniformly in $\gamma\in\Gamma$. We show the third component of $B_{3,nT}(\gamma)$ for example as the others can be shown similarly. By Lemma A.4,

$$\begin{aligned} &\tfrac{(nT)^{\tau-1}}{\sigma_0^2}[\mathbf{Z}+\mathbf{GV}]'\mathbf{D}(\gamma)\mathbf{Q}_{nT}[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20} + \lambda_{20}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}] \\ &= \tfrac{1}{\sigma_0^2 nT}[\mathbf{Z}+\mathbf{GV}]'\mathbf{D}(\gamma)\mathbf{Q}_{nT}[\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\delta_0 + l_0\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}] \\ &= \tfrac{1}{\sigma_0^2 nT}\mathrm{E}[\mathbf{Z}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\delta_0 + \mathbf{V}'\mathbf{G}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{GV}l_0] + o_p(1) = O_p(1). \end{aligned}$$

Similarly, we also have $(nT)^{\tau-1}B_{2,nT}(\gamma) = O_p(1)$ uniformly in $\gamma\in\Gamma$. We show one of the two non-zero elements in $B_{2,nT}(\gamma)$ for example, as the other can be shown similarly. Noting that $\mathbf{G}(\lambda_0,\gamma) - \mathbf{G} = \mathbf{G}(\lambda_0,\gamma)(I_{nT} - \mathbf{A}(\lambda_0,\gamma)\mathbf{A}^{-1}) = \lambda_{20}\mathbf{G}(\lambda_0,\gamma)\mathbf{D}(\gamma,\gamma_0)\mathbf{G}$, one has

$$\begin{aligned} &(nT)^{\tau-1}\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)(\mathbf{G}(\lambda_0,\gamma)-\mathbf{G})) \\ &= \tfrac{l_0}{nT}\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}(\lambda_0,\gamma)\mathbf{D}(\gamma,\gamma_0)\mathbf{G}) = O_p(1). \end{aligned} \tag{B.13}$$

Finally, we show $(nT)^{\tau-1}B_{1,nT}(\gamma) = o_p(1)$. Note that the nonzero elements in $B_{1,nT}(\gamma)$ is either $O(T)$ or $O_p(T)$ so that elements of $(nT)^{\tau-1}B_{1,nT}(\gamma)$ is either $O(\frac{T^\tau}{n^{1-\tau}})$ or $O_p(\frac{T^\tau}{n^{1-\tau}})$. As $\frac{T}{n}\to c<\infty$ and $\tau\in(0,\frac{1}{2})$, $\frac{T^\tau}{n^{1-\tau}} = \frac{c^\tau}{n^{1-2\tau}} = o(1)$. Thus, the desired result holds.

Proof of (b): Note that

$$\tfrac{(nT)^{2\tau}}{nT}[\ell^{*c}_{nT}(\hat{\lambda}_{nT},\hat{\gamma}_{nT}) - \ell^{*c}_{nT}(\lambda_0,\gamma_0)] = -\tfrac{(nT)^{2\tau}}{2}[\ln\hat{\sigma}^2_{nT} - \ln\hat{\sigma}^2_{nT}(\lambda_0,\gamma_0)] + \tfrac{(nT)^{2\tau}}{nT}\ln|\mathbf{A}(\hat{\lambda}_{nT},\hat{\gamma}_{nT})\mathbf{A}^{-1}|.$$

For simplicity, we denote $\hat{\lambda}^\dagger_{nT} = ((\lambda_0-\hat{\lambda}_{nT})', \lambda_{20})'$ and $\hat{\phi}^\dagger_{nT} = ((\lambda_0-\hat{\lambda}_{nT})', \theta_{20}')'$. By (a) and

Assumption F, $\hat{\phi}^{\dagger}_{nT}, \hat{\lambda}^{\dagger}_{nT} = O((nT)^{-\tau})$. Thus, using (B.3) and $\hat{\sigma}^2_{nT} \equiv \hat{\sigma}^2_{nT}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})$, we have

$$\hat{\sigma}^2_{nT} - \hat{\sigma}^2_{nT}(\lambda_0, \gamma_0) = \tfrac{2}{N}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT} + \tfrac{1}{N}\hat{\lambda}^{\dagger\prime}_{nT}\mathbf{K}'_{\mathbf{V}}(\hat{\gamma}_{nT})\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT}$$
$$+ \tfrac{1}{N}\hat{\phi}^{\dagger\prime}_{nT}\mathbf{K}'_{\mathbf{H}}(\hat{\gamma}_{nT})\mathbb{M}(\hat{\gamma}_{nT})\mathbf{K}_{\mathbf{H}}(\hat{\gamma}_{nT})\hat{\phi}^{\dagger}_{nT} + o_p((nT)^{-2\tau}), \tag{B.14}$$

where $\mathbf{K}_{\mathbf{H}}(\gamma) = (\mathbb{Z}(\gamma),\ \mathbf{D}(\gamma_0, \gamma)\mathbf{H})$ and $\mathbf{K}_{\mathbf{V}}(\gamma) = (\mathbb{V}(\gamma), \mathbf{D}(\gamma_0, \gamma)\mathbf{G}\mathbf{V})$.

By Theorem 2.8 of Hall (2015), we have $\ln|\mathbf{A}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})\mathbf{A}^{-1}| = \mathtt{tr}[\ln(\mathbf{A}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})\mathbf{A}^{-1})] = \mathtt{tr}\big[\sum_{m=1}^{\infty}(-1)^{m+1}\frac{(\mathbf{A}(\hat{\lambda}_{nT},\hat{\gamma}_{nT})\mathbf{A}^{-1} - I_{nT})^m}{m}\big]$. By further using $\mathbf{A}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})\mathbf{A}^{-1} = I_n + (\lambda_{10} - \hat{\lambda}_{1,nT})\mathbf{G} + (\lambda_{20} - \hat{\lambda}_{2,nT})\mathbf{D}(\hat{\gamma}_{nT})\mathbf{G} + \lambda_{20}\mathbf{D}(\gamma_0, \hat{\gamma}_{nT})\mathbf{G}$ and $\hat{\lambda}^{\dagger}_{nT} = O((nT)^{-\tau})$, we have

$$\tfrac{1}{nT}\ln|\mathbf{A}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})\mathbf{A}^{-1}| = \mathcal{G}'_{2,nT}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT} - \tfrac{1}{2}\hat{\lambda}^{\dagger\prime}_{nT}\mathbb{G}_{2,nT}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT} + o_p((nT)^{-2\tau}), \tag{B.15}$$

where $\mathcal{G}_{2,nT}(\gamma) = \frac{1}{nT}\big\{\mathtt{tr}(\mathbf{G}), \mathtt{tr}(\mathbf{D}(\gamma)\mathbf{G}), \mathtt{tr}(\mathbf{D}(\gamma_0, \gamma)\mathbf{G})\big\}'$, and

$$\mathbb{G}_{2,nT}(\gamma) = \tfrac{1}{nT}\begin{bmatrix} \mathtt{tr}(\mathbf{G}^2), & \mathtt{tr}(\mathbf{D}(\gamma)\mathbf{G}^2), & \mathtt{tr}(\mathbf{D}(\gamma_0, \gamma)\mathbf{G}^2), \\ \sim, & \mathtt{tr}((\mathbf{D}(\gamma)\mathbf{G})^2), & \mathtt{tr}(\mathbf{D}(\gamma_0, \gamma)\mathbf{G}\mathbf{D}(\gamma)\mathbf{G}), \\ \sim, & \sim, & \mathtt{tr}((\mathbf{D}(\gamma_0, \gamma)\mathbf{G})^2), \end{bmatrix}.$$

Before proceeding with the subsequent derivation, we recall $\mathcal{G}_{1,nT}(\gamma)$ and $\mathbb{G}_{1,nT}(\gamma)$ in (B.9). Note that $\mathcal{G}_{1,nT}(\gamma) = \mathcal{G}_{2,nT}(\gamma) + O_p(\frac{1}{n})$, $(nT)^{\tau}[\mathcal{G}_{1,nT}(\gamma) - \mathcal{G}_{2,nT}(\gamma)] = O_p(\frac{T^{\tau}}{n^{1-\tau}}) = o_p(1)$. Similarly, $\mathbb{G}_{1,nT}(\gamma) = \mathbb{G}_{3,nT}(\gamma) + O_p(\frac{1}{n})$, where

$$\mathbb{G}_{3,nT}(\gamma) = \tfrac{1}{nT}\sum_{t=1}^{T}\begin{bmatrix} \mathtt{tr}(\mathbf{G}\mathbf{G}'), & \mathtt{tr}(\mathbf{D}(\gamma)\mathbf{G}\mathbf{G}'), & \mathtt{tr}(\mathbf{D}(\gamma_0, \gamma)\mathbf{G}\mathbf{G}'), \\ \sim, & \mathtt{tr}(\mathbf{D}(\gamma)\mathbf{G}\mathbf{G}'), & \mathtt{tr}(\mathbf{D}(\gamma_0, \gamma)\mathbf{G}\mathbf{G}'\mathbf{d}(\gamma)), \\ \sim, & \sim, & \mathtt{tr}(\mathbf{D}(\gamma_0, \gamma)\mathbf{G}\mathbf{G}'), \end{bmatrix}.$$

Then, using the Taylor expansion for logarithm and plugging in (B.14) and (B.15), we have

$$\begin{aligned} & -\tfrac{(nT)^{2\tau}}{2}[\ln\hat{\sigma}^2_{nT} - \ln\hat{\sigma}^2_{nT}(\lambda_0, \gamma_0)] + \tfrac{(nT)^{2\tau}}{nT}\ln|\mathbf{A}(\hat{\lambda}_{nT}, \hat{\gamma}_{nT})\mathbf{A}^{-1}| \\ = & -\tfrac{(nT)^{2\tau}}{N\sigma_0^2}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT} - \tfrac{(nT)^{2\tau}}{2N\sigma_0^2}\hat{\lambda}^{\dagger\prime}_{nT}\mathbf{K}'_{\mathbf{V}}(\hat{\gamma}_{nT})\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT} \\ & -\tfrac{(nT)^{2\tau}}{2N\sigma_0^2}\hat{\phi}^{\dagger\prime}_{nT}\mathbf{K}'_{\mathbf{H}}(\hat{\gamma}_{nT})\mathbb{M}(\hat{\gamma}_{nT})\mathbf{K}_{\mathbf{H}}(\hat{\gamma}_{nT})\hat{\phi}^{\dagger}_{nT} + [\tfrac{(nT)^{\tau}}{N\sigma_0^2}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT}]^2 \\ & + (nT)^{2\tau}\mathcal{G}'_{2,nT}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT} - \tfrac{(nT)^{2\tau}}{2}\hat{\lambda}^{\dagger\prime}_{nT}\mathbb{G}_{2,nT}(\hat{\gamma}_{nT})\hat{\lambda}^{\dagger}_{nT} + o_p(1), \end{aligned} \tag{B.16}$$

where we use the facts that $\hat{\sigma}^2_{nT}(\lambda_0, \gamma_0) - \sigma_0^2 = O_p(N^{-1/2})$ and $\hat{\sigma}^2_{nT}(\lambda_0, \gamma_0)$ appears multiplicatively. Consider the combination of the first and fifth terms,

$$\begin{aligned} & -\tfrac{(nT)^{2\tau}}{\sigma_0^2}[\tfrac{1}{N}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\hat{\gamma}_{nT}) - \sigma_0^2\mathcal{G}'_{2,nT}(\hat{\gamma}_{nT})]\hat{\lambda}^{\dagger}_{nT} \\ = & -\tfrac{(nT)^{2\tau}}{\sigma_0^2}[\tfrac{1}{N}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\hat{\gamma}_{nT}) - \sigma_0^2\mathcal{G}'_{1,nT}(\hat{\gamma}_{nT}) + \sigma_0^2(\mathcal{G}'_{1,nT}(\hat{\gamma}_{nT}) - \mathcal{G}'_{2,nT}(\hat{\gamma}_{nT}))]\hat{\lambda}^{\dagger}_{nT} = o_p(1), \end{aligned}$$

as $\frac{1}{N}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\gamma) - \sigma_0^2\mathcal{G}'_{1,nT}(\gamma)$ is $O_p(N^{-1/2})$ uniformly in $\gamma \in \Gamma$, which can be shown similarly to Lemma B.1, $\hat{\lambda}^{\dagger}_{nT} = O_p((nT)^{-\tau})$, and $(nT)^{\tau}[\mathcal{G}_{1,nT}(\gamma) - \mathcal{G}_{2,nT}(\gamma)] = o_p(1)$.

Then, for the second term, we have $\frac{1}{N}\mathbf{K}_{\mathbf{V}}'(\gamma)\mathbf{Q}_{nT}\mathbf{K}_{\mathbf{V}}(\gamma) = \sigma_0^2\mathbb{G}_{1,nT}(\gamma)+o_p(1) = \sigma_0^2\mathbb{G}_{3,nT}(\gamma)+o_p(1)$, uniformly in $\gamma \in \Gamma$. These together lead to

$$\begin{aligned}&\tfrac{(nT)^{2\tau}}{nT}[\ell_{nT}^{*c}(\hat{\lambda}_{nT},\hat{\gamma}_{nT}) - \ell_{nT}^{*c}(\lambda_0,\gamma_0)]\\ =& -\tfrac{(nT)^{2\tau}}{2}\hat{\lambda}_{nT}^{\dagger\prime}[\mathbb{G}_{2,nT}(\hat{\gamma}_{nT}) + \mathbb{G}_{3,nT}(\hat{\gamma}_{nT}) - 2\mathcal{G}_{2,nT}(\hat{\gamma}_{nT})\mathcal{G}_{2,nT}'(\hat{\gamma}_{nT})]\hat{\lambda}_{nT}^{\dagger}\\ &-\tfrac{(nT)^{2\tau}}{2N\sigma_0^2}\hat{\phi}_{nT}^{\dagger\prime}\mathbf{K}_{\mathbf{H}}'(\hat{\gamma}_{nT})\mathbb{M}(\hat{\gamma}_{nT})\mathbf{K}_{\mathbf{H}}(\hat{\gamma}_{nT})\hat{\phi}_{nT}^{\dagger} + o_p(1).\end{aligned} \tag{B.17}$$

First, we note that $\mathbb{G}_{2,nT}(\gamma) + \mathbb{G}_{3,nT}(\gamma) - 2\mathcal{G}_{2,nT}(\gamma)\mathcal{G}_{2,nT}'(\gamma) = \frac{1}{2}\mathbb{C}(\gamma)$, which is p.s.d. as $\frac{1}{2}z'\mathbb{C}(\gamma)z = \frac{1}{2nT}\mathtt{tr}[(z_1\mathbb{C}_1^s + z_2\mathbb{C}_2^s(\gamma) + z_3\mathbb{C}_3^s(\gamma))'(z_1\mathbb{C}_1^s + z_2\mathbb{C}_2^s(\gamma) + z_3\mathbb{C}_3^s(\gamma))] \geq 0$ for all $z = (z_1,z_2,z_3)'$ in $\mathbf{R}^3$. Second, for any comformable vector d, $d'[\frac{1}{N}\mathbf{K}_{\mathbf{H}}'(\gamma)\mathbb{M}(\gamma)\mathbf{K}_{\mathbf{H}}(\gamma)]d$ can be written into the form of $a'Qa$ with some $nT\times 1$ vector a and $nT \times nT$ idempotent matrix Q, so that the second term of (B.17) is also non-positive. Therefore, we have $\lim_{nT\to\infty}\frac{(nT)^{2\tau}}{nT}[\ell_{nT}^{*c}(\hat{\lambda}_{nT},\hat{\gamma}_{nT}) - \ell_{nT}^{*c}(\lambda_0,\gamma_0)] \leqslant 0$. By Theorem 5 of Smith (1992) and under Assumption H(i), we have

$$\rho_{\min}\big(\tfrac{1}{N}\mathbf{K}_{\mathbf{H}}'(\hat{\gamma}_{nT})\mathbb{M}(\hat{\gamma}_{nT})\mathbf{K}_{\mathbf{H}}(\hat{\gamma}_{nT})\big) \geq \rho_{\min}\big(\tfrac{1}{N}\mathcal{H}'(\hat{\gamma}_{nT})\mathbf{Q}_{nT}\mathcal{H}(\hat{\gamma}_{nT})\big) \geq c|\hat{\gamma}_{nT}-\gamma_0|.$$

It follows that

$$\tfrac{(nT)^{2\tau}}{nT}[\ell_{nT}^{*c}(\hat{\lambda}_{nT},\hat{\gamma}_{nT}) - \ell_{nT}^{*c}(\lambda_0,\gamma_0)] \leq -\tfrac{1}{2\sigma_0^2}c|\hat{\gamma}_{nT}-\gamma_0|\|(nT)^{\tau}\hat{\phi}_{nT}^{\dagger}\|^2 + o_p(1).$$

By the definition of $(\hat{\lambda}_{nT},\hat{\gamma}_{nT})$, we have $\frac{(nT)^{2\tau}}{nT}(\ell_{nT}^{*c}(\hat{\lambda}_{nT},\hat{\gamma}_{nT}) - \ell_{nT}^{*c}(\lambda_0,\gamma_0)) \geq 0$. Hence, we must have that $|\hat{\gamma}_{nT}-\gamma_0| = o_p(1)$. Similarly, Assumption H(ii) can also guarantee that $\hat{\gamma}_{nT}-\gamma_0 \xrightarrow{p} 0$. ■

Proof of Theorem 2.3: We first show that $(nT)^{\tau}(\hat{\theta}_{nT}-\theta_0) = o_p(1)$. Given the results ($i$)-($iv$) from the proof of Theorem 2.2, we only need to show that $(nT)^{\tau-1}B_{2,nT}(\lambda_0,\hat{\gamma}_{nT})$ and $(nT)^{\tau-1}B_{3,nT}(\theta_0,\hat{\gamma}_{nT})$ are both $o_p(1)$, which is directly implied by the consistency of $\hat{\gamma}_{nT}$.

Then, let B, k, l and $\mathcal{N}_{nT}$ be defined in Lemma B.4, and $\bar{m} \equiv \max\big(k,l,\|\delta_0\|,|l_0|,1,\sigma_0^2\big)$. Pick $\eta,\iota > 0$ small enough such that $\max(\eta,\iota) < \bar{m}$ and

$$\mathcal{M}_0 \equiv -\tfrac{1}{2}\bar{T}k - \tfrac{k}{\sigma_0^2+\iota} + \tfrac{1}{2}(\bar{m}\eta + 6\bar{m}^3\iota) + \tfrac{1}{\sigma_0^2-\iota}(4\bar{m}\eta + 8\bar{m}^2\eta + 18\bar{m}^3\iota + 4\bar{m}^4\iota) < 0.$$

Let E_{nT} be the joint event such that (1) $|\hat{\gamma}_{nT}-\gamma_0| \leqslant B$, (2) $\frac{T^{\tau}}{n^{1-\tau}} < \iota$, (3) $(nT)^{\tau}|\hat{\theta}_{nT}-\theta_0| \leqslant \iota$, (4) $\inf_{\gamma\in\mathcal{N}_{nT}}\frac{D_{r,nT}(\gamma)}{|\gamma-\gamma_0|} > (1-\eta)k$, (5) $\sup_{\gamma\in\mathcal{N}_{nT}}\frac{\|F_{s,nT}(\gamma)\|}{|\gamma-\gamma_0|} < (1+\eta)l$, (6) $\sup_{\gamma\in\mathcal{N}_{nT}}\frac{|K_{s,nT}(\gamma)|}{|\gamma-\gamma_0|} < (1+\eta)l$, (7) $\sup_{\gamma\in\mathcal{N}_{nT}}\frac{|L_{r,nT}(\gamma)|}{|\gamma-\gamma_0|} < (1+\eta)l$, (8) $\sup_{\gamma\in\mathcal{N}_{nT}}\frac{\|P_{r,nT}(\gamma)\|}{|\gamma-\gamma_0|} < \eta$, (9) $\sup_{\gamma\in\mathcal{N}_{nT}}\frac{\|\mathcal{J}_{s,nT}(\gamma,\gamma_0)\|}{\sqrt{a_{nT}}|\gamma-\gamma_0|} < \eta$, for $s = 1,2$ and $r = 1,2,3$, and (10) will be established later.

To use the result $(nT)^{\tau}(\hat{\theta}_{nT}-\theta_0) = o_p(1)$, we let $\ell_{nT}^{\ddagger}(\gamma) = \ell_{nT}^{*}(\hat{\theta}_{nT},\gamma)$, which must have a unique maximizer, $\gamma = \hat{\gamma}_{nT}$. We have,

$$\frac{1}{c_{nT}}[\ell^{\ddagger}_{nT}(\gamma) - \ell^{\ddagger}_{nT}(\gamma_0)] = \frac{1}{c_{nT}}[\ln|\mathbf{A}(\hat{\lambda}_{nT},\gamma)| - \ln|\mathbf{A}(\hat{\lambda}_{nT},\gamma_0)|]$$
$$- \frac{1}{2\hat{\sigma}^2_{nT}}[\mathbf{V}'(\hat{\phi}_{nT},\gamma)\mathbf{Q}_{nT}\mathbf{V}(\hat{\phi}_{nT},\gamma) - \mathbf{V}'(\hat{\phi}_{nT},\gamma_0)\mathbf{Q}_{nT}\mathbf{V}(\hat{\phi}_{nT},\gamma_0)]. \quad \text{(B.18)}$$

For the first differenced term, we have, by Theorem 2.8 of Hall (2015),

$$\begin{aligned}
&\frac{1}{c_{nT}}[\ln|\mathbf{A}(\hat{\lambda}_{nT},\gamma)| - \ln|\mathbf{A}(\hat{\lambda}_{nT},\gamma_0)|] \\
&= \hat{\lambda}_{2,nT}\bar{T}\mathtt{tr}\big[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}(\hat{\lambda}_{nT},\gamma_0)\big] - \tfrac{1}{2}\hat{\lambda}^2_{2,nT}\bar{T}\mathtt{tr}\big\{[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}(\hat{\lambda}_{nT},\gamma_0)]^2\big\} + R_{em} \\
&= \hat{\lambda}_{2,nT}\bar{T}\mathtt{tr}\big[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\big] - \tfrac{1}{2}\hat{\lambda}^2_{2,nT}\bar{T}\mathtt{tr}\{[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}]^2\} + R_{em}, \\
&= \mathcal{A}_1(\gamma) + \mathcal{A}_2(\gamma) + \mathcal{A}_3(\gamma),
\end{aligned} \quad \text{(B.19)}$$

where $\mathcal{A}_3(\gamma) = R_{em}$ is the remainder term. Noting that $\mathbf{G}(\hat{\lambda}_{nT},\gamma_0) - \mathbf{G} = \mathbf{G}(\hat{\lambda}_{nT},\gamma_0)[(\hat{\lambda}_{1,nT} - \lambda_{10})\mathbf{G} + (\hat{\lambda}_{2,nT} - \lambda_{20})\mathbf{D}(\gamma_0)\mathbf{G}]$, the elements of $[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}(\hat{\lambda}_{nT},\gamma_0)]^r$, $r \geq 1$, are uniformly bounded by Lemma A.1, and $\hat{\lambda}_{2,nT} = O_p((nT)^{-\tau})$, implied by $\hat{\lambda}_{2,nT} - \lambda_{20} = o_p(1)$ and $\lambda_{20} = O((nT)^{-\tau})$, we see $\mathcal{A}_3(\gamma)$ is of smaller order relative to $\mathcal{A}_2(\gamma)$, uniformly in $\gamma \in \Gamma$.

For the second differenced term in (B.18), $\mathbf{Q}_{nT}\mathbf{V}(\hat{\phi}_{nT},\gamma) = \mathbf{Q}_{nT}[\mathbf{A}(\hat{\lambda}_{nT},\gamma)\mathbf{Y} - \mathbb{X}(\gamma)\hat{\beta}_{nT}] = \mathbf{Q}_{nT}[\mathbf{V} + \mathbb{H}(\gamma_0)(\phi_0 - \hat{\phi}_{nT}) + \mathbf{D}(\gamma_0,\gamma)\mathbf{H}\hat{\theta}_{2,nT} + \mathbb{V}(\gamma_0)(\lambda_0 - \hat{\lambda}_{nT}) + \hat{\lambda}_{2,nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}]$. Thus,

$$\begin{aligned}
&\mathbf{V}'(\hat{\phi}_{nT},\gamma)\mathbf{Q}_{nT}\mathbf{V}(\hat{\phi}_{nT},\gamma) - \mathbf{V}'(\hat{\phi}_{nT},\gamma_0)\mathbf{Q}_{nT}\mathbf{V}(\hat{\phi}_{nT},\gamma_0) \\
&= [\mathbf{V}(\hat{\phi}_{nT},\gamma) + \mathbf{V}(\hat{\phi}_{nT},\gamma_0)]'\mathbf{Q}_{nT}[\mathbf{V}(\hat{\phi}_{nT},\gamma) - \mathbf{V}(\hat{\phi}_{nT},\gamma_0)] \equiv \textstyle\sum_{s=1}^9 \mathcal{B}_s(\gamma)
\end{aligned} \quad \text{(B.20)}$$

where $\mathcal{B}_s(\gamma)$, $s = 1,\ldots,9$, respectively, takes the following forms, $2\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\hat{\theta}_{2,nT}$,
$2\hat{\lambda}_{2,nT}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$, $\hat{\theta}'_{2,nT}\mathbf{H}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\hat{\theta}_{2,nT}$,
$2\hat{\lambda}_{2,nT}\hat{\theta}'_{2,nT}\mathbf{H}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$, $2(\phi_0 - \hat{\phi}_{nT})'\mathbb{H}'(\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\hat{\theta}_{2,nT}$,
$2\hat{\lambda}_{2,nT}(\phi_0 - \hat{\phi}_{nT})'\mathbb{H}'(\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$, $2(\lambda_0 - \hat{\lambda}_{nT})'\mathbb{V}'(\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\hat{\theta}_{2,nT}$,
$2\hat{\lambda}_{2,nT}(\lambda_0 - \hat{\lambda}_{nT})'\mathbb{V}'(\gamma_0)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$, and $\hat{\lambda}^2_{2,nT}\mathbf{V}'\mathbf{G}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}$.

From (B.19) and (B.20), we have

$$\frac{\ell^{\ddagger}_{nT}(\gamma) - \ell^{\ddagger}_{nT}(\gamma_0)}{c_{nT}a_{nT}(\gamma-\gamma_0)} \leqslant -\frac{\sum_{s=1}^9 \mathcal{B}_s(\gamma) - 2\hat{\sigma}^2_{nT}\mathcal{A}_1(\gamma)}{2\hat{\sigma}^2_{nT}a_{nT}(\gamma-\gamma_0)} + \frac{\mathcal{A}_2(\gamma)}{a_{nT}(\gamma-\gamma_0)} + \frac{\mathcal{A}_3(\gamma)}{a_{nT}(\gamma-\gamma_0)}. \quad \text{(B.21)}$$

As is shown latter, $\frac{\mathcal{A}_2(\gamma)}{a_{nT}(\gamma-\gamma_0)}$ are uniformly bounded on the set E_{nT}. This implies that $\frac{\mathcal{A}_3(\gamma)}{a_{nT}(\gamma-\gamma_0)}$ will shrink to zero as sample increase. Therefore, we let $\frac{\mathcal{A}_3(\gamma)}{a_{nT}(\gamma-\gamma_0)} \leqslant \iota$ be the event (10) of E_{nT}. Fix $\epsilon > 0$, one can choose $\bar{v}$ for large enough (n,T) such that $P(E_{nT}) \geqslant 1-\epsilon$, by Theorem 2.2, Assumption E, $(nT)^{\tau}(\hat{\theta}_{nT} - \theta_0) = o_p(1)$ shown at the beginning and Lemma B.4. Suppose $\gamma \in [\gamma_0 + \bar{v}/a_{nT}, \gamma_0 + B]$ and E_{nT} holds. Let $\hat{l}_{nT} = (nT)^{\tau}\hat{\lambda}_{2,nT}$ and $\hat{b}_{nT} = (nT)^{\tau}\hat{\beta}_{2,nT}$ so that $\|\hat{\delta}_{nT} - \delta_0\| \leqslant \iota$, where $\hat{\delta}_{nT} = (\hat{l}'_{nT}, \hat{b}_{nT})'$, by event (3). Besides, we have $\sigma_0^2 - \iota \leqslant \sigma_0^2 - \iota(nT)^{-\tau} \leqslant \hat{\sigma}^2_{nT} \leqslant \sigma_0^2 + \iota(nT)^{-\tau} \leqslant \sigma_0^2 + \iota$. Given these, we are going to study each term on the right-hand side of inequality (B.21).

By events (1), (3), (4) and (7), we have

$$\begin{aligned}\frac{\mathcal{A}_2(\gamma)}{a_{nT}(\gamma-\gamma_0)} &= -\frac{\lambda_{20}^2\bar{T}\mathtt{tr}[(\mathbf{D}(\gamma_0,\gamma)\mathbf{G})^2]}{2a_{nT}(\gamma-\gamma_0)} - \frac{(\hat{\lambda}_{2,nT}-\lambda_{20})(\hat{\lambda}_{2,nT}+\lambda_{20})\bar{T}\mathtt{tr}[(\mathbf{D}(\gamma_0,\gamma)\mathbf{G})^2]}{2a_{nT}(\gamma-\gamma_0)}\\ &\leqslant -\frac{\bar{T}D_{3,nT}(\gamma)}{2(\gamma-\gamma_0)} + \frac{|\hat{l}_{nT}-l_0||\hat{l}_{nT}+l_0||L_{2,nT}(\gamma)|}{2(\gamma-\gamma_0)}\\ &\leqslant -\frac{1}{2}\bar{T}(1-\eta)k + \frac{1}{2}\iota(2|l_0|+\iota)(1+\eta)l \leqslant -\frac{1}{2}\bar{T}k + \frac{1}{2}(\bar{m}\eta + 6\bar{m}^3\iota).\end{aligned}$$

By events (1), (2), (3), (7) and (9), we have

$$\begin{aligned}&-\frac{\mathcal{B}_1(\gamma)+\mathcal{B}_2(\gamma)-2\hat{\sigma}_{nT}^2\mathcal{A}_1(\gamma)}{2\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)}\\ &= -\frac{\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\hat{\theta}_{2,nT}}{\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)} - \frac{\hat{\lambda}_{2,nT}\big\{\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}-\sigma_0^2\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}]\big\}}{\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)}\\ &\quad+\frac{\sigma_0^2\bar{T}\hat{\lambda}_{2,nT}\mathtt{tr}[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{J}]}{\hat{\sigma}_{nT}^2na_{nT}(\gamma-\gamma_0)} - \frac{(\sigma_0^2-\hat{\sigma}_{nT}^2)\bar{T}\hat{\lambda}_{2,nT}\mathtt{tr}[\mathbf{D}(\gamma_0,\gamma)\mathbf{G}]}{\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)}\\ &\leqslant \frac{\|\hat{\delta}_{nT}\|\|\mathcal{J}_{1,nT}(\gamma,\gamma_0)\|}{\hat{\sigma}_{nT}^2\sqrt{a_{nT}}(\gamma-\gamma_0)} + \frac{|\hat{l}_{nT}|\|\mathcal{J}_{2,nT}(\gamma,\gamma_0)\|}{\hat{\sigma}_{nT}^2\sqrt{a_{nT}}(\gamma-\gamma_0)} + \frac{T^\tau|\hat{l}_{nT}|\sigma_0^2|L_{3,nT}(\gamma)|}{n^{1-\tau}\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \frac{\iota|\hat{l}_{nT}||L_{1,nT}(\gamma)|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)}\\ &\leqslant \frac{(\|\delta_0\|+\iota)\eta + (|l_0|+\iota)[\eta+\iota\sigma_0^2(1+\eta)l+\iota(1+\eta)l]}{\hat{\sigma}_{nT}^2} \leqslant \frac{4\bar{m}\eta+4\bar{m}^3\iota+4\bar{m}^4\iota}{\sigma_0^2-\iota}.\end{aligned}$$

Next, by events (1), (3), (4), and (5), we have

$$\begin{aligned}-\frac{\mathcal{B}_3(\gamma)}{2\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)} &= -\frac{\theta_{20}'\mathbf{H}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}\theta_{20}}{2\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)}\\ &\quad-\frac{(\hat{\theta}_{2,nT}-\theta_{20})'\mathbf{H}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{H}(\hat{\theta}_{2,nT}+\theta_{20})}{2\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)}\\ &\leqslant -\frac{D_{1,nT}(\gamma)}{2\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \|\hat{\delta}_{nT}-\delta_0\|\|\hat{\delta}_{nT}+\delta_0\|\frac{F_{1,nT}}{2\hat{\sigma}_{nT}^2(\gamma-\gamma_0)}\\ &\leqslant \frac{-(1-\eta)k+\iota(2\|\delta_0\|+\iota)(1+\eta)l}{2\hat{\sigma}_{nT}^2} \leqslant -\frac{k}{2\sigma_0^2+2\iota} + \frac{3\bar{m}^3\iota}{\sigma_0^2-\iota}.\end{aligned}$$

Similarly, we have, by events (1), (3) and (8),

$$\begin{aligned}-\frac{\mathcal{B}_4(\gamma)+\mathcal{B}_6(\gamma)+\mathcal{B}_7(\gamma)}{2\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)} &\leqslant \|\hat{\delta}_{nT}\||\hat{l}_{nT}|\frac{\|P_{1,nT}(\gamma)\|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \iota\frac{|\hat{l}_{nT}|\|P_{2,nT}(\gamma)\|+\|\hat{\delta}_{nT}\|\|P_{3,nT}(\gamma)\|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)}\\ &\leqslant \frac{(\|\delta_0\|+\iota)(|l_0|+\iota)\eta+\iota(|l_0|+\iota)\eta+\iota(\|\delta_0\|+\iota)\eta}{\hat{\sigma}_{nT}^2} \leqslant \frac{8\bar{m}^2\eta}{\sigma_0^2-\iota}.\end{aligned}$$

Then, we show that by events (1), (3), (5) and (6),

$$\begin{aligned}-\frac{\mathcal{B}_5(\gamma)+\mathcal{B}_8(\gamma)}{2\hat{\sigma}_{nT}^2a_{nT}(\gamma-\gamma_0)} &\leqslant \frac{\iota\|\hat{\delta}_{nT}\|\|F_{2,nT}(\gamma)\|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \frac{\iota|\hat{l}_{nT}|\|K_{2,nT}(\gamma)\|}{\hat{\sigma}_{nT}^2(\gamma-\gamma_0)}\\ &\leqslant \frac{\iota[(\|\delta_0\|+\iota)+(|l_0|+\iota)](1+\eta)l}{\hat{\sigma}_{nT}^2} \leqslant \frac{8\bar{m}^3\iota}{\sigma_0^2-\iota}.\end{aligned}$$

Finally, by events (1), (3), (4) and (7), we have

$$
\begin{aligned}
-\frac{\mathcal{B}_9(\gamma)}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} &= -\frac{[\lambda_{20}^2+(\hat{\lambda}_{2,nT}-\lambda_{20})(\hat{\lambda}_{2,nT}+\lambda_{20})]\mathbf{V}'\mathbf{G}'\mathbf{D}(\gamma_0,\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\gamma)\mathbf{G}\mathbf{V}}{2\hat{\sigma}_{nT}^2 a_{nT}(\gamma-\gamma_0)} \\
&\leqslant -\frac{D_{2,nT}(\gamma)}{2\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} + \frac{|\hat{l}_{nT}-l_0||\hat{l}_{nT}+l_0||K_{1,nT}(\gamma)|}{2\hat{\sigma}_{nT}^2(\gamma-\gamma_0)} \\
&\leqslant -\frac{k}{2\sigma_0^2+2\iota} + \frac{\iota(2|l_0|+\iota)(1+\eta)l}{2\sigma_0^2-2\iota} \leqslant -\frac{k}{2\sigma_0^2+2\iota} + \frac{3\bar{m}^3\iota}{\sigma_0^2-\iota}.
\end{aligned}
$$

Together, we show that

$$
\frac{\ell_{nT}^{\ddagger}(\gamma)-\ell_{nT}^{\ddagger}(\gamma_0)}{c_{nT}a_{nT}(\gamma-\gamma_0)} \leqslant \mathcal{M}_0 < 0.
$$

Thus, we have shown that on the set E_{nT} with probability large than $1-\epsilon$, if $\gamma \in [\gamma_0+\bar{v}/a_{nT}, \gamma_0+B]$, then $\ell_{nT}^{\ddagger}(\gamma)-\ell_{nT}^{\ddagger}(\gamma_0)<0$. We can similarly show that if $\gamma \in [\gamma_0-B, \gamma_0-\bar{v}/a_{nT}]$ then $\ell_{nT}^{\ddagger}(\gamma)-\ell_{nT}^{\ddagger}(\gamma_0)<0$. Since $\ell_{nT}^{\ddagger}(\hat{\gamma}_{nT})-\ell_{nT}^{\ddagger}(\gamma_0)\geqslant 0$, this implies that $|\hat{\gamma}_{nT}-\gamma_0|\leqslant \bar{v}/a_{nT}$ is with probability larger than $1-\epsilon$. That is, $a_{nT}(\hat{\gamma}_{nT}-\gamma_0)=O_p(1)$. ■

Proof of Theorem 2.4: Recall $B_{r,nT}(\gamma), r=1,2,3$, defined in (B.12).

Proof of Result (i)**:** Using (B.11) and (B.12), we have

$$
\sqrt{N}(\hat{\theta}_{nT}-\theta_0) = \left[\tfrac{1}{nT}H_{nT}^{*}(\bar{\theta},\hat{\gamma}_{nT})\right]^{-1}\tfrac{1}{\sqrt{N}}[S_{\theta,nT}^{*u}(\theta_0,\hat{\gamma}_{nT})+\textstyle\sum_{r=1}^{3}B_{r,nT}(\hat{\gamma}_{nT})].
$$

From the proof of Theorem 2.2, we see that $\frac{1}{nT}H_{nT}^{*}(\bar{\theta},\hat{\gamma}_{nT})-\Sigma_{nT}=o_p(1)$ as $\hat{\gamma}_{nT}-\gamma_0 \xrightarrow{p} 0$ and $\bar{\theta}-\theta_0 \xrightarrow{p} 0$ implied by $\hat{\theta}_{nT}-\theta_0=o_p(1)$. Lemma A.3 and Lemma B.1 imply that $\frac{1}{\sqrt{N}}S_{\theta,nT}^{*u}(\theta_0,\hat{\gamma}_{nT})$ converges to a mean zero Gaussian distribution with variance $\frac{1}{N}\mathrm{Var}(S_{\theta,nT}^{*u})$, again due to $\hat{\gamma}_{nT}-\gamma_0 \xrightarrow{p} 0$. The derivation of $\mathrm{Var}(S_{\theta,nT}^{*u})$ is straightforward following Lemma B.5 of Yang (2015a), but is complicated when compared with $\Sigma_{nT}(\theta_0,\gamma_0)$ is given in (2.20). Some intermediate results are useful to derive the final expression. We focus on the variance of one specific quadratic term, as the derivations for the other variances or covariances are similar or less difficult. Let κ_4 be the excess kurtosis of the idiosyncratic errors. Hence, the variance of $\frac{1}{\sqrt{N}}\mathbf{V}'\mathbf{G}'\mathbf{Q}_{nT}\mathbf{V}$ can be written as $\frac{\kappa_4\sigma_0^4}{N}\mathrm{E}[\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G})'\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G})]+\frac{\sigma_0^4}{N}\mathrm{E}[\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{G}\mathbf{Q}_{nT}\mathbf{G}+\mathbf{Q}_{nT}\mathbf{G}\mathbf{G}')]$ by Yang (2015a). After some algebra, we have $\frac{1}{N}\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G})'\mathtt{diagv}(\mathbf{Q}_{nT}\mathbf{G}) = \frac{\bar{T}^2}{N}\mathtt{diagv}(\mathbf{G})'\mathtt{diagv}(\mathbf{G})+O_p(\frac{1}{n})$, $\frac{1}{N}\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{G}\mathbf{Q}_{nT}\mathbf{G})=\frac{1}{nT}\mathtt{tr}(\mathbf{G}\mathbf{G})+\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathtt{tr}[(G_t-G_k)G_t]+O_p(\frac{1}{n})$ and $\frac{1}{N}\mathtt{tr}(\mathbf{Q}_{nT}\mathbf{G}\mathbf{G}')=\frac{1}{nT}\sum_{t=1}^{T}\mathtt{tr}(\mathbf{G}\mathbf{G}')+O_p(\frac{1}{n})$. With these results, we have $\frac{1}{N}\mathrm{Var}(S_{\theta,nT}^{*u})=\Omega_{nT}(\theta_0,\gamma_0)+o(1)$, where $\Omega_{nT}(\theta_0,\gamma_0)$ is given in (2.21).

Next, we see that $\frac{1}{\sqrt{N}}B_{1,nT}(\lambda_0,\hat{\gamma}_{nT})-\sqrt{\frac{T}{n}}b_{\theta,nT} \xrightarrow{p} 0$ as $\hat{\gamma}_{nT}-\gamma_0 \xrightarrow{p} 0$. Hence, it is left to show that $\frac{1}{\sqrt{N}}B_{3,nT}(\theta_0,\hat{\gamma}_{nT})=o_p(1)$ and $\frac{1}{\sqrt{N}}B_{2,nT}(\lambda_0,\hat{\gamma}_{nT})=o_p(1)$. For $B_{3,nT}(\theta_0,\hat{\gamma}_{nT})$, by $\theta_{20}=(nT)^{-\tau}\delta_0$ and $\lambda_{20}=(nT)^{-\tau}l_0$, the 3rd component of $\frac{1}{\sqrt{N}}B_{3,nT}(\theta_0,\hat{\gamma}_{nT})$ equals to

$$
(nT)^{\tau-1/2}\sqrt{c_{nT}}\tfrac{a_{nT}}{\sigma_{0nT}^2}[\mathbf{Z}+\mathbf{G}\mathbf{V}]'\mathbf{D}(\hat{\gamma}_{nT})\mathbf{Q}_{nT}[\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{H}\delta_0+\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{G}\mathbf{V}l_0].
$$

Note that these terms in $\frac{a_{nT}}{nT}[\mathbf{Z}+\mathbf{GV}]'\mathbf{D}(\hat{\gamma}_{nT})\mathbf{Q}_{nT}[\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{H}\delta_0+\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{GV}l_0]$ have forms similar to $\mathcal{F}_{nT}(\hat{v}_{nT})$ or $\mathcal{K}_{nT}(\hat{v}_{nT})$ from Lemma B.2 and $\hat{\gamma}_{nT}=\gamma_0+\hat{v}_{nT}/a_{nT}$ by Theorem 2.3, and therefore we can show that they are all $O_p(1)$ following the proof of Lemma B.2. As $(nT)^{\tau-1/2}=o(1)$ by Assumption F, the 3rd component of $\frac{1}{\sqrt{N}}B_{3,nT}(\theta_0,\hat{\gamma}_{nT})$ is $o_p(1)$. Similarly, the 4th component of $\frac{1}{\sqrt{N}}B_{3,nT}(\theta_0,\hat{\gamma}_{nT})$ equals to

$$\begin{aligned}&\tfrac{1}{2\sqrt{N}}\big[\tfrac{a_{nT}}{nT}\delta_0'\mathbf{H}'\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{H}\delta_0+\tfrac{a_{nT}}{nT}l_0^2\mathbf{V}'\mathbf{G}'\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{GV}\\&+\tfrac{2\sqrt{a_{nT}}}{\sqrt{nT}}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{H}\delta_0+\tfrac{2a_{nT}}{nT}l_0\mathbf{V}'\mathbf{G}'\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{H}\delta_0\\&+(nT)^{\tau}\tfrac{2a_{nT}}{nT}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{GV}l_0\big].\end{aligned}$$

The first two terms in the square bracket are $O_p(1)$ by Lemma B.2, the third is $O_p(1)$ by Lemma B.3, and the fourth and the fifth (without the factor $(nT)^{\tau}$) can be shown to be $O_p(1)$ following the proof of Lemma B.2. Therefore, the 4th component of $\frac{1}{\sqrt{N}}B_{3,nT}(\theta_0,\hat{\gamma}_{nT})$ is also $o_p(1)$. Similarly, the other components of $\frac{1}{\sqrt{N}}B_{3,nT}(\theta_0,\hat{\gamma}_{nT})$ are shown to be $o_p(1)$.

Finally, we show all the components of $\frac{1}{\sqrt{N}}B_{2,nT}(\lambda_0,\hat{\gamma}_{nT})$ are also $o_p(1)$. Consider its second non-zero element for example; the proofs of others are simpler. Similar to (B.13),

$$\begin{aligned}&-\tfrac{1}{\sqrt{N}}\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\hat{\gamma}_{nT})(\mathbf{G}(\lambda_0,\hat{\gamma}_{nT})-\mathbf{G})]\\&=\tfrac{(nT)^{-\tau}l_0}{\sqrt{N}}\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\hat{\gamma}_{nT})\mathbf{G}(\lambda_0,\hat{\gamma}_{nT})\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{G})]\\&=(nT)^{\tau-1/2}\sqrt{c_{nT}}\tfrac{l_0a_{nT}}{nT}\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\hat{\gamma}_{nT})\mathbf{G}(\lambda_0,\hat{\gamma}_{nT})\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{G}]=o_p(1),\end{aligned}$$

because $(nT)^{\tau-1/2}=o(1)$, and $\frac{l_0a_{nT}}{nT}\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\hat{\gamma}_{nT})\mathbf{G}(\lambda_0,\hat{\gamma}_{nT})\mathbf{D}(\gamma_0,\hat{\gamma}_{nT})\mathbf{G}]$ has similar form to $\mathcal{L}_{nT}(\hat{v}_{nT})$ from Lemma B.2 and thus can be shown to be $O_p(1)$ in a similar manner. By the continuous mapping theorem (CMT), the result in (i) follows.

Proof of Result (ii)**:** When γ_0 were known, it is easy to see that the AQMLE $\hat{\theta}_{nT}(\gamma_0)$ is consistent to θ_0. Thus, by the mean value theorem, we also have

$$\sqrt{N}(\hat{\theta}_{nT}(\gamma_0)-\theta_0)=\big[\tfrac{1}{nT}H^*_{nT}(\dot{\theta},\gamma_0)\big]^{-1}\tfrac{1}{\sqrt{N}}(S^{*u}_{\theta,nT}+\textstyle\sum_{r=1}^3B_{r,nT}),$$

where $H^*_{nT}(\dot{\theta},\gamma)$ denotes $-\frac{\partial}{\partial\theta'}S^*_{\theta,nT}(\theta,\gamma)\big|_{\theta=\dot{\theta}_r\text{ in }r\text{th row}}$ and $\{\dot{\theta}_r\}$ are on the line segment between $\hat{\theta}_{nT}(\gamma_0)$ and θ_0. As $\dot{\theta}-\theta_0\xrightarrow{p}0$ implied by $\hat{\theta}_{nT}(\gamma_0)-\theta_0=o_p(1)$, $\frac{1}{nT}H^*_{nT}(\dot{\theta},\gamma_0)-\Sigma_{nT}=o_p(1)$. Thus, it is equivalent to showing that $\frac{1}{\sqrt{N}}[S^{*u}_{\theta,nT}(\theta_0,\hat{\gamma}_{nT})-S^{*u}_{\theta,nT}]=o_p(1)$, because $\frac{1}{\sqrt{N}}B_{2,nT}(\lambda_0,\hat{\gamma}_{nT})$ and $\frac{1}{\sqrt{N}}B_{3,nT}(\theta_0,\hat{\gamma}_{nT})$ are both $o_p(1)$, shown in (i) and $\frac{1}{\sqrt{N}}[B_{1,nT}(\lambda_0,\hat{\gamma}_{nT})-B_{1,nT}]=o_p(1)$ is directly implied by $\hat{\gamma}_{nT}-\gamma_0=o_p(1)$. For the non-zero components of $\frac{1}{\sqrt{N}}[S^{*u}_{\theta,nT}(\theta_0,\hat{\gamma}_{nT})-S^{*u}_{\theta,nT}]$, they are $O_p(\frac{1}{\sqrt{a_{nT}}})$ by Lemma B.3, completing the proof. ∎

Proof of Theorem 2.5: Let $\mathcal{Q}_{nT}(v)=\frac{1}{c_{nT}}[\ell^{\ddagger}_{nT}(\gamma_0+v/a_{nT})-\ell^{\ddagger}_{nT}(\gamma_0)]$ and $\mathcal{Q}(v)=\frac{1}{2\sigma_0^2}[-\Xi_1f|v|+2\sqrt{\sigma_0^2\Xi f}W(v)]$. We first show $\mathcal{Q}_{nT}(v)\Rightarrow\mathcal{Q}(v)$ on any compact set $\Upsilon=[-\bar{v},\bar{v}]$.

For ease of presentation, we follow the notations used in the proof of Theorem 2.3 and define $\mathcal{B}_s^*(v) = \mathcal{B}_s(\gamma_0 + v/a_{nT})$, for $s = 1$ to 9, and $\mathcal{A}_m^*(v) = \mathcal{A}_m(\gamma_0 + v/a_{nT})$, for $m = 1, 2, 3$. Recalling the notations defined in Lemma B.4, the proof of Lemma B.2 implies that $F_{2,nT}(\gamma_0 + v/a_{nT})$, $K_{2,nT}(\gamma_0 + v/a_{nT})$, $L_{1,nT}(\gamma_0 + v/a_{nT})$ and $L_{3,nT}(\gamma_0 + v/a_{nT})$ are all $O_p(\frac{1}{a_{nT}})$, and $P_{r,nT}(\gamma_0 + v/a_{nT})$ for $r = 1, 2, 3$ are all $o_p(\frac{1}{a_{nT}})$. Given these, we see that $\mathcal{B}_s^*(v)$ for $s = 4$ to 8, that involve these above quantities, are all $o_p(1)$, since $(nT)^\tau(\hat{\phi}_{nT} - \phi_0) = o_p(1)$ by Theorem 2.4 and $(nT)^\tau\theta_{20} = O(1)$ by Assumption F. Similarly, we have $\mathcal{B}_3^*(v) = \mathcal{F}_{nT}(v) + o_p(1)$, $\mathcal{B}_9^*(v) = \mathcal{K}_{nT}(v) + o_p(1)$, $\mathcal{A}_2^*(v) = -\frac{1}{2}\bar{T}\mathcal{L}_{nT}(v) + o_p(1)$, $\mathcal{A}_3^*(v) = o_p(1)$, and finally

$$
\begin{aligned}
&\mathcal{B}_1^*(v) + \mathcal{B}_2^*(v) - 2\hat{\sigma}_{nT}^2\mathcal{A}_1^*(v) \\
= &\; 2\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma_0, \gamma_0 + v/a_{nT})\mathbf{H}\hat{\theta}_{2,nT} + 2\hat{\lambda}_{2,nT}\mathbf{V}'\mathbf{Q}_{nT}\mathbf{D}(\gamma_0, \gamma_0 + v/a_{nT})\mathbf{G}\mathbf{V} \\
&- 2\hat{\sigma}_{nT}^2\hat{\lambda}_{2,nT}\bar{T}\mathtt{tr}[\mathbf{D}(\gamma_0, \gamma_0 + v/a_{nT})\mathbf{G}] \\
= &\; -2\mathcal{R}_{nT}(v) - 2\hat{l}_{nT}(nT)^\tau(\hat{\sigma}_{nT}^2 - \sigma_0^2)\bar{T}a_{nT}L_{1,nT}(\gamma_0 + v/a_{nT}) \\
&- 2\hat{l}_{nT}\tfrac{T^\tau}{n^{1-\tau}}\hat{\sigma}_{nT}^2 a_{nT}\bar{T}L_{3,nT}(\gamma_0 + v/a_{nT}) + o_p(1) = -2\mathcal{R}_{nT}(v) + o_p(1),
\end{aligned}
$$

where we use $\frac{T^\tau}{n^{1-\tau}} = o(1)$ by Assumption E.

Then, from (B.18), (B.19) and (B.20), we have

$$
\begin{aligned}
\mathcal{Q}_{nT}(v) = &\; -\tfrac{1}{2\hat{\sigma}_{nT}^2}[\textstyle\sum_{s=1}^9 \mathcal{B}_s^*(v) - 2\hat{\sigma}_{nT}^2\mathcal{A}_1^*(v)] + \mathcal{A}_2^*(v) + \mathcal{A}_3^*(v) \\
= &\; -\tfrac{1}{2\hat{\sigma}_{nT}^2}[\mathcal{F}_{nT}(v) + \mathcal{K}_{nT}(v) - 2\mathcal{R}_{nT}(v)] - \tfrac{\bar{T}}{2}\mathcal{L}_{nT}(v) + o_p(1).
\end{aligned}
$$

Using Lemma B.2, Lemma B.3 and $\hat{\sigma}_{nT}^2 - \sigma_0^2 = o_p(1)$, we finally get $\mathcal{Q}_{nT}(v) \Rightarrow \mathcal{Q}(v)$.

By Theorem 2.3, $a_{nT}(\hat{\gamma}_{nT} - \gamma_0) = \underset{v}{\text{argmax}}\ \mathcal{Q}_{nT}(v) = O_p(1)$. The functional $\mathcal{Q}(v)$ is continuous and has a unique maximum; $\lim_{|v|\to\infty} \mathcal{Q}(v) = -\infty$ almost surely since $\lim_{v\to\infty} B(v)/v = 0$ almost surely. Therefore, the conditions of Theorem 2.7 of Kim and Pollard (1990) are satisfied, which implies that

$$
a_{nT}(\hat{\gamma}_{nT} - \gamma_0) \xrightarrow{D} \underset{-\infty<v<\infty}{\text{argmax}}\ \mathcal{Q}(v).
$$

Making a change-of-variable $v = \frac{\sigma_0^2}{f}\frac{\Xi}{\Xi_1^2}r$, the asymptotic distribution is then rewritten as

$$
\begin{aligned}
\underset{-\infty<v<\infty}{\text{argmax}}\ \mathcal{Q}(v) &= \underset{-\infty<v<\infty}{\text{argmax}}\,[-\Xi_1 f|v| + 2\sqrt{\sigma_0^2\Xi f}W(v)] \\
&= \tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}\underset{-\infty<r<\infty}{\text{argmax}}\,[-\tfrac{\sigma_0^2\Xi}{\Xi_1}|r| + 2\sqrt{\sigma_0^2\Xi f}W(\tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}r)] \\
&= \tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}\underset{-\infty<r<\infty}{\text{argmax}}\,[-\tfrac{\sigma_0^2\Xi}{\Xi_1}|r| + 2\tfrac{\sigma_0^2\Xi}{\Xi_1}W(r)] \\
&= \tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}\underset{-\infty<r<\infty}{\text{argmax}}\,[-\tfrac{|r|}{2} + W(r)].
\end{aligned}
$$

$\blacksquare$

Proof of Theorem 2.6: By Theorem 2.3, we can write $\hat{\gamma}_{nT} = \gamma_0 + \frac{\hat{v}_{nT}}{a_{nT}}$. Note that

$$\begin{aligned} LR_{nT}(\gamma_0) &= \tfrac{2}{c_{nT}}[\ell^{*c}_{nT}(\hat{\gamma}_{nT}) - \ell^{*c}_{nT}(\gamma_0)] \\ &= \tfrac{2}{c_{nT}}[\ell^{*}_{nT}(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) - \ell^{*}_{nT}(\hat{\theta}_{nT}(\gamma_0), \gamma_0)] \\ &= \tfrac{2}{c_{nT}}[\ell^{*}_{nT}(\hat{\theta}_{nT}, \hat{\gamma}_{nT}) - \ell^{*}_{nT}(\hat{\theta}_{nT}, \gamma_0)] + o_p(1) \ \ \text{(Theorem 2.4)} \\ &= 2\mathcal{Q}_{nT}(\hat{v}_{nT}) + o_p(1) \xrightarrow{D} 2\sup_v \mathcal{Q}(v). \end{aligned}$$

This limiting distribution equals, by the change-of-variable $v = \frac{\sigma_0^2}{f}\frac{\Xi}{\Xi_1^2}r$,

$$\begin{aligned} \tfrac{1}{\sigma_0^2}\sup_v[-\Xi_1 f|v| + 2\sqrt{\sigma_0^2\Xi f}W(v)] &= \tfrac{1}{\sigma_0^2}\sup_r[-\Xi_1 f|\tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}r| + 2\sqrt{\sigma_0^2\Xi f}W(\tfrac{\sigma_0^2}{f}\tfrac{\Xi}{\Xi_1^2}r)] \\ &= \tfrac{\Xi}{\Xi_1}\sup_r[-|r| + 2W(r)] = \varpi^2\,\mho. \end{aligned}$$

To find the distribution of $\mho$, note that $\mho = 2\max(\mho_1, \mho_2)$, where $\mho_1 = \sup_{r\leqslant 0}[-|r|/2 + W(r)]$ and $\mho_1 = \sup_{r\geqslant 0}[-|r|/2+W(r)]$. $\mho_1$ and $\mho_2$ are iid exponential random variables with distribution function $P(\mho_1 \leqslant x) = 1 - e^{-x}$. It follows that $P(\mho \leqslant x) = P(2\max(\mho_1, \mho_2) \leqslant x) = P(\mho_1 \leqslant x/2)P(\mho_2 \leqslant x/2) = (1 - e^{-x/2})^2$. $\blacksquare$

Proof of Theorem 3.1: Applying the MVT to each element of $S^*_{\theta,nT}(\hat{\theta}_{nT}(\gamma), \gamma)$, one has

$$0 = S^*_{\theta,nT}(\hat{\theta}_{nT}(\gamma), \gamma) = S^*_{\theta,nT}(\theta_0, \gamma) + \big[\tfrac{\partial}{\partial\theta'}S^*_{\theta,nT}(\theta, \gamma)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}\big](\hat{\theta}_{nT}(\gamma) - \theta_0),$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}_{nT}(\gamma)$ and θ_0. In the following arguments, we use $H^*_{nT}(\bar{\theta}, \gamma)$ to denote $-\frac{\partial}{\partial\theta'}S^*_{\theta,nT}(\theta, \gamma)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}$ for simplicity, where the components of $H^*_{nT}(\theta, \gamma)$ are in the proof of Theorem 2.2. Note that $\bar{\theta} - \theta_0 \xrightarrow{p} 0$, as $\hat{\theta}^{\rm bc}_{nT}(\gamma) - \theta_0 \xrightarrow{p} 0$ implied by Theorem 2.4. These together with Lemma A.4 imply that the limit of $\frac{1}{nT}H^*_{nT}(\bar{\theta}, \gamma)$ is equivalent to that of $\Sigma^*_{nT}(\gamma, \gamma)$, where

$$\Sigma^*_{nT}(\gamma_1, \gamma_2) = \begin{bmatrix} \frac{1}{N\sigma_0^2}\mathrm{E}[\mathbb{X}'(\gamma_1)\mathbf{Q}_{nT}\mathbb{X}(\gamma_2)], & \frac{1}{N\sigma_0^2}\mathrm{E}[\mathbb{X}'(\gamma_1)\mathbf{Q}_{nT}\mathbb{Z}(\gamma_2)], & 0_{2k\times 1}, \\ \frac{1}{N\sigma_0^2}\mathrm{E}[\mathbb{Z}'(\gamma_1)\mathbf{Q}_{nT}\mathbb{X}(\gamma_2)], & \Sigma^*_{22,nT}(\gamma_1, \gamma_2), & \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{S}_{nT}(\gamma_1), \\ \sim, & \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{S}'_{nT}(\gamma_2), & \frac{1}{2\sigma_0^4}, \end{bmatrix}, \tag{B.22}$$

$$\begin{aligned} \Sigma^*_{22,nT}(\gamma_1, \gamma_2) &= \tfrac{1}{N\sigma_0^2}\mathrm{E}[\mathbb{Z}'(\gamma_1)\mathbf{Q}_{nT}\mathbb{Z}(\gamma_2)] + \mathrm{E}[\mathbb{S}^*_{nT}(\gamma_1, \gamma_2)],\ \mathcal{S}_{nT}(\gamma) \text{ is in (2.20) and} \\ \mathbb{S}^*_{nT}(\gamma_1, \gamma_2) &= \tfrac{1}{nT}\big\{\mathtt{tr}(\mathbf{G}\mathbf{G}^s),\ \mathtt{tr}[\mathbf{G}^\circ(\gamma_1)\mathbf{G}^s]; \mathtt{tr}[\mathbf{G}^{\circ s}(\gamma_2)\mathbf{G}],\ \mathtt{tr}[\mathbf{G}^\circ(\gamma_1)\mathbf{G}^{\circ s}(\gamma_2)]\big\}. \end{aligned}$$

Thus, by (B.12) and $\frac{1}{\sqrt{N}}B_{1,nT}(\gamma) - \sqrt{\frac{T}{n}}b_{\theta,nT}(\hat{\lambda}_{nT}(\gamma), \gamma) \xrightarrow{p} 0$, we have

$$\sqrt{N}(\hat{\theta}^{\rm bc}_{nT}(\gamma) - \theta_0) = \Sigma^{*-1}_{nT}(\gamma, \gamma)\tfrac{1}{\sqrt{N}}[S^{*u}_{\theta,nT}(\theta_0, \gamma) + B_{2,nT}(\gamma) + B_{3,nT}(\gamma)] + o_p(1). \tag{B.23}$$

By Lemma B.1, we have $\frac{1}{\sqrt{N}}S^{*u}_{\theta,nT}(\theta_0, \gamma) \Rightarrow S_\theta(\gamma)$, a Gaussian process with mean zero and covariance kernel $\lim_{nT\to\infty} \Omega^*_{nT}(\gamma_1,\gamma_2)$, where

$$\Omega^*_{nT}(\gamma_1,\gamma_2) = \Sigma^*_{nT}(\gamma_1,\gamma_2) + \Gamma^*_{nT}(\gamma_1,\gamma_2), \tag{B.24}$$

$$\Gamma^*_{nT}(\gamma_1,\gamma_2) = \begin{bmatrix} 0_{2k\times 2k}, & \frac{\bar{T}\kappa_3}{N\sigma_0}\mathrm{E}[\mathbb{X}'(\gamma_1)\mathbf{Q}_{nT}\mathbb{R}(\gamma_2)], & 0_{2k\times 1} \\ \frac{\bar{T}\kappa_3}{N\sigma_0}\mathrm{E}[\mathbb{R}'(\gamma_1)\mathbf{Q}_{nT}\mathbb{Z}(\gamma_2)], & \Gamma^*_{22,nT}(\gamma_1,\gamma_2), & \frac{\kappa_4\bar{T}^2}{2N\sigma_0^2}\mathrm{E}[\mathbb{R}'(\gamma_1)l_{nT}] \\ \sim, & \frac{\kappa_4\bar{T}^2}{2N\sigma_0^2}\mathrm{E}[l'_{nT}\mathbb{R}(\gamma_2)], & \frac{\kappa_4\bar{T}}{4\sigma_0^4} \end{bmatrix},$$

$\Gamma^*_{22,nT}(\gamma_1,\gamma_2) = \frac{2\kappa_3\bar{T}}{N\sigma_0}\mathrm{E}[\mathbb{Z}'(\gamma_1)\mathbf{Q}_{nT}\mathbb{R}(\gamma_2)] + \frac{\kappa_4\bar{T}^2}{N}\mathrm{E}[\mathbb{R}'(\gamma_1)\mathbb{R}(\gamma_2)] + \mathrm{E}[\mathbb{B}^*_{nT}(\gamma_1,\gamma_2)]$, $\mathbb{B}^*_{nT}(\gamma_1,\gamma_2) = [\mathbb{B}_{11,nT},\ \mathbb{B}_{12,nT}(\gamma_2);\ \mathbb{B}_{21,nT}(\gamma_1),\ \frac{1}{NT^2}\sum_{t=1}^T\sum_{k=1}^T \mathtt{tr}[(d_t(\gamma_1)G_t - d_k(\gamma_1)G_k)d_t(\gamma_2)G_t]$, and $\mathbb{B}_{11,nT}$, $\mathbb{B}_{12,nT}(\gamma)$ and $\mathbb{B}_{21,nT}(\gamma)$ are given below (2.21).

By Lemma A.4 and under the alternatives, one shows that $\frac{\sqrt{nT}}{N}[B_{2,nT}(\gamma) + B_{3,nT}(\gamma)] = [\Sigma^*_{nT}(\gamma,\gamma_0) - \Sigma^*_{nT}(\gamma,\gamma)]\mathbf{L}c + o_p(1)$. With (B.23), $\mathbf{L}'\theta_0 = \theta_{20}$, $\sqrt{nT}\theta_{20} = c$ and $c = \mathbf{L}'\mathbf{L}c$,

$$\begin{aligned}
\sqrt{N}\mathbf{L}'\hat{\theta}^{\mathrm{bc}}_{nT}(\gamma) &= \sqrt{N}\mathbf{L}'\theta_0 + \mathbf{L}'\big[\tfrac{1}{nT}H^*_{nT}(\bar{\theta},\gamma)\big]^{-1}\tfrac{\sqrt{N}}{nT}S^*_{\theta,nT}(\theta_0,\gamma) \\
&= \sqrt{\bar{T}}\mathbf{L}'\mathbf{L}c + \mathbf{L}'\Sigma^{*-1}_{nT}(\gamma,\gamma)\tfrac{1}{\sqrt{N}}S^{*u}_{\theta,nT}(\theta_0,\gamma) \\
&\quad + \sqrt{\bar{T}}\mathbf{L}'\Sigma^{*-1}_{nT}(\gamma,\gamma)[\Sigma^*_{nT}(\gamma,\gamma_0) - \Sigma^*_{nT}(\gamma,\gamma)]\mathbf{L}c + o_p(1) \\
&\Rightarrow \mathbf{L}'\Sigma^{*-1}(\gamma,\gamma)S_\theta(\gamma) + \bar{\Sigma}(\gamma)c.
\end{aligned}$$

Given the uniform convergence of $\hat{\theta}^*_{nT}(\gamma)$ to θ_0, it is also standard to show $\widehat{\mathbb{Q}}_{nT}(\gamma,\gamma) - \mathbb{Q}(\gamma,\gamma) \xrightarrow{p} 0$, based on the proof of Theorem 2.2. Therefore, we have $W_{nT}(\gamma) \Rightarrow W^c(\gamma)$ by the CMT. ■

Proof of Proposition 1: We only need to show that $\widetilde{S}^b_{\theta,nT}(\gamma)$ converges weakly in $\mathcal{F}_{nT}$ to a Gaussian process with covariance $\Omega^*(\gamma_1,\gamma_2)$. For this, we define

$$S^{*b}_{\theta,nT}(\gamma) = \begin{cases} \frac{1}{\sigma_0^2}\mathbb{X}'(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N, \\ \frac{1}{\sigma_0^2}[\eta(\beta_{10},\lambda_{10}) + \mathbf{G}_1\mathbf{S}\tilde{\mathbf{V}}^b_N]'\mathbf{S}\tilde{\mathbf{V}}^b_N - \mathtt{tr}[\mathbf{Q}_{nT}\mathbf{G}_1], \\ \frac{1}{\sigma_0^2}[\eta(\beta_{10},\lambda_{10}) + \mathbf{G}_1\mathbf{S}\tilde{\mathbf{V}}^b_N]'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N - \mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}_1], \\ \frac{1}{2\sigma_0^4}\tilde{\mathbf{V}}^{b\prime}_N\tilde{\mathbf{V}}^b_N - \frac{N}{2\sigma_0^2}. \end{cases}$$

Let P^* and E^* denote the probability and expectation, respectively, under the bootstrap empirical distribution $\mathcal{F}_{nT}$, conditional on the observed data. We use $o_{p^*}(1)$ to denote a sequence of random variables that converges to zero in probability, and $O_{p^*}(1)$ to denote a sequence that is bounded in probability, both under the bootstrap distribution conditional on the original sample. Furthermore, $\Rightarrow^*$ denotes weak convergence under $\mathcal{F}_{nT}$, with respect to the uniform metric on Γ. Thus, it suffices to show that, under the null hypothesis and uniformly in $\gamma\in\Gamma$, the following two conditions hold:

(a) $\frac{1}{\sqrt{N}}\widetilde{S}^b_{\theta,nT}(\gamma) - \frac{1}{\sqrt{N}}S^{*b}_{\theta,nT}(\gamma) = o_{p^*}(1)$ and (b) $\frac{1}{\sqrt{N}}S^{*b}_{\theta,nT}(\gamma) \Rightarrow^* S_\theta(\gamma)$.

To show (a), we observe that the difference between $S^b_{\theta,nT}(\gamma)$ and $\widetilde{S}^{*b}_{\theta,nT}(\gamma)$ arises from the replacement of θ_{10} with $\hat{\theta}_{1,nT}$. We only show the result for the third component as the others can be shown more easily. The proof of Theorem 3.1 implies that $\hat{\theta}_{1,nT}-\theta_{10}=O_{p^*}((N)^{-1/2})$. Thus, as $\hat{\sigma}^{-2}_{nT}-\sigma^{-2}_0=o_{p^*}(1)$ and σ^{-2} appears multiplicatively, we only need show that (i) $\frac{1}{\sqrt{N}}[\eta(\hat{\beta}_{1,nT},\hat{\lambda}_{1,nT})-\eta(\beta_{10},\lambda_{10})]'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N=o_{p^*}(1)$ and (ii) $\frac{1}{\sqrt{N}}\{\frac{1}{\sigma^2_0}[\mathbf{G}_1(\hat{\lambda}_{1,nT})\mathbf{S}\tilde{\mathbf{V}}^b_N-\mathbf{G}_1\mathbf{S}\tilde{\mathbf{V}}^b_N]'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N-\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\gamma)(\mathbf{G}_1(\hat{\lambda}_{1,nT})-\mathbf{G}_1)]\}=o_{p^*}(1)$.

For (i), recall $\eta(\beta_1,\lambda_1)=\mathbf{G}_1(\lambda_1)(\mathbf{P}_{nT}\mathbf{A}_1(\lambda_1)\mathbf{Y}+\mathbf{Q}_{nT}\mathbf{X}\beta_1)$ and note that $\mathbf{G}_1(\hat{\lambda}_{1,nT})-\mathbf{G}_1=\mathbf{G}_1(\hat{\lambda}_{1,nT})(I_{nT}-\mathbf{A}_1(\hat{\lambda}_{1,nT})\mathbf{A}^{-1}_1)=(\hat{\lambda}_{1,nT}-\lambda_{10})\mathbf{G}_1(\hat{\lambda}_{1,nT})\mathbf{G}_1$. Thus, we have,

$$
\begin{aligned}
&\tfrac{1}{\sqrt{N}}[\eta(\hat{\beta}_{1,nT},\hat{\lambda}_{1,nT})-\eta(\beta_{10},\lambda_{10})]'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N\\
=&\tfrac{1}{\sqrt{N}}[(\hat{\lambda}_{1,nT}-\lambda_{10})\mathbf{G}_1(\hat{\lambda}_{1,nT})\mathbf{G}_1(\mathbf{P}_{nT}\mathbf{A}_1\mathbf{Y}+\mathbf{Q}_{nT}\mathbf{X}\beta_{10})]'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N\\
&-\tfrac{1}{\sqrt{N}}[(\hat{\lambda}_{1,nT}-\lambda_{10})\mathbf{P}_{nT}\mathbf{W}\mathbf{Y}-\mathbf{Q}_{nT}\mathbf{X}(\hat{\beta}_{1,nT}-\beta_{10})]'\mathbf{G}'_1(\hat{\lambda}_{1,nT})\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N.
\end{aligned}
$$

We show the first term for example, as the second term can be shown similarly. Firstly, we have $\sqrt{N}(\hat{\lambda}_{1,nT}-\lambda_{10})=O_{p^*}(1)$, implied by the proof of Theorem 3.1. Secondly, we note that $\tilde{\mathbf{V}}^b_N$ is a random sample drawn from centered $\tilde{\mathbf{V}}^*(\hat{\phi}_{nT},\hat{\gamma}_{nT})=\mathbf{S}'\tilde{\mathbf{V}}(\hat{\phi}_{nT},\hat{\gamma}_{nT})$, which is a "consistent" estimator for $\mathbf{S}'\mathbf{V}$ whose elements are iid normal if $\{v_{it}\}$ are iid normal, and are uncorrelated if $\{v_{it}\}$ are iid. Therefore, as B goes to infinity and the value of $\mathbf{Y}$ is treated as constant during bootstrap draws, we have $\frac{1}{N}[\mathbf{G}_1(\mathbf{P}_{nT}\mathbf{A}_1\mathbf{Y}+\mathbf{Q}_{nT}\mathbf{X}\beta_{10})]'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N=o_{p^*}(1)$, given the exogeneity of regressors and the threshold variable in Assumption B, uniformly in $\gamma\in\Gamma$.

For (ii), the left-hand side of the equation can be written as

$$
\sqrt{N}(\hat{\lambda}_{1,nT}-\lambda_{10})\tfrac{1}{N}\{\tfrac{1}{\sigma^2_0}\tilde{\mathbf{V}}^{b\prime}_N\mathbf{S}'\mathbf{G}'_1\mathbf{G}'_1(\hat{\lambda}_{1,nT})\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N-\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}_1(\hat{\lambda}_{1,nT})\mathbf{G}_1]\},
$$

which is also $o_{p^*}(1)$ as $\sqrt{N}(\hat{\lambda}_{1,nT}-\lambda_{10})=O_{p^*}(1)$ and the remain quantity is $o_{p^*}(1)$ implied by Lemma A.4. Thus, we have shown (a).

To prove (b), we first show that $\frac{1}{\sqrt{N}}S^{*b}_{\theta,nT}(\gamma)$ is asymptotically a mean-zero Gaussian process with covariance kernel $\Omega^*(\gamma,\gamma)$ in $\mathcal{F}_{nT}$ under H_0. The result then follows by establishing the tightness of the convergence. For the former result, we express $S^{*b}_{\theta,nT}(\gamma)$ in terms of the linear and quadratic forms of $\tilde{\mathbf{V}}^b_N$. As $\mathbf{A}_1\mathbf{Y}=\mathbf{X}\beta_{10}+\mathbf{C}\psi_0+\mathbf{V}$, the third component of $S^{*b}_{\theta,nT}(\gamma)$ for example can be written as

$$
\begin{aligned}
S^{*b}_{\lambda_2,nT}(\gamma)=&\tfrac{1}{\sigma^2_0}[\mathbf{G}_1(\mathbf{P}_{nT}\mathbf{A}_1\mathbf{Y}+\mathbf{Q}_{nT}\mathbf{X}\beta_{10})+\mathbf{G}_1\mathbf{S}\tilde{\mathbf{V}}^b_N]'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N-\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}_1]\\
=&\tfrac{1}{\sigma^2_0}(\mathbf{G}_1\mathbf{P}_{nT}\mathbf{V})'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N+\{\tfrac{1}{\sigma^2_0}(\mathbf{G}_1\mathbf{S}\tilde{\mathbf{V}}^b_N)'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N-\mathtt{tr}[\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{G}_1]\}\\
&+\tfrac{1}{\sigma^2_0}[\mathbf{G}_1(\mathbf{C}\psi_0+\mathbf{X}\beta_{10})]'\mathbf{D}(\gamma)\mathbf{S}\tilde{\mathbf{V}}^b_N\equiv\Pi_1(\gamma)+\Pi_2(\gamma)+\Pi_3(\gamma).
\end{aligned}
\tag{B.25}
$$

As $\mathbf{Y}$ (or $\mathbf{V}$) are treated as constant during bootstrap draws, we have that $\tilde{\mathbf{V}}^b_N$ are resampled independently of $\mathbf{V}$ under $\mathcal{F}_{nT}$. Therefore, the above expression is simply a linear and quadratic form of $\tilde{\mathbf{V}}_N^b$, which is a "consistent" estimation for $\mathbf{S}'\mathbf{V}$. As the sample size and B go to infinity, we can verify that (B.25) satisfies the conditions of the CLT in Lemma A.3 for each $\gamma \in \Gamma$, as the vectors and matrix corresponding to linear and quadratic terms are bounded as required, by Lemma A.1. Similarly, the other components of $S^{*b}_{\theta,nT}(\gamma)$ satisfy the same conditions. These ensure the asymptotic normality of $\frac{1}{\sqrt{N}}S^{*b}_{\theta,nT}(\gamma)$, which can then be extended to any finite collection of γ to establish the convergence of finite-dimensional distributions.

We next verify that the $\frac{1}{N}\mathrm{E}^*[S^{*b}_{\theta,nT}(\gamma_1)S^{*b\prime}_{\theta,nT}(\gamma_2)] = \Omega^*_{nT}(\gamma_1,\gamma_2)|_{\theta_{20}=0} + o(1)$. For illustration, we focus on verifying $\frac{1}{N}\mathrm{E}^*[S^{*b}_{\lambda_2,nT}(\gamma_1)S^{*b\prime}_{\lambda_2,nT}(\gamma_2)] = \Omega^*_{\lambda_2\lambda_2}(\gamma_1,\gamma_2)|_{\theta_{20}=0}+o(1)$, where $\Omega^*_{\lambda_2\lambda_2}(\gamma_1,\gamma_2)$ represents the λ_2-λ_2 element of the covariance matrix, as the remaining elements can be established similarly. That is to show $\frac{1}{N}[\mathrm{E}^*(\Pi_1(\gamma_1)\Pi_1(\gamma_2)) + \mathrm{E}^*(\Pi_2(\gamma_1)\Pi_2(\gamma_2)) + \mathrm{E}^*(\Pi_3(\gamma_1)\Pi_3(\gamma_2)) + 2\mathrm{E}^*(\Pi_1(\gamma_1)\Pi_2(\gamma_2)) + 2\mathrm{E}^*(\Pi_2(\gamma_1)\Pi_3(\gamma_2)) + 2\mathrm{E}^*(\Pi_1(\gamma_1)\Pi_3(\gamma_2))] = \Omega^*_{\lambda_2,\lambda_2}(\gamma_1,\gamma_2)|_{\theta_{20}=0} + o(1)$. According to the derivations in the proof of Theorem 2.1, $\frac{1}{N}[\mathrm{E}^*(\Pi_1(\gamma_1)\Pi_1(\gamma_2))+\mathrm{E}^*(\Pi_2(\gamma_1)\Pi_2(\gamma_2))] = \frac{1}{nT}\mathrm{E}^*[\mathtt{tr}(\mathbf{G}^\circ(\gamma)\mathbf{G}^{\circ s}(\gamma))]|_{\lambda_{20}=0} + \frac{\kappa_4\bar{T}^2}{N}\mathrm{E}^*[\mathtt{diagv}(\mathbf{G}_1)'\mathbf{D}(\gamma)\mathtt{diagv}(\mathbf{G}_1)] + \mathrm{E}^*[\mathbb{B}_{22,nT}(\gamma)]|_{\lambda_{20}=0} + o(1)$ and $\frac{1}{N}\mathrm{E}^*(\Pi_3(\gamma_1)\Pi_3(\gamma_2)) = \frac{1}{N}\mathrm{E}^*[\mathbf{Z}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathbf{Z}]|_{\theta_{20}=0} + o(1)$. Besides, the covariance terms yield $\frac{2}{N}\mathrm{E}^*(\Pi_2(\gamma)\Pi_3(\gamma)) = \frac{2\kappa_3\bar{T}}{N\sigma_0}\mathrm{E}^*[\mathbf{Z}'\mathbf{D}(\gamma)\mathbf{Q}_{nT}\mathbf{D}(\gamma)\mathtt{diagv}(\mathbf{G}_1)]|_{\theta_{20}=0}+o(1)$, $\frac{2}{N}\mathrm{E}^*(\Pi_1(\gamma)\Pi_2(\gamma)) = o(1)$ and $\frac{2}{N}\mathrm{E}^*(\Pi_1(\gamma)\Pi_3(\gamma)) = o(1)$. Thus, we have shown the desired result. The other elements of $\frac{1}{N}\mathrm{Var}^*[S^{*b}_{\theta,nT}(\gamma)]$ can be shown similarly. Subsequently, the stochastic equicontinuity can be established by similar arguments to the proof of Lemma B.1. This completes the proof of (b).

Finally, as $\widetilde{\Sigma}^*_{nT}(\gamma,\gamma) - \Sigma^*(\gamma,\gamma)|_{\theta_{20}=0} = o_{p^*}(1)$ and $\widetilde{\Omega}^*_{nT}(\gamma,\gamma) - \Omega^*(\gamma,\gamma)|_{\theta_{20}=0} = o_{p^*}(1)$, uniformly in $\gamma \in \Gamma$, shown in the proof of Theorem 3.1, the final result follows from the CMT. ■

Supplementary Material

The `Supplementary Material` contains the detailed proofs of Lemma B.1–B.4, and can be found online at `http://www.mysmu.edu.sg/faculty/zlyang/SubPages/research.htm`.

Acknowledgments

We thank Anil K. Bera, Xi Qu, Xiaodong Liu, Yasumasa Mastuda, and the participants of CMES 2021, and XVIII World Conference of Spatial Econometrics Association, Groningen, the Netherlands, June 2024, for their helpful comments. Zhenlin Yang gratefully acknowledges the financial support from Singapore Management University. Xiaoyu Meng acknowledges with gratitude the financial support from the National Natural Science Foundation of China (Grant No. 72403136).

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*, 821–856.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* The Netherlands: Kluwer Academic Press.

Anselin, L., J. L. Gallo, and H. Jayet (2008). Spatial panel econometrics. In *The econometrics of panel data: Fundamentals and recent developments in theory and practice*, pp. 625–660. Springer.

Arellano, M. and J. Hahn (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs 43*, 381.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics 79*, 551–563.

Baltagi, B. H., S. H. Song, and W. Koh (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics 117*, 123–150.

Baltagi, B. H. and Z. L. Yang (2013a). Heteroskedasticity and non-normality robust lm tests of spatial dependence. *Regional Science and Urban Economics 43*, 725–739.

Baltagi, B. H. and Z. L. Yang (2013b). Standardized lm tests for spatial error dependence in linear or panel regressions. *Econometrics Journal 16*, 103–134.

Caner, M. and B. E. Hansen (2004). Instrumental variable estimation of a threshold model. *Econometric Theory 20*, 813–843.

Card, D., A. Mas, and J. Rothstein (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics 123*, 177–218.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of statistics 21*, 520–533.

Cox, D. R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B (Methodological) 49*, 1–18.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*, 247–254.

Elhorst, J. P. and S. Fréret (2009). Evidence of political yardstick competition in france using a two-regime spatial durbin model with fixed effects. *Journal of Regional Science 49*, 931–951.

Ertur, C. and W. Koch (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics 22*, 1033–1062.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996). Crime and social interactions. *The Quarterly Journal of Economics 111*, 507–548.

Hall, B. C. (2015). *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction.* New York: Springer.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*, 413–430.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics 93*, 345–368.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*, 575–603.

Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and its Interface 4*, 123–127.

Kelejian, H. H. and G. Piras (2016). An extension of the j-test to spatial panel data framework. *Journal of Applied Econometrics 31*, 387–402.

Kim, J. and D. Pollard (1990). Cube root asymptotics. *The Annals of Statistics 18*, 191–219.

Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. *Econometric Theory 18*, 252–277.

Lee, L. F. and J. Yu (2010a). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*, 165–185.

Lee, L. F. and J. Yu (2010b). Some recent developments in spatial panel data models. *Regional Science and Urban Economics 40*, 255–271.

Lee, L. F. and J. Yu (2015). Spatial panel data models. In *The Oxford Handbook of Panel Data*, pp. 363–401. Oxford University Press.

Li, K. (2018). Spatial panel data models with structural change. MPRA Working Paper No. 85388.

Li, K. and W. Lin (2024). Threshold spatial autoregressive model. *Journal of Econometrics 244*(1), 105841.

Liu, S. F. and Z. Yang (2020). Robust estimation and inference of spatial panel data models with fixed effects. *Japanese Journal of Statistics and Data Science 3*, 257–311.

Meng, X. Y. and Z. L. Yang (2024). Genuinely unbalanced spatial panel data models: fixed effects m-estimation and inference. Working paper.

Miao, K., K. Li, and L. Su (2020). Panel threshold models with interactive fixed effects. *Journal of Econometrics 219*, 137–170.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Pesaran, M. H. and A. Pick (2007). Econometric issues in the analysis of contagion. *Journal of Economic Dynamics and Control 31*, 1245–1277.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology 1*, 143–186.

Smith, R. L. (1992). Some interlacing properties of the schur complement of a hermitian matrix. *Linear Algebra and its Applications 177*, 137–144.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Wei, L., C. Zhang, J. J. Su, and L. Yang (2021). Panel threshold spatial durbin models with individual fixed effects. *Economics Letters 201*, 109778.

Wu, J. and Y. Matsuda (2021). A threshold extension of spatial dynamic panel model with fixed effects. *Journal of Spatial Econometrics 2*(1), 3.

Yang, Z. L. (2015a). A general method for third-order bias and variance correction on a nonlinear estimator. *Journal of Econometrics 186*, 178–200.

Yang, Z. L. (2015b). Lm tests of spatial dependence based on bootstrap critical values. *Journal of Econometrics 185*(1), 33–59.

Yang, Z. L. (2018). Unified m-estimation of fixed-effects spatial dynamic models with short panels. *Journal of Econometrics 205*, 423–447.

Yang, Z. L., J. H. Yu, and S. F. Liu (2016). Bias correction and refined inferences for fixed effects spatial panel data models. *Regional Science and Urban Economics 61*, 52–72.

Yao, Y. and M. Zhang (2013). Performance of officials and the promotion tournament: Evidence from chinese cities. *Economic Research Journal 1*, 137–150.

Yao, Y. and M. Zhang (2015). Subnational leaders and economic growth: evidence from chinese cities. *Journal of Economic Growth 20*, 405–436.

Yu, J., L. A. Zhou, and G. Zhu (2016). Strategic interaction in political competition: Evidence from spatial effects across chinese cities. *Regional Science and Urban Economics 57*, 23–37.

Table 1: Empirical bias(sd)[$\hat{sd}$] of the estimators for FE-SPR model with threshold effects; W_t=Queen Contiguity; error = 1(normal), 2(normal mixture), 3(chi-square).

	2SLS	AQML	bc-AQML	2SLS	AQML	bc-AQML
	(a) $(n,T)=(50,5)$			(b) $(n,T)=(50,10)$		
β_1	.0380(2.629)	.0032(.075)[.071]	-.0005(.075)[.071]	.0241(0.084)	-.0002(.049)[.050]	-.0010(.049)[.050]
β_2	-.0313(2.644)	-.0072(.111)[.112]	-.0022(.111)[.112]	-.0561(0.159)	-.0011(.083)[.083]	-.0013(.083)[.083]
λ_1	.1803(15.798)	-.0257(.060)[.059]	-.0083(.060)[.059]	-.1339(0.638)	-.0229(.045)[.042]	-.0044(.044)[.042]
λ_2	-.2001(15.856)	.0125(.052)[.054]	.0116(.052)[.054]	.2380(0.931)	.0048(.037)[.037]	.0038(.037)[.037]
σ^2	.1887(0.127)	-.0208(.093)[.098]	-.0233(.093)[.098]	.1611(0.083)	-.0103(.065)[.067]	-.0126(.065)[.067]
γ	.6503(2.117)	.0093(.147)[- -]	.0093(.147)[- -]	.3965(1.424)	-.0129(.121)[- -]	-.0129(.121)[- -]
β_1	.0436(0.893)	.0015(.071)[.070]	-.0022(.071)[.070]	.0192(0.077)	.0029(.048)[.050]	.0021(.048)[.050]
β_2	-.0472(0.923)	-.0049(.110)[.111]	.0000(.110)[.111]	-.0501(0.146)	-.0047(.082)[.083]	-.0048(.082)[.083]
λ_1	-.0338(9.795)	-.0233(.061)[.059]	-.0059(.061)[.059]	-.1508(0.539)	-.0244(.046)[.042]	-.0059(.046)[.042]
λ_2	.1028(9.870)	.0117(.053)[.054]	.0109(.053)[.054]	.2234(0.839)	.0073(.039)[.037]	.0063(.039)[.037]
σ^2	.1921(0.244)	-.0157(.225)[.206]	-.0183(.224)[.206]	.1573(0.166)	-.0003(.164)[.150]	-.0027(.163)[.150]
γ	.6024(2.135)	-.0120(.155)[- -]	-.0120(.155)[- -]	.4129(1.409)	-.0268(.141)[- -]	-.0268(.141)[- -]
β_1	.0572(0.453)	.0032(.068)[.071]	-.0004(.068)[.071]	.0204(0.077)	.0008(.054)[.050]	.0001(.054)[.050]
β_2	-.0553(0.515)	-.0041(.113)[.111]	.0009(.113)[.111]	-.0518(0.144)	.0015(.089)[.083]	.0014(.089)[.083]
λ_1	.0110(3.326)	-.0225(.060)[.058]	-.0052(.060)[.058]	-.1459(0.524)	-.0203(.045)[.041]	-.0019(.045)[.041]
λ_2	-.0074(3.561)	.0067(.054)[.053]	.0060(.054)[.053]	.2112(0.850)	.0045(.038)[.037]	.0036(.038)[.037]
σ^2	.1830(0.183)	-.0279(.152)[.150]	-.0304(.151)[.150]	.1562(0.120)	-.0109(.111)[.108]	-.0133(.111)[.108]
γ	.5347(2.145)	-.0010(.149)[- -]	-.0010(.149)[- -]	.4080(1.412)	-.0332(.144)[- -]	-.0332(.144)[- -]
	(c) $(n,T)=(50,20)$			(d) $(n,T)=(50,40)$		
β_1	.0134(0.066)	.0017(.032)[.032]	.0009(.032)[.032]	.0080(0.025)	.0013(.021)[.021]	.0007(.021)[.021]
β_2	-.0217(0.097)	-.0020(.054)[.054]	-.0014(.054)[.054]	-.0227(0.052)	.0009(.036)[.037]	.0008(.036)[.037]
λ_1	.0028(0.432)	-.0209(.032)[.031]	-.0016(.032)[.031]	-.0342(0.159)	-.0223(.022)[.022]	-.0028(.022)[.022]
λ_2	.0208(0.616)	.0027(.032)[.032]	.0012(.032)[.032]	.0992(0.305)	.0022(.023)[.024]	.0019(.023)[.024]
σ^2	.1125(0.054)	-.0033(.045)[.046]	-.0056(.045)[.046]	.0800(0.038)	-.0023(.033)[.032]	-.0044(.033)[.032]
γ	-.4050(1.158)	-.0226(.122)[- -]	-.0226(.122)[- -]	-.1565(0.944)	-.0027(.060)[- -]	-.0027(.060)[- -]
β_1	.0095(0.062)	-.0005(.031)[.032]	-.0012(.031)[.032]	.0100(0.030)	.0003(.022)[.021]	-.0003(.022)[.021]
β_2	-.0132(0.103)	-.0011(.054)[.054]	-.0004(.054)[.054]	-.0245(0.051)	.0015(.039)[.037]	.0014(.039)[.037]
λ_1	-.0016(0.374)	-.0242(.033)[.032]	-.0049(.033)[.032]	-.0200(0.314)	-.0196(.023)[.022]	-.0001(.023)[.022]
λ_2	-.0080(0.659)	.0077(.030)[.032]	.0061(.030)[.032]	.0737(0.419)	.0017(.023)[.024]	.0013(.023)[.024]
σ^2	.1124(0.117)	-.0069(.109)[.107]	-.0090(.109)[.107]	.0807(0.084)	-.0033(.071)[.076]	-.0055(.071)[.076]
γ	-.3896(1.094)	-.0135(.128)[- -]	-.0135(.128)[- -]	-.1893(0.991)	.0025(.072)[- -]	.0025(.072)[- -]
β_1	.0107(0.063)	.0015(.032)[.032]	.0007(.032)[.032]	.0087(0.027)	.0006(.021)[.021]	.0000(.021)[.021]
β_2	-.0213(0.091)	.0005(.053)[.054]	.0011(.053)[.054]	-.0223(0.054)	.0009(.039)[.037]	.0009(.039)[.037]
λ_1	.0058(0.427)	-.0227(.032)[.031]	-.0034(.032)[.031]	-.0119(0.356)	-.0223(.023)[.022]	-.0028(.023)[.022]
λ_2	.0192(0.567)	.0050(.029)[.031]	.0034(.029)[.031]	.0637(0.519)	.0048(.024)[.024]	.0044(.024)[.024]
σ^2	.1142(0.086)	-.0046(.078)[.077]	-.0068(.078)[.077]	.0797(0.059)	-.0020(.055)[.055]	-.0041(.055)[.055]
γ	-.3495(1.091)	-.0241(.124)[- -]	-.0241(.124)[- -]	-.1670(0.995)	.0037(.081)[- -]	.0037(.081)[- -]

Note: (i) Error distributions: error = 1, 2, 3, for the three panels under each (n,T);

(ii) True parameter values: $\beta_1 = 1, \lambda_1 = 0.2, \gamma = 0$, and $\beta_2 = \lambda_2 = (nT)^{-0.2}$.

(iii) Empirical bias(sd) for QMLE of σ^2 under the three error distributions:

(a) $\{-.2323(.073), -.2283(.177), -.2379(.119)\}$; (b) $\{-.1271(.057), -.1183(.145), -.1276(.098)\}$

(c) $\{-.0721(.042), -.0754(.102), -.0733(.072)\}$; (d) $\{-.0467(.031), -.0477(.068), -.0464(.053)\}$

Table 1 (Cont'd): Empirical bias(sd)[$\hat{sd}$] of the estimators for FE-SPR model with threshold effects; W_t=Queen Contiguity; error = 1(normal), 2(normal mixture), 3(chi-square).

	2SLS	AQML	bc-AQML	2SLS	AQML	bc-AQML
		(e) $(n,T)=(100,5)$			(f) $(n,T)=(100,10)$	
β_1	.0306(0.221)	-.0010(.047)[.045]	-.0021(.047)[.045]	.0089(0.058)	.0007(.030)[.032]	.0006(.030)[.032]
β_2	-.0094(0.356)	.0029(.086)[.081]	.0029(.086)[.081]	-.0342(0.117)	-.0017(.055)[.055]	-.0025(.055)[.055]
λ_1	-.0986(0.802)	-.0132(.043)[.042]	-.0054(.042)[.042]	-.0402(0.525)	-.0118(.032)[.030]	-.0023(.032)[.030]
λ_2	.1955(1.555)	.0119(.049)[.052]	.0120(.049)[.052]	.0753(0.821)	.0039(.027)[.026]	.0038(.027)[.026]
σ^2	.0781(0.080)	-.0146(.072)[.070]	-.0156(.071)[.070]	.1021(0.054)	-.0083(.046)[.047]	-.0094(.046)[.047]
γ	-.7348(1.432)	-.0438(.204)[- -]	-.0438(.204)[- -]	.3430(1.497)	.0243(.080)[- -]	.0243(.080)[- -]
β_1	-.0160(2.345)	.0011(.045)[.045]	.0000(.045)[.045]	.0096(0.061)	-.0007(.032)[.032]	-.0008(.032)[.032]
β_2	.0317(2.350)	.0037(.083)[.081]	.0037(.083)[.081]	-.0323(0.110)	-.0005(.055)[.055]	-.0013(.055)[.055]
λ_1	-.0078(7.182)	-.0105(.044)[.043]	-.0028(.044)[.043]	-.0489(0.583)	-.0115(.030)[.030]	-.0020(.030)[.030]
λ_2	.0885(7.232)	.0049(.049)[.052]	.0050(.049)[.052]	.0951(0.873)	.0004(.026)[.026]	.0003(.026)[.026]
σ^2	.0736(0.159)	-.0052(.154)[.151]	-.0062(.154)[.151]	.1040(0.117)	-.0098(.108)[.106]	-.0108(.108)[.106]
γ	-.7482(1.460)	-.0262(.202)[- -]	-.0262(.202)[- -]	.2983(1.554)	.0236(.079)[- -]	.0236(.079)[- -]
β_1	.0385(0.555)	-.0023(.045)[.044]	-.0033(.045)[.044]	.0080(0.057)	-.0010(.034)[.032]	-.0011(.034)[.032]
β_2	-.0390(0.596)	.0099(.083)[.081]	.0100(.083)[.081]	-.0323(0.098)	-.0012(.057)[.056]	-.0021(.057)[.056]
λ_1	-.0934(1.867)	-.0115(.044)[.042]	-.0038(.044)[.042]	-.0370(0.547)	-.0087(.029)[.029]	.0008(.029)[.029]
λ_2	.1126(2.126)	.0037(.052)[.052]	.0038(.052)[.052]	.0628(0.825)	.0023(.026)[.026]	.0022(.026)[.026]
σ^2	.0804(0.122)	-.0196(.110)[.107]	-.0205(.110)[.107]	.0963(0.084)	-.0018(.086)[.079]	-.0029(.086)[.079]
γ	-.7584(1.465)	-.0354(.235)[- -]	-.0354(.235)[- -]	.4271(1.516)	.0130(.078)[- -]	.0130(.078)[- -]
		(g) $(n,T)=(200,5)$			(h) $(n,T)=(200,10)$	
β_1	.0008(0.082)	-.0005(.036)[.035]	-.0009(.036)[.035]	-.0127(0.056)	.0006(.022)[.023]	.0005(.022)[.023]
β_2	-.0277(0.210)	.0041(.064)[.060]	.0043(.064)[.060]	-.0016(0.076)	-.0014(.039)[.039]	-.0015(.039)[.039]
λ_1	-.0918(0.529)	-.0050(.033)[.033]	.0003(.033)[.033]	-.0648(0.167)	-.0053(.024)[.022]	-.0001(.024)[.022]
λ_2	.1848(1.113)	.0020(.032)[.032]	.0020(.032)[.032]	.0913(0.210)	.0005(.021)[.021]	.0006(.021)[.021]
σ^2	.0933(0.058)	-.0053(.049)[.050]	-.0059(.049)[.050]	.0850(0.037)	-.0038(.035)[.033]	-.0044(.035)[.033]
γ	.3477(1.342)	-.0167(.114)[- -]	-.0167(.114)[- -]	.2709(0.854)	-.0013(.055)[- -]	-.0013(.055)[- -]
β_1	.0040(0.078)	.0015(.035)[.035]	.0011(.035)[.035]	-.0032(0.136)	.0002(.022)[.023]	.0001(.022)[.023]
β_2	-.0336(0.141)	-.0042(.060)[.060]	-.0040(.060)[.060]	-.0126(0.148)	.0000(.038)[.039]	-.0001(.038)[.039]
λ_1	-.0546(0.589)	-.0084(.033)[.033]	-.0031(.033)[.033]	-.0260(0.801)	-.0072(.022)[.022]	-.0020(.022)[.022]
λ_2	.1443(0.900)	.0052(.031)[.032]	.0052(.031)[.032]	.0511(0.825)	.0027(.020)[.021]	.0029(.020)[.021]
σ^2	.0925(0.118)	-.0093(.112)[.107]	-.0099(.111)[.107]	.0868(0.082)	-.0040(.072)[.076]	-.0046(.073)[.076]
γ	.4237(1.344)	-.0148(.104)[- -]	-.0148(.104)[- -]	.2858(0.936)	-.0013(.044)[- -]	-.0013(.044)[- -]
β_1	-.0001(0.070)	-.0009(.034)[.035]	-.0013(.034)[.035]	-.0041(0.073)	.0010(.024)[.023]	.0008(.024)[.023]
β_2	-.0256(0.130)	-.0002(.058)[.060]	.0000(.058)[.060]	-.0117(0.109)	-.0014(.041)[.039]	-.0016(.040)[.039]
λ_1	-.0943(0.522)	-.0039(.031)[.033]	.0014(.031)[.033]	-.0351(0.366)	-.0059(.022)[.022]	-.0007(.022)[.022]
λ_2	.1759(0.837)	-.0008(.033)[.032]	-.0008(.033)[.032]	.0547(0.433)	.0001(.021)[.021]	.0002(.021)[.021]
σ^2	.0890(0.090)	-.0058(.085)[.079]	-.0064(.085)[.079]	.0855(0.060)	-.0032(.056)[.055]	-.0038(.056)[.055]
γ	.4289(1.338)	-.0194(.132)[- -]	-.0194(.132)[- -]	.2572(0.949)	-.0006(.055)[- -]	-.0006(.055)[- -]

Note: (i) Error distributions: error = 1, 2, 3, for the three panels under each (n,T);

(ii) True parameter values: $\beta_1=1, \lambda_1=0.2, \gamma=0$, and $\beta_2=\lambda_2=(nT)^{-0.2}$.

(iii) Empirical bias(sd) for QMLE of σ^2 under three error distributions:

(e) $\{-.2196(.057), -.2121(.122), -.2235(.087)\}$; (f) $\{-.1164(.041), -.1177(.096), -.1106(.076)\}$

(g) $\{-.2082(.039), -.2114(.089), -.2086(.067)\}$; (h) $\{-.1079(.031), -.1081(.065), -.1074(.050)\}$

Table 2: Confidence interval coverage probability for γ based on LR test at 5% level; W_t=Queen Contiguity.

n	T	Normal error	Normal mixture	Chi-square
50	5	95.20%	95.80%	96.80%
	10	98.40%	95.00%	96.60%
	20	97.20%	97.80%	96.40%
	40	97.60%	96.20%	95.60%
100	5	96.80%	93.60%	95.00%
	10	98.20%	96.00%	96.60%
200	5	96.40%	98.00%	97.80%
	10	97.80%	97.40%	96.00%

Table 3: Rejecting frequency of tests for threshold effects at 0.01, 0.05, and 0.10 levels; W_t = Queen Contiguity; error = 1 (normal), 2 (normal mixture), 3 (chi-square).

error	n	T	$\lambda_2 = \beta_2 = 0$			$\lambda_2 = \beta_2 = 2/\sqrt{nT}$			$\lambda_2 = \beta_2 = 10/\sqrt{nT}$		
			0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
1	50	5	.171	.092	.030	.546	.424	.196	1.000	1.000	1.000
		10	.124	.079	.018	.464	.356	.162	1.000	1.000	1.000
		20	.114	.038	.018	.402	.288	.110	1.000	1.000	1.000
		40	.096	.038	.008	.374	.262	.116	1.000	1.000	1.000
	100	5	.164	.112	.038	.338	.240	.090	1.000	1.000	1.000
		10	.097	.046	.013	.322	.214	.104	1.000	1.000	1.000
	200	5	.104	.064	.017	.398	.276	.114	1.000	1.000	1.000
		10	.096	.034	.006	.356	.266	.108	1.000	1.000	1.000
2	50	5	.229	.150	.069	.624	.502	.266	1.000	1.000	1.000
		10	.176	.122	.054	.496	.390	.230	1.000	1.000	1.000
		20	.158	.100	.032	.414	.310	.164	1.000	1.000	1.000
		40	.138	.078	.024	.430	.294	.138	1.000	1.000	1.000
	100	5	.198	.150	.056	.382	.282	.118	1.000	1.000	1.000
		10	.157	.101	.038	.336	.212	.080	1.000	1.000	1.000
	200	5	.132	.078	.025	.400	.302	.142	1.000	1.000	1.000
		10	.130	.080	.022	.384	.292	.116	1.000	1.000	1.000
3	50	5	.210	.140	.055	.568	.440	.232	1.000	1.000	1.000
		10	.133	.090	.041	.432	.316	.152	1.000	1.000	1.000
		20	.136	.070	.022	.402	.276	.114	1.000	1.000	1.000
		40	.124	.076	.022	.406	.284	.124	1.000	1.000	1.000
	100	5	.140	.096	.038	.352	.272	.104	1.000	1.000	1.000
		10	.122	.072	.027	.322	.228	.086	1.000	1.000	1.000
	200	5	.096	.042	.014	.418	.304	.134	1.000	1.000	1.000
		10	.122	.068	.024	.356	.254	.114	1.000	1.000	1.000

Table 4: Estimates of spatial competitions in government investments among Chinese cities.

Government Investments	Model 1		Model 2	
	Mayor: old vs all		Mayor: old vs old	
Threshold estimate:				
Threshold (γ)	55.33		54.58	
95% confidence interval	[50.00, 57.92]		[54.25, 57.92]	
Spatial effects:				
Base effect (λ_1)	-0.033	(.034)	-0.041	(.036)
Threshold effect (λ_2)	0.105***	(.035)	0.137***	(.038)
Impact of covariates:				
Fiscal revenue	0.308***	(.061)	0.312***	(.061)
Fiscal expenditure	0.129***	(.043)	0.123***	(.043)
Population	0.023	(.035)	0.025	(.035)
manufacturing ratio	0.767***	(.062)	0.769***	(.062)
GDP per capita	-0.131***	(.048)	-0.134***	(.048)
Provincial fiscal revenue	0.259***	(.077)	0.273***	(.077)
Provincial fiscal expenditure	-0.126	(.080)	-0.135*	(.080)
Public capital investment	-0.007	(.015)	-0.013	(.015)

Note: Significance levels: *:10%, **:5%, and ***: 1%. Standard errors are in parentheses.