

Model Selection in Panel Data Models: A Generalization of the Vuong Test

Jinyong Hahn [*]	Zhipeng Liao [†]	Konrad Menzel [‡]	Quang Vuong [§]
UCLA	UCLA	NYU	NYU

January 29, 2026

Abstract

This paper generalizes the classical Vuong (1989) test to panel data models by employing modified profile likelihoods and the Kullback–Leibler information criterion. Unlike the standard likelihood function, the profile likelihood lacks certain regular properties, making modification necessary. We adopt a generalized panel data framework that incorporates group fixed effects for time and individual pairs, rather than traditional individual fixed effects. Applications of our approach include linear models with non-nested specifications of individual-time effects.

JEL Classification: C14, C31, C32

Keywords: Panel data models, Vuong test, Bias correction, Grouped heterogeneity.

1 Introduction

With the advent of big data, modern empirical research increasingly adopts more complex models. These models often involve high-dimensional nuisance parameters, which can introduce substantial biases in standard estimators - manifesting as variations of the incidental parameters problem. Classical references include Neyman and Scott (1948) and Nickell

^{*}Department of Economics, UCLA, Los Angeles, CA 90095, USA. Email: hahn@econ.ucla.edu.

[†]Department of Economics, UCLA, Los Angeles, CA 90095, USA. Email: zhipeng.liao@econ.ucla.edu.

[‡]Department of Economics, NYU, New York, NY 10012, USA. Email: km125@nyu.edu.

[§]Department of Economics, NYU, New York, NY 10012, USA. Email: qvuong@nyu.edu.

(1981) among others. Numerous studies have been conducted to address and mitigate these issues. It is now well established that these issues represent a form of higher-order bias, and a variety of methodological approaches have been proposed to overcome them. See Arellano and Hahn (2007), Arellano and Bonhomme (2009), Bester and Hansen (2009), Carro (2007), Dhaene and Jochmans (2015), Fernández-Val (2009), Fernández-Val and Weidner (2018), Fernández-Val and Weidner (2016a), Hahn and Kuersteiner (2002), and Hahn and Newey (2004), for example.

Although the literature is well developed, its generalizations have only recently found application in substantively meaningful but technically complex contexts, such as network-type models. See Bonhomme, Lamadon, and Manresa (2019), Jochmans and Weidner (2019), Bonhomme, Jochmans, and Weidner (2025). Motivated by these developments, we propose to investigate issues of model specification in nonlinear panel data models. Like most econometric frameworks, panel data models often rely on potentially restrictive assumptions. However, to the best of our knowledge, there is currently no literature that provides a systematic approach for comparing potentially non-nested models in this context. We would like to close this gap in the literature by proposing a panel generalization of the classical Vuong (1989) test.

Specifically, we consider a model characterized by the joint log-likelihood of the data $\{z_{i,t}\}_{i \leq n, t \leq T}$ specified as:

$$L_{n,T}(\phi) \equiv \sum_{i \leq n} \sum_{t \leq T} \log f(z_{i,t}; \theta, \gamma_{g(i),m(t)}), \quad (1)$$

where $f(\cdot)$ is a known function, θ represents unknown parameters that are individual/time-invariant, $\gamma_{g(i),m(t)}$ denotes the individual and time fixed effect which has a grouping structure determined by cluster/group assignment functions $g(\cdot)$ and $m(\cdot)$ with ranges $\mathcal{G} \equiv \{1, \dots, G\}$ and $\mathcal{M} \equiv \{1, \dots, M\}$, respectively, and $\phi \equiv (\theta^\top, (\gamma_{g,m}^\top)_{g \in \mathcal{G}, m \in \mathcal{M}})^\top$ include all unknown parameters specified in the model. For ease of notations, we let I_g for $g \in \mathcal{G}$ and I_m for $m \in \mathcal{M}$ denote the partitions of $\{1, \dots, n\}$ and $\{1, \dots, T\}$ according to the group assignment functions $g(\cdot)$ and $m(\cdot)$, respectively. The pseudo true value ϕ^* is defined as the maximizer of $\mathbb{E}[L_{n,T}(\phi)]$ over the parameter space, and it is often estimated by the maximizer $\hat{\phi}$ of $L_{n,T}(\phi)$ over the same space. Our group structure is inspired the analyses by Bonhomme and Manresa (2015) or Bonhomme, Lamadon, and Manresa (2022), yet it is general enough to include the classical panel models with individual fixed effects.

In practice, different ways of specifying the likelihood function $f(\cdot)$ and the group as-

signment functions $g(\cdot)$ and $m(\cdot)$ lead to different modeling strategies in panel data models, and we often have the situation to compare competing modeling strategies and hope to find the one which is closer (or the closest) to the true data generating process in terms of the Kullback–Leibler (KL) distance. It would be reasonable to compare models through the Vuong (1989) test, which tests the null hypothesis:

$$H_0 : \mathbb{E} [L_{1,n,T}(\phi_1^*)] = \mathbb{E} [L_{2,n,T}(\phi_2^*)],$$

where $L_{j,n,T}(\cdot)$ represents the joint log-likelihood specified by modeling strategy j with ϕ_j^* denoting the pseudo true value for $j = 1, 2$. The alternative hypothesis can be

$$H_1^{2\text{-sided}} : \mathbb{E} [L_{1,n,T}(\phi_1^*)] \neq \mathbb{E} [L_{2,n,T}(\phi_2^*)] \quad \text{or} \quad H_1^{1\text{-sided}} : \mathbb{E} [L_{1,n,T}(\phi_1^*)] > \mathbb{E} [L_{2,n,T}(\phi_2^*)].$$

The classical Vuong test employs the quasi-likelihood ratio (QLR), defined as:

$$QLR_{n,T} \equiv (nT)^{-1/2}(L_{1,n,T}(\hat{\phi}_1) - L_{2,n,T}(\hat{\phi}_2)), \quad (2)$$

to construct the test statistic, where $\hat{\phi}_j$ denotes the estimator of ϕ_j^* in model j . It turns out that in panel models, the log-likelihoods $L_{1,n,T}$ and $L_{2,n,T}$ are biased due to the incidental parameter problems. Panel adaptation of the Vuong test thus requires working with bias-corrected versions of the log likelihoods, which we will call $LM_{1,n,T}^*$ and $LM_{2,n,T}^*$. We also show that there is additional complication because the order of magnitude of the standard error of the test statistic (after such modification) can depend on the nature of the hypothesis. Our technical contribution consists of estimation of $LM_{1,n,T}^*$ and $LM_{2,n,T}^*$ to make the modification feasible as well as estimation of the standard error while avoiding the uniformity issue.

We extend the classical Vuong (1989) test to panel models, contributing to the understanding of model selection complexities in high-dimensional settings. The Vuong test originally compares two non-nested parametric models and formally tests the hypothesis that their KL distances from the true data distribution are equal. Originally designed for parsimoniously parameterized models, this test has been expanded in various studies, see, Rivers and Vuong (2002), Chen, Hong, and Shum (2007), and Shi (2015), for example. In this paper, we demonstrate that the incidental parameters problem in panel data contexts invalidates the classical Vuong test, which compares maximized likelihood ratios between two models. We present a modified test statistic that ensures validity for panel data analysis.

Our modification follows a precedent in Lee and Phillips (2015) and Liao and Shi (2020). Lee and Phillips (2015) explored model selection in panel data settings using nested hypothesis testing. Our work complements theirs by introducing approaches for testing non-nested hypotheses. Liao and Shi (2020) examined the Vuong test for semiparametric models in a cross-sectional setting, whereas we extend the analysis to a panel data context with fixed effects. Our paper and Lee and Phillips (2015) share the common feature in the sense that both analyses utilize application of profiling to the Kullback–Leibler information criterion (KLIC). It is well-known that profile likelihood does not share the standard properties of the genuine likelihood function. Modification of profile likelihood was discussed as a way of overcoming this problem. See Arellano and Hahn (2007), Arellano and Hahn (2016), Cox and Reid (1987), DiCiccio and Stern (1993), DiCiccio, Martin, Stern, and Young (1996), Ferguson, Reid, and Cox (1991), and Pace and Salvan (2006), for example. Following this tradition, our paper as well as Lee and Phillips (2015) propose a modified profile likelihood that provides approximations to the standard KLIC, and provided a valid method of model selection in panel data models with individual fixed effects. Interestingly, we find some close connection between the modification in Liao and Shi (2020) and the panel type modification following Arellano and Hahn (2007), Arellano and Hahn (2016), and Lee and Phillips (2015).

Recently developed methods in panel data allow applied researchers to choose among a broad range of alternative models for separable, interactive, or discretely supported unobserved heterogeneity across units and time. It is widely understood that parameter estimates are often sensitive to the chosen specification, however there is often no clear theoretical guidance on how to choose among alternative models. Our paper makes further contribution in this regard by considering the cases where the definition of “individuals” may not be agreed across different models. In empirical practice, the group/cluster structure may not be obviously agreed among researchers. The competing group/cluster structures may be nested in the sense that one structure is “finer” than the other one, but they may be non-nested. In this situation, extension of the insights in previous papers such as Lee and Phillips (2015) or Liao and Shi (2020) may not be straightforward. We establish that the classical Vuong test is valid after some modification. In doing so, we also make a technical contribution that may be of some independent interest. Fernández-Val and Weidner (2016b) analyzed the asymptotic distribution of the (pseudo) MLE in nonlinear panel data models with individual and time effects. Under the assumption that the objective function is globally concave in all the parameters, they established the asymptotic bias formula. Our paper

contains a result¹ that relaxes this concavity assumption, although it is achieved at the cost of restricting the group structure. Therefore, our result can be argued to complement Fernández-Val and Weidner (2016b).

The remainder of the paper is organized as follows. In Section 2, we introduce the Vuong test for comparing classical panel data models that allow for potential group heterogeneity. Section 3 presents a test for comparing the two-way fixed effects (TWFE) model against a heterogeneous time fixed effects model. Section 4 concludes. The proofs of the main results, along with auxiliary lemmas, are provided in the Appendix. Additionally, Appendix C develops a general asymptotic theory for estimators in panel models with individual and time fixed effects, as well as the asymptotic distribution of the statistic $QLR_{n,T}$ in this broader setting – results that are of independent interest.²

The following notation will be adopted throughout the paper. We use K to denote a generic strictly positive constant that may vary from one instance to another but remains independent of the panel dimensions n and T . We adopt the convention that a summation over an empty set equals zero. We use $a \equiv b$ to indicate that a is defined as b . For real numbers $a_1, \dots, a_m$, $(a_j)_{j \leq m} \equiv (a_1, \dots, a_m)^\top$. For any matrix A , $A^\top$ denotes the transpose of A , and $\|A\|$ denotes the Euclidean norm of A . For any doubly indexed sequence $a_{i,t}$ (where $i = 1, \dots, n$ and $t = 1, \dots, T$), we define $\bar{a}_i \equiv T^{-1} \sum_{t \leq T} a_{i,t}$, $\bar{a}_t \equiv n^{-1} \sum_{i \leq n} a_{i,t}$ and $\bar{a} \equiv (nT)^{-1} \sum_{t \leq T} \sum_{i \leq n} a_{i,t}$. The summation $\sum_{i' \neq i}$ is taken over all i' except i , which means $\sum_{i' \neq i} a_{i',t} = \sum_{i'=1}^{i-1} a_{i',t} + \sum_{i'=i+1}^n a_{i',t}$. For two sequences of positive numbers a_n and b_n , we write $a_n \succ b_n$ if $a_n \geq c_n b_n$ for some strictly positive sequence $c_n \rightarrow \infty$. The notation $\|\cdot\|_p$ denotes the L_p -norm. Finally, for any function $g(\cdot)$, define $\mathbb{E}_T[g(z_{i,t})] \equiv T^{-1} \sum_{t \leq T} \mathbb{E}[g(z_{i,t})]$ and $\hat{\mathbb{E}}_T[g(z_{i,t})] \equiv T^{-1} \sum_{t \leq T} g(z_{i,t})$.

2 Vuong Test for Classical Panel Models

In this section, we investigate the Vuong test for panel models with time-invariant individual effects.³ With a dataset $\{z_{i,t}\}_{i \leq n, t \leq T}$, we compare two models, denoted as model 1 and model 2 respectively. Model 1 represents the classical panel model with individual effects. In contrast, model 2 may specify a different joint likelihood function and/or adopt a distinct

¹See Theorem S1 in the Online Appendix and Theorem A1 in Appendix C, along with the related discussion therein, for further details.

²The technical proofs for this general theory, together with the proofs of the auxiliary lemmas used to establish the main results in Sections 2 and 3, are provided in the Online Appendix.

³A simple version of the test appeared in Hahn and Liu (2022), which later became a chapter in Liu's UCLA dissertation. Our current paper substantially extends and generalizes the earlier simple result.

clustering or grouping structure for the individual effects.

For each model j ($j = 1, 2$), the joint log-likelihood function simplifies from the general specification in (1) to the following form:

$$L_{j,n,T}(\phi_j) \equiv \sum_{i \leq n} \sum_{t \leq T} \log f_j(z_{i,t}; \theta_j, \gamma_{g_j(i)}) = \sum_{g \in \mathcal{G}_j} \sum_{i \in I_{j,g}} \sum_{t \leq T} \psi_j(z_{i,t}; \phi_{j,g}),$$

where $\phi_j \equiv (\theta_j^\top, (\gamma_{j,g})_{g \in \mathcal{G}_j}^\top)^\top$, $\psi_j(z_{i,t}; \phi_{j,g}) \equiv \log f_j(z_{i,t}; \theta_j, \gamma_{j,g})$, $\phi_{j,g} \equiv (\theta_j^\top, \gamma_{j,g})^\top$, $\mathcal{G}_j \equiv \{1, \dots, G_j\}$ defines the range of the grouping function $g_j(\cdot)$, and $I_{j,g}$ ($g \in \mathcal{G}_j$) represents the partition of $\{1, \dots, n\}$ induced by $g_j(\cdot)$. For model 1, it follows that $\mathcal{G}_1 = \{1, \dots, n\}$, implying that each individual has its own group. For model 2, however, the partitions $\{I_{2,g}\}_{g \in \mathcal{G}_2}$ may form any arbitrary grouping of $\{1, \dots, n\}$ allowing for greater flexibility in the specification of individual effects.⁴

As this section is relatively long, we provide a roadmap to facilitate reading. Sections 2.1 - 2.4 provides the theoretical foundation of the infeasible test, while Sections 2.5 - 2.6 presents the actual algorithm for implementation of the feasible version of the test. In Section 2.1, we provide an intuitive overview of the adjustment to the maximized joint likelihood and discuss its importance for model comparison using the Vuong test in the panel data models considered in this paper. In Section 2.2, we present the log likelihood of each model. In Section 2.3, we present an infeasible adjustment that corrects for the bias in the likelihood, and in Section 2.4, we present a theoretical analysis for the asymptotic variance of the modified likelihood. Asymptotic properties of the (infeasible) modified test statistic is presented in Theorem 1. In Sections 2.5 and 2.6, we present actual algorithms for estimation of the modified log likelihoods and the asymptotic variances. Section 2.7 presents Theorem 5 that characterizes the asymptotic properties of the feasible test statistic.

2.1 Incidental Parameter Problem in Panel Data Models

To streamline notation and improve clarity, we focus here on a simplified version of (1) that includes only individual fixed effects, specified as:

$$L_{n,T}(\phi) \equiv \sum_{i \leq n} \sum_{t \leq T} \log f(z_{i,t}; \theta, \gamma_i) = \sum_{i \leq n} \sum_{t \leq T} \psi(z_{i,t}; \phi_i), \quad (3)$$

⁴This means that model 1 is nested within model 2 in terms of the grouping structure, although the likelihood specifications may differ. In many applications, the grouping structure is expected to be identical across models.

where $\psi(z_{i,t}; \phi_i) \equiv \log f(z_{i,t}; \theta, \gamma_i)$, $\phi_i \equiv (\theta^\top, \gamma_i)^\top$ and $\phi \equiv (\theta^\top, (\gamma_i)_{i \leq n}^\top)^\top$. Since QLR statistic defined in (2) involves the maximized joint likelihood $L_{n,T}(\hat{\phi})$, understanding the QLR's properties requires investigating the estimation error of $L_{n,T}(\hat{\phi})$.

Let $\phi^* \equiv (\theta^{*\top}, (\gamma_i^*)_{i \leq n}^\top)^\top$ denote the pseudo-true parameter which maximizes $\mathbb{E}[L_{n,T}(\phi)]$.⁵ Define:

$$\psi(z_{i,t}) \equiv \psi(z_{i,t}; \phi_i^*), \quad \psi_\gamma(z_{i,t}) \equiv \partial \psi(z_{i,t}; \phi_i^*) / \partial \gamma_i \quad \text{and} \quad \psi_{\gamma\gamma}(z_{i,t}) \equiv \partial^2 \psi(z_{i,t}; \phi_i^*) / \partial \gamma_i^2, \quad (4)$$

where $\phi_i^* \equiv (\theta^{*\top}, \gamma_i^{*\top})^\top$. It can be shown⁶ that under the asymptotics with $nT^{-1} \rightarrow \rho$ where $\rho \in (0, \infty)$:

$$L_{n,T}(\hat{\phi}) - \mathbb{E}[L_{n,T}(\phi^*)] = L_{n,T}(\phi^*) - \mathbb{E}[L_{n,T}(\phi^*)] + 2^{-1} \sum_{i \leq n} \tilde{\Psi}_{\gamma,i}^2 + o_p(n^{1/2}), \quad (5)$$

where $L_{n,T}(\phi^*) = \sum_{i \leq n} \sum_{t \leq T} \psi(z_{i,t})$, and for any $i \leq n$,

$$\tilde{\Psi}_{\gamma,i} \equiv T^{-1/2} \sum_{t \leq T} \tilde{\psi}_\gamma^*(z_{i,t}), \quad \text{and} \quad \tilde{\psi}_\gamma^*(z_{i,t}) \equiv \frac{\psi_\gamma(z_{i,t}) - \mathbb{E}[\psi_\gamma(z_{i,t})]}{(\mathbb{E}_T[-\psi_{\gamma\gamma}(z_{i,t})])^{1/2}}. \quad (6)$$

The expansion in (5) shows that the estimation error of $L_{n,T}(\hat{\phi})$ has two leading components. The first component, $L_{n,T}(\phi^*) - \mathbb{E}[L_{n,T}(\phi^*)]$, results from estimating $\mathbb{E}[L_{n,T}(\phi^*)]$ at the given pseudo-true parameter ϕ^* . When scaled by a factor of $(nT)^{-1/2}$, this component converges in distribution to a normal distribution under some regularity conditions. The second component, $2^{-1} \sum_{i \leq n} \tilde{\Psi}_{\gamma,i}^2$, arises from estimating the unknown pseudo-true parameter ϕ^* and is determined by the estimation error of the incidental parameters $(\gamma_i^*)_{i \leq n}$.⁷

From the expansion in (5), it is evident that in panel data models, the $L_{n,T}(\hat{\phi})$ has a first-order bias, stemming from $2^{-1} \sum_{i \leq n} \tilde{\Psi}_{\gamma,i}^2$. To evaluate this bias explicitly, suppose that $\{\tilde{\psi}_\gamma(z_{i,t})\}_{i,t}$ are i.i.d. across $t \leq T$ for any i . Since $\mathbb{E}[\tilde{\psi}_\gamma^*(z_{i,t})] = 0$, it follows that $\mathbb{E}[2^{-1} \sum_{i \leq n} \tilde{\Psi}_{\gamma,i}^2] = 2^{-1} \sum_{i \leq n} \sigma_{\gamma,i}^2$, where $\sigma_{\gamma,i}^2 \equiv \mathbb{E}[\tilde{\psi}_\gamma^*(z_{i,t})^2]$. Assuming that $\sigma_{\gamma,i}^2$ is bounded above and below, the bias from the mean of $2^{-1} \sum_{i \leq n} \tilde{\Psi}_{\gamma,i}^2$ is of the same order as $L_{n,T}(\phi^*) - \mathbb{E}[L_{n,T}(\phi^*)]$. This provides the explanation why the maximized joint likelihood $L_{n,T}(\hat{\phi})$

⁵Potential dependence of the pseudo-true parameters ϕ^* on the sample size is not made explicit for notational simplicity throughout the paper.

⁶See Theorem A2 in the Appendix, specialized to the case $\mathcal{G} = \{1, \dots, n\}$ and $\mathcal{M} = \{1\}$, so that $G = n$ and $M = 1$.

⁷Since the dimension of θ^* is fixed, the contribution of its estimation error to $L_{n,T}(\hat{\phi})$ is asymptotically negligible compared with the two leading components, and is therefore included in the remainder term of (5), represented as $o_p(n^{1/2})$.

should be modified with a bias correction based on the mean of $2^{-1} \sum_{i \leq n} \tilde{\Psi}_{\gamma,i}^2$.

The above discussion demonstrates that the maximized joint likelihood $L_{n,T}(\hat{\phi})$ requires a first-order bias correction in the form of a consistent estimator of $2^{-1} \sum_{i \leq n} \mathbb{E}[\tilde{\Psi}_{\gamma,i}^2]$. As a result, the QLR statistic in the classical Vuong test must also be adjusted with a bias correction to ensure valid inference. This bias correction for the QLR statistic is given by the difference in the bias of $L_{j,n,T}(\hat{\phi}_j)$ ($j = 1, 2$) from the two competing models.

There is an additional complication. Depending on whether the models are nested or overlapping, the first components, $L_{j,n,T}(\phi_j^*) - \mathbb{E}[L_{j,n,T}(\phi_j^*)]$, from the two models may cancel out. See Vuong (1989) for discussion in the cross section context. In such cases, the second components, denoted as $2^{-1} \sum_{i \leq n} \tilde{\Psi}_{j,\gamma,i}^2$ ($j = 1, 2$), can become the primary drivers of the asymptotic distribution of the bias-corrected QLR statistic.⁸ Consequently, the estimation error of the incidental parameters influences not only the bias but also the variance of the QLR statistic.

2.2 Quasi-likelihood Estimator

The pseudo true parameter ϕ_j^* for model j is estimated as the maximizer $\hat{\phi}_j$ of $L_{j,n,T}(\phi_j)$ where $\phi_j \equiv (\theta_j^\top, (\gamma_{j,g})_{g \in \mathcal{G}_j}^\top)^\top$. The estimator $\hat{\phi}_j$ satisfies the following first-order conditions:

$$\sum_{g \in \mathcal{G}_j} \sum_{i \in I_{j,g}} \sum_{t \leq T} \psi_{j,\theta}(z_{i,t}; \hat{\phi}_{j,g}) = 0_{d_\theta \times 1}, \quad (7)$$

and for any $g \in \mathcal{G}_j$,

$$\sum_{i \in I_{j,g}} \sum_{t \leq T} \psi_{j,\gamma}(z_{i,t}; \hat{\phi}_{j,g}) = 0, \quad (8)$$

where $\phi_{j,g} \equiv (\theta_j^\top, \gamma_{j,g})^\top$ and $\psi_{j,a}(z_{i,t}; \phi_{j,g}) \equiv \partial \psi_j(z_{i,t}; \phi_{j,g}) / \partial a$ for $a \in \{\theta_j, \gamma_{j,g}\}$. The estimator $\hat{\phi}_j$ is subsequently used to compute the maximized joint likelihood $L_{j,n,T}(\hat{\phi}_j)$ for model j .

Following the discussion in the previous subsection, the maximized joint quasi-likelihood for each model must be adjusted to account for the respective first-order bias, which can be derived from the asymptotic expansion of $L_{j,n,T}(\hat{\phi}_j)$. To present this expansion in a more general setting than previously considered, we first update the notations introduced earlier. Let $\psi_j(z_{i,t})$, $\psi_{j,\gamma}(z_{i,t})$, $\psi_{j,\gamma\gamma}(z_{i,t})$ and $\tilde{\Psi}_{j,\gamma,i}$ represent the counterparts to their definitions in

⁸In the classical nested case, where the functional form of the likelihood f is identical across the two competing models, the second component of the test statistic cancels out. In this setting, bias correction of the estimator itself, along with further refinements of (5), is needed. This case was previously studied by Lee and Phillips (2015), and we do not pursue it further in this paper.

(4) and (6), with $\phi_{j,i} \equiv \phi_{j,g}$,

$$\psi_{j,\gamma}^*(z_{i,t}) \equiv \psi_{j,\gamma}(z_{i,t}) - \mathbb{E}[\psi_{j,\gamma}(z_{i,t})] \quad \text{and} \quad \tilde{\psi}_{j,\gamma}^*(z_{i,t}) \equiv \frac{\psi_{j,\gamma}^*(z_{i,t})}{(n_{j,g}^{-1} \sum_{i \in I_{j,g}} \mathbb{E}_T[-\psi_{j,\gamma}(z_{i,t})])^{1/2}},$$

for any $i \in I_{j,g}$ and any $g \in \mathcal{G}_j$.⁹

2.3 Infeasible Modified QLR Statistic - Bias

With the updated notations, we can obtain¹⁰ the following expansion:

$$L_{j,n,T}(\hat{\phi}_j) = T^{1/2} \sum_{i \leq n} (\tilde{\Psi}_{j,i} + \tilde{V}_{j,i} + \tilde{U}_{j,i}) + o_p(G_j^{1/2}), \quad (9)$$

where the components are defined¹¹ as:

$$\tilde{\Psi}_{j,i} \equiv T^{-1/2} \sum_{t \leq T} \psi_j(z_{i,t}; \phi_{j,i}^*), \quad \tilde{V}_{j,i} \equiv \frac{\tilde{\Psi}_{j,\gamma,i}^2}{2n_{j,g}T^{1/2}} \quad \text{and} \quad \tilde{U}_{j,i} \equiv \sum_{\{i' \in I_{j,g}: i' < i\}} \frac{\tilde{\Psi}_{j,\gamma,i} \tilde{\Psi}_{j,\gamma,i'}}{n_{j,g}T^{1/2}}.$$

It is reasonable to expect $\mathbb{E}[\tilde{U}_{j,i}] = 0$ for $j = 1, 2$ and for any $i \leq n$,¹² which implies that the first-order bias arises from the mean of $\tilde{V}_{j,i}$. To account for this bias, we define the (infeasible) modified maximized joint likelihood as:

$$LM_{j,n,T}^*(\hat{\phi}_j) \equiv L_{j,n,T}(\hat{\phi}_j) - T^{1/2} \sum_{i \leq n} \mathbb{E}[\tilde{V}_{j,i}]. \quad (10)$$

Using this expression, we define the (infeasible) modified QLR statistic, whose asymptotic expansion follows from (9):

$$MQLR_{n,T}^* \equiv (nT)^{-1/2} \left(LM_{1,n,T}^*(\hat{\phi}_1) - LM_{2,n,T}^*(\hat{\phi}_2) \right) = n^{-1/2} \sum_{i \leq n} (\tilde{\Psi}_i + \tilde{V}_i^* + \tilde{U}_i) + o_p(T^{-1/2}), \quad (11)$$

⁹The demeaned version of $\psi_{j,\gamma}(z_{i,t})$, i.e., $\psi_{j,\gamma}^*(z_{i,t})$ is introduced here because the first-order condition for ϕ_j^* only ensures that $\sum_{i \in I_{j,g}} \sum_{t \leq T} \mathbb{E}[\psi_{j,\gamma}(z_{i,t})] = 0$. This condition does not guarantee that $\mathbb{E}[\psi_{j,\gamma}(z_{i,t})] = 0$. If $\psi_{j,\gamma}(z_{i,t})$ is stationary across t for each i , the first-order condition for ϕ_j^* is simplified to $\sum_{i \in I_{j,g}} \mathbb{E}[\psi_{j,\gamma}(z_{i,t})] = 0$. In this case, $\mathbb{E}[\psi_{1,\gamma}(z_{i,t})] = 0$ holds for model 1, but is still not guaranteed for model 2.

¹⁰See Theorem A2 together with the decomposition in (99) in the Appendix.

¹¹The $\tilde{\Psi}_{j,\gamma,i}$ is not the derivative of $\tilde{\Psi}_{j,i}$, as can be seen from its definition (6).

¹²For example, it is implied by Assumption 1 in the Appendix.

where $\tilde{\Psi}_i \equiv \tilde{\Psi}_{1,i} - \tilde{\Psi}_{2,i}$, $\tilde{V}_i^* \equiv \tilde{V}_{1,i} - \tilde{V}_{2,i} - \mathbb{E}[\tilde{V}_{1,i} - \tilde{V}_{2,i}]$ and $U_i \equiv \tilde{U}_{1,i} - \tilde{U}_{2,i}$.

2.4 Infeasible Modified QLR Statistic - Asymptotic Distribution

After bias correction, the main remaining components arising from the estimation errors of the incidental parameters, i.e., $\tilde{V}_i^* + \tilde{U}_i$ primarily affect the variance of the modified QLR statistic $MQLR_{n,T}^*$. This effect becomes particularly significant when the two models are overlapping. In such cases, the first component on the right-hand side of (9), i.e., $n^{-1/2} \sum_{i \leq n} \tilde{\Psi}_i$, can become arbitrarily small or even vanish, leaving the asymptotic distribution of $MQLR_{n,T}^*$ predominantly determined by the estimation errors of the incidental parameters, specifically $n^{-1/2} \sum_{i \leq n} (\tilde{V}_i^* + \tilde{U}_i)$.

To facilitate the discussion of the variance and asymptotic distribution of $MQLR_{n,T}^*$, we follow the literature (see, e.g., Vuong (1989), Shi (2015) and Liao and Shi (2020)) and formally define the relationships between models in the panel setting below.

Definition 1. (i) Model 1 and model 2 are **strictly non-nested** if there do not exist $\phi_{1,i} \in \Phi_1$ and $\phi_{2,i} \in \Phi_{12}$ such that

$$\psi_1(z; \phi_{1,i}) = \psi_2(z; \phi_{2,i}), \quad (12)$$

for any i and any z in the support of $z_{i,t}$; (ii) the two models are **overlapping** if they are not strictly non-nested; (iii) model 1 and model 2 are said to be **nested** if, for any $\phi_{j,i} \in \Phi_j$, there exists a $\phi_{j',i} \in \Phi_{j'}$ (where $j, j' = 1, 2$ and $j \neq j'$) such that (12) holds.

When the two models are strictly non-nested, the variance of $n^{-1/2} \sum_{i \leq n} \tilde{\Psi}_i$ is bounded away from zero. In this case, the component $n^{-1/2} \sum_{i \leq n} (\tilde{V}_i^* + \tilde{U}_i)$ on the right-hand side of (9) becomes asymptotically negligible, as it is of order $O_p(T^{-1/2})$ under certain regularity conditions.¹³ Consequently, the asymptotic distribution of $MQLR_{n,T}^*$ is primarily determined by $n^{-1/2} \sum_{i \leq n} \tilde{\Psi}_i$. Conversely, when the two models are nested, $\tilde{\Psi}_i = 0$ under the null hypothesis. In this scenario, the asymptotic distribution of $MQLR_{n,T}^*$ is determined by $n^{-1/2} \sum_{i \leq n} (\tilde{V}_i^* + \tilde{U}_i)$. As a result, deriving the asymptotic distribution of $MQLR_{n,T}^*$ is relatively straightforward, when the models are known to be either strictly non-nested or nested. In both cases, the dominating term on the right-hand side of (9) is identifiable, and the asymptotic distribution of $MQLR_{n,T}^*$ can be deduced from that of the dominating term.

¹³This order follows from an approximation of $n^{-1} \sum_{i \leq n} (\text{Var}(\tilde{V}_i) + \text{Var}(\tilde{U}_i))$ (see (S238) in the Online Appendix, which appears in the proof of Lemma S16), together with an upper bound on the variance of $\tilde{\psi}_{j,\gamma}^*(z_{i,t})$ for $j = 1, 2$.

The asymptotic distribution of $QLR_{n,T}^*$ is more challenging to derive when the two models are overlapping. Because in this case, it is in general unclear which components between $n^{-1/2} \sum_{i \leq n} \tilde{\Psi}_i$ and $n^{-1/2} \sum_{i \leq n} (\tilde{V}_i^* + \tilde{U}_i)$ could be the dominating term, and both could be equally important. To address this issue, we follow Liao and Shi (2020) to consider both terms together when deriving the asymptotic distribution of $MQLR_{n,T}^*$.

Theorem 1. *Under Assumptions 1 and 2 in the Appendix, we have*

$$\frac{MQLR_{n,T}^* - \overline{QLR}_{n,T}}{\omega_{n,T}} \rightarrow_d N(0, 1), \quad (13)$$

where

$$\omega_{n,T}^2 \equiv n^{-1} \sum_{i \leq n} \text{Var}(\tilde{\Psi}_i + \tilde{V}_i + \tilde{U}_i) \quad \text{and} \quad \overline{QLR}_{n,T} \equiv (nT)^{-1/2} \mathbb{E}[L_{1,n,T}(\phi_1^*) - L_{1,n,T}(\phi_2^*)]. \quad (14)$$

Remark 1. *Theorem 1 provides the asymptotic distribution of $MQLR_{n,T}^*$ under both the null and alternative hypotheses. Since $\overline{QLR}_{n,T} = 0$ under the null hypothesis, we can construct a test for the null by using a consistent estimator of $\omega_{n,T}$, along with a feasible version of the modified QLR statistic. This feasible statistic replaces the unknown bias correction terms in $MQLR_{n,T}^*$ with their consistent estimators.¹⁴*

2.5 Estimation of Bias

We first discuss the estimation of the bias correction terms. Under Assumption 3 in the Appendix, the definition of $\tilde{V}_{j,i}$ implies that:

$$T^{1/2} \sum_{i \leq n} \mathbb{E}[\tilde{V}_{j,i}] = \sum_{g \in \mathcal{G}_j} \sum_{i \in I_{j,g}} (2n_{j,g})^{-1} \sigma_{j,\gamma,i}^2, \quad \text{where } \sigma_{j,\gamma,i} \equiv (\mathbb{E}_T[\tilde{\psi}_{j,\gamma}^*(z_{i,t})^2])^{1/2}. \quad (15)$$

This motivates a straightforward plug-in estimator of the bias correction term for model j , relying on a consistent estimator of $\sigma_{j,\gamma,i}^2$ for any $i \in I_{j,g}$ and any $g \in \mathcal{G}_j$:

$$\hat{\sigma}_{j,\gamma,i}^2 \equiv \frac{\hat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})^2] - (\hat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})])^2}{|\hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})|}, \quad (16)$$

¹⁴We impose Assumption 3 to simplify the estimation of the bias correction and the variance of $MQLR_{n,T}^*$. All assumptions are collected in the Appendix.

where

$$\hat{\psi}_{j,\gamma}(z_{i,t}) \equiv \psi_{j,\gamma}(z_{i,t}; \hat{\phi}_{j,g}) \quad \text{and} \quad \hat{\Psi}_{j,\gamma\gamma}(\phi_{j,g}) \equiv n_{j,g}^{-1} \sum_{i \in I_{j,g}} \hat{\mathbb{E}}_T[\psi_{j,\gamma\gamma}(z_{i,t}; \phi_{j,g})]. \quad (17)$$

The bias correction term for model j is then estimated as

$$R_{j,n,T}(\hat{\phi}_j) \equiv \sum_{g \in \mathcal{G}_j} \sum_{i \in I_{j,g}} (2n_{j,g})^{-1} \hat{\sigma}_{j,\gamma,i}^2. \quad (18)$$

Using this, the feasible modified QLR statistic is defined as:

$$MQLR_{n,T} \equiv (nT)^{-1/2} \left(LM_{1,n,T}(\hat{\phi}_1) - LM_{2,n,T}(\hat{\phi}_2) \right), \quad (19)$$

where

$$LM_{j,n,T}(\hat{\phi}_j) \equiv L_{j,n,T}(\hat{\phi}_j) - R_{j,n,T}(\hat{\phi}_j)$$

denotes the modified maximized joint likelihood for model j .

Theorem 2. *Under Assumptions 1, 2 and 3 in the Appendix, we have*

$$\frac{R_{j,n,T}(\hat{\phi}_j) - T^{1/2} \sum_{i \leq n} \mathbb{E}[\tilde{V}_{j,i}]}{(nT)^{1/2} \omega_{n,T}} = O_p(T^{-1/2}).$$

Theorem 2 establishes the consistency of $R_{j,n,T}(\hat{\phi}_j)$ as an estimator of the bias of the maximized joint likelihood $L_{j,n,T}(\hat{\phi}_j)$ in model j . Combined with the result in (13), this directly leads to the asymptotic normality of the modified QLR statistic $QLR_{n,T}$:

$$\frac{MQLR_{n,T} - \overline{QLR}_{n,T}}{\omega_{n,T}} \rightarrow_d N(0, 1). \quad (20)$$

To apply this result for inference under the null hypothesis, it remains to construct a consistent estimator for $\omega_{n,T}^2$, which we now proceed to address.

2.6 Estimation of Variance

A consistent estimator of $\omega_{n,T}^2$, which was defined in (14), can be constructed using the sample analog of the variance of $n^{-1/2} \sum_{i \leq n} \tilde{\Psi}_i$, and a variance correction term to account for estimation errors of the incidental parameters. Define $\sigma_{n,T}^2 \equiv n^{-1} \sum_{i \leq n} \text{Var}(\tilde{\Psi}_i)$ and $\Delta\psi(z_{i,t}, \phi_i) \equiv \psi_1(z_{i,t}; \phi_{1,i}) - \psi_2(z_{i,t}; \phi_{2,i})$. Under Assumption 3, a natural estimator for $\sigma_{n,T}^2$

is given by:

$$\hat{\sigma}_{n,T}^2 \equiv (nT)^{-1} \sum_{i \leq n} \sum_{t \leq T} \Delta\psi(z_{i,t}; \hat{\phi}_i)^2 - (nT)^{-1} (MQLR_{n,T})^2. \quad (21)$$

The theoretical property of $\hat{\sigma}_{n,T}^2$ is provided in the lemma below.¹⁵

Theorem 3. *Under Assumptions 1, 2 and 3 in the Appendix, we have:*

$$\frac{\hat{\sigma}_{n,T}^2 - (\sigma_{n,T}^2 + 2\sigma_{S,n,T}^2)}{\omega_{n,T}^2} = O_p(T^{-1/2}), \quad (22)$$

where

$$\sigma_{S,n,T}^2 \equiv (2nT)^{-1} \sum_{g \in \mathcal{G}_2} \sum_{i \in I_{2,g}} \left(\sigma_{1,\gamma,i}^4 + n_{2,g}^{-2} \sigma_{2,\gamma,i}^2 \sum_{i' \in I_{2,g}} s_{2,\gamma,i'}^2 - 2n_{2,g}^{-1} \sigma_{12,\gamma,i}^2 \right), \quad (23)$$

$s_{2,\gamma,i}^2 \equiv \mathbb{E}_T[\tilde{\psi}_{2,\gamma}(z_{i,t})^2]$ and $\sigma_{12,\gamma,i} \equiv \mathbb{E}_T[\tilde{\psi}_{1,\gamma}^*(z_{i,t}) \tilde{\psi}_{2,\gamma}^*(z_{i,t})]$ for any $i \in I_{2,g}$ and any $g \in \mathcal{G}_2$.¹⁶

Theorem 3 shows that $\hat{\sigma}_{n,T}^2$ tends to overestimate the true variance $\sigma_{n,T}^2$ by $2\sigma_{S,n,T}^2$. It is reasonable to expect the latter to be of order $O(T^{-1})$, which is formalized by Assumptions 2(iii) and 9. On the other hand, it can be shown¹⁷ that

$$\frac{\omega_{n,T}^2 - (\sigma_{n,T}^2 + \sigma_{U,n,T}^2)}{\omega_{n,T}^2} = O(T^{-1/2}), \quad (24)$$

where

$$\sigma_{U,n,T}^2 \equiv (2nT)^{-1} \sum_{g \in \mathcal{G}_2} \sum_{i \in I_{2,g}} \left(\sigma_{1,\gamma,i}^4 + n_{2,g}^{-2} \sigma_{2,\gamma,i}^2 \sum_{i' \in I_{2,g}} \sigma_{2,\gamma,i'}^2 - 2n_{2,g}^{-1} \sigma_{12,\gamma,i}^2 \right), \quad (25)$$

which represents the leading term in the variance of $n^{-1/2} \sum_{i \leq n} (\tilde{V}_i^* + \tilde{U}_i)$.

Comparing their expressions in (23) and (25), it is evident that $\sigma_{S,n,T}^2 \geq \sigma_{U,n,T}^2$ in general, with the inequality becoming strict when $\mathbb{E}[\tilde{\psi}_{j,\gamma}(z_{i,t})] \neq 0$. From (22) and (24), it follows

¹⁵The sample variance $\hat{\sigma}_{n,T}^2$ will be used below as a building block for variance estimation and therefore requires $\{\Delta\psi(z_{i,t}, \phi_i^*)\}_t$ to be serially uncorrelated, as assumed in Assumption 3. If serial correlation is present, an autocorrelation-consistent estimator is required. Since variance estimation is mainly needed under the null for size control, Assumption 3 is imposed only under the null; Lemma 1 in the Appendix shows that consistency of the test does not rely on this assumption.

¹⁶Since $\mathbb{E}_T[\tilde{\psi}_{1,\gamma}(z_{i,t})] = 0$ by the first-order condition for ϕ_1^* , we have $\sigma_{12,\gamma,i} = \mathbb{E}_T[\tilde{\psi}_{1,\gamma}(z_{i,t}) \tilde{\psi}_{2,\gamma}(z_{i,t})]$ for any $i \leq n$.

¹⁷See Lemma S16 in the Online Appendix.

that $\hat{\sigma}_{n,T}^2$ is a consistent estimator of $\omega_{n,T}^2$ when $\sigma_{n,T}^2$ dominates $\sigma_{S,n,T}^2$. Since $\sigma_{S,n,T}^2 = O(T^{-1})$, this dominance is guaranteed when the two models are strictly non-nested, as $\sigma_{n,T}^2$ is bounded away from zero in such cases. Conversely, if $\sigma_{n,T}^2$ is of the same or even smaller order than $\sigma_{S,n,T}^2$, a scenario that arises when the two models are (nearly) nested or overlapping, the result in (22) implies that $\hat{\sigma}_{n,T}^2$ overestimates $\omega_{n,T}^2$ by $2\sigma_{S,n,T}^2 - \sigma_{U,n,T}^2$, and becomes an inconsistent estimator of $\omega_{n,T}^2$.

The over-estimation issue in $\hat{\sigma}_{n,T}^2$ can be adjusted through consistent estimators of $\sigma_{U,n,T}^2$ and $\sigma_{S,n,T}^2$. From its expression in (25), $\sigma_{U,n,T}^2$ can be estimated by replacing $\sigma_{j,\gamma,i}^2$ with $\hat{\sigma}_{j,\gamma,i}^2$ defined in (16), and replacing $\sigma_{12,\gamma,i}^2$ by

$$\hat{\sigma}_{12,\gamma,i}^2 \equiv \frac{(\hat{\mathbb{E}}_T[\hat{\psi}_{1,\gamma}(z_{i,t})\hat{\psi}_{2,\gamma}(z_{i,t})])^2}{|\hat{\Psi}_{1,\gamma\gamma}(\hat{\phi}_{1,g})\hat{\Psi}_{2,\gamma\gamma}(\hat{\phi}_{2,g})|}.$$

Specifically, the estimator of $\sigma_{U,n,T}^2$ is defined as

$$\hat{\sigma}_{U,n,T}^2 \equiv (2nT)^{-1} \sum_{g \in \mathcal{G}_2} \sum_{i \in I_{2,g}} \left(\hat{\sigma}_{1,\gamma,i}^4 + n_{2,g}^{-2} \hat{\sigma}_{2,\gamma,i}^2 \sum_{i' \in I_{2,g}} \hat{\sigma}_{2,\gamma,i'}^2 - 2n_{2,g}^{-1} \hat{\sigma}_{12,\gamma,i}^2 \right). \quad (26)$$

Similarly, we define the estimator of $\sigma_{S,n,T}^2$ as

$$\hat{\sigma}_{S,n,T}^2 \equiv (2nT)^{-1} \sum_{g \in \mathcal{G}_2} \sum_{i \in I_{2,g}} \left(\hat{\sigma}_{1,\gamma,i}^4 + n_{2,g}^{-2} \hat{\sigma}_{2,\gamma,i}^2 \sum_{i' \in I_{2,g}} \hat{s}_{2,\gamma,i'}^2 - 2n_{2,g}^{-1} \hat{\sigma}_{12,\gamma,i}^2 \right). \quad (27)$$

Here, for any $i \in I_{2,g}$ and any $g \in \mathcal{G}_2$:

$$\hat{s}_{2,\gamma,i}^2 \equiv \frac{\hat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})^2]}{|\hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})|}. \quad (28)$$

Thus, $\omega_{n,T}^2$ can be consistently estimated in a general setting as $\hat{\sigma}_{n,T}^2 + \hat{\sigma}_{U,n,T}^2 - 2\hat{\sigma}_{S,n,T}^2$, as indicated by Theorem 4 below.

Theorem 4. *Under Assumptions 1, 2 and 3 in the Appendix, we have:*

$$\frac{\hat{\sigma}_{U,n,T}^2 - \sigma_{U,n,T}^2}{\omega_{n,T}^2} = O_p(n^{-1/2}) \quad \text{and} \quad \frac{\hat{\sigma}_{S,n,T}^2 - \sigma_{S,n,T}^2}{\omega_{n,T}^2} = O_p(n^{-1/2}). \quad (29)$$

One drawback of using $\hat{\sigma}_{n,T}^2 - 2\hat{\sigma}_{S,n,T}^2 + \hat{\sigma}_{U,n,T}^2$ as a variance estimator is that it may take

negative values in finite samples. To address this issue, we follow Liao and Shi (2020) and propose the following hybrid variance estimator:

$$\hat{\omega}_{n,T}^2 = \max \{ \hat{\sigma}_{n,T}^2 + \hat{\sigma}_{U,n,T}^2 - 2\hat{\sigma}_{S,n,T}^2, \hat{\sigma}_{U,n,T}^2 \}. \quad (30)$$

Remark 2. By definition, $\hat{\sigma}_{U,n,T}^2$ can be expressed as

$$\begin{aligned} \hat{\sigma}_{U,n,T}^2 &= (2nT)^{-1} \sum_{g \in \mathcal{G}_2} \sum_{i \in I_{2,g}} (\hat{\sigma}_{1,\gamma,i}^4 - 2n_{2,g}^{-1} \hat{\sigma}_{12,\gamma,i}^2 + n_{2,g}^{-2} \hat{\sigma}_{2,\gamma,i}^4) \\ &\quad + (2nT)^{-1} \sum_{g \in \mathcal{G}_2} n_{2,g}^{-2} \sum_{i \in I_{2,g}} \sum_{i' \in I_{2,g}, i' \neq i} \hat{\sigma}_{2,\gamma,i}^2 \hat{\sigma}_{2,\gamma,i'}^2. \end{aligned} \quad (31)$$

Since $\hat{\sigma}_{12,\gamma,i}^2 \leq \hat{\sigma}_{1,\gamma,i}^2 \hat{\sigma}_{2,\gamma,i}^2$ for any i , the first term on the right-hand side of (31) is non-negative. Additionally, the second term after the equality in (31) is also non-negative by definition. Therefore, $\hat{\sigma}_{U,n,T}^2$ is guaranteed to be non-negative, and by construction, $\hat{\omega}_{n,T}^2$ is also non-negative. This ensures the practical reliability of $\hat{\omega}_{n,T}^2$ as a variance estimator.

2.7 Feasible Modified QLR Statistic and Generalized Vuong Test

Given the variance estimator $\hat{\omega}_{n,T}^2$, the two-sided and one-sided Vuong tests are defined as:

$$\begin{aligned} \varphi_{n,T}^{2\text{-side}}(p) &\equiv 1 \{ |MQLR_{n,T}| > \hat{\omega}_{n,T} z_{1-p/2} \}, \\ \varphi_{n,T}^{1\text{-side}}(p) &\equiv 1 \{ MQLR_{n,T} > \hat{\omega}_{n,T} z_{1-p} \}, \end{aligned} \quad (32)$$

respectively, where z_{1-p} is the $1 - p$ quantile of the standard normal distribution, and $p \in (0, 1)$ denotes the significance level.

Theorem 5. Suppose that Assumptions 1, 2 and 3 in the Appendix hold. Further suppose that as $n, T \rightarrow \infty$,

$$\frac{\overline{QLR}_{n,T}}{\omega_{n,T}} \rightarrow c, \quad (33)$$

for some constant c in the extended real line. Then as $n, T \rightarrow \infty$,

$$\mathbb{E}[\varphi_{n,T}^{2\text{-side}}(p)] \rightarrow 2 - \Phi(z_{1-p/2} - c) - \Phi(z_{1-p/2} + c), \quad (34)$$

and

$$\mathbb{E}[\varphi_{n,T}^{1\text{-side}}(p)] \rightarrow 1 - \Phi(z_{1-p} - c), \quad (35)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal.

Remark 3. Theorem 5 establishes both the size and power properties of the Vuong test defined in (32). Under the null hypothesis where $\overline{QLR}_{n,T} = 0$, (33) holds with $c = 0$. In this scenario, (34) and (35) imply that

$$\mathbb{E}[\varphi_{n,T}^{2-side}(p)] \rightarrow 2(1 - \Phi(z_{1-p/2})) = p \quad \text{and} \quad \mathbb{E}[\varphi_{n,T}^{1-side}(p)] \rightarrow p, \quad (36)$$

which establishes the size control over data generating processes in which the two models being compared are either nested or overlapping. Furthermore, (34) demonstrates that the two-sided Vuong test has power against local alternatives with $|c| \in (0, \infty)$, and is consistent against any fixed alternatives, while the one-sided Vuong test share the similar power properties with $c \in (0, \infty)$.

3 Testing Heterogeneous Time Effects vs. TWFE

The general result in the previous section was based on Theorems A1 – A3 in Appendix C. As discussed in the introduction, the result established in Appendix C is a technical contribution that relaxes the concavity assumption in Fernández-Val and Weidner (2016b), but it was done at the cost of restricting the group structure. In this section, we consider potentially more complicated group structure but we do so in the context of the linear model, which satisfies the concavity assumption. Because of substantive economic interest associated with the linear model, in particular, the literature initiated by Bonhomme and Manresa (2015), we believe that the linear model may deserve a special attention.

Bonhomme and Manresa (2015) have recently proposed a new model that extends beyond the scope of traditional panel data models. In light of its computational demands, one may naturally question whether a standard TWFE specification could offer a practical alternative. Obviously the two models are not nested, so it is of interest to find a Vuong test for this comparison.

We consider the linear panel model

$$y_{i,t} = x_{i,t}^\top \beta + \gamma_{g(i),m(t)} + \varepsilon_{i,t},$$

where $y_{i,t}$ is the dependent variable, $x_{i,t}$ are the observed regressors, and $\varepsilon_{i,t}$ is i.i.d. across i and t and with mean zero and variance σ^2 . Here, the $g(\cdot)$ and $m(\cdot)$ are *known* cluster/group assignment functions with range $\mathcal{G} = \{1, \dots, G\}$ and $\mathcal{M} = \{1, \dots, M\}$ respectively, and

$\gamma_{g(i),m(t)}$ denotes the unknown fixed effect.¹⁸ In the main text of the paper, we focus on the Vuong test comparing the heterogeneous time effects model (as considered by Bonhomme and Manresa (2015)) and the TWFE model, denoted as model 1 and model 2, respectively.

Specifically, model 1 defines the joint log-likelihood as

$$\sum_{g \in \mathcal{G}_1} \sum_{i \in I_g} \sum_{t \leq T} \psi(z_{i,t}; \theta_1, \gamma_{1,g,t}), \quad (37)$$

where $\mathcal{G}_1 \equiv \{1, \dots, G_1\}$ and G_1 is a positive integer. The function $\psi(z_{i,t}; \theta, \gamma)$ is given by

$$\psi(z_{i,t}; \theta, \gamma) = (-2)^{-1} (y_{i,t} - x_{i,t}^\top \theta - \gamma)^2. \quad (38)$$

In contrast, model 2 assumes that the joint log-likelihood is

$$\sum_{i \leq n} \sum_{t \leq T} \psi(z_{i,t}; \theta_2, \gamma_{2,i} + \gamma_{2,t}), \quad (39)$$

where $\gamma_{2,i}$ represents the time-invariant individual effect, and $\gamma_{2,t}$ captures the homogeneous time effect. We impose the following normalization condition on model 2:

$$\sum_{t \leq T} \gamma_{2,t} = 0, \quad (40)$$

to ensure unique identification of the pseudo-true parameters.

In the following, we introduce the algorithm for calculating the modified QLR statistic and explain its asymptotic properties. Even though the technical details are more involved and use more complex notation, the results in this section are presented in a way quite similar to those in the previous one. To make things easier to follow, we have organized this section in a way that mirrors the structure of the (second half of the) previous section.

¹⁸The $\gamma_{g(i),m(t)}$ takes different forms under various model specifications. For example, in models involving time invariant individual fixed effects only, we have $\gamma_{g(i),m(t)} = \gamma_i$, e.g., with $\mathcal{G} = \{1, \dots, n\}$, and $\mathcal{M} = \{1\}$. In models involving only time fixed effects, we have $\gamma_{g(i),m(t)} = \gamma_t$, e.g., with $\mathcal{G} = \{1\}$, and $\mathcal{M} = \{1, \dots, T\}$. In the Section 4 of the Online Appendix, we consider non-nested hypothesis testing comparing several variants of the linear model.

3.1 Estimation under Group-Time Effects and TWFE

The pseudo true parameters θ_1^* and $\gamma_{1,g,t}^*$ (for $g \in \mathcal{G}_1$ and $t \leq T$) for model 1, defined as the maximizers of the population version of the objective in (37), take the following forms:

$$\theta_1^* = \frac{\Sigma_{\dot{x}}^{-1}}{nT} \sum_{i \leq n} \sum_{t \leq T} \mathbb{E}[\dot{x}_{i,t}^* y_{i,t}] \quad \text{and} \quad \gamma_{1,g,t}^* = \mathbb{E}[\bar{y}_{g,t}] - \mathbb{E}[\bar{x}_{g,t}^\top] \theta_1^*,$$

where $\Sigma_{\dot{x}} \equiv (nT)^{-1} \sum_{i \leq n} \sum_{t \leq T} \mathbb{E}[\dot{x}_{i,t}^* \dot{x}_{i,t}^{*\top}]$, $\dot{x}_{i,t}^* = x_{i,t} - \mathbb{E}[\bar{x}_{g,t}]$, and for any $i \in I_g$ and any $g \in \mathcal{G}_1$, $\bar{x}_{g,t} \equiv n_g^{-1} \sum_{i \in I_g} x_{i,t}$ and $\bar{y}_{g,t} \equiv n_g^{-1} \sum_{i \in I_g} y_{i,t}$. The estimators of θ_1^* and $\gamma_{1,g,t}^*$ are their sample analogs:

$$\hat{\theta}_1 = \frac{\hat{\Sigma}_{\dot{x}}^{-1}}{nT} \sum_{i \leq n} \sum_{t \leq T} \dot{x}_{i,t} y_{i,t} \quad \text{and} \quad \hat{\gamma}_{1,g,t} = \bar{y}_{g,t} - \bar{x}_{g,t}^\top \hat{\theta}_1, \quad (41)$$

where $\dot{x}_{i,t} \equiv x_{i,t} - \bar{x}_{g,t}$ and $\hat{\Sigma}_{\dot{x}} \equiv (nT)^{-1} \sum_{i \leq n} \sum_{t \leq T} \dot{x}_{i,t} \dot{x}_{i,t}^\top$.

Applying (40), we can solve the pseudo true parameters θ_2^* , $(\gamma_{2,i}^*)_{i \leq n}$ and $(\gamma_{2,t}^*)_{t \leq T}$ in model 2 as

$$\theta_2^* = \frac{\Sigma_{\ddot{x}}^{-1}}{nT} \sum_{i \leq n} \sum_{t \leq T} \mathbb{E}[\ddot{x}_{i,t}^* y_{i,t}], \quad \gamma_{2,i}^* = \mathbb{E}[\bar{y}_i] - \mathbb{E}[\bar{x}_i^\top] \theta_2^* \quad \text{and} \quad \gamma_{2,t}^* = \mathbb{E}[\bar{y}_t - \bar{y}] - \mathbb{E}[\bar{x}_t^\top - \bar{x}^\top] \theta_2^*,$$

where $\Sigma_{\ddot{x}} \equiv (nT)^{-1} \sum_{i \leq n} \sum_{t \leq T} \mathbb{E}[\ddot{x}_{i,t}^* \ddot{x}_{i,t}^{*\top}]$ and $\ddot{x}_{i,t}^* \equiv x_{i,t} - \mathbb{E}[\bar{x}_i - \bar{x}_t + \bar{x}]$. The estimators of the pseudo true parameters are their empirical counterparts:

$$\hat{\theta}_2 = \frac{\hat{\Sigma}_{\ddot{x}}^{-1}}{nT} \sum_{i \leq n} \sum_{t \leq T} \ddot{x}_{i,t} y_{i,t}, \quad \hat{\gamma}_{2,i} = \bar{y}_i - \bar{x}_i^\top \hat{\theta}_2 \quad \text{and} \quad \hat{\gamma}_{2,t} = (\bar{y}_t - \bar{y}) - (\bar{x}_t - \bar{x})^\top \hat{\theta}_2, \quad (42)$$

where $\ddot{x}_{i,t} \equiv x_{i,t} - \bar{x}_i - \bar{x}_t + \bar{x}$ and $\hat{\Sigma}_{\ddot{x}} \equiv (nT)^{-1} \sum_{i \leq n} \sum_{t \leq T} \ddot{x}_{i,t} \ddot{x}_{i,t}^\top$.

3.2 Asymptotic Distribution of the QLR Statistic

Using the estimators of the pseudo true parameters from both models, we obtain the QLR statistic for model comparison as

$$QLR_{n,T} = (nT)^{-1/2} \left(\sum_{i \leq n} \sum_{t \leq T} \frac{\hat{\varepsilon}_{1,i,t}^2}{-2} - \sum_{i \leq n} \sum_{t \leq T} \frac{\hat{\varepsilon}_{2,i,t}^2}{-2} \right)$$

where $\hat{\varepsilon}_{j,i,t} \equiv y_{i,t} - x_{i,t}^\top \hat{\theta}_j - \hat{\gamma}_{j,i,t}$ for $j = 1, 2$, $\hat{\gamma}_{1,i,t} \equiv \hat{\gamma}_{1,g,t}$ for any $i \in I_g$ and any $g \in \mathcal{G}_1$, and $\hat{\gamma}_{2,i,t} \equiv \hat{\gamma}_{2,i} + \hat{\gamma}_{2,t}$ for any $i \leq n$ and any $t \leq T$.

We next derive the asymptotic distribution of $QLR_{n,T}$. Some notation are needed. Let $\varepsilon_{j,i,t} \equiv y_{i,t} - x_{i,t}^\top \theta_j^* - \gamma_{j,i,t}^*$ and $\varepsilon_{j,i,t}^* \equiv \varepsilon_{j,i,t} - \mathbb{E}[\varepsilon_{j,i,t}]$ for $j = 1, 2$, where $\gamma_{1,i,t}^* \equiv \gamma_{1,g,t}^*$ for any $i \in I_g$ and any $g \in \mathcal{G}_1$, and $\gamma_{2,i,t}^* \equiv \gamma_{2,i}^* + \gamma_{2,t}^*$ for any $i \leq n$ and any $t \leq T$. Without loss of generality, we suppose that for any $g \in \mathcal{G}_1$, $I_g = \{N_g + 1, \dots, N_g + n_g\}$ where $N_g \equiv \sum_{g' < g} n_{g'}$.

For model 1, we define

$$\tilde{V}_{1,i} = \frac{\sum_{t \leq T} \varepsilon_{1,i,t}^{*2}}{2n_g T^{1/2}} \quad \text{and} \quad \tilde{U}_{1,i} = \frac{\sum_{i'=N_g+1}^{i-1} \sum_{t \leq T} \varepsilon_{1,i',t}^* \varepsilon_{1,i,t}^*}{n_g T^{1/2}},$$

for any $g \in \mathcal{G}_1$ and any $i \in I_g$. For model 2, we let

$$\tilde{V}_{2,i} = \frac{(\sum_{t \leq T} \varepsilon_{2,i,t}^*)^2}{2T^{3/2}} + \frac{\sum_{t \leq T} \varepsilon_{2,i,t}^{*2}}{2n T^{1/2}} \quad \text{and} \quad \tilde{U}_{2,i} = \frac{\sum_{i'=1}^{i-1} \sum_{t \leq T} \varepsilon_{2,i',t}^* \varepsilon_{2,i,t}^*}{n T^{1/2}}.$$

It can be shown¹⁹ that

$$QLR_{n,T} = n^{-1/2} \sum_{i \leq n} (\tilde{\Psi}_i + \tilde{V}_i + \tilde{U}_i) + o_p(n^{-1/2}), \quad (43)$$

where $\tilde{\Psi}_i = (2T^{1/2})^{-1} \sum_{t \leq T} (\varepsilon_{2,i,t}^2 - \varepsilon_{1,i,t}^2)$, $\tilde{V}_i = \tilde{V}_{1,i} - \tilde{V}_{2,i}$ and $\tilde{U}_i = \tilde{U}_{1,i} - \tilde{U}_{2,i}$. The variance $\omega_{n,T}^2$ of the $QLR_{n,T}$ is defined as in (14), with $\tilde{\Psi}_i$, $\tilde{V}_i$ and $\tilde{U}_i$ taking the specific forms given above. Below is the asymptotic property of the infeasible test statistic that mirrors Theorem 1:

Theorem 6. *Under Assumptions 1, 4 and 5 in the Appendix, we have*

$$\frac{QLR_{n,T} - \mathbb{E}[S_{n,T}] - \overline{QLR}_{n,T}}{\omega_{n,T}} \rightarrow_d N(0, 1),$$

where $\mathbb{E}[S_{n,T}] = (2nT)^{-1/2} \sum_{g \in \mathcal{G}_1} \sum_{i \in I_g} \mathbb{E} [\sum_{t \leq T} (n_g^{-1} \varepsilon_{1,i,t}^{*2} - n^{-1} \varepsilon_{2,i,t}^{*2}) - T^{-1} (\sum_{t \leq T} \varepsilon_{2,i,t}^*)^2]$.

3.3 Bias/Variance Estimation and the Feasible Test

Following the discussion in Section 2, we present a feasible modified QLR procedure by estimating the bias $\mathbb{E}[S_{n,T}]$ and the variance $\omega_{n,T}^2$. Consistent estimation of these quantities is required under the null hypothesis to ensure correct size control of the test. For this

¹⁹See Lemma S37 and Lemma S43 in the Online Appendix.

purpose, we assume that $(\varepsilon_{1,i,t}, \varepsilon_{2,i,t})$ are independent across t with time-invariant first and second moments; this condition is imposed in Assumption 6.²⁰

Under this condition, we have

$$\mathbb{E}[S_{n,T}] = (2nT)^{-1/2} \sum_{g \in \mathcal{G}_1} \sum_{i \in I_g} [Tn_g^{-1} \sigma_{1,i}^2 - (1 + Tn^{-1}) \sigma_{2,i}^2]$$

where $\sigma_{j,i}^2 \equiv \text{Var}(\varepsilon_{j,i,t})$ can be estimated by $\hat{\sigma}_{j,i}^2 \equiv T^{-1} \sum_{t \leq T} \hat{\varepsilon}_{j,i,t}^2$. Therefore, the bias term can be estimated by

$$\widehat{\mathbb{E}[S_{n,T}]} = (2nT)^{-1/2} \sum_{g \in \mathcal{G}_1} \sum_{i \in I_g} [Tn_g^{-1} \hat{\sigma}_{1,i}^2 - (1 + Tn^{-1}) \hat{\sigma}_{2,i}^2], \quad (44)$$

which leads to the modified QLR statistic:

$$MQLR_{n,T} = QLR_{n,T} - \widehat{\mathbb{E}[S_{n,T}]}. \quad (45)$$

It can be shown²¹ that $\omega_{n,T}^2$ can be approximated by $\sigma_{n,T}^2 + \sigma_{U,n,T}^2$, where

$$\sigma_{n,T}^2 \equiv (4n)^{-1} \sum_{i \leq n} \text{Var}(\varepsilon_{2,i,t}^2 - \varepsilon_{1,i,t}^2) \quad (46)$$

and

$$\begin{aligned} \sigma_{U,n,T}^2 &\equiv (2nT)^{-1} \sum_{i \leq n} \sigma_{2,i}^4 + (2n)^{-1} \sum_{g \in \mathcal{G}_1} n_g^{-2} \left(\sum_{i \in I_g} \sigma_{1,i}^2 \right)^2 \\ &\quad + (2n^3)^{-1} \left(\sum_{i \leq n} \sigma_{2,i}^2 \right)^2 - n^{-2} \sum_{g \in \mathcal{G}_1} n_g^{-1} \left(\sum_{i \in I_g} \sigma_{1,2,i} \right)^2. \end{aligned} \quad (47)$$

We consider the sample variance

$$\hat{\sigma}_{n,T}^2 \equiv (4nT)^{-1} \sum_{i \leq n} \sum_{t \leq T} (\hat{\varepsilon}_{2,i,t}^2 - \hat{\varepsilon}_{1,i,t}^2)^2 - (nT)^{-1} (MQLR_{n,T})^2, \quad (48)$$

as an estimator of $\sigma_{n,T}^2$. The variance $\sigma_{U,n,T}^2$ from estimating the incidental parameters is

²⁰The independence of $(\varepsilon_{1,i,t}, \varepsilon_{2,i,t})$ over t is imposed primarily to simplify the form of $\omega_{n,T}^2$ and its consistent estimator. This condition is only required under the null hypothesis; see Lemma 6 in the Appendix for consistency of the test without this assumption.

²¹See Lemma 7 in the Appendix.

then estimated by

$$\begin{aligned}\hat{\sigma}_{U,n,T}^2 &\equiv (2nT)^{-1} \sum_{i \leq n} \hat{\sigma}_{2,i}^4 + (2n)^{-1} \sum_{g \in \mathcal{G}_1} n_g^{-2} \left(\sum_{i \in I_g} \hat{\sigma}_{1,i}^2 \right)^2 \\ &\quad + (2n^3)^{-1} \left(\sum_{i \leq n} \hat{\sigma}_{2,i}^2 \right)^2 - n^{-2} \sum_{g \in \mathcal{G}_1} n_g^{-1} \left(\sum_{i \in I_g} \hat{\sigma}_{1,2,i}^2 \right)^2,\end{aligned}\tag{49}$$

where $\hat{\sigma}_{1,2,i} \equiv T^{-1} \sum_{t \leq T} \hat{\varepsilon}_{1,i,t} \hat{\varepsilon}_{2,i,t}$.

Proceeding as in the previous section, our estimate of $\omega_{n,T}^2$ is thus

$$\hat{\omega}_{n,T}^2 \equiv \max \left\{ \hat{\sigma}_{n,T}^2 - \hat{\sigma}_{U,n,T}^2, \hat{\sigma}_{U,n,T}^2 \right\},\tag{50}$$

and the Vuong test follows (32) with $MQLR_{n,T}$ and $\hat{\omega}_{n,T}^2$ constructed in (45) and (50), respectively.²² Mirroring Theorem 5, we have:

Theorem 7. *Suppose that Assumptions 1, 4, 5 and 6 in the Appendix hold. Further suppose that as $n, T \rightarrow \infty$,*

$$\frac{(nT)^{-1/2} \sum_{i \leq n} \sum_{t \leq T} \mathbb{E}[\varepsilon_{2,i,t}^2 - \varepsilon_{1,i,t}^2]}{\omega_{n,T}} \rightarrow c.\tag{51}$$

Then we have as $n, T \rightarrow \infty$,

$$\mathbb{E}[\varphi_{n,T}^{2-side}(p)] \rightarrow 2 - \Phi(z_{1-p/2} - c) - \Phi(z_{1-p/2} + c),$$

and

$$\mathbb{E}[\varphi_{n,T}^{1-side}(p)] \rightarrow 1 - \Phi(z_{1-p} - c),$$

where $\varphi_{n,T}^{2-side}$ and $\varphi_{n,T}^{1-side}$ follow the identical form as (32).

4 Conclusion

This paper extends the classical Vuong (1989) test to panel data models with fixed effects, addressing challenges that arise from high-dimensional nuisance parameters. By modifying the profile likelihood and applying bias correction techniques, we develop a valid procedure

²²In Lemma 8 of the Appendix, we show that $\hat{\sigma}_{U,n,T}^2$ is positive in finite samples, which implies that the estimator of $\omega_{n,T}^2$ constructed in (50) below is also positive.

for comparing non-nested panel data models. Our analysis emphasizes the need for variance adjustments in the presence of incidental parameter problems and shows how these modifications support more reliable model selection in high-dimensional settings.

We contribute to the literature by generalizing the Vuong test to accommodate grouped heterogeneity in both individual and time effects. Furthermore, we propose a methodological framework for model selection when competing specifications define cross-sectional units differently—a common issue in empirical work. The theoretical results in this paper highlight the critical role of bias correction in panel data analysis and complement existing research on non-nested hypothesis testing.

Appendix

Throughout the appendix, K denotes a generic finite constant that may vary from line to line but does not depend on n and T . For brevity, proofs of lemmas are omitted here and are provided in the Online Appendix.

A Assumptions and Proofs for Section 2

We first state sufficient conditions for the main results in Section 2, namely, Theorems 1, 2, 3, 4 and 5, and then provide their proofs. Auxiliary lemmas used in these proofs are collected in Subsection A.1 below.

Assumption 1. (i) $n, T \rightarrow \infty$ such that $n/T \rightarrow \rho$, where $0 < \rho < \infty$; (ii) for each i , $\{z_{i,t}\}_{t \geq 1}$ is an alpha-mixing process; (iii) $\{z_{i,t}\}_{t \geq 1}$ are independent across i ; (iv) $\sup_i |\alpha_i(j)| \leq Ka^j$ for some a such that $a \in (0, 1)$, where

$$\alpha_i(j) \equiv \sup_t \sup_{A \in \mathcal{A}_t^i, B \in \mathcal{B}_{t+j}^i} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where $\mathcal{A}_t^i$ and $\mathcal{B}_t^i$ denote the sigma-fields generated by $\{z_{i,s}\}_{s \leq t}$ and $\{z_{i,s'}\}_{s' \geq t}$, respectively.

Assumption 2. (i) Assumptions 7, 8 and 9 in Section C hold for both models 1 and 2; (ii) for some $\delta \in (0, 1)$, $\mathbb{E}[(\Delta\psi(z_{i,t}, \phi_i^*)/\omega_{n,T})^{4+\delta}] < K$; (iii) $\mathbb{E}[\psi_{j,\gamma}(z_{i,t})^8 + M(z_{i,t})^8] < K$, where $M(z_{i,t})$ is defined in Assumption 8; (iv) $G_2^{-1} \max_{g \in \mathcal{G}_2} n_g^{-1} = o(1)$; (v) $n\omega_{n,T}^2 \geq K^{-1}$.

Assumption 3. For each $i = 1, \dots, n$, the process $\{\Delta\psi(z_{i,t}, \phi_i^*)\}_{t \geq 1}$ is serially uncorrelated. Moreover $\mathbb{E}[\psi_{j,\gamma}^*(z_{i,t+1}) | \{(\psi_{j,\gamma}^*(z_{i,s}), \psi_{2,\gamma}^*(z_{i,s}))\}_{s \leq t}] = 0$ for $j = 1, 2$ and any $t \geq 1$.

Assumption 1 specifies conditions governing the joint asymptotics and the cross-sectional and time-series dependence structure of the data. Together with Assumption 1, Assumption 2(i) is used to obtain the expansion of $MQLR_{n,T}^*$ in (11). Assumptions 2(ii, iii) ensure the Lindeberg condition required for the martingale central limit theorem. Assumption 2(iv) requires that, under model 2, the number of groups and/or the size of the smallest group diverges. Assumption 2(v) guarantees that the remainder term $o_p(n^{-1/2})$ in the expansion of the (infeasible) modified QLR statistic in (11) becomes asymptotically negligible after scaling by $\omega_{n,T}$. Finally, Assumption 3 is imposed to establish the consistency of the bias and variance estimators for the QLR statistic.

PROOF OF THEOREM 1. To establish the claim, we apply Theorem A3 in Appendix C. To this end, it suffices to verify its sufficient conditions.

First, note that Assumption 1 and Assumptions 7, 8, 9 in Appendix C are maintained for both models. Moreover, Assumption 11(i) in Appendix C is implied by Assumption 2. It therefore remains to verify that Assumptions 10 and 11(ii) in Appendix C also hold. Since the conditions in Assumption 12 in Appendix C are subsumed by Assumptions 2(ii)–(iv), it follows from Theorem A4 in Appendix C that Assumption 11(ii) in Appendix C is satisfied.

It remains to verify Assumption 10 in Appendix C. For model $j = 1, 2$, we have $\mathcal{G}_j \equiv \{1, \dots, G_j\}$ and $\mathcal{M}_j \equiv \{1\}$. Consequently, $G_j \leq n$, $M_j = 1$ and $T_m = T$. Therefore, we can establish the following inequality:

$$(G_j M_j)^{2/p} \max_{g \in \mathcal{G}_j} (n_{j,g} T_m)^{-1/2} \leq n^{2/p} T^{-1/2} = o(1), \quad (52)$$

where the final equality follows from Assumption 1(i) and $p > 4$. Similarly,

$$(G_j M_j)(nT)^{-1/2} \leq n^{1/2} T^{-1/2} = O(1). \quad (53)$$

Combining (52) with (53), we conclude that Assumption 10 in Appendix C holds for both models 1 and 2. Q.E.D.

PROOF OF THEOREM 2. For any $i \in I_{j,g}$ and any $g \in \mathcal{G}_j$, Lemma 5(i) implies that, with probability approaching 1 (wpa1),

$$\hat{\sigma}_{j,\gamma,i}^2 - \sigma_{j,\gamma,i}^2 = \frac{\widehat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}^2(z_{i,t})]}{-\hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})} - \frac{\mathbb{E}_T[\psi_{j,\gamma}^2(z_{i,t})]}{-\Psi_{j,\gamma\gamma,g}} - \frac{(\widehat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})])^2}{-\hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})} + \frac{(\mathbb{E}_T[\psi_{j,\gamma}(z_{i,t})])^2}{-\Psi_{j,\gamma\gamma,g}},$$

where $\Psi_{j,\gamma\gamma,g} \equiv n_{j,g}^{-1} \sum_{i \in I_{j,g}} \mathbb{E}_T [\psi_{j,\gamma\gamma}(z_{i,t})]$. Summing over i and g yields

$$\begin{aligned} \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{j,\gamma,i}^2 - \sigma_{j,\gamma,i}^2) &= - \sum_{g \in \mathcal{G}_j} \sum_{i \in I_{j,g}} \left(\frac{\widehat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}^2(z_{i,t})]}{n_{j,g} \hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})} - \frac{\mathbb{E}_T[\psi_{j,\gamma}^2(z_{i,t})]}{n_{j,g} \Psi_{j,\gamma\gamma,g}} \right) \\ &\quad + \sum_{g \in \mathcal{G}_j} \sum_{i \in I_{j,g}} \left(\frac{(\widehat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})])^2}{n_{j,g} \hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})} - \frac{(\mathbb{E}_T[\psi_{j,\gamma}(z_{i,t})])^2}{n_{j,g} \Psi_{j,\gamma\gamma,g}} \right), \end{aligned} \quad (54)$$

which together with (73) and (74) in Lemma 2 implies

$$R_{j,n,T}(\hat{\phi}_j) - \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} \sigma_{j,\gamma,i}^2 = O_p(1). \quad (55)$$

Finally, under Assumption 3, $T^{1/2} \sum_{i \leq n} \mathbb{E}[\tilde{V}_{j,i}] = 2 \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} \sigma_{j,\gamma,i}^2$. This identity, combined with Assumption 2(v), establishes the claim of the lemma. *Q.E.D.*

PROOF OF THEOREM 3. From (75) in Lemma 2, (77) in Lemma 3 and Lemma 5(ii), we have

$$\begin{aligned} \hat{\sigma}_{n,T}^2 &\equiv (nT)^{-1} \sum_{i \leq n} \sum_{t \leq T} \Delta\psi(z_{i,t}, \hat{\phi}_i)^2 - (nT)^{-1} (MQLR_{n,T})^2 \\ &= n^{-1} \sum_{i \leq n} \mathbb{E}_T[\Delta\psi(z_{i,t})^2] - (nT)^{-1} (\overline{QLR}_{n,T})^2 + 2\sigma_{S,n,T}^2 + O_p(T^{-3/2} + \omega_{n,T}(nT)^{-1/2}) \\ &= \sigma_{n,T}^2 + 2\sigma_{S,n,T}^2 + O_p(T^{-3/2} + \omega_{n,T}(nT)^{-1/2}), \end{aligned} \quad (56)$$

where the last equality follows from the definition of $\sigma_{n,T}^2$ and Assumption 2(i). Thus,

$$\frac{\hat{\sigma}_{n,T}^2 - (\sigma_{n,T}^2 + 2\sigma_{S,n,T}^2)}{\omega_{n,T}^2} = O_p(\omega_{n,T}^{-2} T^{-3/2} + \omega_{n,T}^{-1} (nT)^{-1/2}) = O_p(T^{-1/2}), \quad (57)$$

where the second equality holds by Assumptions 1(i) and 2(v). The claim of the lemma follows from (57). *Q.E.D.*

PROOF OF THEOREM 4. From (76) in Lemma 3 and (79) in Lemma 4, we have:

$$\begin{aligned} \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{j,\gamma,i}^2 - \sigma_{j,\gamma,i}^2)^2 &\leq 2 \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} \left(\frac{(\widehat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})])^2}{\hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})} - \frac{(\mathbb{E}_T[\psi_{j,\gamma}(z_{i,t})])^2}{\Psi_{j,\gamma\gamma,g}} \right)^2 \\ &\quad + 2 \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} (\hat{s}_{j,\gamma,i}^2 - s_{j,\gamma,i}^2)^2 = O_p(1). \end{aligned} \quad (58)$$

This along with the triangle inequality implies that

$$\begin{aligned}
\left| \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{j,\gamma,i}^4 - \sigma_{j,\gamma,i}^4) \right| &\leq \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{j,\gamma,i}^2 - \sigma_{j,\gamma,i}^2)^2 \\
&\quad + 2 \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} |\hat{\sigma}_{j,\gamma,i}^2 - \sigma_{j,\gamma,i}^2| \sigma_{j,\gamma,i}^2 \\
&= 2 \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} |\hat{\sigma}_{j,\gamma,i}^2 - \sigma_{j,\gamma,i}^2| \sigma_{j,\gamma,i}^2 + O_p(1). \tag{59}
\end{aligned}$$

From Assumptions 2(iii) and 9, it follows that

$$\max_{i \leq n} s_{j,\gamma,i}^2 = \frac{\max_{i \leq n} T^{-1} \sum_{t \leq T} \mathbb{E}[\psi_{j,\gamma}(z_{i,t})^2]}{-\Psi_{j,\gamma\gamma,g}} \leq K. \tag{60}$$

Thus, by the Cauchy-Schwarz inequality and (58),

$$\sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} |\hat{\sigma}_{j,\gamma,i}^2 - \sigma_{j,\gamma,i}^2| \sigma_{j,\gamma,i}^2 \leq K G_j^{1/2} \left(\sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{j,\gamma,i}^2 - \sigma_{j,\gamma,i}^2)^2 \right)^{1/2} = O_p(G_j^{1/2}).$$

Together with (59), this implies

$$n^{-1} \sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{j,\gamma,i}^4 - \sigma_{j,\gamma,i}^4) = O_p(n^{-1/2}). \tag{61}$$

Similarly, we can apply the Cauchy-Schwarz inequality and (80) in Lemma 4 to show that

$$n^{-1} \sum_{g \in \mathcal{G}_j} n_{2,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{12,\gamma,i}^2 - \sigma_{12,\gamma,i}^2) = O_p(n^{-1/2}). \tag{62}$$

By the definitions of $\hat{\sigma}_{U,n,T}^2$ and $\sigma_{U,n,T}^2$, we obtain the following expression:

$$\begin{aligned}
\hat{\sigma}_{U,n,T}^2 - \sigma_{U,n,T}^2 &= (2nT)^{-1} \sum_{i=1}^n (\hat{\sigma}_{1,\gamma,i}^4 - \sigma_{1,\gamma,i}^4) - (nT)^{-1} \sum_{g \in \mathcal{G}_j} n_{2,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{12,\gamma,i}^2 - \sigma_{12,\gamma,i}^2) \\
&\quad + (2nT)^{-1} \sum_{g \in \mathcal{G}_2} \left(\left(n_{2,g}^{-1} \sum_{i \in I_{2,g}} \hat{\sigma}_{2,\gamma,i}^2 \right)^2 - \left(n_{2,g}^{-1} \sum_{i \in I_{2,g}} \sigma_{2,\gamma,i}^2 \right)^2 \right). \tag{63}
\end{aligned}$$

From (61) and (62), and Assumption 2(v), it follows that

$$\frac{(2nT)^{-1} \sum_{i=1}^n (\hat{\sigma}_{1,\gamma,i}^4 - \sigma_{1,\gamma,i}^4)}{\omega_{n,T}^2} = O_p(n^{-1/2}), \quad (64)$$

and

$$\frac{(nT)^{-1} \sum_{g \in \mathcal{G}_j} n_{2,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{12,\gamma,i}^2 - \sigma_{12,\gamma,i}^2)}{\omega_{n,T}^2} = O_p(n^{-1/2}). \quad (65)$$

The first claim in (29) follows from (63), (64), (65) and (81) in Lemma 4.

Next, by the definitions of $\hat{\sigma}_{S,n,T}^2$ and $\sigma_{S,n,T}^2$, we obtain

$$\begin{aligned} \hat{\sigma}_{S,n,T}^2 - \sigma_{S,n,T}^2 &= (2nT)^{-1} \sum_{i=1}^n (\hat{\sigma}_{1,\gamma,i}^4 - \sigma_{1,\gamma,i}^4) - (nT)^{-1} \sum_{g \in \mathcal{G}_j} n_{2,g}^{-1} \sum_{i \in I_{j,g}} (\hat{\sigma}_{12,\gamma,i}^2 - \sigma_{12,\gamma,i}^2) \\ &\quad + (2nT)^{-1} \sum_{g \in \mathcal{G}_2} n_{2,g}^{-2} \sum_{i \in I_{2,g}} \left(\hat{\sigma}_{2,\gamma,i}^2 \sum_{i' \in I_{2,g}} \hat{s}_{2,\gamma,i'}^2 - \sigma_{2,\gamma,i}^2 \sum_{i' \in I_{2,g}} s_{2,\gamma,i'}^2 \right). \end{aligned} \quad (66)$$

Therefore, the second claim in (29) follows from (64) and (65), and (82) in Lemma 4.Q.E.D.

PROOF OF THEOREM 5. (a) By the definition of $\hat{\omega}_{n,T}^2$, we have

$$\hat{\sigma}_{n,T}^2 + \hat{\sigma}_{U,n,T}^2 - 2\hat{\sigma}_{S,n,T}^2 \leq \hat{\omega}_{n,T}^2 \leq \max \{ \hat{\sigma}_{n,T}^2 + \hat{\sigma}_{U,n,T}^2 - 2\hat{\sigma}_{S,n,T}^2, \hat{\sigma}_{U,n,T}^2 + \sigma_{n,T}^2 \},$$

which implies:

$$\frac{|\hat{\omega}_{n,T}^2 - \omega_{n,T}^2|}{\omega_{n,T}^2} \leq \frac{|\hat{\sigma}_{n,T}^2 + \hat{\sigma}_{U,n,T}^2 - 2\hat{\sigma}_{S,n,T}^2 - \omega_{n,T}^2|}{\omega_{n,T}^2} + \frac{|\hat{\sigma}_{U,n,T}^2 + \sigma_{n,T}^2 - \omega_{n,T}^2|}{\omega_{n,T}^2}. \quad (67)$$

Using Theorem 3, Theorem 4, (78) in Lemma 3, and the triangle inequality, the first term after the inequality in (67) satisfies:

$$\begin{aligned} \frac{|\hat{\sigma}_{n,T}^2 + \hat{\sigma}_{U,n,T}^2 - 2\hat{\sigma}_{S,n,T}^2 - \omega_{n,T}^2|}{\omega_{n,T}^2} &\leq \frac{|\hat{\sigma}_{n,T}^2 - (\sigma_{n,T}^2 + 2\sigma_{S,n,T}^2)|}{\omega_{n,T}^2} + \frac{|\hat{\sigma}_{U,n,T}^2 - \sigma_{U,n,T}^2|}{\omega_{n,T}^2} \\ &\quad + \frac{2|\hat{\sigma}_{S,n,T}^2 - \sigma_{S,n,T}^2|}{\omega_{n,T}^2} + \frac{|\omega_{n,T}^2 - (\sigma_{n,T}^2 + \sigma_{U,n,T}^2)|}{\omega_{n,T}^2} \\ &= O_p(T^{-1/2} + n^{-1/2} + \omega_{n,T}^{-1} T^{-1}) = O_p(T^{-1/2}) \end{aligned} \quad (68)$$

where the second equality holds by Assumptions 1(i) and 2(v). For the second term after

the inequality in (67), we can bound it as

$$\frac{|\hat{\sigma}_{U,n,T}^2 + \sigma_{n,T}^2 - \omega_{n,T}^2|}{\omega_{n,T}^2} \leq \frac{|\hat{\sigma}_{U,n,T}^2 - \sigma_{U,n,T}^2|}{\omega_{n,T}^2} + \frac{|\omega_{n,T}^2 - \sigma_{n,T}^2 - \sigma_{U,n,T}^2|}{\omega_{n,T}^2} = O_p(T^{-1/2}), \quad (69)$$

where the equality follows from Theorem 3, (78) in Lemma 3, and Assumptions 1(i) and 2(v). Combining the results of (67), (68) and (69), we conclude:

$$\frac{\hat{\omega}_{n,T}^2}{\omega_{n,T}^2} = 1 + O_p(T^{-1/2}). \quad (70)$$

Using (20), (33) and (70), we apply Slutsky's theorem to deduce:

$$\begin{aligned} \frac{|MQLR_{n,T}| - \hat{\omega}_{n,T} z_{1-p/2}}{\omega_{n,T}} &= \left| \frac{MQLR_{n,T} - \overline{QLR}_{n,T}}{\omega_{n,T}} + \frac{\overline{QLR}_{n,T}}{\omega_{n,T}} \right| - \frac{\hat{\omega}_{n,T}}{\omega_{n,T}} z_{1-p/2} \\ &\rightarrow_d |Z + c| - z_{1-p/2}, \end{aligned} \quad (71)$$

where Z denotes a standard normal random variable. Therefore, as $n, T \rightarrow \infty$,

$$\begin{aligned} \mathbb{E}[\varphi_{n,T}^{2\text{-side}}(p)] &= \mathbb{P}(|MQLR_{n,T}| > \hat{\omega}_{n,T} z_{1-p/2}) \\ &= \mathbb{P}\left(\frac{|MQLR_{n,T}| - \hat{\omega}_{n,T} z_{1-p/2}}{\omega_{n,T}} > 0\right) \rightarrow \mathbb{P}(|Z + c| > z_{1-p/2}), \end{aligned} \quad (72)$$

which shows (34). The second claim of the theorem, stated in (35), follows by the similar arguments and its proof is omitted. Q.E.D.

A.1 Auxiliary Results and Lemmas

This subsection presents the power analysis of the proposed test without Assumption 3, and collects the auxiliary lemmas used in the proofs of Theorems 1-5.

Lemma 1. *Suppose that Assumptions 1 and 2 hold. Then:*

- (a) *If $|\overline{QLR}_{n,T}| \succ G_1 + G_2$, we have $\mathbb{E}[\varphi_{n,T}^{2\text{-side}}(p)] \rightarrow 1$ as $n, T \rightarrow \infty$.*
- (b) *If $\overline{QLR}_{n,T} \succ G_1 + G_2$, we have $\mathbb{E}[\varphi_{n,T}^{1\text{-side}}(p)] \rightarrow 1$ as $n, T \rightarrow \infty$.*

Lemma 2. Under Assumptions 1 and 2, we have

$$\sum_{g \in \mathcal{G}_j} \sum_{i \in I_{j,g}} \left(\frac{\widehat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})^2]}{n_{j,g} \hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})} - \frac{\mathbb{E}_T[\psi_{j,\gamma}(z_{i,t})^2]}{n_{j,g} \Psi_{j,\gamma\gamma,g}} \right) = O_p(1), \quad (73)$$

$$\sum_{g \in \mathcal{G}_j} \sum_{i \in I_{j,g}} \left(\frac{(\widehat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})])^2}{n_{j,g} \hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})} - \frac{(\mathbb{E}_T[\psi_{j,\gamma}(z_{i,t})])^2}{n_{j,g} \Psi_{j,\gamma\gamma,g}} \right) = O_p(1), \quad (74)$$

$$\frac{(MQLR_{n,T})^2 - (\overline{QLR}_{n,T})^2}{(nT)^{1/2} \omega_{n,T}^2} = O_p(1). \quad (75)$$

$$\sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} \left(\frac{(\widehat{\mathbb{E}}_T[\hat{\psi}_{j,\gamma}(z_{i,t})])^2}{\hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})} - \frac{(\mathbb{E}_T[\psi_{j,\gamma}(z_{i,t})])^2}{\Psi_{j,\gamma\gamma,g}} \right)^2 = O_p(1). \quad (76)$$

Lemma 3. Under Assumptions 1, 2 and 3, we have

$$\frac{n^{-1} \sum_{i \leq n} (\widehat{\mathbb{E}}_T[\Delta\psi(z_{i,t}, \hat{\phi}_i)^2] - \mathbb{E}_T[\Delta\psi(z_{i,t})^2]) - 2\sigma_{S,n,T}^2}{\omega_{n,T}(nT)^{-1/2} + T^{-3/2}} = O_p(1), \quad (77)$$

$$\frac{T(\omega_{n,T}^2 - \sigma_{n,T}^2 - \sigma_{U,n,T}^2)}{\omega_{n,T}} = O(1). \quad (78)$$

Lemma 4. Under Assumptions 1, 2 and 3, we have

$$\sum_{g \in \mathcal{G}_j} n_{j,g}^{-1} \sum_{i \in I_{j,g}} (\hat{s}_{j,\gamma,i}^2 - s_{j,\gamma,i}^2)^2 = O_p(1), \quad (79)$$

$$\sum_{g \in \mathcal{G}_2} n_{2,g}^{-1} \sum_{i \in I_{2,g}} (\hat{\sigma}_{12,\gamma,i}^2 - \sigma_{12,\gamma,i}^2)^2 = O_p(1), \quad (80)$$

$$\frac{\sum_{g \in \mathcal{G}_2} \left(\left(n_{2,g}^{-1} \sum_{i \in I_{2,g}} \hat{\sigma}_{2,\gamma,i}^2 \right)^2 - \left(n_{2,g}^{-1} \sum_{i \in I_{2,g}} \sigma_{2,\gamma,i}^2 \right)^2 \right)}{(n^{1/2}T) \omega_{n,T}^2} = O_p(1), \quad (81)$$

$$\frac{\sum_{g \in \mathcal{G}_2} n_{2,g}^{-2} \sum_{i \in I_{2,g}} \left(\hat{\sigma}_{2,\gamma,i}^2 \sum_{i' \in I_{2,g}} \hat{s}_{2,\gamma,i'}^2 - \sigma_{2,\gamma,i}^2 \sum_{i' \in I_{2,g}} s_{2,\gamma,i'}^2 \right)}{(n^{1/2}T) \omega_{n,T}^2} = O_p(1). \quad (82)$$

Lemma 5. Under Assumptions 1 and 2, we have: (i) $\min_{g \in \mathcal{G}_j} (-\hat{\Psi}_{j,\gamma\gamma}(\hat{\phi}_{j,g})) \geq K^{-1}$; and (ii) $\omega_{n,T}^2 \leq K$.

B Assumptions and Proofs for Section 3

Assumption 4. (i) $\lambda_{\min}(\Sigma_{\ddot{x}}) \geq K^{-1}$ and $\lambda_{\min}(\Sigma_{\dot{x}}) \geq K^{-1}$; (ii) $\max_{i,t} \mathbb{E}[y_{i,t}^8 + \|x_{i,t}\|^8] \leq K$.

Assumption 5. (i) $n\omega_{n,T}^2 \geq K^{-1}$; (ii) $\mathbb{E}[\omega_{n,T}^{-4} |\varepsilon_{2,i,t}^2 - \varepsilon_{1,i,t}^2|^4] < K$; (iii) $\max_{g \in \mathcal{G}_1} n_g^{-1} = o(1)$.

Assumption 6. (i) The vectors $(\varepsilon_{1,i,t}, \varepsilon_{2,i,t})$ are independent across t , and $\mathbb{E}[\varepsilon_{j,i,t}]$, $\mathbb{E}[\varepsilon_{j,i,t}^2]$ and $\mathbb{E}[\varepsilon_{1,i,t}\varepsilon_{2,i,t}]$ are invariant in t ; (ii) $G_1 n^{-1/2} + n^{1/2} \sum_{g \in \mathcal{G}_1} n_g^{-1} = o(1)$.

Assumptions 4 and 5 are imposed to study the properties of the estimators in (41) and (42), and to derive the asymptotic distribution of $QLR_{n,T}$ for comparing the heterogeneous time effects model with the TWFE model. Assumption 4 ensures the pseudo true parameters are uniquely identified (and well-defined). Assumptions 5(i, ii) are the counterparts of Assumptions 11(i) and 12(i) in Appendix C, respectively, and are used to establish the asymptotic distribution of $QLR_{n,T}$. Finally, Assumption 6 is imposed primarily to simplify the variance of the QLR statistic and to ensure the consistency of the proposed variance estimator.

PROOF OF THEOREM 6. We begin by observing that

$$\begin{aligned} \frac{\mathbb{E} \left[\left| \sum_{i=1}^n (\tilde{V}_{1,i} - \mathbb{E}[\tilde{V}_{1,i}]) \right|^2 \right]}{n\omega_{n,T}^2} &= \frac{\sum_{i=1}^n \mathbb{E}[(\tilde{V}_{1,i} - \mathbb{E}[\tilde{V}_{1,i}])^2]}{n\omega_{n,T}^2} \\ &\leq K \sum_{g \in \mathcal{G}_1} n_g^{-2} \sum_{i \in I_g} \mathbb{E} \left[\left(T^{-1/2} \sum_{t \leq T} (\varepsilon_{1,i,t}^{*2} - \mathbb{E}[\varepsilon_{1,i,t}^{*2}]) \right)^2 \right] \\ &\leq K \sum_{g \in \mathcal{G}_1} n_g^{-2} \sum_{i \in I_g} \max_{t \leq T} \|\varepsilon_{1,i,t}^{*2}\|_{2+\delta}^2 \leq KG_1 \max_{g \in \mathcal{G}_1} n_g^{-1} = o(1), \quad (83) \end{aligned}$$

where the first inequality follows from Assumption 5(i). The second inequality holds for any $\delta \in (0, 1)$ by the covariance inequality for strong mixing processes together with Assumption 1(iv). The third inequality follows from $\mathbb{E}[\varepsilon_{1,i,t}^8] \leq K$, which is implied by Assumptions 1 and 4 (see, Lemma S34(ii) in the Online Appendix). The final equality follows from Assumption 5 (iii). Similarly, we can establish that

$$\frac{\mathbb{E} \left[\left| (nT^{1/2})^{-1} \sum_{i=1}^n \sum_{t \leq T} (\varepsilon_{2,i,t}^{*2} - \mathbb{E}[\varepsilon_{2,i,t}^{*2}]) \right|^2 \right]}{n\omega_{n,T}^2} \leq Kn^{-1} = o(1), \quad (84)$$

which along with (43), (83) and Markov's inequality implies that

$$\frac{\widehat{QLR}_{n,T} - \mathbb{E}[S_{n,T}] - \overline{QLR}_{n,T}}{\omega_{n,T}} = \sum_{i \leq n} \xi_{nT,i} + o_p(1), \quad (85)$$

where $\xi_{nT,i} = (n^{1/2}\omega_{n,T})^{-1}(\tilde{\Psi}_i^* + \tilde{V}_i^* + \tilde{U}_i)$ and $\tilde{\Psi}_i^* = \tilde{\Psi}_i - \mathbb{E}[\tilde{\Psi}_i]$. It can be shown²³ that sufficient conditions of the martingale CLT for $\sum_{i \leq n} \xi_{nT,i}$ are satisfied, which establishes the asymptotic normality of (85). Q.E.D.

PROOF OF THEOREM 7. Under Assumptions 1, 4, 5 and 6, we have

$$\begin{aligned} \frac{MQLR_{n,T}}{\hat{\omega}_{n,T}} &= \frac{\omega_{n,T}}{\hat{\omega}_{n,T}} \left(\frac{QLR_{n,T} - \mathbb{E}[S_{n,T}] - \overline{QLR}_{n,T}}{\omega_{n,T}} + \frac{\overline{QLR}_{n,T}}{\omega_{n,T}} - \frac{\widehat{\mathbb{E}[S_{n,T}]} - \mathbb{E}[S_{n,T}]}{\omega_{n,T}} \right) \\ &= (1 + o_p(1)) \left(\frac{QLR_{n,T} - \mathbb{E}[S_{n,T}] - \overline{QLR}_{n,T}}{\omega_{n,T}} + \frac{\overline{QLR}_{n,T}}{\omega_{n,T}} + o_p(1) \right) \\ &\rightarrow_d Z + c, \end{aligned} \quad (86)$$

where $\overline{QLR}_{n,T} = (nT)^{-1/2} \sum_{i \leq n} \sum_{t \leq T} \mathbb{E}[\varepsilon_{2,i,t}^2 - \varepsilon_{1,i,t}^2]$. The first equality follows from the definition of $MQLR_{n,T}$, while the second equality is implied by Lemma 9. The stated asymptotic distribution follows from Theorem 6, (51) and Slutsky's theorem. The claim of the theorem then follows directly from (86) together with arguments analogous to those used in the proof of Theorem 5. Q.E.D.

B.1 Auxiliary Results and Lemmas

Lemma 6. *Suppose that Assumptions 1, 4 and 5 hold. Then:*

- (a) *If $|(nT)^{-1/2} \sum_{i \leq n} \sum_{t \leq T} \mathbb{E}[\varepsilon_{2,i,t}^2 - \varepsilon_{1,i,t}^2]| \succ G_1$, then $\mathbb{E}[\varphi_{n,T}^{2-side}(p)] \rightarrow 1$, as $n, T \rightarrow \infty$.*
- (b) *If $(nT)^{-1/2} \sum_{i \leq n} \sum_{t \leq T} \mathbb{E}[\varepsilon_{2,i,t}^2 - \varepsilon_{1,i,t}^2] \succ G_1$, then $\mathbb{E}[\varphi_{n,T}^{1-side}(p)] \rightarrow 1$, as $n, T \rightarrow \infty$.*

Lemma 7. *Under Assumptions 1, 4, 5 and 6, we have $\omega_{n,T}^2 = \sigma_{n,T}^2 + \sigma_{U,n,T}^2 + o(\omega_{n,T}^2)$.*

Lemma 8. *We have $\hat{\sigma}_{U,n,T}^2 \geq (2nT)^{-1} \sum_{i \leq n} \hat{\sigma}_{2,i}^4 + n^{-3} \sum_{g=2}^{G_1} \sum_{g'=1}^{g-1} (\sum_{i \in I_g} \hat{\sigma}_{2,i}^2) (\sum_{i \in I_{g'}} \hat{\sigma}_{2,i}^2)$.*

Lemma 9. *Suppose Assumptions 1, 4, 5 and 6 hold. Then, under the null hypothesis,*

$$\widehat{\mathbb{E}[S_{n,T}]} - \mathbb{E}[S_{n,T}] = o_p(\omega_{n,T}) \quad \text{and} \quad \hat{\omega}_{n,T}^2 / \omega_{n,T}^2 = 1 + o_p(1). \quad (87)$$

²³See Lemma S21 in the Online Appendix 2.

Lemma 10. *Suppose Assumptions 1, 4 and 5 hold. Then $\hat{\omega}_{n,T}^2 = O_p(1)$ and $\widehat{\mathbb{E}[S_{n,T}]} = O_p(G_1)$. Moreover, $|\mathbb{E}[S_{n,T}]| \leq KG_1$ and $\omega_{n,T}^2 \leq K$.*

C General Theory for Nonlinear Panel Likelihood Models

In this section, we present general results on the properties of estimators for nonlinear panel models and derive the asymptotic distribution of the QLR statistic for comparing different panel models. The proofs of these results are provided in the Online Appendix of the paper.

C.1 General Framework

We consider a general nonlinear panel model with joint likelihood specified in (1). Recall that I_g for $g \in \mathcal{G}$ and I_m for $m \in \mathcal{M}$ denote the partitions of $\{1, \dots, n\}$ and $\{1, \dots, T\}$, respectively, induced by the group assignment functions $g(\cdot)$ and $m(\cdot)$. For notational convenience, we assume that $m(\cdot)$ is nondecreasing. This implies that for any two consecutive time groups m and $m+1$, we have $\min I_{m+1} = \max I_m + 1$.

Under this notation, the joint likelihood in (1) can be written as

$$\sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} \sum_{i \in I_g} \sum_{t \in I_m} \psi(z_{i,t}; \theta, \gamma_{g,m}),$$

where $\psi(z_{i,t}; \theta, \gamma_{g,m}) \equiv \log f(z_{i,t}; \theta, \gamma_{g,m})$. Let $(\theta^{*\top}, \gamma_{g,m}^{*\top})^\top$ denote the pseudo true value defined as

$$(\theta^{*\top}, \gamma_{g,m}^{*\top})^\top = \arg \max_{(\theta^\top, \gamma_{g,m}^\top)^\top \in \Phi} \Psi_{g,m}(\theta, \gamma_{g,m}),$$

where $\Psi_{g,m}(\theta, \gamma_{g,m}) \equiv (n_g T_m)^{-1} \sum_{i \in I_g} \sum_{t \in I_m} \mathbb{E}[\psi(z_{i,t}; \theta, \gamma_{g,m})]$, Φ denotes the parameter space including $(\theta^{*\top}, \gamma_{g,m}^{*\top})^\top$ in its interior, and n_g and T_m denote the sizes of I_g and I_m respectively. The unknown pseudo true parameters are estimated through M-estimation, i.e.,

$$(\hat{\theta}^\top, \hat{\gamma}_{\mathcal{G},1}^\top, \dots, \hat{\gamma}_{\mathcal{G},M}^\top)^\top \equiv \arg \max_{\theta, \{\gamma_{g,m}\}_{m \in \mathcal{M}}} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} \sum_{i \in I_g} \sum_{t \in I_m} \psi(z_{i,t}; \theta, \gamma_{g,m}),$$

where $\gamma_{g,m} = (\gamma_{g,m}^\top)_{g \in \mathcal{G}}^\top$ for $m = 1, \dots, M$. The M-estimator $(\hat{\theta}^\top, \hat{\gamma}_{\mathcal{G},1}^\top, \dots, \hat{\gamma}_{\mathcal{G},M}^\top)^\top$ satisfies

the first-order conditions

$$\sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} \sum_{i \in I_g} \sum_{t \in I_m} \frac{\partial \psi(z_{i,t}; \hat{\theta}, \hat{\gamma}_{g,m})}{\partial \theta} = 0_{d_\theta \times 1}, \quad (88)$$

$$\sum_{i \in I_g} \sum_{t \in I_m} \frac{\partial \psi(z_{i,t}; \hat{\theta}, \hat{\gamma}_{g,m})}{\partial \gamma} = 0_{d_\gamma \times 1}, \quad (89)$$

and for any $g \in \mathcal{G}$ and $m \in \mathcal{M}$.

C.2 Asymptotic Theory of the M-Estimator

We establish the asymptotic properties of the M-estimator $(\hat{\theta}^\top, \hat{\gamma}_{\mathcal{G},1}^\top, \dots, \hat{\gamma}_{\mathcal{G},M}^\top)^\top$, together with the asymptotic expansion of the maximized joint likelihood in this subsection. We begin by stating sufficient conditions under which these results hold.

Assumption 7. $\min_{g \in \mathcal{G}, m \in \mathcal{M}} \left[\Psi_{g,m}(\phi_{g,m}^*) - \sup_{\{\phi_{g,m} \in \Phi: \|\phi_{g,m} - \phi_{g,m}^*\| > \eta\}} \Psi_{g,m}(\phi_{g,m}) \right] > 0$ for each $\eta > 0$, where $\Psi_{g,m}(\phi_{g,m}) \equiv \Psi_{g,m}(\theta, \gamma_{g,m})$ and $\phi_{g,m} \equiv (\theta^\top, \gamma_{g,m}^\top)^\top$.

Assumption 8. Let $\psi(z_{i,t}; \tilde{\phi})$ be a function indexed by the parameter $\tilde{\phi} = (\tilde{\phi}_j)_{j \leq k} \in \text{int } \Phi$, where $k \equiv \dim(\tilde{\phi})$ and Φ is a compact, convex subset of $\mathbb{R}^k$. Let $\nu \equiv (\nu_1, \dots, \nu_k)^\top$ be a vector of non-negative integers, $|\nu| = \sum_{j=1}^k \nu_j$ and $D^\nu \psi(z_{i,t}, \tilde{\phi}) = \partial^{|\nu|} \psi(z_{i,t}, \tilde{\phi}) / (\partial \tilde{\phi}_1^{\nu_1} \cdots \partial \tilde{\phi}_k^{\nu_k})$. Then: (i) there exists a function $M(z_{i,t})$ such that

$$\left| D^\nu \psi(z_{i,t}, \tilde{\phi}_1) - D^\nu \psi(z_{i,t}, \tilde{\phi}_2) \right| \leq M(z_{i,t}) \|\tilde{\phi}_1 - \tilde{\phi}_2\|$$

for all $\tilde{\phi}_1, \tilde{\phi}_2 \in \Phi$ and $|\nu| \leq 4$; (ii) the function $M(z_{i,t})$ satisfies $\sup_{i,t} \mathbb{E}[|M(z_{i,t})|^{p+\delta}] < K$ for some integer $p > \max\{4, k + \delta\}$ and for some $\delta > 0$; (iii) for all multi-indices ν with $|\nu| \leq 4$, $\max_{g \in \mathcal{G}, m \in \mathcal{M}} \sup_{i,t} \mathbb{E} \left[|D^\nu \psi(z_{i,t}, \phi_{g,m}^*)|^{4+\delta} \right] < K$ where $\phi_{g,m}^* \equiv (\theta^{*\top}, \gamma_{g,m}^{*\top})^\top$.

Let $\hat{\Psi}_{ab,g,m} \equiv (n_g T_m)^{-1} \sum_{i \in I_g} \sum_{t \in I_m} \psi_{ab}(z_{i,t}; \phi_{g,m}^*)$ and $\Psi_{ab,g,m} \equiv \mathbb{E}[\hat{\Psi}_{ab,g,m}]$ for $a, b \in \{\gamma, \theta\}$, where $\psi_{ab}(z_{i,t}; \phi) \equiv \frac{\partial \psi(z_{i,t}; \phi)}{\partial a \partial b^\top}$.

Assumption 9. The smallest eigenvalues of $-\Psi_{\gamma\gamma,g,m}$ and $\Psi_{\theta\gamma,g,m} \Psi_{\gamma\gamma,g,m} \Psi_{\gamma\theta,g,m} - \Psi_{\theta\theta,g,m}$ are bounded below by K^{-1} uniformly over g, m, n and T .

Assumption 7 is the identification uniqueness condition to ensure the consistency of the M-estimator. Assumption 8 imposes smoothness and uniform moment bounds on the quasi-likelihood function and its derivatives. Assumption 9 provides local identification conditions for θ^* and $\gamma_{g,m}^*$.

Assumption 10. (i) $(GM)^{2/p} \max_{g \in \mathcal{G}, m \in \mathcal{M}} (n_g T_m)^{-1/2} = o(1)$; (ii) $(GM)(nT)^{-1/2} = O(1)$.

Assumption 10(i) is used to establish the consistency of the M-estimator. It accommodates a wide range of grouping structures, including Examples 1–4 in Section 4 of the Online Appendix as special cases. To better illustrate the restriction imposed by Assumption 10(i) on the number of groups, let's consider a simple scenario where the sizes of the groups I_g ($g \in \mathcal{G}$) are proportional to each other and the sizes of I_m ($m \in \mathcal{G}$) are also proportional to each other, i.e.,

$$K^{-1} \leq \min \left\{ \frac{\min_{g \in \mathcal{G}} n_g}{\max_{g \in \mathcal{G}} n_g}, \frac{\min_{m \in \mathcal{M}} T_m}{\max_{m \in \mathcal{M}} T_m} \right\} \leq \max \left\{ \frac{\max_{g \in \mathcal{G}} n_g}{\min_{g \in \mathcal{G}} n_g}, \frac{\max_{m \in \mathcal{M}} T_m}{\min_{m \in \mathcal{M}} T_m} \right\} \leq K,$$

where K denotes a generic finite constant. In this case

$$K^{-1} \leq n(Gn_g)^{-1} \leq K \quad \text{and} \quad K^{-1} \leq T(MT_m)^{-1} \leq K \quad (90)$$

for any $g \in \mathcal{G}$ and any $m \in \mathcal{M}$. Under (90), Assumption 10(i) becomes

$$(GM)^{1/2+2/p}(nT)^{-1/2} = o(1) \quad (91)$$

which requires that the number of groups GM is of a smaller order of nT . Assumption 10(ii) is used to establish the $(nT)^{-1/2}$ -consistency of $\hat{\theta}$.

Let $\hat{\Psi}_{\gamma,g,m} \equiv (n_g T_m)^{-1} \sum_{i \in I_g} \sum_{t \in I_m} \psi_{\gamma}(z_{i,t})$, where $\psi_{\gamma}(z_{i,t}) \equiv \partial \psi(z_{i,t}; \phi_{g,m}^*) / \partial \gamma$. Define $\hat{U}_{\gamma,g,m} \equiv \hat{\Psi}_{\theta\gamma,g,m} - \Psi_{\theta\gamma,g,m} \Psi_{\gamma\gamma,g,m}^{-1} \hat{\Psi}_{\gamma\gamma,g,m}$, and for $j = 1, \dots, d_{\gamma}$,

$$\Psi_{\gamma_j\gamma\gamma,g,m} \equiv (n_g T_m)^{-1} \sum_{i \in I_g} \sum_{t \in I_m} \mathbb{E} \left[\frac{\partial^2}{\partial \gamma \partial \gamma^\top} \left(\frac{\partial \psi(z_{i,t}; \phi_{g,m}^*)}{\partial \gamma_j} \right) \right].$$

The matrix $\Psi_{\theta_j\gamma\gamma,g,m}$ is defined in the similar way for $j = 1, \dots, d_{\theta}$. Theorem A1 below provides the second order expansion of the estimation error in $\hat{\theta}$.

Theorem A1. Under Assumptions 1, 7, 8, 9 and 10, we have

$$\begin{aligned} \hat{\theta} - \theta^* &= H_{\theta\theta}^{-1}(nT)^{-1} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) (\hat{\Psi}_{\theta,g,m} - \Psi_{\theta\gamma,g,m} \Psi_{\gamma\gamma,g,m}^{-1} \hat{\Psi}_{\gamma,g,m}) \\ &\quad + H_{\theta\theta}^{-1}(nT)^{-1} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) B_{\theta,g,m} + o_p((nT)^{-1/2}), \end{aligned}$$

where

$$\begin{aligned}
H_{\theta\theta} &\equiv -(nT)^{-1} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) (\Psi_{\theta\theta,g,m} - \Psi_{\theta\gamma,g,m} \Psi_{\gamma\gamma,g,m}^{-1} \Psi_{\gamma\theta,g,m}), \\
B_{\theta,g,m} &\equiv 2^{-1} \mathbb{E} \left[\left(\hat{\Psi}_{\gamma,g,m}^\top \Psi_{\gamma\gamma,g,m}^{-1} \Psi_{\theta_j\gamma,g,m} \Psi_{\gamma\gamma,g,m}^{-1} \hat{\Psi}_{\gamma,g,m} \right)_{j \leq d_\theta} \right] - \mathbb{E} \left[\hat{U}_{\gamma,g,m} \Psi_{\gamma\gamma,g,m}^{-1} \hat{\Psi}_{\gamma,g,m} \right] \\
&\quad - 2^{-1} \Psi_{\theta\gamma,g,m} \Psi_{\gamma\gamma,g,m}^{-1} \mathbb{E} \left[\left(\hat{\Psi}_{\gamma,g,m}^\top \Psi_{\gamma\gamma,g,m}^{-1} \Psi_{\gamma_j\gamma,g,m} \Psi_{\gamma\gamma,g,m}^{-1} \hat{\Psi}_{\gamma,g,m} \right)_{j \leq d_\gamma} \right].
\end{aligned}$$

Theorem A2 establishes a second-order expansion of the likelihood function evaluated at the M-estimator $\hat{\phi} \equiv (\hat{\theta}^\top, \hat{\gamma}_{\mathcal{G},1}^\top, \dots, \hat{\gamma}_{\mathcal{G},M}^\top)^\top$.

Theorem A2. *Under Assumptions 1, 7, 8, 9 and 10, we have*

$$L_{n,T}(\hat{\phi}) = L_{n,T}(\phi^*) + 2^{-1} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) \hat{\Psi}_{\gamma,g,m}^\top (-\Psi_{\gamma\gamma,g,m})^{-1} \hat{\Psi}_{\gamma,g,m} + o_p((GM)^{1/2}),$$

where $L_{n,T}(\hat{\phi}) = \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} \sum_{i \in I_g} \sum_{t \in I_m} \psi(z_{i,t}; \hat{\theta}, \hat{\gamma}_{g,m})$, and $L_{n,T}(\phi^*)$ is defined analogously with $\hat{\phi}$ replaced by the pseudo true parameter ϕ^* .

Let $\hat{\Psi}_{g,m}(\phi_{g,m}) \equiv (n_g T_m)^{-1} \sum_{i \in I_g} \sum_{t \in I_m} \psi(z_{i,t}; \theta, \gamma_{g,m})$. Then the joint likelihood evaluated at the pseudo true parameter can be written as $L_{n,T}(\phi^*) = \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) \hat{\Psi}_{g,m}(\phi_{g,m}^*)$. Define

$$R_{n,T}^* \equiv \frac{(nT)^{-1/2}}{2} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) \mathbb{E}[\hat{\Psi}_{\gamma,g,m}^\top (-\Psi_{\gamma\gamma,g,m}^{-1}) \hat{\Psi}_{\gamma,g,m}].$$

With this notation, Theorem A2 implies the following expansion for the maximized joint likelihood:

$$\begin{aligned}
&(nT)^{-1/2} L_{n,T}(\hat{\phi}) - R_{n,T}^* \\
&= (nT)^{-1/2} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) \mathbb{E}[\hat{\Psi}_{g,m}(\phi_{g,m}^*)] \tag{92}
\end{aligned}$$

$$+ (nT)^{-1/2} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) (\hat{\Psi}_{g,m}(\phi_{g,m}^*) - \mathbb{E}[\hat{\Psi}_{g,m}(\phi_{g,m}^*)]) \tag{93}$$

$$- (4nT)^{-1/2} \sum_{g \in \mathcal{G}} \sum_{m \in \mathcal{M}} (n_g T_m) (\hat{\Psi}_{\gamma,g,m}^\top \Psi_{\gamma\gamma,g,m}^{-1} \hat{\Psi}_{\gamma,g,m} - \mathbb{E}[\hat{\Psi}_{\gamma,g,m}^\top \Psi_{\gamma\gamma,g,m}^{-1} \hat{\Psi}_{\gamma,g,m}]) \tag{94}$$

$$+ o_p((GM)^{1/2} (nT)^{-1/2}), \tag{95}$$

where the first term (92) is the population likelihood function evaluated at the pseudo true parameter, the second term (93) and the third term (94) are the first-order term and the

second-order term respectively, (95) represents the higher order term dominated by the term (94). The second term (93) plus the third term (94) after self-normalization/studentization can be approximated by a standard normal random variable, which as we shall see in the next section, enables us to achieve valid model comparison inference uniformly over different model structures.

C.3 Generic QLR: Expansion and Limit Theory

Suppose that, for the same data set $\{z_{i,t}\}_{i \leq n, t \leq T}$, we consider two models $j = 1, 2$. For model j , the joint log-likelihood function is

$$\sum_{t \leq T} \sum_{i \leq n} \log f_j(z_{i,t}; \theta_j, \gamma_{j,i,t}),$$

where $f_j(\cdot)$ is known, θ_j collects individual- and time-invariant parameters, and $\gamma_{j,i,t} \equiv \gamma_{g_j(i), m_j(t)}$ denotes fixed effects indexed by known group assignment function $g_j(\cdot)$ and $m_j(\cdot)$ with ranges $\mathcal{G}_j = \{1, \dots, G_j\}$ and $\mathcal{M}_j = \{1, \dots, M_j\}$ respectively. For any i and t with $g_j(i) = g$ and $m_j(t) = m$, let

$$\psi_j(z_{i,t}; \theta_j, \gamma_{j,g,m}) \equiv \log f_j(z_{i,t}; \theta_j, \gamma_{j,i,t}).$$

Then the joint log-likelihood function can be written as

$$\sum_{g \in \mathcal{G}_j} \sum_{m \in \mathcal{M}_j} \sum_{i \in I_g} \sum_{t \in I_m} \psi_j(z_{i,t}; \theta_j, \gamma_{j,g,m}) \equiv \sum_{g \in \mathcal{G}_j} \sum_{m \in \mathcal{M}_j} (n_g T_m) \hat{\Psi}_{j,g,m}(\phi_{j,g,m}),$$

where $\hat{\Psi}_{j,g,m}(\phi_{j,g,m}) \equiv (n_g T_m)^{-1} \sum_{i \in I_g} \sum_{t \in I_m} \psi_j(z_{i,t}; \theta_j, \gamma_{j,g,m})$ and $\phi_{j,g,m} \equiv (\theta_j^\top, \gamma_{j,g,m}^\top)^\top$. The two models may differ in their specifications of the likelihood functions $f_j(\cdot)$ and/or the group assignment functions $g_j(\cdot)$ and/or $m_j(\cdot)$.

Let $\phi_{j,i,t}^* \equiv (\theta_j^{*\top}, \gamma_{j,i,t}^{*\top})^\top$ denote the pseudo true parameters of model j . The Vuong test compares two models by testing

$$H_0 : \sum_{t \leq T} \sum_{i \leq n} \mathbb{E}[\Delta \psi(z_{i,t})] = 0, \text{ where } \Delta \psi(z_{i,t}) \equiv \psi_1(z_{i,t}; \phi_{1,i,t}^*) - \psi_2(z_{i,t}; \phi_{2,i,t}^*) \quad (96)$$

using the QLR statistic

$$QLR_{n,T} \equiv (nT)^{-1/2} \left(\sum_{g \in \mathcal{G}_1} \sum_{m \in \mathcal{M}_1} (n_g T_m) \hat{\Psi}_{1,g,m}(\hat{\phi}_{1,g,m}) - \sum_{g \in \mathcal{G}_2} \sum_{m \in \mathcal{M}_2} (n_g T_m) \hat{\Psi}_{2,g,m}(\hat{\phi}_{2,g,m}) \right), \quad (97)$$

where $\hat{\phi}_{j,g,m} \equiv (\hat{\theta}_j^\top, \hat{\gamma}_{j,g,m}^\top)^\top$ denotes the estimator of $\phi_{j,g,m}^*$ in model j .

Let $\tilde{\Psi}_{j,i} \equiv T^{-1/2} \sum_{t \leq T} \psi_j(z_{i,t}; \phi_{j,i,t}^*)$ and $\hat{\Psi}_{j,\gamma,g,m} \equiv (n_g T_m)^{-1} \sum_{i \in I_g} \sum_{t \in I_m} \psi_{j,\gamma}(z_{i,t}; \phi_{j,g,m}^*)$. To study the properties of the QLR statistic, we suppose that Assumptions 1, 7, 8, 9 and 10 hold for model 1 and model 2, which enables us to apply the expansion of the likelihood function listed in (92)-(95) to both models to obtain

$$QLR_{n,T} = n^{-1/2} \sum_{i \leq n} (\tilde{\Psi}_{1,i} - \tilde{\Psi}_{2,i}) + S_{n,T} + o_p(n^{-1/2}), \quad (98)$$

where

$$S_{n,T} \equiv \sum_{j=1,2} \sum_{g \in \mathcal{G}_j} \sum_{m \in \mathcal{M}_j} \frac{(-1)^j (n_g T_m) \hat{\Psi}_{j,\gamma,g,m}^\top \Psi_{j,\gamma,g,m}^{-1} \hat{\Psi}_{j,\gamma,g,m}}{2(nT)^{1/2}}.$$

Since $z_{i,t}$ are independent across i for all t , $\hat{\Psi}_{1,i} - \hat{\Psi}_{2,i}$ are independent across i . Consequently, the asymptotic distribution of $n^{-1/2} \sum_{i \leq n} (\tilde{\Psi}_{1,i} - \tilde{\Psi}_{2,i})$ is straightforward to derive. However, when the two models are nested or overlapping non-nested, its variance may be zero or arbitrarily close to zero, invalidating inference based solely on this leading term. To conduct valid inference for (92) in a more general setting, the second term $S_{n,T}$ on the right-hand side of (98) must therefore be taken into account when deriving the asymptotic distribution of the QLR statistic.

For this purpose, we define

$$\tilde{\Psi}_{j,\gamma,i,m} \equiv T_m^{-1/2} \sum_{t \in I_m} \tilde{\psi}_{j,\gamma}(z_{i,t}; \phi_{j,g,m}^*),$$

where $\tilde{\psi}_{j,\gamma}(z_{i,t}; \phi_{j,g,m}^*) \equiv (-\Psi_{j,\gamma,g,m})^{-1/2} \psi_{j,\gamma}(z_{i,t}; \phi_{j,g,m}^*)$ for any $i \in I_g$ and $t \in I_m$, and for any $g \in \mathcal{G}_j$ and $m \in \mathcal{M}_j$. Then for any $g \in \mathcal{G}_j$ and $m \in \mathcal{M}_j$, we obtain²⁴

$$\frac{(n_g T_m) \hat{\Psi}_{j,\gamma,g,m}^\top \Psi_{j,\gamma,g,m}^{-1} \hat{\Psi}_{j,\gamma,g,m}}{2T^{1/2}} = \sum_{i \in I_g} \frac{\tilde{\Psi}_{j,\gamma,i,m}^\top \tilde{\Psi}_{j,\gamma,i,m}}{2n_g T^{1/2}} + \sum_{i \in I_g} \sum_{\{i' \in I_g: i' < i\}} \frac{\tilde{\Psi}_{j,\gamma,i,m}^\top \tilde{\Psi}_{j,\gamma,i',m}}{n_g T^{1/2}}. \quad (99)$$

²⁴We adopt the convention that the summation over an empty set is zero. Therefore under this condition, the summation $\sum_{\{i' \in I_g: i' < i\}} \tilde{\Psi}_{j,\gamma,i,m}^\top \tilde{\Psi}_{j,\gamma,i',m} = 0$ if the set $\{i' \in I_g: i' < i\}$ is empty.

The decomposition in (99) is useful to study $QLR_{n,T}$ with $S_{n,T}$ taken into account, since by the independence of $z_{i,t}$ across i for all t , $\tilde{\Psi}_{j,\gamma,i,m}^\top \tilde{\Psi}_{j,\gamma,i,m}$ are independent across i for all m , and $\sum_{\{i' \in I_g: i' < i\}} \tilde{\Psi}_{j,\gamma,i,m}^\top \tilde{\Psi}_{j,\gamma,i',m}$ are uncorrelated across i for all m .

To employ the decomposition in (99) in derivation of the asymptotic distribution of $QLR_{n,T}$, we define

$$\tilde{V}_{j,i} \equiv \frac{\sum_{m \in \mathcal{M}_j} \tilde{\Psi}_{j,\gamma,i,m}^\top \tilde{\Psi}_{j,\gamma,i,m}}{2n_{g_j(i)} T^{1/2}} \quad \text{and} \quad \tilde{U}_{j,i} \equiv \frac{\sum_{\{i' \in I_{g_j(i)}: i' < i\}} \sum_{m \in \mathcal{M}_j} \tilde{\Psi}_{j,\gamma,i,m}^\top \tilde{\Psi}_{j,\gamma,i',m}}{n_{g_j(i)} T^{1/2}} \quad (100)$$

for $j = 1, 2$ and for any $i = 1, \dots, n$. Then

$$\sum_{g \in \mathcal{G}_j} \sum_{m \in \mathcal{M}_j} \frac{(n_g T_m) \hat{\Psi}_{j,\gamma,g,m}^\top \Psi_{j,\gamma,g,m}^{-1} \hat{\Psi}_{j,\gamma,g,m}}{2T^{1/2}} = \sum_{g \in \mathcal{G}_j} \sum_{i \in I_g} (\tilde{V}_{j,i} + \tilde{U}_{j,i}) = \sum_{i \leq n} (\tilde{V}_{j,i} + \tilde{U}_{j,i}),$$

and the expansion of the QLR statistic in (98) can be written as

$$QLR_{n,T} = n^{-1/2} \sum_{i \leq n} (\tilde{\Psi}_i + \tilde{V}_i + \tilde{U}_i) + o_p(n^{-1/2}) \quad (101)$$

where

$$\tilde{\Psi}_i \equiv \tilde{\Psi}_{1,i} - \tilde{\Psi}_{2,i}, \quad \tilde{V}_i \equiv \tilde{V}_{1,i} - \tilde{V}_{2,i}, \quad \tilde{U}_i \equiv \tilde{U}_{1,i} - \tilde{U}_{2,i}. \quad (102)$$

Let $\mathcal{F}_{nT,i}$ denote the sigma field generated by $\{z_{i't}\}_{t \leq T, i' \leq i}$. Then by the independence of $\{z_{i,t}\}_{t \leq T}$ across i , we have

$$\mathbb{E}[\tilde{\Psi}_i^* + \tilde{V}_i^* + \tilde{U}_i | \mathcal{F}_{nT,i-1}] = 0 \quad (103)$$

for any $i = 1, \dots, n$, where $\tilde{\Psi}_i^* \equiv \tilde{\Psi}_i - \mathbb{E}[\tilde{\Psi}_i]$ and $\tilde{V}_i^* \equiv \tilde{V}_i - \mathbb{E}[\tilde{V}_i]$. This shows that $\{\tilde{\Psi}_i^* + \tilde{V}_i^* + \tilde{U}_i\}_{i \leq n}$ forms a martingale difference array. In view of the expansion of the QLR statistic in (101) and the martingale property in (103), an inference procedure which is valid uniformly over the model relationship can be established by the martingale CLT.

Let $\omega_{n,T} \equiv (n^{-1} \sum_{i \leq n} \text{Var}(\tilde{\Psi}_i + \tilde{V}_i + \tilde{U}_i))^{1/2}$. The following assumption is imposed to derive the asymptotic distribution of $QLR_{n,T}$.

Assumption 11. (i) $n\omega_{n,T}^2 \geq K^{-1}$; (ii) $(n\omega_{n,T}^2)^{-1/2} \sum_{i=1}^n (\tilde{\Psi}_i^* + \tilde{V}_i^* + \tilde{U}_i) \rightarrow_d N(0, 1)$ as $n, T \rightarrow \infty$.

Assumption 11(i) ensures that the second term in the expansion of the QLR statistic later in (101) is asymptotically negligible, while Assumption 11(ii) requires that the leading term in the expansion given in (11) admits an asymptotic normal distribution. Assumption

11(ii) can be derived by verifying the sufficient conditions of the martingale CLT (e.g., Corollary 3.1 of Hall and Heyde (1980)). Specifically, let

$$\xi_{nT,i} \equiv (n\omega_{n,T}^2)^{-1/2}(\tilde{\Psi}_i^* + \tilde{V}_i^* + \tilde{U}_i). \quad (104)$$

Then Assumption 11(ii) holds under the following conditions:

$$\sum_{i \leq n} \mathbb{E}[\xi_{nT,i}^2 | \mathcal{F}_{nT,i-1}] \rightarrow_p 1, \quad (105)$$

and for any $\varepsilon > 0$,

$$\sum_{i \leq n} \mathbb{E}[\xi_{nT,i}^2 I\{|\xi_{nT,i}| > \varepsilon\} | \mathcal{F}_{nT,i-1}] \rightarrow_p 0. \quad (106)$$

Details on verification of Assumption 11(ii) under $d_\gamma = 1$ and some low level conditions are provided in Theorem A4 below.

Theorem A3. *Suppose that Assumptions 1, 7, 8, 9 and 10 hold for model 1 and model 2. Then under Assumption 11, we have*

$$\frac{QLR_{n,T} - \mathbb{E}[S_{n,T}] - \overline{QLR}_{n,T}}{\omega_{n,T}} \rightarrow_d N(0, 1) \quad (107)$$

where $\overline{QLR}_{n,T} \equiv n^{-1/2} \sum_{i \leq n} \mathbb{E}[\tilde{\Psi}_i]$.

PROOF OF THEOREM A3. Since Assumptions 1, 7, 8, 9 and 10 hold for both models, we can invoke (101) and then apply Assumption 11(i) to get

$$\frac{QLR_{n,T} - \mathbb{E}[S_{n,T}] - QLR_{n,T}^*}{\omega_{n,T}} = (n\omega_{n,T}^2)^{-1/2} \sum_{i \leq n} (\tilde{\Psi}_i^* + \tilde{V}_i^* + \tilde{U}_i) + o_p(1)$$

which together with Assumption 11(ii) and Slutsky's theorem shows (107). Q.E.D.

We conclude this section by presenting low-level sufficient conditions for verifying Assumption 11(ii) in the case $d_\gamma = 1$. Let $A_{j,i} \equiv (\tilde{\Psi}_{j,\gamma,i,m})_{m \in \mathcal{M}_j}$ for $i = 1, \dots, n$. Then

$$\tilde{V}_{j,i} = \frac{A_{j,i}^\top A_{j,i}}{2n_{g_j(i)} T^{1/2}} \quad \text{and} \quad \tilde{U}_{j,i} = \sum_{\{i' \in I_{g_j(i)} : i' < i\}} \frac{A_{j,i}^\top A_{j,i'}}{n_{g_j(i)} T^{1/2}}.$$

Assumption 12. (i) $\mathbb{E}[(\Delta\psi(z_{i,t}))^4 \omega_{n,T}^{-4}] < K$; (ii) $\mathbb{E}[|\psi_{j,\gamma}(z_{i,t}; \phi_{j,i,t}^*)|^{4+\delta}] < K$; (iii) $G_j^{-1} \max_{g \in \mathcal{G}_j} n_g^{-1} = o(1)$.

Assumptions 12(i, ii) are used to show the Lindeberg condition for the martingale CLT and to bound the moments of $\tilde{\Psi}_{j,\gamma,i,m}$. Assumption 12(iii) requires that under both model 1 and model 2, the number of the groups and/or the size of the smallest group across i should go to infinity.

Theorem A4. *Assumption 11(ii) holds under Assumptions 1, 10(ii), 11(i) and 12.*

The proof of this theorem is provided in the Online Appendix.

References

- ARELLANO, M., AND S. BONHOMME (2009): “Robust Priors in Nonlinear Panel Data Models,” *Econometrica*, 77(2), 489–536.
- ARELLANO, M., AND J. HAHN (2007): *Understanding Bias in Nonlinear Panel Models: Some Recent Developments* vol. 3 of *Econometric Society Monographs*, p. 381–409. Cambridge University Press.
- (2016): “A likelihood-Based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects,” *Global Economic Review*, 45(3), 251–274.
- BESTER, C. A., AND C. HANSEN (2009): “A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects,” *Journal of Business & Economic Statistics*, 27(2), 131–148.
- BONHOMME, S., K. JOCHMANS, AND M. WEIDNER (2025): “A Neyman-Orthogonalization Approach to the Incidental Parameter Problem,” .
- BONHOMME, S., T. LAMADON, AND E. MANRESA (2019): “A Distributional Framework for Matched Employer Employee Data,” *Econometrica*, 87(3), 699–739.
- (2022): “Discretizing Unobserved Heterogeneity,” *Econometrica*, 90(2), 625–643.
- BONHOMME, S., AND E. MANRESA (2015): “Grouped Patterns of Heterogeneity in Panel Data,” *Econometrica*, 83(3), 1147–1184.
- CARRO, J. M. (2007): “Estimating dynamic panel data discrete choice models with fixed effects,” *Journal of Econometrics*, 140(2), 503–528.

- CHEN, X., H. HONG, AND M. SHUM (2007): “Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models,” *Journal of Econometrics*, 141(1), 109–140, Semiparametric methods in econometrics.
- COX, D. R., AND N. REID (1987): “Parameter Orthogonality and Approximate Conditional Inference,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(1), 1–39.
- DHAENE, G., AND K. JOCHMANS (2015): “Split-panel Jackknife Estimation of Fixed-effect Models,” *The Review of Economic Studies*, 82(3), 991–1030.
- DiCICCIO, T. J., M. A. MARTIN, S. E. STERN, AND G. A. YOUNG (1996): “Information Bias and Adjusted Profile Likelihoods,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 189–203.
- DiCICCIO, T. J., AND S. E. STERN (1993): “An Adjustment to Profile Likelihood Based on Observed Information,” *Technical Report, Department of Statistics, Stanford University*.
- FERGUSON, H., N. REID, AND D. R. COX (1991): “Estimating equations from modified profile likelihood,” in *Estimating Functions*. Oxford University Press.
- FERNÁNDEZ-VAL, I. (2009): “Fixed effects estimation of structural parameters and marginal effects in panel probit models,” *Journal of Econometrics*, 150(1), 71–85.
- FERNÁNDEZ-VAL, I., AND M. WEIDNER (2016a): “Individual and time effects in nonlinear panel models with large N, T,” *Journal of Econometrics*, 192(1), 291–312.
- (2016b): “Individual and time effects in nonlinear panel models with large N, T,” *Journal of Econometrics*, 192(1), 291–312.
- (2018): “Fixed Effects Estimation of Large-T Panel Data Models,” *Annual Review of Economics*, 10, 109–138.
- HAHN, J., AND G. KUERSTEINER (2002): “Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large,” *Econometrica*, 70(4), 1639–1657.
- HAHN, J., AND X. LIU (2022): “A Vuong Test for Panel Data Models with Fixed Effects,” unpublished manuscript.
- HAHN, J., AND W. NEWEY (2004): “Jackknife and Analytical Bias Reduction for Nonlinear Panel Models,” *Econometrica*, 72(4), 1295–1319.

- HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. Academic Press, Inc. New York.
- JOCHMANS, K., AND M. WEIDNER (2019): “Fixed-Effect Regressions on Network Data,” *Econometrica*, 87(5), 1543–1560.
- LEE, Y., AND P. C. PHILLIPS (2015): “Model selection in the presence of incidental parameters,” *Journal of Econometrics*, 188(2), 474–489, Heterogeneity in Panel Data and in Nonparametric Analysis in honor of Professor Cheng Hsiao.
- LIAO, Z., AND X. SHI (2020): “A nondegenerate Vuong test and post selection confidence intervals for semi/nonparametric models,” *Quantitative Economics*, 11(3), 983–1017.
- NEYMAN, J., AND E. L. SCOTT (1948): “Consistent Estimates Based on Partially Consistent Observations,” *Econometrica*, 16(1), 1–32.
- NICKELL, S. (1981): “Biases in Dynamic Models with Fixed Effects,” *Econometrica*, 49(6), 1417–1426.
- PACE, L., AND A. SALVAN (2006): “Adjustments of the profile likelihood from a new perspective,” *Journal of Statistical Planning and Inference*, 136(10), 3554–3564.
- RIVERS, D., AND Q. VUONG (2002): “Model Selection Tests for Nonlinear Dynamic Models,” *Econometrics Journal*, 5, 1–39.
- SHI, X. (2015): “A Nondegenerate Vuong Test,” *Quantitative Economics*, 6, 85–121.
- VUONG, Q. H. (1989): “Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses,” *Econometrica*, 57, 307–33.