A Vuong Test for Panel and Network Data Models with Fixed Effects

Xueyuan Liu

February 23, 2022

Abstract

This paper generalizes the Vuong test to nonlinear panel models where the dimension of nuisance parameters grows with the sample size. We allow for disagreements about incidental parameters and obtain a specification test for strictly non-nested, overlapping, and nested models based on a modified likelihood function. The test statistic for panel data models is identical to that in nonparametric models, since the incidental parameters in panel data models are similar to nuisance parameters in nonparametric models. It is also shown the test can accommodate two-way fixed effects, which are widely used in network data models.

1 Introduction

When researchers working with panel data seek to select among models in terms of a few parameters of interest, they can turn to model selection tests. The classical test is the Vuong test. However, incidental parameters—high-dimensional parameters that affect the unbiasednesss of the parameters of interest—are also important for panel data models as they capture unobserved heterogeneity. In the presence of incidental parameters, we cannot easily apply the classical Vuong test to select a panel data model.

An advantage of panel models is to deal with unobserved heterogeneity, which can be modeled as time-invariant individual-specific effects. With the fixed effects approaches to panel data models, we do not need to impose distributional assumptions on the unobserved effects, thereby allowing the unobserved effects to be arbitrarily related with the observed covariates. The fixed effect estimators treat the unobserved effects as parameters to be estimated. It is well-known that maximum likelihood estimates of the panel data models suffer from the incidental parameters problem as noted by Neyman and Scott (1948); that is, the estimation of the fixed effects when the time dimension is short can be severely biased. Some approaches circumvent the inconsistency problem based on the intuition that there are clever ways to avoid estimating the fixed effects; for instance, in linear models, fixed effects are numerically equivalent to the within-group estimator that removes the individual effects by taking differences within each individual, so that first differencing yields models free of fixed effects. However, such estimators generally only apply to specific models, and the existence of such estimators seems to be quite rare when we want to analyze fixed effects estimation of average partial effects (APEs), which are averages of functions of the data, parameters, and unobserved effects. Moreover, fixed effects are unavoidable if we are interested in settings where individuals are clustered at different levels. For example, students may be clustered at class, school, and district level; or observations are clustered according to household, county, and state codes. Therefore, dealing with the incidental parameter problem coming from the noise of estimation of fixed effects is important for panel data models. From this perspective, it is natural to propose a test for model selection among various panel data models in the presence of incidental parameters.

The fixed-T approximation limitations can be overcome by an alternative asymptotic approximation that considers sequences of panels where both N and T increase to infinity (see Arellano and Hahn (2007); Hahn and Kuersteiner (2011); Hahn and Newey (2004)). This large-T approximation makes the incidental parameter problem of fixed effects estimation become an asymptotic bias problem that is easier to tackle. These bias-corrected estimators are designed to remove the O(1/T) bias and generally applicable. There are several ways to achieve this goal in the literature: Hahn and Newey (2004) and Hahn and Kuersteiner (2011) construct an analytical or numerical bias correction of a fixed-effects estimator for nonlinear panel data models. Another approach is to consider estimators from bias-corrected moment equations (see Woutersen (2002), Arellano (2003), Carro (2007), and Fernández-Val (2005)).

In addition, Pace and Salvan (2006) and Arellano and Hahn (2007) propose estimation from a bias-corrected objective function relative to some target criterion. Another strand of literature focuses on a modified objective function where the correction term is designed to remove the O(1/T) bias of the resulting estimator (Bester and Hansen (2009); Arellano and Hahn (2016)). The correction term could be trace-based or determinant-based.

A vast literature assumes that incidental parameters are correctly specified. But often this may not be the case. For example, a researcher may incorrectly specify incidental parameters if she clusters her data at zip code, city, or country level when in fact the data are actually clustered at the individual level. In fact, the specification problem in panel data models is more severe than in cross sectional data because incidental parameters change over time, thereby generating a different set of potential model specifications for every period of time.

At present, we do not have a test for model selection for these situations where there are many incidental parameters. Without a test specifically designed for panel data with incidental parameters, the misspecifications of discrepancy in the structure of incidental parameters could have serious consequences for model selection. Because it affects the unbiasedness of the low-dimensional parameters of interests. If we ignore the specifications of incidental parameters and only select models based on low-dimensional parameters, we may choose the wrong model.

This paper proposes a new model selection test for panel data models by extending the classical Vuong test (Vuong (1989)), which selects from two parametric likelihood models based on their Kullback–Leibler information criterion (KLIC). Suppose there are two panel models, for example, panel probit and panel logit. We do not know whether both or one of the two models would be misspecified because the true model is unknown. But it is of interest to know whether one of the two models is superior to the other. Following Vuong (1989), we derive an LR based test of the null hypothesis that two models are equivalent in terms of their distances to the true model. Under the null hypothesis, both models are equally close to the true data distribution in terms of the Kullback–Leibler (KL) divergence. When the null does not hold, the tests direct the researcher to the model closer to the true distribution with probability approaching one. An important feature of panel data models is the specification of incidental parameters or fixed effects, which is different from the classical Vuong test. We exploit a modified objective function to deal with infinite-dimensional nuisance parameters. Our paper is different from Lee and Phillips (2015) in that they assume the parameter space of fixed effects is common across the candidate models. Therefore their tests choose the model that best fits the data generating process when only a subset of the parameters is of central interest, which is a special case in this paper.

This paper provides a test to select a better model from two competing nonlinear panel models with incidental parameters. It may seem that we can easily extend the criterion function in classical Vuong test to panel data models. Indeed, when there is no incidental parameter, the classical Vuong test allows the researcher to select between two parametric likelihood models based on their Kullback–Leibler information criterion (KLIC) . However, with incidental parameters, the estimators for panel data models are severely biased. This is called the incidental parameters problem, as noted by Neyman and Scott (1948).

In order to extend classical Vuong test to panel data with incidental parameters, we propose three new test statistics based on a new criterion function. In classical Vuong test, the object function is maximized at pseudo-true values. However, the expectation of the concentrated likelihood for panel data with incidental parameters is not maximized at the true value of the parameter. To be consistent with classical Vuong test, we use a modified likelihood function as the new criterion function. The purpose is to generate a closer approximation to the target likelihood function.

The discrepancy in incidental parameters could have serious consequences for model selection; for example, as noted by MacKinnon, Nielsen, and Webb (2019a), there is a vast literature on cluster-robust inference that assumes the structure of the clusters is correctly specified, which is often violated. An interesting case investigated in their paper is a test for the appropriate clustering level in linear regression models. They show that clustering at either the classroom or school level is better than no clustering using data from the Tennessee Student Teacher Achievement Ratio (STAR) experiment. More generally, if we have observations taken from individuals in different geographical locations, there could possibly be clustering at the zip-code, city, county, state, or country level. Even in this simple cross-sectional setting, we need to choose one model among many different specifications of fixed effects. In panel data models, it highly possible that the fixed effects change over time, thereby generating different model specifications for every period of time. This paper's main goal is to provide a way to select a better model from two competing nonlinear panel models with fixed effects, which allows for disagreements about both parameters of interests and incidental parameters.

We offer three different test statistics for researchers who need to deal with all possible relationships between candidate models: overlapping models,nested models, and strictly nonnested models. These three model relationships are classified according to the structure of low-dimensional parameter of interest and high-dimensional incidental parameters. It is shown that these three test statistics have different convergence rates. Users can choose one test according to the specific model relationship.

Since both finite-dimensional parameters of interest and infinite-dimensional incidental parameters present in the model, the non-nested hypotheses is different from the literature featuring Cox (1961), Atkinson (1970), M Hashem Pesaran (1974), Mohammad Hashem Pesaran and Deaton (1978), Mizon and Jean-Francois Richard (1986), Gourieroux and Monfort (1995), Ramalho and Smith (2002), Bontemps, Florens, and Jean-François Richard (2008), among others. There are three possible situations in which two models are strictly non-nested. First, they share the same structure of incidental parameters but have different parameters of interest, which are non-nested. For example, panel logit and panel probit with identical individual-level fixed effects. Second, they have the same parameters of interest but different specifications of incidental parameters, such as panel logit models clustered at different levels. Third, both the parameters of interest and incidental parameters are different in those two models. Continue with the example, panel logit and panel probit clustered at different levels.

The classical Vuong test suffers from a discontinuity problem in the asymptotic distribution of the test statistic, which means that the asymptotics depends on whether the models are nested, non-nested, or overlapping, as noted by Shi (2015), Hsu and Shi (2017), Liao and Shi (2020), Liu and Lee (2019) and so on. Shi (2015) shows that the classical Vuong tests either have severe size distortion or poor power due to this discontinuity problem and propose a one-step nondegenerate Vuong-type test for moment-based models. Liao and Shi (2020) then extend the test to semi/nonparametric models. Hsu and Shi (2017) introduce some additional randomness into the test statistic and derive a one-step test for model selection between conditional moment restriction models. Liu and Lee (2019) show that their intuition carries over to spatial models. This paper follows the manner of Shi (2015) and Liao and Shi (2020), to construct bias-corrected test statistics for panel model selection. The test achieves uniformly asymptotic size control and is consistent regardless of the true DGPs for non-nested, nested, and overlapping models.

This paper is organized as follows. Section 2 sets up the objective function for model selection and compares it with the classical Vuong test. Section 3 conducts the test for non-nested models, nested models, and overlapping models. Section 4 discusses grouped fixed effects in linear models, Section 5 is an extension to two-way fixed effects on network data, Section 6 shows simulation results. Useful lemmas and some important proofs are in the Appendix.

2 The classical Vuong test and extension to panel data models

2.1 The classical Vuong test

When there is no incidental parameter, the classical Vuong test allows the researcher to select between two parametric likelihood models based on their Kullback–Leibler information criterion (KLIC) .

Assume that we have two parametric models $\mathcal{F}$ and $\mathcal{G}$ to choose from, their densities are $y_i \sim f(y;\theta) : \theta \in \Theta \subset R^{d_\theta}$ and $y_i \sim g(y;\gamma) : \gamma \in \Gamma \subset R^{d_\gamma}$ respectively. Since we do not know the true underlining data generating process (DGP), we are interested in the comparison between these two models: which one is closer to the truth. The classical Vuong test (Vuong (1989)) looks at their distances to the DGP in terms of the Kullback and Leibler (1951) Information Criterion (KLIC). Let $\psi_i(\phi)$ be the logarithm of the ratio of the two p.d.f.s:

$$\sum_{i=1}^{n} \psi_i(\phi) = \sum_{i=1}^{n} \log f(y_i;\theta) - \sum_{i=1}^{n} \log g(y_i;\gamma)$$

$\phi_0 = (\theta_0', \gamma_0')$ is the concatenated vector of the pseudo-true values that maximizes the expectation under the density functions $f(y_i;\theta)$ and $g(y_i;\gamma)$:

$$\theta_0 = \arg\max_{\theta\in\Theta} E\left[\sum_{i=1}^{n} \log f(y_i;\theta)\right]$$
$$\gamma_0 = \arg\max_{\gamma\in\Gamma} E\left[\sum_{i=1}^{n} \log g(y_i;\gamma)\right]$$

The distance between the DGP and these two models is measured by the minimum KLIC among distributions in the model. And we would like to select the best model among a collection of competing models is the one that is closest to the true model. In order to do that, Vuong (1989) adopts the following test hypotheses:

$$H_0 : E\left[\sum_{i=1}^{n} \psi_i(\phi_0)\right] = 0$$
$$H_f : E\left[\sum_{i=1}^{n} \psi_i(\phi_0)\right] > 0$$
$$H_g : E\left[\sum_{i=1}^{n} \psi_i(\phi_0)\right] < 0$$

Under H_0, $\mathcal{F}$ and $\mathcal{G}$ are equally good since they are equally distant from true distribution in the Kullback-Leibler sense. If $\mathcal{F}$ is a "better" model, $E[\sum_{i=1}^{n} \psi_i(\phi_0)]$ is expected to be "big", therefore under H_f, f is favored since it is closer to the true distribution. Under H_g, g is "better" as $E[\sum_{i=1}^{n} \psi_i(\phi_0)]$ is small. The test statistics $LR_n = \sum_{i=1}^{n} \psi_i(\hat{\phi}_n)$ is the sample analogue of $LR_0 = \sum_{i=1}^{n} \psi_i(\phi_0)$, where $\widehat{\phi}_n = \left(\widehat{\theta}'_n, \widehat{\gamma}'_n\right)$, and that

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log f(y_i;\theta)$$
$$\widehat{\gamma}_n = \arg\max_{\gamma\in\Gamma} \sum_{i=1}^{n} \log g(y_i;\gamma)$$

If these two models are strictly non-nested: $\mathcal{F} \cap \mathcal{G} = \varnothing$, which is the easiest case, we have

$$n^{-1/2} LR_n = n^{-1/2} \sum_{i=1}^{n} \psi_i\left(\hat{\phi}_n\right) \xrightarrow{d} N\left(0, \omega^2\right)$$

The estimator for the variance term is

$$\hat{\omega}_n^2 = n^{-1} \sum_{i=1}^{n} \left[\psi_i\left(\hat{\phi}_n\right)\right]^2 \xrightarrow{p} \omega^2$$

Thus we have the following asymptotics:

$$H_0 : n^{-1/2}LR_n/\hat{\omega}_n \xrightarrow{d} N(0,1)$$
$$H_f : n^{-1/2}LR_n/\hat{\omega}_n \xrightarrow{\text{a.s.}} +\infty$$
$$H_g : n^{-1/2}LR_n/\hat{\omega}_n \xrightarrow{\text{a.s.}} -\infty$$

If these two models are nested, for example, $\mathcal{F}$ nests $\mathcal{G}$: $\mathcal{G} \subset \mathcal{F}$. Assume that $\dim(\theta) \geq \dim(\gamma)$, under H_0:

$$2LR_n \xrightarrow{d} \chi^2(\dim(\theta) - \dim(\gamma))$$

H_0 is rejected if $2LR_n > c\left(\hat{Q}_n, 1-\alpha\right)$, $c\left(\hat{Q}_n, 1-\alpha\right)$ is $1-\alpha$ quantile of $Y'QY$. $\hat{Q}_n$ is a consistent estimator of Q.

2.2 Incidental parameter problem in panel data models

With incidental parameters, the estimators for panel data models are severely biased. This is called the incidental parameters problem, as noted by Neyman and Scott (1948). Since it violates the unbiasedness properties in classical Vuong test, we need to tackle with it for model selection tests.

In this part, we review the bias-corrected estimator and explain why the modified likelihood function behaves more like a genuine likelihood function. Consider panel observations $\{y_{it}\}$ for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$. The density function is $y_{it} \sim f(y|\theta_0, \alpha_{i0})$, where θ_0 is the common parameter of interest and α_{i0} denotes individual fixed effect, which are considered to be non-stochastic constants. A maximization estimator is defined by

$$\left(\hat{\theta}, \hat{\alpha}_1, \ldots, \hat{\alpha}_n\right) = \underset{\theta,\gamma_1,\ldots,\gamma_n}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{t=1}^{T} \log f\left(x_{it}; \theta, \alpha_i\right)$$

We assume that if n is fixed and $T \to \infty$, the estimator $\left(\hat{\theta}, \hat{\alpha}_1, \ldots, \hat{\alpha}_n\right)$ is consistent for $(\theta_0, \alpha_{10}, \ldots, \alpha_{n0})$ .To simplify notation, we assume $\dim(\alpha_i) = 1$, concentrating out the α_i leads to the characterization

$$\hat{\theta} \equiv \underset{\theta}{\operatorname{argmax}} \sum_i \sum_t \log f\left(y_{it} \mid \theta, \widehat{\alpha}_i(\theta)\right)$$

$$
\widehat{\alpha}_i(\theta) \equiv \underset{\alpha}{\operatorname{argmax}} \sum_t \log f\left(y_{it} \mid \theta, \alpha\right)
$$

The $\widehat{\alpha}_i(\theta)$ on the data only through the i-th observation so there are only T observations to estimate α_i. Let

$$
\theta_T \equiv \underset{\theta}{\operatorname{argmax}} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left[\sum_{t=1}^{T} \log f\left(y_{it} \mid \theta, \widehat{\alpha}_i(\theta)\right)\right]
$$

is biased in general $\theta_T \neq \theta_0$ (Neyman and Scott (1948)'s incidental parameters problem). Before proceeding to the bias-corrected estimator, it will be useful to define some notation:

$$
\begin{aligned}
u_{it}(\theta,\alpha) \equiv& \frac{\partial}{\partial\theta} \log f\left(y_{it} \mid \theta, \alpha\right) \\
v_{it}(\theta,\alpha) \equiv& \frac{\partial}{\partial\alpha} \log f\left(y_{it} \mid \theta, \alpha\right) \\
U_{it}(\theta,\alpha) \equiv& u_{it}(\theta,\alpha) - v_{it}(\theta,\alpha) E\left[v_{it}^{\alpha_i}\right]^{-1} E\left[u_{it}^{\alpha_i}\right] \\
b_i\left(\theta_0\right) =& -\left(\frac{E\left[v_{it}U_{it}^{\alpha_i}\right]}{E\left[v_{it}^{\alpha_i}\right]} - \frac{E\left[U_{it}^{\alpha_i\alpha_i}\right]E\left[v_{it}^{2}\right]}{2\left(E\left[v_{it}^{\alpha_i}\right]\right)^2}\right) \\
\mathcal{I}_i \equiv& -E\left[\frac{\partial U_{it}\left(\theta_0,\alpha_0\right)}{\partial\theta}\right] \\
B =& \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} b_i\left(\theta_0\right)
\end{aligned}
$$

We use the short-hand notation $u_{it} \equiv u_{it}\left(\theta_0,\alpha_0\right), v_{it} \equiv v_{it}\left(\theta_0,\alpha_0\right), U_{it} \equiv U_{it}\left(\theta_0,\alpha_0\right)$ and we denote by v_{it}^{α} and $v_{it}^{\alpha\alpha}$ the first and second derivatives of v_{it} with respect to α_i. We denote $\widetilde{\theta}$ as biased-corrected MLE estimator, which is

$$
\tilde{\theta} \equiv \widehat{\theta} - \frac{\widehat{B}}{T} \tag{1}
$$

$\widehat{B}$ is an estimator of the bias term, which can be a sample analogue of B.[1] The idea behind this method is to expand the incidental parameters bias of the estimator on the

[1] According to Arellano and Hahn (2007), we can estimate the bias term using $\widehat{B}(\widehat{\theta}) = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i\right)^{-1} \frac{1}{n}\sum_{i=1}^{n}\widehat{b}_i(\widehat{\theta})$, $\widehat{\mathcal{I}}_i = -\left(\widehat{E}_T\left[\widehat{u}_{it}^{\theta}\right] - \widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i}\right]\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]^{-1}\widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i\prime}\right]\right)$, and

$$
\widehat{b}_i(\widehat{\theta}) = \left(\frac{-\widehat{E}_T\left[\widehat{v}_{it}^{2}\right]}{\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]}\right)\left\{-\frac{1}{\left(-\widehat{E}_T\left[\widehat{v}_{it}^{2}\right]\right)}\left(\widehat{E}_T\left[\widehat{v}_{it}\widehat{u}_{it}^{\alpha_i}\right] - \widehat{E}_T\left[\widehat{v}_{it}\widehat{v}_{it}^{\alpha_i}\right]\frac{\widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i}\right]}{\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]}\right)\frac{1}{2\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]}\left(\widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i\alpha_i}\right] - \widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i\alpha_i}\right]\frac{\widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i}\right]}{\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]}\right)\right\}
$$

where $\widehat{E}_T(.) = \sum_{t=1}^{T}(.)/T, \widehat{u}_{it}^{\theta} = u_{it}^{\theta}\left(\widehat{\theta},\widehat{\alpha}_i(\widehat{\theta})\right), \widehat{u}_{it}^{\alpha_i} = u_{it}^{\alpha_i}\left(\widehat{\theta},\widehat{\alpha}_i(\widehat{\theta})\right)$, etc.

order of magnitude T, and to subtract an estimate of the leading term of the bias from the estimator. The bias stems from the fact that we use $\sum_i \sum_t \log f\left(y_{it} \mid \theta, \widehat{\alpha}_i(\theta)\right)$ rather than $\sum_i \sum_t \log f(y_{it}; \theta_0, \alpha_i(\theta_0))$ to estimate θ.

2.3 The new criterion function

It may seem that we can easily extend the criterion function in classical Vuong test to panel data models, but the incidental parameter problem induces a difficulty. Using a modified likelihood function as the new criterion function, we extend the classical Vuong test to panel data models with incidental parameters.

Intuitively, we can treat the likelihood function which is evaluated at bias-corrected estimator $\tilde{\theta}$ in equation (1) as criterion function $\psi_f\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right)$:

$$\sum_i \sum_t \psi_f\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right) \equiv \sum_i \sum_t \log f\left(y_{it}; \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right)$$

In classical Vuong test, the object function is maximized at pseudo-true values. However, the expectation of the concentrated likelihood $\psi_f\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right)$ is not maximized at the true value of the parameter. If we use plug-in estimator of incidental parameters $\widehat{\alpha}_i(\theta)$ instead of $\alpha_i(\theta_0)$, the first-order derivative of $\sum_i \sum_t \log f\left(y_{it} \mid \theta, \widehat{\alpha}_i(\theta)\right)$ with respect to θ is not centered at zero when $\theta = \widetilde{\theta}$. In order to be consistent with classical Vuong test, we use a modified likelihood function instead:

$$LM_{nT}(\widetilde{\theta}) \equiv \sum_i \sum_t \psi_f\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right) - \sum_i \widehat{R}^*_{fi}(\widetilde{\theta})$$

The purpose of including an extra modification term $\sum_i \widehat{R}^*_{fi}(\widetilde{\theta})$, which is defined in equation (3), is to generate a closer approximation to the target likelihood function $\sum_i \sum_t \psi_f\left(\theta_0, \alpha_i(\theta_0)\right)$. For example, Bester and Hansen (2009) , Arellano and Hahn (2016), Lee and Phillips (2015) consider trace-based correction term $\widehat{R}^*_{fi}(\widetilde{\theta})$, which depends exclusively on the Hessian and the outer product of the scores of the fixed effects. In the literature, it is shown that the modified likelihood function has zero first-order derivative at $\tilde{\theta}$. We will discuss the formula of $\widehat{R}^*_{fi}(\widetilde{\theta})$ in different models.

3 Test statistics

Based on the modified likelihood functions, we offer three different test statistics for researchers who need to deal with all possible relationships between candidate models.

3.1 The non-nested models

There are three possible situations in which two models are strictly non-nested. First, they share the same structure of incidental parameters but have different parameters of interest, which are non-nested. For example, panel logit and panel probit with identical individual-level fixed effects.

Second, they have the same parameters of interest but different specifications of incidental parameters, such as panel logit models clustered at different levels, e.g. the zip-code, city, county, state, or country level.

Third, both the parameters of interest and incidental parameters are different in those two models. For instance, panel logit clustered at individual level and panel probit clustered at zip-code level.

Suppose there are two panel models f and g which are strictly non-nested but have same structure of fixed effects, for example, the binary panel model with fixed effects is characterized by $y_{it} = 1(\gamma_{i0} + z_{it}'\theta_0 + \varepsilon_{it} \geq 0)$,where ε_{it} conditional on z_{it} either has a logistic or standard normal distribution. The former one is defined by model f and the latter one is model g, they are considered as non-nested cases. It is necessary to consider both finite-dimensional parameters of interest and infinite-dimensional incidental parameters when determining model relationships. Continue with the previous example in the introduction, we have observations taken from individuals in different geographical locations, there could possibly be clustering at the zip-code, city, county, state, or country level.

The main assumption we make about nonnested models is as follows:

Assumption 1 *Suppose that f and g have different parameters of interest and structure of incidental parameters but they are non-nested, i.e., there is no $(\theta_0, \gamma_0, \alpha_{i0}, \lambda_{i0}) \in \Theta \times \Upsilon \times A \times \Lambda$ such that $f(y_{it}, \theta_0, \alpha_i) = g(y_{it}, \gamma_0, \lambda_i)\, \forall y_{it} \in Y$.*

We consider the strictly non-nested models and denote the modified likelihood function evaluated at $\widetilde{\theta}$ in equation (1) as (take model f as an example):

$$LM_{nT}(\widetilde{\theta}) \equiv \sum_i \sum_t \psi_f(y_{it}, \widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})) - \sum_i \widehat{R}^*_{fi}(\widetilde{\theta}) \tag{2}$$

Without the modification term $\sum_i \widehat{R}^*_{fi}(\widetilde{\theta})$, $\frac{1}{nT}\sum_i \sum_t \psi_f\left(y_{it}, \widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right)$ does not have zero first-order derivatives at $\tilde{\theta}$, which violates the property of objective function in the classical Vuong test. Expanding equation (2) around $(\theta_0, \widehat{\alpha}_i(\theta_0))$, the Taylor's Theorem implies that for $\widetilde{\widetilde{\theta}}$ in between θ_0 and $\widetilde{\theta}$:

$$\begin{aligned}
&\sum_i \sum_t \psi_f\left(y_{it}, \widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \sum_i \widehat{R}^*_{fi}(\widetilde{\theta}) \\
=&\sum_i \sum_t \psi_f(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \sum_i \widehat{R}^*_{fi}(\theta_0) && \text{(i)} \\
+&\frac{\partial\left[\sum_i \sum_t \psi_f(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \sum_i \widehat{R}^*_{fi}(\theta_0)\right]}{\partial\theta}(\tilde{\theta} - \theta_0) && \text{(ii)} \\
+&\frac{1}{2}(\tilde{\theta} - \theta_0)' \frac{\partial^2([\sum_i \sum_t \psi_f(y_{i,t}; \widetilde{\widetilde{\theta}}, \widehat{\alpha}_i(\widetilde{\widetilde{\theta}})) - \sum_i R_{fi}(\widetilde{\widetilde{\theta}})])}{\partial\theta\partial\theta'}(\tilde{\theta} - \theta_0) && \text{(iii)}
\end{aligned}$$

where

$$\begin{aligned}
\widehat{R}^*_{fi}(\theta) &\equiv -\frac{1}{2}\frac{\frac{1}{T}\sum_t v^2_{fi,t}(\theta, \alpha_i(\theta))}{E\left[v^{\alpha}_{fi,t}(\theta, \alpha_i(\theta))\right]} \\
\widehat{R}^*_{fi}(\theta_0) &\equiv -\frac{1}{2}\frac{\frac{1}{T}\sum_t v^2_{fi,t}}{E\left[v^{\alpha}_{fi,t}\right]} \\
\widehat{R}^*_{fi}(\widetilde{\theta}) &\equiv -\frac{1}{2}\frac{\frac{1}{T}\sum_t \widehat{v}^2_{fi,t}}{\frac{1}{T}\sum_t \widehat{v}^{\alpha}_{fi,t}}
\end{aligned} \tag{3}$$

and $v \equiv v(\theta_0, \alpha_i(\theta_0))$, $v^{\alpha}_{fi,t} \equiv v^{\alpha}_{fi,t}(\theta_0, \alpha_i(\theta_0))$,$\widehat{v}_{fi,t} \equiv v_{fi,t}[\tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})]$, $\widehat{v}^{\alpha}_{fi,t} \equiv v^{\alpha}_{i,t}[\tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})]$. According to Proposition 3 in the appendix, we prove that in the first term (i):

$$\sum_i \sum_t \psi_f(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \sum_i R_{fi}(\theta_0) = \sum_i \sum_t \psi_f(y_{it}, \theta_0, \alpha_i(\theta_0)) + o_p(1) \tag{4}$$

where $R_{fi}(\theta_0) \equiv -\frac{1}{2}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{fi,t}\right)^2}{E\left[v^{\alpha}_{fi,t}\right]}$. Therefore the first term (i) can be written as:

$$\sum_i \sum_t \psi_f(y_{it}, \theta_0, \alpha_i(\theta_0)) + \sum_i R_{fi}(\theta_0) - \sum_i \widehat{R}^*_{fi}(\theta_0) \tag{5}$$

In Proposition 2, the second term (ii) is shown to be

$$\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)+o_p(1) \tag{6}$$

where $\mathcal{I}_f \equiv \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_{fi}$. We deduce that the third term (iii) becomes:

$$-\frac{1}{2}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)+o_p(1) \tag{7}$$

Combining equations (5), (6) and (7):

$$\begin{aligned}
&\sum_i\sum_t\psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\sum_i\widehat{R}^*_{fi}(\widetilde{\theta}) && (8)\\
=&\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)+\sum_i R_{fi}(\theta_0)-\sum_i\widehat{R}^*_{fi}(\theta_0) && (a)\\
&+\frac{1}{2}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right) && (b)
\end{aligned}$$

The first term (a) is $O_p\left(\sqrt{nT}\right)$ since we know that from Proposition 5 and 8 in the appendix:

$$\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}(\theta_0)-\widehat{R}^*_{fi}(\theta_0)\right]\xrightarrow{d}N(0,\sigma^2_{U_f})$$

$$\frac{1}{\sqrt{nT}}\sum_i\left[\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)-E\left(\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)\right)\right]\xrightarrow{d}N(0,\omega_f^2)$$

where

$$\sigma^2_{U_f}\equiv\lim_{n\to\infty}\frac{1}{2n}\sum_i\frac{\left(E\left[v^2_{fi,t}\right]\right)^2}{\left(E\left[v^{\alpha}_{fi,t}\right]\right)^2} \tag{9}$$

$$\omega_f^2\equiv\lim_{n\to\infty}\frac{1}{n}\sum_i E\left[\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)-E\left(\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)\right)\right]^2 \tag{10}$$

The second term (b) is $O_p(1)$. We divide the equation (8) by $\sqrt{nT}$ at both sides, we have:

$$\begin{aligned}
\frac{1}{\sqrt{nT}}\left[\sum_i\sum_t\psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\sum_i\widehat{R}^*_{fi}(\widetilde{\theta})\right]=&\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)\\
&+\frac{1}{\sqrt{nT}}\sum_i\left[R_{fi}(\theta_0)-\widehat{R}^*_{fi}(\theta_0)\right]+o_p(1)\\
=&\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)+o_p(1)
\end{aligned}$$

and hence

$$\frac{1}{\sqrt{nT}}\sum_i\sum_t \psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\frac{1}{\sqrt{nT}}\sum_i \widehat{R}^*_{fi}(\widetilde{\theta})=\frac{1}{\sqrt{nT}}\sum_i\sum_t \psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)+o_p(1) \tag{11}$$

we further denote:

$$LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)\equiv\left(\sum_i\sum_t\psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\sum_i\widehat{R}^*_{fi}(\widetilde{\theta})\right)-\left(\sum_i\sum_t\psi_g\left(y_{it},\widetilde{\gamma},\widehat{\alpha}_i(\widetilde{\gamma})\right)-\sum_i\widehat{R}^*_{gi}(\widetilde{\gamma})\right)$$

$$LM_{nT}\left(\theta_0,\gamma_0\right)\equiv\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)-\sum_i\sum_t\psi_g\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)$$

Equation (11) becomes:

$$\frac{1}{\sqrt{nT}}LM_{nT}(\tilde{\theta},\tilde{\gamma})=\frac{1}{\sqrt{nT}}LM_{nT}\left(\theta_0,\gamma_0\right)+o_p(1) \tag{12}$$

Theorem 1 *If assumption 1 and condition 1 hold, then under* $H_0 : E\left[LM_{nT}\left(\theta_0,\gamma_0\right)\right]=0:$

$$\frac{\frac{1}{\sqrt{nT}}LM_{nT}(\tilde{\theta},\tilde{\gamma})}{\widehat{\omega}_n}\xrightarrow{d}N(0,1)$$

$H_f : E\left[LM_{nT}\left(\theta_0,\gamma_0\right)\right]>0:$

$$\frac{\frac{1}{\sqrt{nT}}LM_{nT}(\tilde{\theta},\tilde{\gamma})}{\widehat{\omega}_n}\xrightarrow{a.s.}+\infty$$

$H_g : E\left[LM_{nT}\left(\theta_0,\gamma_0\right)\right]<0:$

$$\frac{\frac{1}{\sqrt{nT}}LM_{nT}(\tilde{\theta},\tilde{\gamma})}{\widehat{\omega}_n}\xrightarrow{a.s.}-\infty$$

where $\widehat{\omega}^2_n$ *is an estimator for* ω^2*, and* $\widehat{\omega}^2_n=\frac{1}{nT}\sum_i\sum_t[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\psi_g\left(\tilde{\gamma},\widehat{\alpha}_i(\tilde{\gamma})\right)]^2$, $\omega^2=\lim_{n\to\infty}\frac{1}{n}\sum_i E\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\alpha_i(\gamma_0)\right)\right]^2$.

Proof. See Proposition 6, 8 and appendix 7.4. ■

3.2 The nested models

When we consider the nested cases, there are two possibilities: (i) there is no disagreement about low-dimensional parameters; (ii) the high-dimensional incidental parameters is unchanging over these specifications. For the first case, the test converges at root n, which is slower than the classical Vuong test. This is because the test is driven by the discrepancy in high-dimensional incidental parameters. For the second case, we have chi-sqaure distribution based on modified criterion function, which is identical to classical Vuong test.

Example 1 *(Different cluster levels) As noted by MacKinnon, Nielsen, and Webb (2019a), there is a vast literature on cluster-robust inference that assumes the structure of the clusters is correctly specified, which is often violated. An interesting case investigated in their paper is a test for the appropriate clustering level in linear regression models. They show that clustering at either the classroom or school level is better than no clustering using data from the Tennessee Student Teacher Achievement Ratio (STAR) experiment. More generally, if we have observations taken from individuals in different geographical locations, there could possibly be clustering at the zip-code, city, county, state, or country level.*

Consider the panel model with fixed effects with known variance of the error term, but the cluster levels change over time:

$$\psi_f(y_{i,t};\theta,\alpha_i) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{\left(y_{it}-x_{it}\theta-\alpha_i\right)^2}{\sigma^2} \quad \text{for} \quad t=1,2,\ldots,2J$$

$$\psi_g(y_{i,t};\theta,\alpha_i) = \begin{cases} -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{\left(y_{it}-x_{it}\theta-\alpha_{i1}\right)^2}{\sigma^2} & \text{for} \quad t=1,2,\ldots,J \\ -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{\left(y_{it}-x_{it}\theta-\alpha_{i2}\right)^2}{\sigma^2} & \text{for} \quad t=J+1,2,\ldots,2J \end{cases}$$

For model f, the fixed effects α_i is constant across different time periods, while for model g, α_i changes with time. For instance, observations are clustered at individual level at first half period of time, but clustered at county level later on. We can treat α_i as different cluster levels of observations. Model f is nested in model g since f is equivalent to g when $\alpha_{i1}=\alpha_{i2}$. Assume that $\sigma^2=1$, and the initial value is taken from a stationary distribution, and we obtain

$$\psi_f(y_{i,t};\theta,\alpha_i) = -\frac{1}{2}\left(y_{it}-x_{it}\theta-\alpha_i\right)^2$$

We note that for model f:

$$\begin{aligned} \alpha_i(\theta) &= E\left(y_{it}-x_{it}\theta\right), & \widehat{\alpha}_i\left(\widetilde{\theta}\right) &= \textstyle\sum_{t=1}^{t=2J}\left(y_{it}-x_{it}\widetilde{\theta}\right). \\ R_{fi}(\theta_0) &= -\tfrac{1}{2}\left[\tfrac{1}{2J}\textstyle\sum_{t=1}^{t=2J}\left(y_{it}-x_{it}\theta_0-\alpha_i\right)\right]^2, & \widehat{R}_{fi}\left(\widetilde{\theta}\right) &= -\tfrac{1}{2}\left[\tfrac{1}{2J}\textstyle\sum_{t=1}^{t=2J}\left(y_{it}-x_{it}\widetilde{\theta}-\widehat{\alpha}_i\left(\widetilde{\theta}\right)\right)\right]^2. \\ R^*_{fi}(\theta_0) &= -\tfrac{1}{2}E\left[\tfrac{1}{2J}\textstyle\sum_{t=1}^{t=2J}\left(y_{it}-x_{it}\theta_0-\alpha_i\right)^2\right], & \widehat{R}^*_{fi}\left(\widetilde{\theta}\right) &= -\tfrac{1}{2}\left[\tfrac{1}{2J}\textstyle\sum_{t=1}^{t=2J}\left[y_{it}-x_{it}\widetilde{\theta}-\widehat{\alpha}_i\left(\widetilde{\theta}\right)\right]^2\right]. \end{aligned}$$

We can also compute the modification terms for model g for $t = 1, 2, \ldots, J$ and $t = J+1, 2, \ldots, 2J$, respectively.

3.2.1 Case (i) no disagreement about parameters of interest

We consider the first case and define the nested relationship as follows:

Assumption 2 *Suppose that f and g share the parameters of interest θ_0, but g nests f, i.e., there exists a function $\phi_\alpha(.)$ from A^F to A^G such that for any $\alpha_{fi}(\theta_0)$ in A^F: $f(y_{it}, \theta_0, \alpha_{fi}(\theta_0)) = g(y_{it}, \theta_0, \phi(\alpha_{gi}(\theta_0)))\, \forall y_{it} \in Y$.*

Divide equation (8) by $\sqrt{n}$ at both sides, since the term (b) is $O_p(1)$, we have the following expansion for model g:

$$\frac{1}{\sqrt{n}} \sum_i \sum_{t=1}^{t=2J} \psi_f\left(y_{i,t}; \theta_0, \alpha_i(\theta_0)\right) + \sum_i \frac{1}{\sqrt{n}} \left[R_{fi}(\theta_0) - \widehat{R}^*_{fi}(\theta_0)\right] \tag{13}$$

$$= \frac{1}{\sqrt{n}} \sum_i \sum_{t=1}^{t=2J} \psi_f\left(y_{it}, \widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \frac{1}{\sqrt{n}} \sum_i \widehat{R}^*_{fi}(\widetilde{\theta}) + o_p(1) \tag{14}$$

where $\widehat{R}^*_{fi}(\widetilde{\theta}) \equiv -\frac{1}{2} \frac{\frac{1}{2J}\sum_{t=1}^{t=2J} \widehat{v}^2_{fi,t}}{\frac{1}{2J}\sum_{t=1}^{t=2J} \widehat{v}^\alpha_{fi,t}}$. Similarly, for model model g, we compute the statistic for two different time periods separately, the corresponding modification terms are:

$$\widehat{R}^*_{1gi}(\widetilde{\theta}) \equiv -\frac{1}{2} \frac{\frac{1}{J}\sum_{t=1}^{t=J} \widehat{v}^2_{1gi,t}}{\frac{1}{J}\sum_{t=1}^{t=J} \widehat{v}^\alpha_{1gi,t}}, \quad \widehat{R}^*_{2gi}(\widetilde{\theta}) \equiv -\frac{1}{2} \frac{\frac{1}{J}\sum_{t=J+1}^{t=2J} \widehat{v}^2_{2gi,t}}{\frac{1}{J}\sum_{t=J+1}^{t=2J} \widehat{v}^\alpha_{2gi,t}}.$$

We define

$$\begin{aligned} LM_{nT}\left(\tilde{\theta}, \widehat{\alpha}_i, \widehat{\alpha}_{i1}, \widehat{\alpha}_{i2}\right) \equiv & \sum_i \left\{ \sum_{t=1}^{t=2J} \psi_f\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right) - \sum_{t=1}^{t=J} \psi_g\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_{i1}(\tilde{\theta})\right) \right. \\ & \left. - \sum_{t=J+1}^{t=2J} \psi_g\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_{i2}(\tilde{\theta})\right) \right\} - \sum_i \left[\widehat{R}^*_{fi}(\tilde{\theta}) - \widehat{R}^*_{1gi}(\tilde{\theta}) - \widehat{R}^*_{2gi}(\tilde{\theta})\right] \end{aligned}$$

Theorem 2 *If assumption 2 and condition 1 hold, then under $H_0 : E\left[LM_{nT}\left(\theta_0, \alpha_i, \alpha_{i1}, \alpha_{i2}\right)\right] = 0$:*

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}(\widetilde{\theta}, \widehat{\alpha}_i, \widehat{\alpha}_{i1}, \widehat{\alpha}_{i2})}{\widehat{\sigma}_{U^{nested}}} \xrightarrow{d} N(0,1)$$

Under $H_g : E\left[LM_{nT}\left(\theta_0, \alpha_i, \alpha_{i1}, \alpha_{i2}\right)\right] < 0$ *:*

$$\frac{\frac{1}{\sqrt{n}} LR_{nT}(\widetilde{\theta}, \widehat{\alpha}_i, \widehat{\alpha}_{i1}, \widehat{\alpha}_{i2})}{\widehat{\sigma}_{Unested}} \overset{a.s.}{\to} -\infty$$

where $\widehat{\sigma}^2_{Unested}$ *is an estimator for* $\sigma^2_{Unested}$*:*

$$\widehat{\sigma}^2_{Unested} \equiv \frac{J-1}{2nJ} \sum_i \frac{\left(\frac{1}{J}\sum_{t=1}^{t=J} \widehat{v}^2_{1gi,t}\right)^2}{\left(\frac{1}{J}\sum_{t=1}^{t=J} \widehat{v}^{\alpha}_{1gi,t}\right)^2}$$

$$\sigma^2_{Unested} \equiv \lim_{n\to\infty} \frac{1}{2n} \sum_i \frac{\left(E\left[v^2_{1gi,t}\right]\right)^2}{\left(E\left[v^{\alpha}_{1gi,t}\right]\right)^2}$$

If the information matrix identity holds, $\sigma^2_{Unested} = \frac{1}{2}$*.*

Proof. See Proposition 7 and proof of Theorem 7.4 in the appendix. ∎

3.2.2 Case (ii) no disagreement about incidental parameters

We assume that the parameter of interest θ is different across candidate models while the incidental parameters α_i is unchanging over these specifications. Take the following case as an example:

$$f: \quad y_{it} = \alpha_i + x_{it}\theta_1 + \varepsilon_{it}$$

$$g: \quad y_{it} = \alpha_i + x_{it1}\theta_1 + x_{it2}\theta_2 + \varepsilon_{it}$$

Similar to nested case (i), model f is a special case for model g when $\theta_2 = 0$. A formal definition is as follows:

Assumption 3 *Suppose that* f *and* g *have same structure of incidental parameters but* g *nests* f*, i.e., there exists a function* $\phi(.)$ *from* Θ *to* Γ *such that for any* θ *in* Θ*:* $f\left(y_{it}, \theta_0, \alpha_i(\theta_0)\right) = g\left(y_{it}, \phi(\theta_0), \alpha_i\left(\phi(\theta_0)\right)\right) \forall y_{it} \in Y$*.*

Under the null hypothesis, $R_{fi}\left(\theta_0\right) = R_{gi}(\gamma_0)$ and $\widehat{R}^*_{fi}(\theta_0) = \widehat{R}^*_{gi}(\gamma_0)$ since these two models have same structure of fixed effects. We obtain the following expansion from equation (8):

$$LM_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right) = LM_{nT}\left(\theta_0, \gamma_0\right) + \frac{NT}{2}(\widetilde{\theta} - \theta_0)'\mathcal{I}_f(\widetilde{\theta} - \theta_0) - \frac{NT}{2}(\widetilde{\gamma} - \gamma_0)'\mathcal{I}_g(\widetilde{\gamma} - \gamma_0) + o_p(1)$$

therefore under null assumption, assume that $\dim(\gamma) \geq \dim(\theta)$ (see proof in the appendix 7.4),

$$-2LM_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right) \xrightarrow{d} \chi^2(\dim(\gamma) - \dim(\theta))$$

Theorem 3 *If assumption 3, the information matrix identity holds for model f, condition 1 holds, then under $H_0 : E\left[LM_{nT}(\theta_0, \gamma_0)\right] = 0$, for any $x > 0$,*

$$\Pr\left(-2LM_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right) < x\right) - c\left(\hat{Q}_n, 1 - \alpha\right) \xrightarrow{a.s.} 0$$

Under $H_g : E\left[LM_{nT}(\theta_0, \gamma_0)\right] < 0$:

$$-2LM_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right) \xrightarrow{a.s.} +\infty$$

where $c(Q, 1-\alpha)$ is the $1-\alpha$ quantile of $\chi^2(\dim(\gamma) - \dim(\theta))$.

Proof. *See the appendix 7.4.* ■

Lee and Phillips (2015) assume that the parameter space of fixed effects is common across the candidate models, their tests choose the model that best fits the data generating process when only a subset of the parameters is of central interest, which is equivalent to our results if the information matrix identity holds (see a proof of equivalence in the appendix 7.4).

3.3 The overlapping models

Here we consider two models that are overlapping and they are not non-nested. We consider the case when the variance term might be zero, and the converge rate of the test is root n. This case is different from classical Vuong test as there is no uniform formula for overlapping models in the literature.

Assumption 4 *Suppose that f and g are overlapping and they are not non-nested.*

The test statistic for overlapping models are similar to case (i) in nested models, since the asymptotic distribution is mostly driven by differences in incidental parameters. It is pointed out by Shi (2015) that the high-order bias may dominate the leading terms in LR_n and result in size distortion if we follow Vuong (1989)'s framework and construct a two-step test for overlapping nonnested models. Some recent papers propose one-step nondegenerate test for different data models: Shi (2015) shows that the classical Vuong tests either have severe size distortion or poor power due to this discontinuity problem and propose a one-step nondegenerate Vuong-type test for moment-based models, Liao and Shi (2020) then extend the test to semi/nonparametric models. Hsu and Shi (2017) introduce some additional randomness into the test statistic and derive a one-step test for model selection between conditional moment restriction models, Liu and Lee (2019) show that their intuition carries over to spatial models. This paper follows the manner of Shi (2015) and Liao and Shi (2020), to construct bias-corrected test statistics for panel model selection. The test achieves uniformly asymptotic size control and is consistent regardless of the true DGPs for non-nested, nested, and overlapping models.

Consider a simple case when $LM_{nT}(\theta_0, \gamma_0) = 0$, and the degeneracy problem is considered under the assumption that f and g share the parameters of interest θ_0, while they have different structures of incidental parameters, which corresponds to previous analysis for nested case (i):

$$\begin{aligned}\frac{1}{\sqrt{n}} LM_{nT}\left(\tilde{\theta}, \widetilde{\gamma}\right) =& \frac{1}{\sqrt{n}} \sum_i \left[R_{fi}(\theta_0) - R_{gi}(\gamma_0) - \left(R_{fi}^* - R_{gi}^*\right)\right] + o_p(1) \\ &\xrightarrow{d} N\left(0, \sigma_U^2\right)\end{aligned}$$

where σ_U^2 is defined in equation (15), the proof is included in Proposition 6 in the appendix.

Theorem 4 *If assumption 4 and condition 1 hold, then under* $H_0 : E\left[LM_{nT}(\theta_0, \gamma_0)\right] = 0$:

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}(\tilde{\theta}, \widetilde{\gamma})}{\widehat{\sigma}_U} \xrightarrow{d} N(0,1)$$

under $H_f : E\left[LM_{nT}(\theta_0, \gamma_0)\right] > 0$:

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}(\tilde{\theta}, \widetilde{\gamma})}{\widehat{\sigma}_U} \xrightarrow{a.s.} +\infty$$

under $H_g : E\left[LM_{nT}(\theta_0, \gamma_0)\right] < 0$:

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}(\tilde{\theta}, \widetilde{\gamma})}{\widehat{\sigma}_U} \xrightarrow{a.s.} -\infty$$

where $\widehat{\sigma}_U^2$ *is an estimator for* σ_U^2*, and that*

$$\widehat{\sigma}_U^2 \equiv \frac{T-1}{2nT}\sum_i \left[\frac{\frac{1}{T}\sum_{t=1}^{t=T}\widehat{v}_{fi,t}^2}{\frac{1}{T}\sum_{t=1}^{t=T}\widehat{v}_{fi,t}^{\alpha}} - \frac{\frac{1}{T}\sum_{t=1}^{t=T}\widehat{v}_{gi,t}^2}{\frac{1}{T}\sum_{t=1}^{t=T}\widehat{v}_{gi,t}^{\alpha}}\right]^2$$

$$\sigma_U^2 \equiv \lim_{n\to\infty}\frac{1}{2n}\sum_i E\left[\frac{E\left[v_{fi,t}^2\right]}{E\left[v_{fi,t}^{\alpha}\right]} - \frac{E\left[v_{gi,t}^2\right]}{E\left[v_{gi,t}^{\alpha}\right]}\right]^2 \tag{15}$$

3.4 Dealing with discontinuity problem: an application of Liao and Shi (2020)

Consider a general case when $LM_{nT}(\theta_0,\gamma_0) \neq 0$, the asymptotic distribution follows from previous analysis for nested case (i):

$$\frac{1}{\sqrt{n}}LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right) = \frac{1}{\sqrt{n}}LM_{nT}(\theta_0,\gamma_0) + \frac{1}{\sqrt{n}}\sum_i\left[R_{fi}(\theta_0) - R_{gi}(\gamma_0) - \left(R_{fi}^* - R_{gi}^*\right)\right] + o_p(1)$$

The intuition behinds non-degenerate test is as follows:

$$\begin{aligned}\frac{1}{\sqrt{n}}LM_{nT}(\theta_0,\gamma_0) + \frac{1}{\sqrt{n}}\sum_i\left[R_{fi}(\theta_0) - R_{gi}(\gamma_0) - \left(R_{fi}^* - R_{gi}^*\right)\right]\\ = \sqrt{T}N_1 + N_2 \xrightarrow{d} N\left(0,\sigma_W^2\right)\end{aligned}$$

It is shown in Proposition 9 in the appendix that:

$$\sigma_W^2 \equiv T\omega^2 + \sigma_U^2$$

The discontinuity problem stems from the fact that $\omega^2 = 0$ in nested models, implying that σ_W^2 is nonzero even in the degenerate case. Therefore we propose an estimator of the bias-corrected variance term as follows:

$$\widehat{\sigma}_W^2 \equiv \frac{1}{n}\sum_i\sum_t\left[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\widetilde{\gamma},\widehat{\lambda}_i(\widetilde{\gamma})\right)\right]^2 - \widehat{\sigma}_U^2$$

Theorem 5 *If assumption 4 and condition 1 hold, then under* $H_0 : E\left[LM_{nT}(\theta_0,\gamma_0)\right] = 0 :$

$$\frac{\frac{1}{\sqrt{n}}LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)}{\widehat{\sigma}_W} \xrightarrow{d} N(0,1)$$

under $H_f : E\left[LM_{nT}(\theta_0,\gamma_0)\right] > 0 :$

$$\frac{\frac{1}{\sqrt{n}}LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)}{\widehat{\sigma}_W} \xrightarrow{a.s.} +\infty$$

under $H_g : E\left[LM_{nT}\left(\theta_0, \gamma_0\right)\right] < 0$:

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right)}{\widehat{\sigma}_W} \overset{a.s.}{\to} -\infty$$

where $\widehat{\sigma}_W^2$ *is an estimator for* σ_W^2*:*

$$\widehat{\sigma}_W^2 \equiv \frac{1}{n}\sum_i \sum_t [\psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma}, \widehat{\lambda}_i(\tilde{\gamma})\right)]^2 - \frac{T-1}{2nT}\sum_i \left[\frac{\frac{1}{T}\sum_{t=1}^{t=T} \widehat{v}_{fi,t}^2}{\frac{1}{T}\sum_{t=1}^{t=T} \widehat{v}_{fi,t}^{\alpha}} - \frac{\frac{1}{T}\sum_{t=1}^{t=T} \widehat{v}_{gi,t}^2}{\frac{1}{T}\sum_{t=1}^{t=T} \widehat{v}_{gi,t}^{\alpha}}\right]^2$$

and that

$$\sigma_W^2 \equiv \lim_{n\to\infty}\{\frac{1}{n}\sum_i E\left[\sum_t \left(\psi_f\left(\theta_0, \alpha_i(\theta_0)\right) - \psi_g\left(\gamma_0, \lambda_i(\gamma_0)\right)\right)\right]^2 - \frac{1}{2n}\sum_i \left[\frac{E\left[v_{fi,t}^2\right]}{E\left[v_{fi,t}^{\alpha}\right]} - \frac{E\left[v_{gi,t}^2\right]}{E\left[v_{gi,t}^{\alpha}\right]}\right]^2\}$$

Proof. *See the appendix 7.4.* ■

3.5 Comparison with nonparametric models

Compared with Liao and Shi (2021), it is shown that the test statistics for panel data models is identical to that in nonparametric models. This is because the incidental parameters in panel data models are similar to nuisance parameters in nonparametric models.

The incidental parameters in panel data models are similar to nuisance parameters in nonparametric models. There are many ways to estimate nonparametric components, for example, local estimators and global estimators. The former one includes kernel smoothing, k-nearest neighborhood (k-NN), local polynomials etc., while the latter one also enjoy widespread popularity, such as series (sieve). All of these aforementioned estimators have convergence rate slower than $\sqrt{n}$, which is similar to the slower rate of α_i in panel data models. In this section, we prove the equivalence between nonparametric models and panel data models using series estimators.The number of series terms k equals to the number of individuals in panel data models n, both of them go to infinity. For series estimators, we follow Liao and Shi (2020), if we directly apply the variance term in their test statistics to panel data models, the corresponding estimator of variance is:

$$\sum_i \sum_t [\psi_f\left(\widetilde{\theta}, \alpha_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma}, \alpha_i(\tilde{\gamma})\right)]^2 - \widehat{\sigma}_U^2$$

the discrepancy between their variance term and our variance term is due to the fact that we use time series observations to estimate each individual fixed effects α_i, more specifically (see Proposition 3 in the appendix),

$$\sum_t [\psi_f(y_{it}, \theta_0, \widehat{\alpha}_i(\theta_0)) - \psi_f(y_{it}, \theta_0, \alpha_i(\theta_0))] = R_{fi}(\theta_0) + o_p(1)$$

$$R_{fi}(\theta_0) = -\frac{1}{2} \frac{\left(\frac{1}{\sqrt{T}} \sum_t v_{fi,t}\right)^2}{E\left[v_{fi,t}^{\alpha}\right]}$$

In a nonparametric setting, Liao and Shi 2020 relies on a second-order expansion, for each series term k:

$$\sum_i [\psi_f(\widehat{\alpha}_k) - \psi_f(\alpha_k)] = -\frac{1}{2} \frac{\left(\frac{1}{\sqrt{n}} \sum_i v_{fk,i}\right)^2}{E\left[v_{fk,i}^{\alpha}\right]} + o_p(1)$$

and the numerator is

$$\sum_i \sum_t \left[\psi_f\left(\widetilde{\theta}, \alpha_i(\widetilde{\theta})\right) - \psi_g(\tilde{\gamma}, \alpha_i(\tilde{\gamma}))\right] - \sum_i \left[\widehat{R}^*_{fi}(\widetilde{\theta}) - \widehat{R}^*_{gi}(\widetilde{\gamma})\right]$$

Therefore the test statistics become

$$\frac{\frac{1}{\sqrt{n}} \sum_i \sum_t \left[\psi_f\left(\widetilde{\theta}, \alpha_i(\widetilde{\theta})\right) - \psi_g(\tilde{\gamma}, \alpha_i(\tilde{\gamma}))\right] - \frac{1}{\sqrt{n}} \sum_i \left[\widehat{R}^*_{fi}(\widetilde{\theta}) - \widehat{R}^*_{gi}(\widetilde{\gamma})\right]}{\left\{\sum_i \sum_t [\psi_f\left(\widetilde{\theta}, \alpha_i(\widetilde{\theta})\right) - \psi_g(\tilde{\gamma}, \alpha_i(\tilde{\gamma}))]^2 - \widehat{\sigma}_U^2\right\}^{\frac{1}{2}}}$$

Compare it with Theorem 5, we conclude that the test statistics can be formed in a way identical to Liao and Shi (2020).

4 Grouped Fixed Effects in Linear Models

Bonhomme and Manresa (2015) consider the following linear model with time-varying grouped patterns of heterogeneity:

$$y_{it} = x_{it}'\theta + \alpha_{g_i t} + v_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T \tag{16}$$

where the covariates x_{it} are uncorrelated with v_{it}, but may be arbitrarily correlated with the group-specific unobservables $\alpha_{g_i t}$. The group membership variables are denoted as $g_i \in \{1, \ldots, G\}$, which may change over time. It can be shown that The Vuong test we propose here is able to select a group classification which is closer to the truth. Continue with MacKinnon, Nielsen, and Webb (2019a)'s example, there are two kinds of group memberships, i.e., clustering at the classroom or clustering at the school level. The estimator $\widehat{\phi}_W = \left(\widehat{\theta}'_W, \widehat{\alpha}_{gt}\right)$ is defined as follows:

$$\widehat{\phi}_W = \underset{\phi_W \in \Phi_W}{\operatorname{argmax}} - \sum_i \sum_t \sum_g \left(y_{it} - x'_{it}\theta - \alpha_{gt}\right)^2 I\left\{g_i = g\right\} \tag{17}$$

Since $\widehat{\phi}_W$ is unbiased. We define $\psi_f\left(y_{it}, x_{it}, \phi_W\right) = -\sum_g \left(y_{it} - x'_{it}\theta - \alpha_{gt}\right)^2 I\left\{g_i = g\right\}$, which gives that

$$\sum_i \sum_t \psi_f\left(y_{it}, x_{it}, \widehat{\phi}_W\right) = \sum_i \sum_t \psi_f\left(y_{it}, x_{it}, \phi_{W0}\right) \tag{L.i}$$

$$+\frac{\partial\left[\sum_i \sum_t \psi_f\left(y_{it}, x_{it}, \phi_{W0}\right)\right]}{\partial\theta}(\widehat{\phi}_W - \phi_{W0}) \tag{L.ii}$$

$$+\frac{1}{2}(\widehat{\phi}_W - \phi_{W0})'\frac{\partial^2\left[\sum_i \sum_t \psi_f\left(y_{it}, x_{it}, \widetilde{\widetilde{\phi_W}}\right)\right]}{\partial\theta\partial\theta'}(\widehat{\phi}_W - \phi_{W0}) \tag{L.iii}$$

It seems that we can apply classical Vuong test if the number of groups if finite and group size goes to infinite. If group size is finite, but the number of groups goes to infinte, we need adopt within estimators $\widehat{\phi}_{WG}$ is defined as follows:

$$\widehat{\phi}_{WG} = \underset{\phi_{WG} \in \Phi_{WG}}{\operatorname{argmax}} - \sum_g \sum_t \sum_i \left(\tilde{y}_{git} - \tilde{x}_{git}\theta\right)^2 \tag{18}$$

where $\tilde{x}_{git} = (x_{it} - \bar{x}_{gt})I\left\{g_{it} = g_t\right\}$ with $\bar{x}_{gt} = \frac{1}{M_g}\sum_i x_{it}$, and $\tilde{y}_{git} = (y_{it} - \bar{y}_{gt})I\left\{g_{it} = g_t\right\}$ with $\bar{y}_{gt} = \frac{1}{M_g}\sum_i y_{it}$. M_g is group size, which is assumed to be a constant. Since $\widehat{\phi}_{WG}$ is also unbiased and we define $\psi_f\left(\tilde{y}_{git}, \tilde{x}_{git}, \Phi_{WG}\right) = -\sum_i \left(\tilde{y}_{git} - \tilde{x}_{git}\theta\right)^2 I\left\{g_i = g\right\}$, which gives that

$$\sum_g \sum_t \psi_f\left(\tilde{y}_{git}, \tilde{x}_{git}, \widehat{\phi}_{WG}\right) = \sum_g \sum_t \psi_f\left(\tilde{y}_{git}, \tilde{x}_{git}, \widehat{\phi}_{WG}\right) \tag{G.i}$$

$$+\frac{\partial\left[\sum_g \sum_t \psi_f\left(\tilde{y}_{git}, \tilde{x}_{git}, \widehat{\phi}_{WG}\right)\right]}{\partial\phi}(\widehat{\phi}_{WG} - \phi_{WG0}) \tag{G.ii}$$

$$+\frac{1}{2}(\widehat{\phi}_{WG} - \phi_{WG0})'\frac{\partial^2\left[\sum_i \sum_t \psi_f\left(\tilde{y}_{git}, \tilde{x}_{git}, \widetilde{\widetilde{\phi_W}}\right)\right]}{\partial\phi\partial\phi'}(\widehat{\phi}_{WG} - \phi_{WG0}) \tag{G.iii}$$

It seems that we can apply classical Vuong test.

5 Two-way Fixed Effects on Network Data

Fernandez-Val and Weidner (2016) consider two-way heterogeneity. Our goal is to generalize Vuong test to network model. We introduce some notations: The data consist of $N \times T$ observations $\{(Y_{it}, X_{it}')' : 1 \leq i \leq \ N, 1 \leq t \leq T\}$, for a scalar outcome variable of interest Y_{it} and a vector of explanatory variables X_{it}. We assume that the outcome for individual i at time t is generated by the sequential process:

$$Y_{it} \mid X_i^t, \alpha, \gamma, \beta \sim f_Y\left(\cdot \mid X_{it}, \alpha_i, \gamma_t, \beta\right) \tag{19}$$

where $X_i^t = \left(X_{i1}, \ldots, X_{it}\right), \alpha = \left(\alpha_1, \ldots, \alpha_N\right), \gamma = \left(\gamma_1, \ldots, \gamma_T\right)$, f_Y is a known probability function, and β is a finite dimensional parameter vector. The variables α_i and γ_t are unobserved individual and time effects that in economic applications capture individual heterogeneity and aggregate shocks, respectively. We collect all these unobserved effects in the vector $\phi_{NT} = (\alpha_1, \ldots, \alpha_N, \gamma_1, \ldots, \gamma_T)'$. The true values of the parameters, denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, \ldots, \alpha_N^0, \gamma_1^0, \ldots, \gamma_T^0)'$, are the solution to the population conditional maximum likelihood problem

$$\max_{(\beta, \phi_{NT}) \in \mathbb{R}^{\dim \beta + \dim \phi_{NT}}} \mathbb{E}_\phi\left[\mathcal{L}_{NT}\left(\beta, \phi_{NT}\right)\right] \tag{20}$$

$$\mathcal{L}_{NT}\left(\beta, \phi_{NT}\right) = (NT)^{-1/2}\left\{\sum_{i,t} \log f_Y\left(Y_{it} \mid X_{it}, \alpha_i, \gamma_t, \beta\right) - b\left(v_{NT}'\phi_{NT}\right)^2/2\right\} \tag{21}$$

for every N, T, where $\mathbb{E}_\phi$ denotes the expectation with respect to the distribution of the data conditional on the unobserved effects and initial conditions including strictly exogenous variables, $b > 0$ is an arbitrary constant, $v_{NT} = (1_N', -1_T')'$, and 1_N and 1_T denote vectors of ones with dimensions N and T. We estimate the parameters by solving the sample analog of problem, i.e.

$$\max_{(\beta, \phi_{NT}) \in \mathbb{R}^{\dim \beta + \dim \phi_{NT}}} \mathcal{L}_{NT}\left(\beta, \phi_{NT}\right) \tag{22}$$

as $N, T \to \infty$,

$$\bar{\beta}_{NT} = \beta^0 + \bar{B}_\infty^\beta/T + \bar{D}_\infty^\beta/N + o_P\left(T^{-1} \vee N^{-1}\right) \tag{23}$$

for some $\bar{B}_\infty^\beta$ and $\bar{D}_\infty^\beta$, where $a \vee b := \max(a, b)$.

$$\sqrt{NT}\left(\widehat{\beta}_{NT} - \bar{\beta}_{NT}\right) \to_d \mathcal{N}\left(0, \bar{V}_\infty\right) \tag{24}$$

for some $\bar{V}_\infty$.

$$\sqrt{NT}\left(\widehat{\beta}_{NT}-\beta^0\right)=\sqrt{NT}\left(\widehat{\beta}_{NT}-\bar{\beta}_{NT}\right)+\sqrt{NT}\left(\bar{B}^\beta_\infty/T+\bar{D}^\beta_\infty/N+o_P\left(T^{-1}\vee N^{-1}\right)\right)$$
$$\to_d \mathcal{N}\left(\kappa\bar{B}^\beta_\infty+\kappa^{-1}\bar{D}^\beta_\infty,\bar{V}_\infty\right) \tag{25}$$

where $N/T\to\kappa^2$. We consider panel models with scalar individual and time effects that enter the likelihood function additively through $\pi_{it}=\alpha_i+\gamma_t$. In these models the dimension of the incidental parameters is $\dim\phi_{NT}=N+T$. The parametric part of our panel models takes the form

$$\log f_Y\left(Y_{it}\mid X_{it},\alpha_i,\gamma_t,\beta\right)=:\ell_{it}\left(\beta,\pi_{it}\right) \tag{26}$$

We denote the derivatives of the log-likelihood function ℓ_{it} by

$$\partial_\beta\ell_{it}(\beta,\pi):=\partial\ell_{it}(\beta,\pi)/\partial\beta \tag{27}$$

$$\partial_{\beta\beta'}\ell_{it}(\beta,\pi):=\partial^2\ell_{it}(\beta,\pi)/(\partial\beta\,\partial\beta') \tag{28}$$

$$\partial_{\pi^q}\ell_{it}(\beta,\pi):=\partial^q\ell_{it}(\beta,\pi)/\partial\pi^q, q=1,2,3,\text{ etc.} \tag{29}$$

We drop the arguments β and π when the derivatives are evaluated at the true parameter values β^0 and $\pi^0_{it}:=\alpha^0_i+\gamma^0_t$, e.g. $\partial_{\pi^q\ell_{it}}:=\partial_{\pi^q}\ell_{it}\left(\beta^0,\pi^0_{it}\right)$ We also drop the dependence on NT from all the sequences of functions and parameters, e.g. we use $\mathcal{L}$ for $\mathcal{L}_{NT}$ and ϕ for ϕ_{NT}.

$$\overline{H}=\mathbb{E}_\phi\left[-\partial_{\phi\phi'}\mathcal{L}\right]=\begin{pmatrix}\overline{H}^*_{(\alpha\alpha)} & \overline{H}^*_{(\alpha\gamma)}\\ \left[\overline{H}^*_{(\alpha\gamma)}\right]' & \overline{H}^*_{(\gamma\gamma)}\end{pmatrix}+\frac{b}{\sqrt{NT}}vv' \tag{30}$$

where $v=v_{NT}=(1'_N,-1'_T)'$, $\overline{H}^*_{(\alpha\alpha)}=\operatorname{diag}\left(\frac{1}{\sqrt{NT}}\sum_t\mathbb{E}_\phi\left[-\partial_{\pi^2}\ell_{it}\right]\right)$, $\overline{H}^*_{(\alpha\gamma)it}=\frac{1}{\sqrt{NT}}\mathbb{E}_\phi\left[-\partial_{\pi^2}\ell_{it}\right]$, $\overline{H}^*_{(\gamma\gamma)}=\operatorname{diag}\left(\frac{1}{\sqrt{NT}}\sum_i\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}]\right)$. Since

$$\left\|\overline{H}^{-1}-\operatorname{diag}\left(\overline{H}^*_{(\alpha\alpha)},\overline{H}^*_{(\gamma\gamma)}\right)^{-1}\right\|_{\max}=\mathcal{O}_P\left((NT)^{-1/2}\right) \tag{31}$$

$$\begin{aligned}\widehat{\phi}(\hat{\beta})-\phi^0=&\mathcal{H}^{-1}\mathcal{S}+\mathcal{H}^{-1}\left[\partial_{\phi\beta'}\mathcal{L}\right]\left(\hat{\beta}-\beta^0\right)\\&+\frac{1}{2}\mathcal{H}^{-1}\sum_{g=1}^{\dim\phi}\left[\partial_{\phi\phi'\phi_g}\mathcal{L}\right]\mathcal{H}^{-1}\mathcal{S}\left[\mathcal{H}^{-1}\mathcal{S}\right]_g+R^\phi(\beta)\end{aligned} \tag{32}$$

$$
\mathcal{S} = \begin{pmatrix} \left[\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \partial_\pi \ell_{it}\right]_{i=1,\ldots,N} \\ \left[\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \partial_\pi \ell_{it}\right]_{t=1,\ldots,T}^{T} \end{pmatrix} \tag{33}
$$

$$
\begin{aligned}
\mathcal{L}_{NT}\left(\beta^0, \hat{\phi}\right) &= \mathcal{L}_{NT}\left(\beta^0, \phi^0\right) + \mathcal{S}'\left(\hat{\phi} - \phi^0\right) - \frac{1}{2}\left(\hat{\phi} - \phi^0\right)' \mathcal{H}\left(\hat{\phi} - \phi^0\right) + o_p(1) \\
&= \mathcal{L}_{NT}\left(\beta^0, \phi^0\right) + \frac{1}{2}\mathcal{S}'\mathcal{H}^{-1}\mathcal{S} + o_p(1) \\
&= \mathcal{L}_{NT}\left(\beta^0, \phi^0\right) + \frac{1}{2}\mathcal{S}' \operatorname{diag}\left(\overline{H}^*_{(\alpha\alpha)}, \overline{H}^*_{(\gamma\gamma)}\right)^{-1} \mathcal{S} + o_p(1)
\end{aligned} \tag{34}
$$

$$
\begin{aligned}
&\mathcal{L}_{NT}\left(\beta^0, \hat{\phi}\right) - \mathcal{L}_{NT}\left(\beta^0, \phi^0\right) \\
=&\frac{1}{2}\mathcal{S}' \operatorname{diag}\left(\overline{H}^*_{(\alpha\alpha)}, \overline{H}^*_{(\gamma\gamma)}\right)^{-1} \mathcal{S} + o_p(1) \\
=&\frac{1}{2}\sum_i \left[\left(\frac{1}{\sqrt{NT}} \sum_t \mathbb{E}_\phi\left[-\partial_{\pi^2}\ell_{it}\right]\right)^{-1} \left(\frac{1}{\sqrt{NT}} \sum_t \partial_\pi \ell_{it}\right)^2\right] \\
&+ \frac{1}{2}\sum_t \left[\left(\frac{1}{\sqrt{NT}} \sum_i \mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}]\right)^{-1} \left(\frac{1}{\sqrt{NT}} \sum_i \partial_\pi \ell_{it}\right)^2\right] + o_p(1)
\end{aligned} \tag{35}
$$

$$
\begin{aligned}
&\sqrt{NT}\mathcal{L}_{NT}\left(\beta^0, \hat{\phi}\right) - \sqrt{NT}\mathcal{L}_{NT}\left(\beta^0, \phi^0\right) \\
=&\sqrt{NT}\frac{1}{2}\mathcal{S}' \operatorname{diag}\left(\overline{H}^*_{(\alpha\alpha)}, \overline{H}^*_{(\gamma\gamma)}\right)^{-1} \mathcal{S} + o_p(1) \\
=&\frac{1}{2}\sum_i \left[\left(\mathbb{E}_\phi\left[-\partial_{\pi^2}\ell_{it}\right]\right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_t \partial_\pi \ell_{it}\right)^2\right] \\
&+ \frac{1}{2}\sum_t \left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}]\right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_i \partial_\pi \ell_{it}\right)^2\right] + o_p(1)
\end{aligned} \tag{36}
$$

In order to construct the test statistics, we follow previous notation and define:

$$
LM_{nT}(\widetilde{\theta}) \equiv \sum_i \sum_t \psi_f(y_{it}, \tilde{\theta}, \widehat{\pi}_{it}(\tilde{\theta})) - \sum_i \widehat{R}^*_{f\pi i}(\widetilde{\theta}) - \sum_t \widehat{R}^*_{f\pi t}(\widetilde{\theta}) \tag{37}
$$

where $\sum_i \sum_t \psi_f(y_{it}, \tilde{\theta}, \widehat{\pi}_{it}(\tilde{\theta})) = \sqrt{NT}\mathcal{L}_{NT}\left(\tilde{\theta}, \widehat{\pi}_{it}(\tilde{\theta})\right)$, we further denote:

$$
\begin{aligned}
LM_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right) \equiv &\left(\sum_i \sum_t \psi_f(y_{it}, \tilde{\theta}, \widehat{\pi}_{it}(\tilde{\theta})) - \sum_i \widehat{R}^*_{f\pi i}(\widetilde{\theta}) - \sum_t \widehat{R}^*_{f\pi t}(\widetilde{\theta})\right) \\
&- \left(\sum_i \sum_t \psi_g(y_{it}, \tilde{\gamma}, \widehat{\pi}_{it}(\tilde{\gamma})) - \sum_i \widehat{R}^*_{g\pi i}(\widetilde{\gamma}) - \sum_t \widehat{R}^*_{g\pi t}(\widetilde{\gamma})\right)
\end{aligned}
$$

$$LM_{nT}\left(\theta_0,\gamma_0\right)\equiv\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\pi_{it}(\theta_0)\right)-\sum_i\sum_t\psi_g\left(y_{it},\theta_0,\pi_{it}(\theta_0)\right)$$

where

$$\widehat{R}^*_{f\pi i}\left(\theta_0\right)\equiv-\frac{1}{2}\left(\mathbb{E}_\phi\left[\partial_{\pi^2}\ell_{fit}\right]\right)^{-1}\left[\frac{1}{T}\sum_t\left(\partial_\pi\ell_{fit}\right)^2\right] \tag{38}$$

$$R_{f\pi i}(\theta_0)\equiv-\frac{1}{2}\left(\mathbb{E}_\phi\left[\partial_{\pi^2}\ell_{fit}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_t\partial_\pi\ell_{fit}\right)^2 \tag{39}$$

$$\widehat{R}^*_{f\pi i}(\widetilde{\theta})\equiv-\frac{1}{2}\left(\frac{1}{T}\sum_t\left[\partial_{\pi^2}\widehat{\ell}_{fit}\right]\right)^{-1}\left[\frac{1}{T}\sum_t\left(\partial_\pi\widehat{\ell}_{fit}\right)^2\right] \tag{40}$$

$$\widehat{R}^*_{f\pi t}\left(\theta_0\right)\equiv-\frac{1}{2}\left(\mathbb{E}_\phi\left[\partial_{\pi^2}\ell_{fit}\right]\right)^{-1}\left[\frac{1}{N}\sum_i\left(\partial_\pi\ell_{fit}\right)^2\right] \tag{41}$$

$$R_{f\pi t}(\theta_0)\equiv-\frac{1}{2}\left(\mathbb{E}_\phi\left[\partial_{\pi^2}\ell_{fit}\right]\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i\partial_\pi\ell_{fit}\right)^2 \tag{42}$$

$$\widehat{R}^*_{f\pi t}(\widetilde{\theta})\equiv-\frac{1}{2}\left(\frac{1}{N}\sum_i\left[\partial_{\pi^2}\widehat{\ell}_{fit}\right]\right)^{-1}\left[\frac{1}{N}\sum_i\left(\partial_\pi\widehat{\ell}_{fit}\right)^2\right] \tag{43}$$

For the nonnested case, high-order terms like $\widehat{R}^*_{f\pi t}(\widetilde{\theta})$ do not drive the asymptotic distribution, that is

$$\frac{1}{\sqrt{nT}}LM_{nT}(\tilde{\theta},\tilde{\gamma})=\frac{1}{\sqrt{nT}}LM_{nT}\left(\theta_0,\gamma_0\right)+o_p(1) \tag{44}$$

So that we can directly borrow the Theorem 1. But for nested model in which there is no disagreement about parameters of interest, high-order terms affect the distribution, we can easily generalize Theorem 2 to the two-way network models (assume that time effects are the same across candidate models, that is $\gamma_{t1}=\gamma_{t2}=\gamma_t$):

$$\frac{1}{\sqrt{n}}\sum_i\sum_{t=1}^{t=2J}\psi_f\left(y_{it};\theta_0,\pi_{it}\left(\theta_0\right)\right)+\sum_i\frac{1}{\sqrt{n}}\left[R_{f\pi i}(\theta_0)-\widehat{R}^*_{f\pi i}(\theta_0)\right] \tag{45}$$

$$=\frac{1}{\sqrt{n}}\sum_i\sum_{t=1}^{t=2J}\psi_f\left(y_{it},\widetilde{\theta},\widehat{\pi}_{it}(\widetilde{\theta})\right)-\frac{1}{\sqrt{n}}\sum_i\widehat{R}^*_{f\pi i}(\widetilde{\theta})+o_p(1) \tag{46}$$

We define

$$\begin{aligned}LM_{nT}\left(\tilde{\theta},\widehat{\pi}_{it},\widehat{\pi}_{it1},\widehat{\pi}_{it2}\right)\equiv&\sum_i\left\{\sum_{t=1}^{t=2J}\psi_f\left(y_{it},\tilde{\theta},\widehat{\pi}_{it}(\tilde{\theta})\right)-\sum_{t=1}^{t=J}\psi_g\left(y_{it},\tilde{\theta},\widehat{\pi}_{it1}(\tilde{\theta})\right)\right.\\&\left.-\sum_{t=J+1}^{t=2J}\psi_g\left(y_{it},\tilde{\theta},\widehat{\pi}_{it2}(\tilde{\theta})\right)\right\}-\sum_i\left[\widehat{R}^*_{f\pi i}(\tilde{\theta})-\widehat{R}^*_{1g\pi i}(\tilde{\theta})-\widehat{R}^*_{2g\pi i}(\tilde{\theta})\right]\end{aligned}$$

Theorem 6 *If assumption 2 and condition 1 hold, then under* $H_0 : E\left[LM_{nT}\left(\theta_0, \pi_{it}, \pi_{it1}, \pi_{it2}\right)\right] = 0 :$

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}\left(\tilde{\theta}, \widehat{\pi}_{it}, \widehat{\pi}_{it1}, \widehat{\pi}_{it2}\right)}{\widehat{\sigma}_{U^{nested}}} \xrightarrow{d} N(0,1)$$

Under $H_g : E\left[LM_{nT}\left(\theta_0, \pi_{it}, \pi_{it1}, \pi_{it2}\right)\right] < 0 :$

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}\left(\tilde{\theta}, \widehat{\pi}_{it}, \widehat{\pi}_{it1}, \widehat{\pi}_{it2}\right)}{\widehat{\sigma}_{U^{nested}}} \xrightarrow{a.s.} -\infty$$

where $\widehat{\sigma}^2_{U^{nested}}$ *is an estimator for* $\sigma^2_{U^{nested}}$.

Proof. We denote $H_{fi} = \frac{1}{\sqrt{2J}} \sum_{t=1}^{t=2J} \partial_\pi \ell_{fi,t}$, $H_{1gi} = \frac{1}{\sqrt{J}} \sum_{t=1}^{t=J} \partial_\pi \ell_{1gi,t}$, $H_{2gi} = \frac{1}{\sqrt{J}} \sum_{t=J+1}^{t=2J} \partial_\pi \ell_{2gi,t}$, $O_{fi} = E\left(\frac{1}{\sqrt{2J}} \sum_{t=1}^{t=2J} \partial_\pi \ell_{fi,t}\right)^2$, $O_{1gi} = E(\frac{1}{\sqrt{J}} \sum_{t=1}^{t=J} \partial_\pi \ell_{1gi,t})^2$, $O_{2gi} = E\left(\frac{1}{\sqrt{J}} \sum_{t=J+1}^{t=2J} \partial_\pi \ell_{2gi,t}\right)^2$, under null hypothesis, $\alpha_{i1} = \alpha_{i2} = \alpha_i$, $E\left[\partial_{\pi^2} \ell_{fi,t}\right] = E\left[\partial_{\pi^2} \ell_{1gi,t}\right] = E\left[\partial_{\pi^2} \ell_{2gi,t}\right]$, $E\left[\left(\partial_\pi \ell_{2gi,t}\right)^2\right] = E\left[\left(\partial_\pi \ell_{1gi,t}\right)^2\right] = E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]$, $H_{ft} = \frac{1}{\sqrt{2}}\left(H_{1gi} + H_{2gi}\right)$.

$TH_{ft} = \frac{1}{\sqrt{2K}} \sum_{i=1}^{i=2K} \partial_\pi \ell_{fi,t}$, $TH_{1gt} = \frac{1}{\sqrt{K}} \sum_{i=1}^{i=K} \partial_\pi \ell_{1gi,t}$, $TH_{2gt} = \frac{1}{\sqrt{K}} \sum_{i=K+1}^{i=2K} \partial_\pi \ell_{2gi,t}$, $TO_{ft} = E\left(\frac{1}{\sqrt{2K}} \sum_{i=1}^{i=2K} \partial_\pi \ell_{fi,t}\right)^2$, $TO_{1gt} = E(\frac{1}{\sqrt{K}} \sum_{t=1}^{t=K} \partial_\pi \ell_{1gi,t})^2$, $TO_{2gt} = E\left(\frac{1}{\sqrt{K}} \sum_{i=K+1}^{i=2K} \partial_\pi \ell_{2gi,t}\right)^2$, under null hypothesis, $\gamma_{t1} = \gamma_{t2} = \gamma_t$, $E\left[\partial_{\pi^2} \ell_{fi,t}\right] = E\left[\partial_{\pi^2} \ell_{1gi,t}\right] = E\left[\partial_{\pi^2} \ell_{2gi,t}\right]$, $E\left[\left(\partial_\pi \ell_{2gi,t}\right)^2\right] = E\left[\left(\partial_\pi \ell_{1gi,t}\right)^2\right] = E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]$, $TH_{ft} = \frac{1}{\sqrt{2}}\left(TH_{1gt} + TH_{2gt}\right)$, and that

Lemma 1

$$\begin{aligned}\sum_{i=1}^{i=2K}\left[\left(\mathbb{E}_\phi\left[-\partial_{\pi^2} \ell_{fi,t}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}} \sum_t \partial_\pi \ell_{fi,t}\right)^2\right] =& \sum_{i=1}^{i=K}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2} \ell_{1gi,t}]\right)^{-1}\left(\frac{1}{\sqrt{T}} \sum_t \partial_\pi \ell_{1gi,t}\right)^2\right] \\ &+ \sum_{i=K+1}^{i=2K}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2} \ell_{2gi,t}]\right)^{-1}\left(\frac{1}{\sqrt{T}} \sum_t \partial_\pi \ell_{2gi,t}\right)^2\right]\end{aligned} \tag{47}$$

Proof.

$$
\begin{aligned}
&\sum_{i=1}^{i=2K}\left[\left(\mathbb{E}_\phi\left[-\partial_{\pi^2}\ell_{fi,t}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_t \partial_\pi \ell_{fi,t}\right)^2\right] \\
&=\sum_{i=1}^{i=2K}\left[\left(\mathbb{E}_\phi\left[-\partial_{\pi^2}\ell_{fi,t}\right]\right)^{-1}\left(\frac{1}{T}\sum_t \left(\partial_\pi \ell_{fi,t}\right)^2\right)\right] \\
&=\sum_{i=1}^{i=2K}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right]+\sum_{i=1}^{i=2K}\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\left[\left(\frac{1}{T}\sum_t\left(\partial_\pi \ell_{fi,t}\right)^2\right)-E\left[(\partial_\pi \ell_{fi,t})^2\right]\right] \\
&=\sum_{i=1}^{i=K}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right]+\sum_{i=K+1}^{i=2K}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right] \\
&\quad+\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\sum_{i=1}^{i=K}\left[\left(\frac{1}{T}\sum_t\left(\partial_\pi \ell_{fi,t}\right)^2\right)-E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right] \\
&\quad+\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\sum_{i=K+1}^{i=2K}\left[\left(\frac{1}{T}\sum_t\left(\partial_\pi \ell_{fi,t}\right)^2\right)-E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right] \\
&=\sum_{i=1}^{i=K}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{1gi,t}]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_t \partial_\pi \ell_{fi,t}\right)^2\right]+\sum_{i=K+1}^{i=2K}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{2gi,t}]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_t \partial_\pi \ell_{fi,t}\right)^2\right]
\end{aligned}
\tag{48}
$$

■

Lemma 2

$$
\begin{aligned}
\sum_{t=1}^{t=2J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i \partial_\pi \ell_{fi,t}\right)^2\right]=&\sum_{t=1}^{t=J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{1gi,t}]\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i \partial_\pi \ell_{1gi,t}\right)^2\right] \\
&+\sum_{t=J+1}^{t=2J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{2gi,t}]\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i \partial_\pi \ell_{2gi,t}\right)^2\right]
\end{aligned}
\tag{49}
$$

Proof.

$$
\begin{aligned}
&\sum_{t=1}^{t=2J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i \partial_\pi \ell_{fi,t}\right)^2\right]\\
&=\sum_{t=1}^{t=2J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\left(\frac{1}{N}\sum_i \left(\partial_\pi \ell_{fi,t}\right)^2\right)\right]\\
&=\sum_{t=1}^{t=2J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right]+\sum_{t=1}^{t=2J}\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\left[\left(\frac{1}{N}\sum_i \left(\partial_\pi \ell_{fi,t}\right)^2\right)-E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right]\\
&=\sum_{t=1}^{t=J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right]+\sum_{t=J+1}^{t=2J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right]\\
&\quad+\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\sum_{t=1}^{t=J}\left[\left(\frac{1}{N}\sum_i \left(\partial_\pi \ell_{fi,t}\right)^2\right)-E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right]\\
&\quad+\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{fi,t}]\right)^{-1}\sum_{t=J+1}^{t=2J}\left[\left(\frac{1}{N}\sum_i \left(\partial_\pi \ell_{fi,t}\right)^2\right)-E\left[\left(\partial_\pi \ell_{fi,t}\right)^2\right]\right]\\
&=\sum_{t=1}^{t=J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{1gi,t}]\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i \partial_\pi \ell_{1gi,t}\right)^2\right]+\sum_{t=J+1}^{t=2J}\left[\left(\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{2gi,t}]\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i \partial_\pi \ell_{2gi,t}\right)^2\right]
\end{aligned}
\tag{50}
$$

■

■

6 Monte Carlo Study

We examined the size and power of Vuong test in this section. We consider the following data generating process:

$$
\begin{aligned}
y_{it} &= \mathbf{1}\left(x_{it}\theta_0+\alpha_i+\varepsilon_{it}>0\right), \quad \alpha_i \sim N(0,1)\\
x_{it} &= t/10 + x_{i,t-1}/2 + u_{it}, \quad x_{i0}=u_{i0}, \quad u_{it}=U(-1/2,1/2)
\end{aligned}
$$

If $\varepsilon_{it} \sim N(0,1)$, we call the model "probit". We consider two kinds of fixed effects: individual fixed effects and grouped fixed effects. There are two competing model to choose from: panel logit and panel probit.

Table 1: Simulation designs for power of the test

	DGP	Model I	Model II
Design 1	logit with individual FE	true model	probit with individual FE
Design 2	probit with individual FE	true model	probit with grouped FE
Design 3	logit with grouped FE	true model	probit with grouped FE
Design 4	probit with grouped FE	logit with individual FE	probit with individual FE

We report the empirical powers of the testin Table 2, which is the percentage of rejecting H_0 in favor of H_f as well as the percentage of rejecting H_0 in favor of H_g. In design 1, 2 and 3, the tests never reject H_0 in favor of H_g, because H1 is the true model, we find that the power is close to 1 because these two competing models are sufficiently far away from each other, so that it is very likely to reject the H_0.

Table 2: Empirical power of the test

N	T	Design 1	Design 2	Design 3	Design 4
20	20	0.9014	0.9520	0.9800	0.9224
50	50	0.9735	0.9851	0.9780	0.9730
100	100	0.9150	1.0000	1.0000	1.0000
500	6	0.9990	0.9023	1.0000	1.0000
500	8	1.0000	0.9620	1.0000	1.0000
500	12	1.0000	1.0000	1.0000	1.0000
500	18	1.0000	1.0000	1.0000	1.0000

7 Appendix

7.1 Regularity Conditions

Condition 1

1. *$n, T \to \infty$ such that $(n/T) \to \rho$,where $0 < \rho < \infty$.*

2. *(i) The function $\log f(.;\theta,\alpha)$ is continuous in $(\theta,\alpha) \in \mathcal{Y}$; (ii) the parameter space $\mathcal{Y}$ is*

compact; (iii) there exists a function such that $|\partial \log f(y_{it};\theta,\alpha_i)| \leq M(y_{it})$,

$$\left|\frac{\partial \log f(y_{it};\theta,\alpha_i)}{\partial(\theta,\alpha_i)}\right| \leq M(y_{it})$$

and $\sup_i E\left[M(y_{it})^{33}\right] < \infty$.

3. *For each* $\eta > 0$,

$$\inf_i \left[G_{(i)}(\theta_0,\alpha_{i0}) - \sup_{|(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_0)|>\eta|} G_{(i)}(\theta,\alpha)\right] > 0$$

where

$$\widehat{G}_{(i)}(\theta,\alpha_i) \equiv T^{-1}\sum_{t=1}^{T}\log f(y_{it};\theta,\alpha_i) \equiv T^{-1}\sum_{t=1}^{T} g(y_{it};\theta,\alpha_i)$$

$$G_{(i)}(\theta,\alpha_i) \equiv E\left[\log f(y_{it};\theta,\alpha_i)\right]$$

Let $\mathcal{I}_i \equiv E\left[U_{it}U_{it}'\right]$.

4. *(i) There exists some* $M(y_{it})$ *such that*

$$\left|\frac{\partial^{m_1+m_2}\log f(z_{it};\theta,\alpha_i)}{\partial\theta^{m_1}\partial\alpha_i^{m_2}}\right| \leq M(z_{it}), \quad 0 \leq m_1+m_2 \leq 1,\ldots,6$$

and $\sup_i E\left[M(y_{it})^{Q}\right] < \infty$ *for some* $Q > 64$; *(ii)* $\bar{E}\left[\mathcal{I}_i\right] > 0$; *(iii)* $\min_i E\left[v_{it}^2\right] > 0$.

5. $\sup_i\left[\frac{1}{T}\left|\frac{\partial S_n(\theta)}{\partial\theta'}\right|\right] = o_p(1)$, *where*

$$S_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty}E\left[U_i^{\alpha_i}\tilde{V}_{it-l}\right] + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\alpha_i\alpha_i}\right]\mathrm{vec}\left(\sum_{l=-\infty}^{\infty}E\left[\tilde{V}_{it}\tilde{V}_{it-l}'\right]\right) + o_p(1)$$

and $\frac{1}{T}\sum_{t=1}^{T}\tilde{V}_{it} = -\left(E\left[\frac{\partial v_i}{\partial\alpha_i}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)$.

7.2 Lemmas

Lemma 3 *(Arellano and Hahn (2016),Theorem 2) Under condition 1,*

$$\sqrt{nT}(\tilde{\theta}-\theta_0) = \mathcal{I}_f^{-1}\left\{\frac{1}{\sqrt{nT}}\sum_i\sum_t U_{fi,t}(\theta_0,\alpha_{i0})\right\} + o_p(1)$$

and that

$$\sqrt{nT}(\tilde{\theta}-\theta_0) \xrightarrow{d} N\left(0,\mathcal{I}_f^{-1}\right)$$

Lemma 4 *(Hahn and Kuersteiner (2011), Lemma 7) Assume that* $\{Q_t, t = 1, 2...\}$ *is a stationary, mixing sequence with* $E[Q_t] = 0$ *and* $E\left[|Q_t|^{2r+\delta}\right] < \infty$ *for any positive integer* r*, some* $\delta > 0$ *and all* t*. Let* $\mathcal{A}_t \equiv \sigma(Q_t, Q_{t-1}, Q_{t-2}, \ldots), \mathcal{B}_t \equiv \sigma(Q_t, Q_{t+1}, Q_{t+2}, \ldots)$*, and*

$$\alpha(m) \equiv \sup_t \sup_{A \in \mathcal{A}_t, B \in \mathcal{B}_{t+m}} |P(A \cap B) - P(A)P(B)| \tag{51}$$

Then, for any m *such that* $1 \leq m < C(r)\, n$*,*

$$E\left[\left(\sum_{i=1}^{n} Q_i\right)^{2r}\right] \leq C(r)\, |Q_t|^{2r+\delta} \left[n^r m^{2r}\right] \tag{52}$$

Lemma 5 *(Hahn and Kuersteiner (2011), Lemma 1) Suppose that, for each* i*,* $\{\xi_{it}, t = 1, 2, ...\}$ *is a mixing sequence with* $E[\xi_{it}] = 0$ *for all* i, t*. Let* $\mathcal{A}_t^{\xi i} \equiv \sigma(\xi_{it}, \xi_{it-1}, \xi_{it-2}, \ldots), \mathcal{B}_t^{\xi i} \equiv \sigma(x_{it}, x_{it+1}, x_{it+2}, \ldots)$*, and*

$$\alpha_i(m) \equiv \sup_t \sup_{A \in \mathcal{A}_t^{\xi i}, B \in \mathcal{B}_{t+m}^{\xi i}} |P(A \cap B) - P(A)P(B)| \tag{53}$$

Assume that $\sup_i |\alpha_i(m)| \leq C a^m$ *for some* a *such that* $0 < a < 1$ *and some* $0 < C < \infty$*. We assume that* $\{\xi_{it}, t = 1, 2, 3, ...\}$ *are independent across* i*. We also assume that* $n = O(T)$*. Finally, assume that* $E\left[|\xi_{it}|^{6+\delta}\right] < \infty$ *for some* $\delta > 0$*. We then have*

$$P\left[\max_{1 \leq i \leq n} \left|\frac{1}{T} \sum_{t=1}^{T} \xi_{it}\right| > \eta\right] = o\left(T^{-1}\right) \tag{54}$$

for every $\eta > 0$*. Now assume that* $E\left[|\xi_{it}|^{10q+12+\delta}\right] < \infty$ *for some* $\delta > 0$ *and some integer* $q \geq 1$*. Then,*

$$P\left[\max_{1 \leq i \leq n} \left|\frac{1}{T} \sum_{t=1}^{T} \xi_{it}\right| > \eta T^{\frac{1}{10} - v}\right] = o\left(T^{-q}\right) \tag{55}$$

for every $\eta > 0$ *and* $0 < v < (100q + 120)^{-1}$*.*

Lemma 6 *(Hahn and Kuersteiner (2011), Lemma 3) Assume that* y_{it} *satisfies condition 1, and let* $\xi(y_{it}, \phi)$ *be a function indexed by the parameter* $\phi \in int\Phi$*, where* Φ *is a convex subset of* $\mathbb{R}^p$*. For any sequence* $\phi_i \in int\Phi$*, assume that* $E[\xi(y_{it}, \phi_i)] = 0$*. Further assume that* $\sup_\phi \|\xi(y_{it}, \phi)\| \leq \mathbf{M}(y_{it})$ *for some* $\mathbf{M}(y_{it})$ *such that* $E\left[\mathbf{M}(y_{it})^4\right] < \infty$*. Let* $\Sigma_{nT} = \sum_{i=1}^{n} \Sigma_{iT}^{\xi\xi}$

with $\Sigma_{iT}^{\xi\xi} = Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(y_{it},\phi_i\right)\right)$. *Denote the smallest eigenvalue of* $\Sigma_{iT}^{\xi\xi}$ *by* λ_{iT}^{ξ}, *and assume that* $\inf_i \inf_T \lambda_{iT}^{\xi} > 0$. *Then,*

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\xi\left(y_{it},\phi_i\right) \xrightarrow{d} N\left(0, f^{\xi\xi}\right), \text{ and } \sup_i \left\|\Sigma_{iT}^{\xi\xi} - f_i^{\xi\xi}\right\| \to 0, \tag{56}$$

where $f^{\xi\xi} \equiv \lim \frac{1}{n}\sum_{i=1}^{n} f_i^{\xi\xi}$, *and* $f_i^{\xi\xi} \equiv \sum_{j=-\infty}^{\infty} E\left[\xi(y_{it},\phi_i)\xi\left(y_{it-j},\phi_i\right)'\right]$.

Lemma 7 *(Hahn and Kuersteiner (2011),Theorem 4)* $P\left[\max_{1\le i\le n}\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\widehat{\alpha}_i\left(\epsilon\right) - \alpha_{i0}\right| \ge \eta\right] = o\left(T^{-1}\right)$ *for every* $\eta > 0$.

Lemma 8 *(i)* $\alpha_i^{\theta} = -\frac{E\left[v_{it}^{\theta}\right]}{E\left[v_{it}^{\alpha}\right]}$; *(ii)* $\frac{\partial\widehat{\alpha}_i}{\partial\theta} = -\frac{\sum_{t=1}^{T} v_{it}^{\theta}}{\sum_{t=1}^{T} v_{it}^{\alpha}}$.

Proof. *Consider* $\alpha_i\left(\theta, F_i\left(\varepsilon\right)\right)$ *solves the estimating equation*

$$\int v_i\left[\theta, \alpha_i\left(\theta, F_i\left(\varepsilon\right)\right)\right] dF_i\left(\varepsilon\right) = 0 \tag{57}$$

Differentiating the LHS with respect to θ *and* ε, *we obtain*

$$0 = \int_i^{\theta} dF_i\left(\varepsilon\right) + \alpha_i^{\theta}\int v_i^{\alpha} dF_i\left(\varepsilon\right) \tag{58}$$

$$0 = \alpha_i^{\varepsilon}\int_i^{\alpha_i} dF_i\left(\varepsilon\right) + \int_i d\Delta_{iT} \tag{59}$$

where

$$\Delta_{iT} = \frac{dF_i\left(\varepsilon\right)}{d\varepsilon} = \sqrt{T}\left(\widehat{F}_i - F_i\right) \tag{60}$$

We solve for these equations and evaluate them at $\varepsilon = 0$ *gives (i):*

$$\alpha_i^{\theta} = -E\left[v_{it}^{\alpha}\right]^{-1} E\left[v_{it}^{\theta}\right] = O_p\left(1\right) \tag{61}$$

For (ii), by the definition of $\widehat{\alpha}_i\left(\theta\right)$,

$$\sum_{t=1}^{T} v_{it}\left(\theta, \widehat{\alpha}_i\left(\theta\right)\right) = 0 \tag{62}$$

Differentiating the LHS with respect to θ, *we obtain*

$$\sum_{t=1}^{T} v_{it}^{\theta} + \sum_{t=1}^{T} v_{it}^{\alpha}\left[\frac{\partial\widehat{\alpha}_i}{\partial\theta}\right] = 0 \tag{63}$$

it follows that

$$\frac{\partial\widehat{\alpha}_i}{\partial\theta} = -\frac{\sum_{t=1}^{T} v_{it}^{\theta}}{\sum_{t=1}^{T} v_{it}^{\alpha}} \tag{64}$$

■

Lemma 9

$$\frac{1}{nT}\sum_i \sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)=-\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+O_p\left(\frac{1}{T\sqrt{T}}\right) \tag{65}$$

$$\frac{1}{nT}\sum_i(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right))^2=\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{(E\left[v_{i,t}^{\alpha}\right])^2}+O_p\left(\frac{1}{T\sqrt{T}}\right) \tag{66}$$

$$\frac{1}{nT}\sum_i(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right))^3=-\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{(E\left[v_{i,t}^{\alpha}\right])^3}+o_p\left(\frac{1}{T\sqrt{T}}\right) \tag{67}$$

Proof.

Since we have

$$\frac{1}{T}\sum_{t=1}^{T} v\left(y_{i,t};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)=0$$

Let $F \equiv (F_1,\ldots,F_n)$ *denote the collection of marginal distribution functions of* y_{it}.*Let* $\hat{F}_i$ *denote the empirical distribution function for the observation* i. *Define* $F_i(\epsilon) \equiv F_i + \epsilon\sqrt{T}\left(\hat{F}_i - F_i\right)$ *for* $\epsilon \in \left[0, T^{-1/2}\right]$. *For each fixed* θ *and* ϵ, *let* $\alpha_i(\epsilon)$ *be the solution to the estimating equation*

$$0=\int v\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)$$

By a Taylor series expansion, we have

$$\widehat{\alpha}_i\left(\theta_0\right)-\alpha_{i0}=\alpha_i\left(\frac{1}{\sqrt{T}}\right)-\alpha_i(0)=\frac{1}{\sqrt{T}}\alpha_i^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})$$

where $\alpha_i^{\epsilon}(\epsilon) \equiv d\alpha_i(\epsilon)/d\epsilon, \alpha_i^{\epsilon\epsilon}(\epsilon) \equiv d^2\alpha_i(\epsilon)/d\epsilon^2, \ldots,$ *and* $\widetilde{\epsilon}$ *is somewhere in between 0 and* $\frac{1}{\sqrt{T}}$.

Let $h_i(\cdot,\epsilon) \equiv v\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]$, *the first order condition could be written as*

$$0=\int h_i(\cdot,\epsilon)dF_i(\epsilon)$$

Differentiating repeatedly with respect to ϵ,*we obtain*

$$0=\int \frac{dh_i(\cdot,\epsilon)}{d\epsilon}dF_i(\epsilon)+\int h_i(\cdot,\epsilon)d\Delta_{iT} \tag{A.1}$$

$$0=\int \frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}dF_i(\epsilon)+2\int \frac{dh_i(\cdot,\epsilon)}{d\epsilon}d\Delta_{iT} \tag{A.2}$$

$$0=\int \frac{d^3h_i(\cdot,\epsilon)}{d\epsilon^3}dF_i(\epsilon)+3\int \frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}d\Delta_{iT} \tag{A.3}$$

where $\Delta_{iT} \equiv \sqrt{T}\left(\hat{F}_i - F_i\right)$.

(A.1) $\alpha_i^\epsilon(0)$ Evaluating (A.1)

$$0 = \left(\int v^\alpha\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)\right)\alpha_i^\epsilon(\epsilon) + \int v\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]d\Delta_{iT}$$

at $\epsilon = 0$, and noting that $E\left[v_{i,t}\right] = 0$, we obtain

$$0 = \left(\int v^\alpha\left[\cdot;\theta_0,\alpha_i(0)\right]dF_i\right)\alpha_i^\epsilon(0) + \int v\left[\cdot;\theta_0,\alpha_i(0)\right]d\Delta_{iT}$$

so $\alpha_i^\epsilon(0) = -\left(E\left[v_{i,t}^\alpha\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T v_{i,t}\right)$.

(A.2) $\alpha_i^{\epsilon\epsilon}(0)$ Evaluating (A.2)

$$\begin{aligned}0 =& \left(\int v^{\alpha\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)\right)\left(\alpha_i^\epsilon(\epsilon)\right)^2 + \left(\int v^\alpha\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)\right)\alpha_i^{\epsilon\epsilon}(\epsilon)\\ &+ 2\left(\int v^\alpha\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]d\Delta_{iT}\right)\alpha_i^\epsilon(\epsilon)\end{aligned}$$

at $\epsilon = 0$, we obtain

$$0 = E\left[v_{i,t}^{\alpha\alpha}\right]\left(\alpha_i^\epsilon(0)\right)^2 + E\left[v_{i,t}^\alpha\right]\alpha_i^{\epsilon\epsilon}(0) + 2\left(\int v^\alpha\left[\cdot;\theta_0,\alpha_i(0)\right]d\Delta_{iT}\right)\alpha_i^\epsilon(0)$$

so that

$$\begin{aligned}\alpha_i^{\epsilon\epsilon}(0) =& -\left(E\left[v_{i,t}^\alpha\right]\right)^{-1}E\left[v_{i,t}^{\alpha\alpha}\right]\left(\alpha_i^\epsilon(0)\right)^2 - 2\left(E\left[v_{i,t}^\alpha\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(v_{i,t}^\alpha - E\left[v_{i,t}^\alpha\right]\right)\right)\alpha_i^\epsilon(0)\\ =& -\left(E\left[v_{i,t}^\alpha\right]\right)^{-1}E\left[v_{i,t}^{\alpha\alpha}\right]\left(E\left[v_{i,t}^\alpha\right]\right)^{-2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T v_{i,t}\right)^2\\ &+ 2\left(E\left[v_{i,t}^\alpha\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(v_{i,t}^\alpha - E\left[v_{i,t}^\alpha\right]\right)\right)\left(E\left[v_{i,t}^\alpha\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T v_{i,t}\right)\\ =& \left(E\left[v_{i,t}^\alpha\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T v_{i,t}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^T\left[2 * \frac{v_{i,t}^\alpha - E\left[v_{i,t}^\alpha\right]}{E\left[v_{i,t}^\alpha\right]} - \frac{E\left[v_{i,t}^{\alpha\alpha}\right]v_{i,t}}{\left(E\left[v_{i,t}^\alpha\right]\right)^2}\right]\\ =& 2\left(E\left[v_{i,t}^\alpha\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T w_{i,t}\right)\end{aligned}$$

where $w_{i,t} \equiv \frac{v_{i,t}^\alpha - E\left[v_{i,t}^\alpha\right]}{E\left[v_{i,t}^\alpha\right]} - \frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^\alpha\right]\right)^2}v_{i,t}$.

(A.3) $\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)$ Evaluating (A.3)

$$0 = \left(\int v^{\alpha\alpha\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)\right)\left(\alpha_i^\epsilon(\epsilon)\right)^3 + \left(\int v^\alpha\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)\right)\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)$$

$$+3\left(\int v^{\alpha\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]d\Delta_{iT}\right)\left(\alpha_i^{\epsilon}(\epsilon)\right)^2+2\left(\int v^{\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]d\Delta_{iT}\right)\alpha_i^{\epsilon\epsilon}(\epsilon)$$

we see that

$$P\left[\max_i\left|\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)\right|>C\right]=o\left(\frac{1}{T}\right)$$

so that

$$P\left[\frac{1}{n}\sum_i\left|\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)\right|>C\right]\leq P\left[\max_i\left|\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)\right|>C\right]=o\left(\frac{1}{T}\right)$$

We therefore have:

$$\begin{aligned}\frac{1}{nT}\sum_i\sqrt{T}\left[\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right]=&\frac{1}{nT}\sum_i\left[\alpha_i^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]\\
=&\frac{1}{nT}\sum_i\left\{-\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_{i,t}\right)\right\}\\
&+\frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\right)\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_{i,t}\right)\\
&+\frac{1}{nT}\sum_i\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]\\
=&-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+O_p\left(\frac{1}{T\sqrt{T}}\right)\end{aligned}$$

and that

$$\begin{aligned}\frac{1}{nT}\sum_i\left[\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right]^2=&\frac{1}{nT}\sum_i\left[\alpha_i^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^2\\
=&\frac{1}{nT}\sum_i[\alpha_i^{\epsilon}(0)]^2+\frac{1}{4}\left(\frac{1}{T}\right)\frac{1}{nT}\sum_i[\alpha_i^{\epsilon\epsilon}(0)]^2+\frac{1}{36}\left(\frac{1}{T}\right)^2\frac{1}{nT}\sum_i[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2\\
&+\frac{1}{nT\sqrt{T}}\sum_i\alpha_i^{\epsilon}(0)\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\frac{1}{nT}\sum_i\alpha_i^{\epsilon}(0)\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\\
&+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\frac{1}{nT}\sum_i\alpha_i^{\epsilon\epsilon}(0)\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\\
=&\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}+O_p\left(\frac{1}{T\sqrt{T}}\right)\end{aligned}$$

also,

$$\frac{1}{nT}\sum_i\left[\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right]^3=\frac{1}{nT}\sum_i\left[\alpha_i^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{T}\right)\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^3$$

$$
\begin{aligned}
=&\frac{1}{nT}\sum_i[\alpha_i^{\epsilon}(0)]^3+\frac{1}{nT^2\sqrt{T}}\sum_i[\frac{1}{2}\alpha_i^{\epsilon\epsilon}(0)]^3+\frac{1}{nT^4}\sum_i\left[\frac{1}{6}\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^3\\
&+3\frac{1}{nT\sqrt{T}}\sum_i[\alpha_i^{\epsilon}(0)]^2\frac{1}{2}\alpha_i^{\epsilon\epsilon}(0)+3\frac{1}{nT^2}\sum_i\alpha_i^{\epsilon}(0)\frac{1}{4}\alpha_i^{\epsilon\epsilon}(0)\\
&+3\frac{1}{nT^2}\sum_i[\alpha_i^{\epsilon}(0)]^2\frac{1}{6}\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})+3\frac{1}{nT^3}\sum_i\alpha_i^{\epsilon}(0)\frac{1}{36}[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2\\
&+3\frac{1}{nT^3\sqrt{T}}\sum_i\frac{1}{2}\alpha_i^{\epsilon\epsilon}(0)\frac{1}{36}[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2+3\frac{1}{nT^3}\sum_i\frac{1}{4}[\alpha_i^{\epsilon\epsilon}(0)]^2\frac{1}{6}\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\\
&+6\frac{1}{nT^2\sqrt{T}}\sum_i\alpha_i^{\epsilon}(0)\frac{1}{2}\alpha_i^{\epsilon\epsilon}(0)\frac{1}{6}\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\\
=&-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}
$$

■

Lemma 10 *(Arellano and Hahn (2016),Theorem 5):*

$$
\frac{\partial\left[\frac{1}{n}\sum_i\widehat{R}_{fi}^{*}(\widetilde{\theta})\right]}{\partial\theta}=\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i}\tilde{V}_{it}\right]+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i\alpha_i}\right]E\left[\left(\tilde{V}_{it}\right)^2\right]+o_p(1)
$$

where $\frac{1}{T}\sum_{t=1}^{T}\tilde{V}_{it}=-\left(E\left[\frac{\partial v_i}{\partial\alpha_i}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)$.

Lemma 11 *(Rao and Mitra (1971), Theorem 9.2.1) Let* $Y\sim N_p(\mu,\Sigma)$*, where* Σ *may be singular, then* $Y'AY\xrightarrow{d}\chi^2(k)$ *iff*

$$
\Sigma A\Sigma A\Sigma=\Sigma A\Sigma \tag{68}
$$

where $k=tr(A\Sigma)$.

Lemma 12 *Under condition 1 in the appendix,*

$$
\widetilde{\widetilde{A}}_{f,n}=\frac{\partial^2(\frac{1}{nT}[\sum_i\sum_t\log f(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))-\sum_i R_{fi}(\widetilde{\widetilde{\theta}})])}{\partial\theta\partial\theta'}=-\frac{1}{n}\sum_i\mathcal{I}_i
$$

Proof. To prove $\widetilde{\widetilde{A}}_{f,n}=-\frac{1}{n}\sum_i\mathcal{I}_i$, it suffices to prove $-\frac{\partial^2(\frac{1}{nT}[\sum_i\sum_t\log f(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))])}{\partial\theta\partial\theta'}=\frac{1}{n}\sum_i\mathcal{I}_i$ and $\frac{\partial^2(\frac{1}{nT}[\sum_i R_i(\widetilde{\widetilde{\theta}})])}{\partial\theta\partial\theta'}=o_p(1)$. Since

$$
\frac{\partial^2(\frac{1}{nT}[\sum_i\sum_t\log f(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))])}{\partial\theta\partial\theta'}=\frac{1}{nT}\sum_i\sum_t\frac{\partial u(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))}{\partial\theta}
$$

$$
\begin{aligned}
=&\frac{1}{nT}\sum_i\sum_t\left[u^\theta(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))-v^\theta(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))\left(\frac{\partial\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})}{\partial\theta}\right)\right]\\
=&\frac{1}{nT}\sum_i\sum_t\left[u^\theta(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))-v^\theta(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))\left(\frac{\frac{1}{T}\sum_t v_{it}^\theta}{\frac{1}{T}\sum_t v_{it}^{\alpha_i}}\right)\right]\\
=&\frac{1}{nT}\sum_i\sum_t U_{it}^\theta\left(\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})\right)+o_p(1)
\end{aligned}
$$

the first and second equalities hold by the definition of $u\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)$, the third equality follows from Lemma 8, the fourth equality hold by the definition of $U(y_{it};\theta,\alpha(\theta))$, and we have

$$
\left|\frac{1}{nT}\sum_i\sum_t u_{it}^\theta\left(\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})\right)-\frac{1}{nT}\sum_i\sum_t u_{it}^\theta\left(\theta_0,\alpha_i\left(\theta_0\right)\right)\right| \tag{69}
$$

$$
\leq\left(\max_i\frac{1}{T}\sum_t M\left(y_{i,t}\right)\right)\times\left(|\widetilde{\widetilde{\theta}}-\theta_0|+\max_i\left|\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})-\alpha_i\left(\theta_0\right)\right|\right) \tag{70}
$$

and that

$$
\left|\frac{1}{nT}\sum_i\sum_t u_{it}^\alpha\left(\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})\right)-\frac{1}{nT}\sum_i\sum_t u_{it}^\alpha\left(\theta_0,\alpha_i\left(\theta_0\right)\right)\right| \tag{71}
$$

$$
\leq\left(\max_i\frac{1}{T}\sum_t M\left(y_{i,t}\right)\right)\times\left(|\widetilde{\widetilde{\theta}}-\theta_0|+\max_i\left|\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})-\alpha_i\left(\theta_0\right)\right|\right) \tag{72}
$$

follows from 4 in Condition 1, we have $\max_i\frac{1}{T}\sum_t M\left(y_{i,t}\right)=O_p(1)$. Because of Lemma 3,

$$
P\left(|\widetilde{\widetilde{\theta}}-\theta_0|\geq\eta\right)=o\left(T^{-1}\right)
$$

for $\widetilde{\widetilde{\theta}}$ lies in between $\widetilde{\theta}$ and θ_0. $P\left[\max_{1\leq i\leq n}|\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})-\alpha_i\left(\theta_0\right)|\geq\eta\right]=o\left(T^{-1}\right)$ is proved in Lemma 9. Therefore we conclude that

$$
\begin{aligned}
\frac{1}{nT}\sum_i\sum_t U_{it}^\theta\left(\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})\right)=&\frac{1}{nT}\sum_i\sum_t U_{it}^\theta\left(\theta_0,\alpha_i\left(\theta_0\right)\right)+o_p(1) && (73)\\
=&\frac{1}{n}\sum_i E\left[U_{it}^\theta\right]+o_p(1)=-\frac{1}{n}\sum_i\mathcal{I}_i+o_p(1) && (74)
\end{aligned}
$$

The second equality follows from Lemma 6. As it is shown in Lemma 10, $\frac{1}{n}\sum_i\frac{\partial R_{fi}(\theta_0)}{\partial\theta}=S_{fn}\left(\theta_0\right)$. According to Condition 1, we have

$$
\frac{\partial^2\frac{1}{nT}\sum_i R_i(\widetilde{\widetilde{\theta}})}{\partial\theta\partial\theta'}=o_p(1)
$$

given $(n/T) \to \rho$. $\blacksquare$

Lemma 13 *Suppose that*

$$K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,\epsilon)\right) = \frac{\partial^{m_1+m_2}\psi\left(y_{it};\theta_0,\alpha_i(\theta_0,\epsilon)\right)}{\partial\alpha_i^m} \tag{75}$$

for some $m \leq 1,...,5$. Then for any $\eta > 0$, we have

$$P\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\frac{1}{n}\sum_{i=1}^{n}\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,\epsilon)\right)dF_i(\epsilon) - \frac{1}{n}\sum_{i=1}^{n}E\left[K_i\left(y_{it};\theta_0,\alpha_{i0}\right)\right]\right| > \eta\right] = o\left(T^{-1}\right) \tag{76}$$

and

$$P\left[\max_i\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,\epsilon)\right)dF_i(\epsilon) - E\left[K_i\left(y_{it};\theta_0,\alpha_{i0}\right)\right]\right| > \eta\right] = o\left(T^{-1}\right) \tag{77}$$

Also,

$$P\left[\max_i\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,\epsilon)\right)d\Delta_{iT}\right| > CT^{\frac{1}{10}-v}\right] = o\left(T^{-1}\right) \tag{78}$$

for some constant $C > 0$ and $0 < v < (100q+120)^{-1}$.

Proof. Note that

$$\left\|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon) - \int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i\right\| \tag{79}$$

$$\leq \left\|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon) - \int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i(\epsilon)\right\| \tag{80}$$

$$+\left\|\int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i(\epsilon) - \int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i\right\| \tag{81}$$

$$\leq \int M\left(y_{it}\right)\left(\left|\alpha_i(\theta_0,F_i(\epsilon)) - \alpha_{i0}\right|\right)d\left|F_i(\epsilon)\right| \tag{82}$$

$$+\,\epsilon\sqrt{T}\left\|\int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)d\left(\hat{F}_i - F_i\right)\right\| \tag{83}$$

Therefore, we have

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon) - \int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i\right\| \tag{84}$$

$$\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_i(\theta_0,F_i(\epsilon)) - \alpha_{i0}\right)^2\right)^{\frac{1}{2}}\left(\frac{1}{n}\sum_{i=1}^{n}\left(E\left[M\left(y_{it}\right)\right] + \frac{1}{T}\sum_{t=1}^{T}M\left(y_{it}\right)\right)^2\right)^{\frac{1}{2}} \tag{85}$$

$$+\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}K_i\left(y_{it};\theta_0,\alpha_{i0}\right)-E\left[K_i\left(y_{it};\theta_0,\alpha_{i0}\right)\right]\right)\right\| \tag{86}$$

the RHS of which can be bounded by using Lemmas 5 and 7 in absolute value by some $\eta > 0$ with probability $1-o\left(T^{-1}\right)$. Because

$$\left|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon)-E\left[K_i\left(y_{it};\theta_0,\alpha_{i0}\right)\right]\right| \tag{87}$$

$$\leq\left|\alpha_i(\theta_0,F_i(\epsilon))-\alpha_i\right|\cdot\left(E\left[M\left(y_{it}\right)\right]+\frac{1}{T}\sum_{t=1}^{T}M\left(y_{it}\right)\right) \tag{88}$$

$$+\left|\frac{1}{T}\sum_{t=1}^{T}M\left(y_{it}\right)-E\left[M\left(y_{it}\right)\right]\right| \tag{89}$$

we can bound

$$\max_i\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon)-E\left[K_i\left(y_{it};\theta_0,\alpha_{i0}\right)\right]\right| \tag{90}$$

in absolute value by some $\eta > 0$ with probability $1-o\left(T^{-1}\right)$. Using 4 in Condition 1 and Lemma 5, we also deduce that $\max_i\left|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)d\Delta_{iT}\right|$ can be bounded by in absolute value by $CT^{\frac{1}{10}-v}$ for some constant $C > 0$ and $0 < v < \frac{1}{160}$ with probability $1-o\left(T^{-1}\right)$. ■

Lemma 14 *Arellano and Hahn (2016)(Lemma 14)*

$$P\left[\max_i\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\alpha_i^{\epsilon}\left(\epsilon\right)\right|>CT^{\frac{1}{10}-v}\right]=o\left(T^{-1}\right) \tag{91}$$

$$P\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\mu^{\epsilon}(\epsilon)\right|>CT^{\frac{1}{10}-v}\right]=o\left(T^{-1}\right) \tag{92}$$

$$P\left[\max_i\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\alpha_i^{\epsilon\epsilon}\left(\epsilon\right)\right|>C\left(T^{\frac{1}{10}-v}\right)^2\right]=o\left(T^{-1}\right) \tag{93}$$

$$P\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\mu^{\epsilon\epsilon}(\epsilon)\right|>C\left(T^{\frac{1}{10}-v}\right)^2\right]=o\left(T^{-1}\right) \tag{94}$$

$$P\left[\max_i\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\alpha_i^{\epsilon\epsilon\epsilon}\left(\epsilon\right)\right|>C\left(T^{\frac{1}{10}-v}\right)^3\right]=o\left(T^{-1}\right) \tag{95}$$

$$P\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\mu^{\epsilon\epsilon\epsilon}(\epsilon)\right|>C\left(T^{\frac{1}{10}-v}\right)^3\right]=o\left(T^{-1}\right) \tag{96}$$

for some constant $C > 0$ and $0 < v < (100q+120)^{-1}$.

Lemma 15

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(y_{it};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}+\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\alpha_i}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)$$
$$+\sqrt{\frac{n}{T}}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i\alpha_i}\right]\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)^2+o_p\left(1\right)$$

where $\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)=-\left(E\left[\frac{\partial v_i}{\partial \alpha_i}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)+o_p\left(\frac{1}{\sqrt{T}}\right)=\frac{1}{T}\sum_{t=1}^{T}\tilde{V}_{it}+o_p\left(\frac{1}{\sqrt{T}}\right)$.

Proof. Let $F\equiv\left(F_1,\ldots,F_n\right)$ denote the collection of marginal distribution functions of y_{it}.Let $\hat{F}_i$ denote the empirical distribution function for the observation i. Define $F_i(\epsilon)\equiv F_i+\epsilon\sqrt{T}\left(\hat{F}_i-F_i\right)$ for $\epsilon\in\left[0,T^{-1/2}\right]$. For each fixed θ and ϵ, let $\alpha_i(\epsilon)$ be the solution to the estimating equation

$$0=\int v_i\left[\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right]dF_i(\epsilon)$$

and let $\mu\left(F\left(\epsilon\right)\right)$ be the solution to the estimating equation

$$0=\sum_{i=1}^{n}\int\left[U_i\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)-\mu\left(F\left(\epsilon\right)\right)\right]dF_i(\epsilon)\tag{97}$$

Note that $\mu\left(F\left(0\right)\right)=0$, and

$$\mu\left(\widehat{F}\right)\equiv\mu\left(F\left(\frac{1}{\sqrt{T}}\right)\right)=\frac{1}{n}\sum_{i=1}^{n}U_i\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i\left(\frac{1}{\sqrt{T}}\right)\right)\right)\tag{98}$$
$$=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(y_{it};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)\tag{99}$$

By a Taylor series expansion, we have

$$\mu\left(\widehat{F}\right)-\mu\left(F\right)=\frac{1}{\sqrt{T}}\mu^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\mu^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\mu^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\tag{100}$$

where $\mu^{\epsilon}(\epsilon)\equiv d\mu(F(\epsilon))/d\epsilon,\mu^{\epsilon\epsilon}(\epsilon)\equiv d^2\mu(F(\epsilon))/d\epsilon^2,\ldots$, and $\widetilde{\epsilon}$ is somewhere in between 0 and $\frac{1}{\sqrt{T}}$.

(C.1) $\mu^{\epsilon}(0)$ Let

$$h_i^{\mu}(\cdot,\epsilon)\equiv U_i\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)-\mu(F(\epsilon))\tag{101}$$

The first order condition could be written as

$$0=\frac{1}{n}\sum_{i=1}^{n}\int h_i^{\mu}(\cdot,\epsilon)dF_i(\epsilon)$$

Differentiating repeatedly with respect to ϵ, we obtain

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i^{\mu}(\cdot,\epsilon)}{d\epsilon}dF_i(\epsilon) + \frac{1}{n}\sum_{i=1}^{n}\int h_i^{\mu}(\cdot,\epsilon)d\Delta_{iT} \tag{C.1}$$

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2h_i^{\mu}(\cdot,\epsilon)}{d\epsilon^2}dF_i(\epsilon) + 2\frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i^{\mu}(\cdot,\epsilon)}{d\epsilon}d\Delta_{iT} \tag{C.2}$$

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^3h_i^{\mu}(\cdot,\epsilon)}{d\epsilon^3}dF_i(\epsilon) + 3\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2h_i^{\mu}(\cdot,\epsilon)}{d\epsilon^2}d\Delta_{iT} \tag{C.3}$$

where $\Delta_{iT} \equiv \sqrt{T}\left(\hat{F}_i - F_i\right)$. Evaluating (C.1)

$$0 =\frac{1}{n}\sum_{i=1}^{n}\int \left[U_i^{\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)\alpha_i^{\epsilon}\left(\theta_0,F_i(\epsilon)\right) - \mu^{\epsilon}(F(\epsilon))\right]dF_i(\epsilon) \tag{102}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\int \left[U_i\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right) - \mu(F(\epsilon))\right]d\Delta_{iT} \tag{103}$$

at $\epsilon = 0$, and noting that $E\left[U_i^{\alpha_i}\right] = 0$, combining Lemma 9, we obtain

$$\mu^{\epsilon}(0) = \frac{1}{n}\sum_{i=1}^{n}\int U_i d\Delta_{iT} \tag{104}$$

(C.2) $\mu^{\epsilon\epsilon}(0)$ Evaluating (C.2)

$$\begin{aligned}
0 =& -\frac{1}{n}\sum_{i=1}^{n}\int \mu^{\epsilon\epsilon}(F(\epsilon))dF_i(\epsilon)\\
&+ \frac{1}{n}\sum_{i=1}^{n}\int U_i^{\alpha_i\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)\left(\alpha_i^{\epsilon}(\theta_0,F_i(\epsilon))\right)^2 dF_i(\epsilon)\\
&+ \frac{1}{n}\sum_{i=1}^{n}\int U_i^{\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)\alpha_i^{\epsilon\epsilon}(\theta_0,F_i(\epsilon))dF_i(\epsilon)\\
&+ \frac{2}{n}\sum_{i=1}^{n}\int \left[U_i^{\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)\alpha_i^{\epsilon}(\theta_0,F_i(\epsilon)) - \mu^{\epsilon}(F(\epsilon))\right]d\Delta_{iT}
\end{aligned}$$

at $\epsilon = 0$, and noting that $E\left[U_i^{\alpha_i}\right] = 0$, combining Lemma 9, we obtain

$$\begin{aligned}
\mu^{\epsilon\epsilon}(0) =& \frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\alpha_i\alpha_i}\right]\left(\alpha_i^{\epsilon}\right)^2 + \frac{2}{n}\sum_{i=1}^{n}(\int U_i^{\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0\right)\right)d\Delta_{iT})\alpha_i^{\epsilon}(\theta_0,F_i(0))\\
=& \frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\alpha_i\alpha_i}\right]\left[\left(E\left[\frac{\partial v_i}{\partial \alpha_i}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_{it}\right)\right]^2\\
&- \frac{2}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\alpha_i}\right)\left(E\left[\frac{\partial v_i}{\partial \alpha_i}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_{it}\right)
\end{aligned}$$

(C.3) We can ignore $\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3 \mu^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})$ according to Lemma 14.

■

Lemma 16 $\frac{1}{nT}\sum_i \left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right) = -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} + O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof. From Lemma 9, we deduce that

$$\begin{aligned}
\frac{1}{nT}\sum_i \left(\sum_t v_{i,t}\right)\left[\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right] =& \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\sqrt{T}\left[\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right] \\
=& \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}\right. \\
&+ \frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]} + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})] \\
=& \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}\right] \qquad \text{(B1)} \\
&+ \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}\right] \\
& \qquad \text{(B2)} \\
&+ \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] \qquad \text{(B3)}
\end{aligned}$$

It can be shown that

$$\begin{aligned}
\text{(B1)} &= -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} \\
\text{(B2)} &= O_p\left(\frac{1}{T\sqrt{T}}\right) \\
\text{(B3)} &= o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}$$

Since $\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right) = O_p(1)$, and $\frac{1}{n}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T}\right)$, $\frac{1}{nT}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T^2}\right)$,

$$\frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] = O_p(1) * \frac{1}{T^2}\frac{1}{n}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T\sqrt{T}}\right)$$

Therefore we have

$$\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)=-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+O_p\left(\frac{1}{T\sqrt{T}}\right)$$

■

Lemma 17 $\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)=-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof. From Lemma 9, we deduce that

$$\begin{aligned}
\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left[\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right]=&\frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\sqrt{T}\left[\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right]\\
=&\frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}\right.\\
&\left.+\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]\\
=&\frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}\right] && \text{(B1)}\\
&+\frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}\right]\\
& && \text{(B2)}\\
&+\frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] && \text{(B3)}
\end{aligned}$$

It can be shown that

$$\begin{aligned}
\text{(B1)}&=-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}\\
\text{(B2)}&=O_p\left(\frac{1}{T\sqrt{T}}\right)\\
\text{(B3)}&=o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}$$

Since $\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)=O_p(1)$, and $\frac{1}{n}\sum_i\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})=o_p\left(\frac{1}{T}\right)$, $\frac{1}{nT}\sum_i\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})=o_p\left(\frac{1}{T^2}\right)$,

$$\frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]=O_p(1)*\frac{1}{T^2}\frac{1}{n}\sum_i\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})=o_p\left(\frac{1}{T\sqrt{T}}\right)$$

Therefore we have

$$\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)=-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v^{\alpha}_{i,t}\right]}+O_p\left(\frac{1}{T\sqrt{T}}\right)$$

■

Lemma 18 $\frac{1}{2}\frac{1}{nT}\sum_i\left(\sum_t v^{\alpha}_{i,t}\right)\left(\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2=\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v^{\alpha}_{i,t}\right]}+O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof.

$$\frac{1}{T}\sum_t v^{\alpha}_{i,t}=E\left[v^{\alpha}_{i,t}\right]+\frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_t\left(v^{\alpha}_{i,t}-E\left[v^{\alpha}_{i,t}\right]\right)$$

where

$$\frac{1}{\sqrt{T}}\sum_t\left(v^{\alpha}_{i,t}-E\left[v^{\alpha}_{i,t}\right]\right)=O_p(1)$$

according to condition 1. Then we have

$$\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v^{\alpha}_{i,t}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{105}$$

$$=\frac{1}{nT}\sum_i\left(E\left[v^{\alpha}_{i,t}\right]+\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_t\left(v^{\alpha}_{i,t}-E\left[v^{\alpha}_{i,t}\right]\right)\right)\right) \tag{106}$$

$$*\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v^{\alpha}_{i,t}\right]}+\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v^{\alpha}_{i,t}\right]}+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^2 \tag{107}$$

Writing

$$\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v^{\alpha}_{i,t}\right]}+\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\frac{1}{\sqrt{T}}\sum_t w_{i,t})}{E\left[v^{\alpha}_{i,t}\right]}+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^2 \tag{108}$$

$$=\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{\left(E\left[v^{\alpha}_{i,t}\right]\right)^2}-\frac{2}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{\left(E\left[v^{\alpha}_{i,t}\right]\right)^2}+\frac{1}{T}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)^2}{\left(E\left[v^{\alpha}_{i,t}\right]\right)^2} \tag{109}$$

$$+\frac{1}{36}\left(\frac{1}{T}\right)^2[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2-\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v^{\alpha}_{i,t}\right]} \tag{110}$$

$$+\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v^{\alpha}_{i,t}\right]} \tag{111}$$

We obtain

$$\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v^{\alpha}_{i,t}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{112}$$

$$=\frac{1}{nT}\sum_i \left[E\left[v_{i,t}^{\alpha}\right]+O_p(\frac{1}{\sqrt{T}})\right] \tag{113}$$

$$*\left[\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}-\frac{2}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}+\frac{1}{T}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}\right. \tag{114}$$

$$+\frac{1}{36}\left(\frac{1}{T}\right)^2[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2-\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]} \tag{115}$$

$$\left.+\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}\right] \tag{116}$$

According to Lemma 9, we have

$$\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{117}$$

$$=\frac{1}{nT}\sum_i\left[\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}-\frac{2}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{T}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}\right. \tag{118}$$

$$+\frac{1}{36}\left(\frac{1}{T}\right)^2[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2E\left[v_{i,t}^{\alpha}\right]-\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\left[\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right] \tag{119}$$

$$+\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right) \tag{120}$$

$$\left.+\frac{1}{nT}\sum_i O_p\left(\frac{1}{\sqrt{T}}\right)\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}+o_p\left(\frac{1}{T\sqrt{T}}\right)\right] \tag{121}$$

Therefore,

$$\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{122}$$

$$=\frac{1}{nT}\sum_i\left[\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)-2E\left[v_{i,t}^{\alpha}\right]w_{i,t}\right)\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}\right] \tag{123}$$

$$+o_p\left(\frac{1}{T\sqrt{T}}\right) \tag{124}$$

Because

$$\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)-2E\left[v_{i,t}^{\alpha}\right]w_{i,t}=\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)-2E\left[v_{i,t}^{\alpha}\right]\left(\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^{2}}v_{i,t}\right)$$

$$=\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)-2\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}$$

$$=-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}$$

We further write

$$\frac{1}{nT}\sum_{i}\left(\frac{1}{T}\sum_{t}v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_{i}\left(\theta_{0}\right)-\alpha_{i}\left(\theta_{0}\right)\right)\right)^{2} \tag{125}$$

$$=\frac{1}{nT}\sum_{i}\left[\frac{\left(\frac{1}{\sqrt{T}}\sum_{t}v_{i,t}\right)^{2}}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_{t}v_{i,t}\right)^{2}\left(\frac{1}{\sqrt{T}}\sum_{t}\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^{2}}\right] \tag{126}$$

$$+o_{p}\left(\frac{1}{T\sqrt{T}}\right) \tag{127}$$

It follows that

$$\frac{1}{2}\frac{1}{nT}\sum_{i}\left(\sum_{t}v_{i,t}^{\alpha}\right)\left(\left(\widehat{\alpha}_{i}\left(\theta_{0}\right)-\alpha_{i}\left(\theta_{0}\right)\right)\right)^{2}=\frac{1}{2}\frac{1}{nT}\sum_{i}\left(\frac{1}{T}\sum_{t}v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_{i}\left(\theta_{0}\right)-\alpha_{i}\left(\theta_{0}\right)\right)\right)^{2} \tag{128}$$

$$=\frac{1}{2}\frac{1}{nT}\sum_{i}\frac{\left(\frac{1}{\sqrt{T}}\sum_{t}v_{i,t}\right)^{2}}{E\left[v_{i,t}^{\alpha}\right]}+O_{p}\left(\frac{1}{T\sqrt{T}}\right) \tag{129}$$

■

Lemma 19 $\frac{1}{6}\frac{1}{nT}\sum_{i}\left(\sum_{t}v_{i,t}^{\alpha\alpha}\left(y_{i,t};\theta_{0},\widehat{\alpha}_{i}\left(\theta_{0}\right)\right)\right)\left(\widehat{\alpha}_{i}\left(\theta_{0}\right)-\alpha_{i}\left(\theta_{0}\right)\right)^{3}=-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_{i}E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_{t}v_{i,t}\right)^{3}}{(E\left[v_{i,t}^{\alpha}\right])^{3}}+o_{p}\left(\frac{1}{T\sqrt{T}}\right)$

Proof.

$$\frac{1}{T}\sum_{t}v_{i,t}^{\alpha\alpha}(y_{i,t};\theta_{0},\widehat{\widehat{\alpha}}_{i}\left(\theta_{0}\right))=E\left[v_{i,t}^{\alpha\alpha}\right]+O_{p}\left(\frac{1}{\sqrt{T}}\right)$$

Follows from Lemma 9, we write

$$\frac{1}{6}\frac{1}{nT}\sum_{i}\left(\sum_{t}v_{i,t}^{\alpha\alpha}\left(y_{i,t};\theta_{0},\widehat{\alpha}_{i}\left(\theta_{0}\right)\right)\right)\left(\widehat{\alpha}_{i}\left(\theta_{0}\right)-\alpha_{i}\left(\theta_{0}\right)\right)^{3} \tag{130}$$

$$=\frac{1}{6nT\sqrt{T}}\sum_i(\frac{1}{T}\sum v_{i,t}^{\alpha\alpha}(y_{i,t};\theta_0,\widehat{\widehat{\alpha}}_i\left(\theta_0\right)))\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^3 \tag{131}$$

$$=\frac{1}{6nT\sqrt{T}}\sum_i\left(E\left[v_{i,t}^{\alpha\alpha}\right]+O_p\left(\frac{1}{\sqrt{T}}\right)\right)\times\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}+O_p\left(\frac{1}{\sqrt{T}}\right)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^3 \tag{132}$$

$$=-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{(E\left[v_{i,t}^{\alpha}\right])^3}+o_p\left(\frac{1}{T\sqrt{T}}\right) \tag{133}$$

■

Lemma 20 *(Hahn and Newey (2004), Lemma 22, Lemma 24, Lemma 25)*

$$E\left[\left(\frac{1}{T}\sum_{t\neq t'}^{T}v_{fi,t}v_{fi,t'}\right)\right]^2=O_p\left(1\right) \tag{134}$$

$$E\left[\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'}^{T}v_{fi,t}^2v_{fi,t'}\right)\right]^2=O_p\left(1\right) \tag{135}$$

$$E\left[\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'\neq t''\neq t}^{T}v_{fi,t}v_{fi,t'}v_{fi,t''}\right)\right]^2=O_p\left(1\right) \tag{136}$$

7.3 Propositions

Proposition 1 $\sum_i\sum_t[\psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)]=O_p(1)$

Proof. Because $\widehat{\alpha}_i(\theta)=\arg\max_a\sum_t\log f\left(y_{i,t}\mid\theta,a\right)$, we have for all θ:

$$\frac{\partial\sum_t\psi_f\left(y_{i,t};\theta,\widehat{\alpha}_i(\theta)\right)}{\partial\alpha}=0$$

and the envelope theorem implies that $\frac{\partial\sum_t\psi(y_{i,t};\theta,\widehat{\alpha}_i(\theta))}{\partial\theta}=\frac{\partial\sum_t\log f(y_{i,t};\theta,\widehat{\alpha}_i(\theta))}{\partial\theta}=\sum_t u\left(y_{i,t};\theta,\widehat{\alpha}_i(\theta)\right)$, therefore we should have

$$\sum_i\sum_t[\psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)]=\sum_i\sum_t u\left(y_{i,t};\tilde{\tilde{\theta}},\widehat{\alpha}_i(\tilde{\tilde{\theta}})\right)\left(\tilde{\theta}-\theta_0\right)$$

for some $\tilde{\tilde{\theta}}$ in between θ_0 and $\widetilde{\theta}$, because $\widehat{\theta},\widehat{\alpha}_1(\theta),\ldots,\widehat{\alpha}_n(\theta)$ maximizes the likelihood, we should have

$$\sum_i\sum_t u\left(y_{i,t};\widehat{\theta},\widehat{\alpha}_i(\widehat{\theta})\right)=\frac{\partial\sum_i\sum_t\log f\left(y_{i,t},\widehat{\theta},\widehat{\alpha}_i(\widehat{\theta})\right)}{\partial\theta}=0$$

and hence

$$\sum_i \sum_t u\left(y_{i,t}; \tilde{\tilde{\theta}}, \widehat{\alpha}_i(\tilde{\tilde{\theta}})\right) = \sum_i \sum_t u\left(y_{i,t}; \widehat{\theta}, \widehat{\alpha}_i(\widehat{\theta})\right) + \left(\sum_i \sum_t \frac{\partial u\left(y_{i,t}, \tilde{\tilde{\tilde{\theta}}}, \widehat{\alpha}_i(\tilde{\tilde{\tilde{\theta}}})\right)}{\partial \theta}\right)(\tilde{\tilde{\theta}} - \widehat{\theta})$$

$$= \left(\sum_i \sum_t \frac{\partial u\left(y_{i,t}; \tilde{\tilde{\tilde{\theta}}}, \widehat{\alpha}_i(\tilde{\tilde{\tilde{\theta}}})\right)}{\partial \theta}\right)(\tilde{\tilde{\theta}} - \widehat{\theta})$$

for some $\tilde{\tilde{\tilde{\theta}}}$ in between $\tilde{\tilde{\theta}}$ and $\widehat{\theta}$. Under Condition 1, we have:

$$\left(\sum_i \sum_t \frac{\partial u\left(y_{i,t}; \tilde{\tilde{\tilde{\theta}}}, \widehat{\alpha}_i(\tilde{\tilde{\tilde{\theta}}})\right)}{\partial \theta}\right)(\tilde{\tilde{\theta}} - \widehat{\theta})\left(\tilde{\theta} - \theta_0\right) = O_p(nT) O_p\left(\frac{1}{\sqrt{nT}}\right) O_p\left(\frac{1}{\sqrt{nT}}\right) = O_p(1)$$

■

Proposition 2

$$\frac{\partial\left[\sum_i \sum_t \psi_f(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \sum_i \widehat{R}^*_{fi}(\theta_0)\right]}{\partial \theta}(\tilde{\theta} - \theta_0) = \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T U_{it}\right) \mathcal{I}_f^{-1} \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T U_{it}\right) + o_p(1)$$

Proof. We can The first order derivative can be written as:

$$\begin{aligned}
&\frac{\partial\left[\frac{1}{\sqrt{nT}} \sum_i \sum_t \psi_f(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \frac{1}{\sqrt{nT}} \sum_i \widehat{R}^*_{fi}(\theta_0)\right]}{\partial \theta} \\
=&\frac{1}{\sqrt{nT}} \sum_i \sum_t u(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \sqrt{\frac{n}{T}} S_{fn}(\theta_0) \\
=&\frac{1}{\sqrt{nT}} \sum_i \sum_t \left[u(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - v(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) \frac{E\left[u_{it}^{\alpha_i}\right]}{E\left[v_{it}^{\alpha_i}\right]}\right] - \sqrt{\frac{n}{T}} S_{fn}(\theta_0) \\
=&\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T U(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \sqrt{\frac{n}{T}} S_{fn}(\theta_0)
\end{aligned}$$

The first equality holds by the definition of $u(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0))$ and $S_{fn}(\theta_0)$, the second equality follows from $v(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) = 0$, the third equality hold by the definition of $U(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0))$. According to Lemma 15, we deduce that

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T U\left(y_{it}; \theta_0, \widehat{\alpha}_i(\theta_0)\right) = \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T U_{it} + \sqrt{\frac{n}{T}} \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T U_{it}^{\alpha_i}\right)\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \widetilde{V}_{it}\right)$$

$$+\sqrt{\frac{n}{T}}\frac{1}{2n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i\alpha_i}\right]\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)^2+o_p\left(1\right)$$

where

$$\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)=-\left(E\left[\frac{\partial v_i}{\partial\alpha_i}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)+o_p\left(\frac{1}{\sqrt{T}}\right)=\frac{1}{T}\sum_{t=1}^{T}\tilde{V}_{it}+o_p\left(\frac{1}{\sqrt{T}}\right)$$

according to Lemma 10,

$$\begin{aligned}\frac{\partial\left[\frac{1}{\sqrt{nT}}\sum_i\widehat{R}^*_{fi}(\widetilde{\theta})\right]}{\partial\theta}=&\sqrt{\frac{n}{T}}S_{fn}\left(\theta_0\right)\\=&\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i}\tilde{V}_{it}\right]+\sqrt{\frac{n}{T}}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i\alpha_i}\right]E\left[\left(\tilde{V}_{it}\right)^2\right]+o_p\left(1\right)\end{aligned}$$

we therefore have

$$\frac{\partial\left[\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-\frac{1}{\sqrt{nT}}\sum_i\widehat{R}^*_{fi}(\theta_0)\right]}{\partial\theta}=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}+o_p\left(1\right)$$

according to Lemma 3, we conclude that

$$\frac{\partial\left[\sum_i\sum_t\psi_f(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-\sum_i\widehat{R}^*_{fi}(\theta_0)\right]}{\partial\theta}(\tilde{\theta}-\theta_0)=\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)+o_p\left(1\right)$$

■

Proposition 3 $\frac{1}{nT}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]=\frac{1}{nT}\sum_i R_{fi}\left(\theta_0\right)+O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof. We decompose the left hand side of the above equation into a sum of three terms. For some $\overline{\alpha}_i\left(\theta_0\right)$ on the line segment adjoining $\widehat{\alpha}_i\left(\theta_0\right)$ and $\alpha_i\left(\theta_0\right)$,

$$\begin{aligned}&\frac{1}{nT}\sum_i\sum_t\left(\psi\left(y_{i,t};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)-\psi\left(y_{i,t};\theta_0,\alpha_i\left(\theta_0\right)\right)\right)\\=&\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\\&+\frac{1}{2}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)^2\\&+\frac{1}{6}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha\alpha}\left(y_{i,t};\theta_0,\overline{\alpha}_i\left(\theta_0\right)\right)\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)^3\end{aligned}$$

By Lemma 16, Lemma 18 and Lemma 19, we conclude that

$$
\begin{aligned}
&\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]\\
=&-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+o_p\left(\frac{1}{T\sqrt{T}}\right)\\
+&\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha,a}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right)}{(E\left[v_{i,t}^{\alpha}\right])^2}\\
-&\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}
$$

and that

$$
\begin{aligned}
&\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]\\
=&-\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}\\
&+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left\{\frac{1}{\sqrt{T}}\sum_t\left[w_{i,t}+\frac{1}{2E\left[v_{i,t}^{\alpha}\right]}\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right]\right\}}{E\left[v_{i,t}^{\alpha}\right]}\\
&-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}
$$

We write

$$
\begin{aligned}
&w_{i,t}+\frac{1}{2E\left[v_{i,t}^{\alpha}\right]}\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)-\frac{1}{6}\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}\\
=&\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}-\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{2E\left[v_{i,t}^{\alpha}\right]}+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}-\frac{1}{6}\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}\\
=&\frac{1}{2}\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{6\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}
\end{aligned}
$$

We obtain

$$
\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]
$$

$$
\begin{aligned}
&= -\frac{1}{2}\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} \\
&\quad + \frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(\frac{1}{2}\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{6\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}\right)\right)}{E\left[v_{i,t}^{\alpha}\right]} + o_p\left(\frac{1}{T\sqrt{T}}\right) \\
&= \frac{1}{nT}\sum_i R_{fi}\left(\theta_0\right) + O_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}
$$

■

Proposition 4 $\frac{1}{nT}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)] = \frac{1}{nT}\sum_i R_{fi}\left(\theta_0\right)+O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof. We decompose the left hand side of the above equation into a sum of three terms. For some $\overline{\alpha}_i\left(\theta_0\right)$ on the line segment adjoining $\widehat{\alpha}_i\left(\theta_0\right)$ and $\alpha_i\left(\theta_0\right)$,

$$
\begin{aligned}
&\frac{1}{nT}\sum_i\sum_t\left(\psi\left(y_{i,t};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)-\psi\left(y_{i,t};\theta_0,\alpha_i\left(\theta_0\right)\right)\right) \\
=&\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right) \\
&+\frac{1}{2}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)^2 \\
&+\frac{1}{6}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha\alpha}\left(y_{i,t};\theta_0,\overline{\alpha}_i\left(\theta_0\right)\right)\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)^3
\end{aligned}
$$

By Lemma 16, Lemma 18 and Lemma 19, we conclude that

$$
\begin{aligned}
&\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)] \\
&= -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} + \frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]} + o_p\left(\frac{1}{T\sqrt{T}}\right) \\
&+\frac{1}{2}\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} + \frac{1}{\sqrt{T}}\frac{1}{2}\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha,a}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} \\
&-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3} + o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}
$$

and that

$$\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]$$

$$=-\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}$$

$$+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left\{\frac{1}{\sqrt{T}}\sum_t\left[w_{i,t}+\frac{1}{2E\left[v_{i,t}^{\alpha}\right]}\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right]\right\}}{E\left[v_{i,t}^{\alpha}\right]}$$

$$-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)$$

We write

$$w_{i,t}+\frac{1}{2E\left[v_{i,t}^{\alpha}\right]}\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)-\frac{1}{6}\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}$$

$$=\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}-\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{2E\left[v_{i,t}^{\alpha}\right]}+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}-\frac{1}{6}\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}$$

$$=\frac{1}{2}\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{6\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}$$

We obtain

$$\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]$$

$$=-\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}$$

$$+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(\frac{1}{2}\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{6\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}\right)\right)}{E\left[v_{i,t}^{\alpha}\right]}+o_p\left(\frac{1}{T\sqrt{T}}\right)$$

$$=\frac{1}{nT}\sum_i R_{fi}\left(\theta_0\right)+O_p\left(\frac{1}{T\sqrt{T}}\right)$$

■

Proposition 5 *$\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}(\theta_0)-\widehat{R}_{fi}^{*}(\theta_0)\right]\xrightarrow{d}N(0,\sigma_{U_f}^2)$, where $\sigma_{U_f}^2=\lim_{n\to\infty}\frac{1}{2n}\sum_i\frac{\left(E\left[v_{fi,t}^2\right]\right)^2}{\left(E\left[v_{fi,t}^{\alpha}\right]\right)^2}$.*

Proof. We denote

$$V_{fi,T} = -\frac{1}{2\sqrt{n}} \frac{\left(\frac{1}{\sqrt{T}} \sum_t v_{fi,t}\right)^2}{E\left[v_{fi,t}^{\alpha}\right]} + \frac{1}{2} \frac{\frac{1}{T} \sum_t v_{fi,t}^2}{E\left[v_{fi,t}^{\alpha}\right]} = -\frac{1}{T\sqrt{n}} \frac{\sum_{t=2}^{T} v_{fi,t} \left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]}$$

from which we have

$$\frac{1}{\sqrt{n}} \sum_i \left[R_{fi}\left(\theta_0\right) - \widehat{R}_{fi}^{*}(\theta_0)\right] = \sum_i V_{fi,T}$$

Observe that, for $t \geq 2$, let $N_{ft,t',i} = \frac{v_{fi,t} v_{fi,t'}}{E\left[v_{fi,t}^{\alpha}\right]}$. Then,

$$\begin{aligned} \frac{1}{\sigma_{U_f}^2} \sum_i E\left[V_{fi,T}^2\right] =& \frac{1}{nT^2\sigma_{U_f}^2} \sum_i E\left[\left(\frac{\sum_{t=2}^{T} v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]}\right)^2\right] \\ =& \frac{1}{nT^2\sigma_{U_f}^2} \sum_i E\left[\sum_{t=2}^{T} \sum_{t'=1}^{t-1} N_{ft,t',i}^2\right] \\ &+ \frac{1}{nT^2\sigma_{U_f}^2} \sum_i E\left[\sum_{t\neq t'\neq t''\neq t}^{T} N_{ft,t',i} N_{ft,t'',i}\right] \\ &+ \frac{1}{nT^2\sigma_{U_f}^2} C \sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T} N_{ft,t',i} N_{ft'',t''',i}\right] \end{aligned}$$

where C is some positive integers which is not important for the analysis. The first equality holds by the definition of $V_{fi,T}$. The second equality holds because for any $t \neq t' \neq t'' \neq t$, $E\left[N_{ft,t',i}.N_{ft,t'',i}\right] = 0$ and for any $t \neq t' \neq t'' \neq t''' \neq t$, $E\left[N_{ft,t',i} N_{ft'',t''',i}\right] = 0$. Consider

$$E\left[\frac{1}{nT^2\sigma_{U_f}^2} \sum_i \sum_{t=2}^{T} \sum_{t'=1}^{t-1} N_{ft,t',i}^2\right] = \frac{1}{nT^2\sigma_{U_f}^2} \sum_{t=1}^{T-1} (T-t) \sum_i \frac{\left(E\left[v_{fi,t}^2\right]\right)^2}{\left(E\left[v_{fi,t}^{\alpha}\right]\right)^2} = 1$$

Also consider:

$$\begin{aligned} & E\left[\frac{1}{nT^2\sigma_{U_f}^2} \sum_i \sum_{t=1}^{T-1} (T-t)\left(N_{ft,t',i}^2 - E\left(N_{ft,t',i}^2\right)\right)\right]^2 \\ =& \frac{1}{n^2T^4\sigma_{U_f}^4} \sum_i \sum_{t=1}^{T-1} (T-t)^2 \left(E\left[N_{ft,t',i}^4\right] - \left[E\left(N_{ft,t',i}^2\right)\right]^2\right) \\ \leq& \frac{1}{n^2T\sigma_{U_f}^4} E\left[\sum_i \left[\frac{\left(\frac{1}{T}\sum_t v_{fi,t}^2\right)^4}{\left(E\left[v_{fi,t}^{\alpha}\right]\right)^4}\right]\right] \\ =& o_p(1) \end{aligned}$$

the first equality holds under condition 1, the first inequality is because that $\left[E\left(N_{ft,t',i}^2\right)\right]^2 \geq 0$, and $\sum_{t=1}^{T-1}(T-t)^2 \leq T^3$. Then, we use Lyapunov Central Limit Theorem to show that

$$\frac{1}{\sqrt{n}}\sum_i \left[R_{fi}(\theta_0) - \widehat{R}^*_{fi}(\theta_0)\right] \xrightarrow{d} N(0, \sigma^2_{U_f}) \tag{137}$$

for this purpose, it is sufficient to verify the following Lyapunov condition: as $n \to \infty$,

$$\sigma_{U_f}^{-3}\sum_i E\left(|V_{fi,T}|^3\right) \to 0 \tag{138}$$

We deduce that

$$\begin{aligned}
\sigma_{U_f}^{-3}\sum_i E\left(|V_{fi,T}|^3\right) =& n^{-\frac{3}{2}}T^{-3}\sigma_{U_f}^{-3}\sum_i E\left[\left|\frac{\sum_{t=2}^{T} v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]}\right|^3\right] \\
\leq & \frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}\sum_i E\left[\sum_{t=2}^{T}\sum_{t'=1}^{t-1}|N_{ft,t',i}|^3\right] \\
&+ \frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}C_1\sum_i E\left[\sum_{t\neq t'\neq t''\neq t}^{T}|N_{ft,t',i}|\,N^2_{ft,t'',i}\right] \\
&+ \frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}C_2\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T}|N_{ft,t',i}N_{ft,t'',i}N_{ft,t''',i}|\right] \\
&+ \frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}C_3\sum_i E\left[\sum_{t\neq t'\neq t''\neq t}^{T}|N_{ft,t',i}N_{ft,t'',i}N_{ft',t'',i}|\right] \\
&+ \frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}C_4\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T}\left|N^2_{ft,t',i}N_{ft'',t''',i}\right|\right] \\
&+ \frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}C_5\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t''''\neq t}^{T}|N_{ft,t',i}N_{ft,t'',i}N_{ft''',t'''',i}|\right] \\
&+ \frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}C_6\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t''''\neq t'''''\neq t}^{T}|N_{ft,t',i}N_{ft'',t''',i}N_{ft'''',t''''',i}|\right]
\end{aligned}$$

The inequality follows from the property of absolute value. By the definition of $N_{ft,t',i}$,

$$\frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}\sum_i E\left[\sum_{t=2}^{T}\sum_{t'=1}^{t-1}|N_{ft,t',i}|^3\right] \leq \frac{1}{n^{\frac{3}{2}}T^3\sigma^3_{U_f}}\sum_i\sum_{t=2}^{T}\sum_{t'=1}^{t-1}\sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3 \tag{139}$$

$$\leq \frac{1}{n^{\frac{3}{2}}T\sigma^3_{U_f}}\sum_i E\left[M^3\left(y_{i,t}\right)\right] = o_p(1) \tag{140}$$

The first inequality holds under condition 1. The second inequality follows from $\sum_{t=1}^{T-1}(T-t) \leq T^2$, we also have

$$
\begin{aligned}
\frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E\left[\sum_{t\neq t'\neq t''\neq t}^{T}|N_{ft,t',i}|\,N_{ft,t'',i}^2\right] \leq & \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i \sum_{t\neq t'\neq t''\neq t}^{T}\sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3 \\
= & \frac{T\left(T-1\right)\left(T-2\right)}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i \sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3 = o_p\left(1\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E[\sum_{t\neq t'\neq t''\neq t}^{T}|N_{ft,t',i}N_{ft,t'',i}N_{ft',t'',i}|] \leq & \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i \sum_{t\neq t'\neq t''\neq t}^{T}\sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3 \\
= & \frac{T\left(T-1\right)\left(T-2\right)}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i \sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3 = o_p\left(1\right)
\end{aligned}
$$

According to Lemma 20, we know that

$$
E\left[\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'}^{T}v_{fi,t}^2 v_{fi,t'}\right)\right]^2 = O_p\left(1\right) \tag{141}
$$

$$
E\left[\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'\neq t''\neq t}^{T}v_{fi,t}v_{fi,t'}v_{fi,t''}\right)\right]^2 = O_p\left(1\right) \tag{142}
$$

we deduce that

$$
\begin{aligned}
& \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T}|N_{ft,t',i}N_{ft,t'',i}N_{ft,t''',i}|\right] \\
= & \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i \sum_{t\neq t'\neq t''\neq t'''\neq t}^{T} E\left[|N_{ft,t',i}N_{ft,t'',i}N_{ft,t''',i}|\right] \\
\leq & \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i \frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6}E\left[\left|\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'}^{T}v_{fi,t}^2 v_{fi,t'}\right)\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t''\neq t'''\neq t}^{T}v_{fi,t}v_{fi,t''}v_{fi,t'''}\right)\right|\right] \\
\leq & \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i \frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'}^{T}v_{fi,t}^2 v_{fi,t'}\right)\right]^2\right)^{\frac{1}{2}}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t''\neq t'''\neq t}^{T}v_{fi,t}v_{fi,t''}v_{fi,t'''}\right)\right]^2\right)^{\frac{1}{2}} \\
= & o_p\left(1\right)
\end{aligned}
$$

The first inequality holds by the definition of $N_{ft,t',i}$. The second inequality follows from Holder's inequality. Likewise, we have

$$
\frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T}|N_{ft,t',i}^2 N_{ft'',t''',i}|\right]
$$

$$
\begin{aligned}
\leq & \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i \frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6} E\left[\left|\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'\neq t''\neq t''\neq t'}^{T} v_{fi,t'}v_{fi,t''}v_{fi,t'''}\right)\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'}^{T} v_{fi,t}^2 v_{fi,t'}\right)\right|\right] \\
\leq & \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i \frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6} \left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'}^{T} v_{fi,t}^2 v_{fi,t'}\right)\right]^2\right)^{\frac{1}{2}} \left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'\neq t''\neq t''\neq t'}^{T} v_{fi,t'}v_{fi,t''}v_{fi,t'''}\right)\right]^2\right)^{\frac{1}{2}} \\
= & o_p(1)
\end{aligned}
$$

and that

$$
\begin{aligned}
& \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t''''\neq t}^{T} \left|N_{ft,t',i}N_{ft,t'',i}N_{ft''',t'''',i}\right|\right] \\
\leq & \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i \frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6} E\left[\left|\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'\neq t''\neq t'''\neq t'}^{T} v_{fi,t'}v_{fi,t''}v_{fi,t'''}\right)\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t''''}^{T} v_{fi,t}^2 v_{fi,t''''}\right)\right|\right] \\
\leq & \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i \frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6} \left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t''''}^{T} v_{fi,t}^2 v_{fi,t''''}\right)\right]^2\right)^{\frac{1}{2}} \left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'\neq t''\neq t'''\neq t'}^{T} v_{fi,t'}v_{fi,t''}v_{fi,t'''}\right)\right]^2\right)^{\frac{1}{2}} \\
= & o_p(1)
\end{aligned}
$$

we also have

$$
\begin{aligned}
& \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t''''\neq t'''''\neq t}^{T} \left|N_{ft,t',i}N_{ft'',t''',i}N_{ft'''',t''''',i}\right|\right] \\
\leq & \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i \frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6} E\left[\left|\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'\neq t''\neq t}^{T} v_{fi,t}v_{fi,t'}v_{fi,t''}\right)\right.\right. \\
& * \left.\left.\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'''\neq t''''\neq t'''''\neq t'''}^{T} v_{fi,t'''}v_{fi,t''''}v_{fi,t'''''}\right)\right|\right] \\
\leq & \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i \frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6} \left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'''\neq t''''\neq t'''''\neq t'''}^{T} v_{fi,t'''}v_{fi,t''''}v_{fi,t'''''}\right)\right]^2\right)^{\frac{1}{2}} \\
& * \left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'\neq t''\neq t}^{T} v_{fi,t}v_{fi,t'}v_{fi,t''}\right)\right]^2\right)^{\frac{1}{2}} \\
= & o_p(1)
\end{aligned}
$$

It follows that

$$\sigma_{U_f}^{-3} \sum_i E\left(|V_{fi,T}|^3\right) = o_p(1) \tag{143}$$

Therefore we conclude that

$$\frac{1}{\sqrt{n}} \sum_i \left[R_{fi}(\theta_0) - \widehat{R}^*_{fi}(\theta_0)\right] = \sum_i V_{fi,T} \xrightarrow{d} N(0, \sigma_{U_f}^2)$$

■

Proposition 6 *$\frac{1}{\sqrt{n}} \sum_i \left[R_{fi}(\theta_0) - R_{gi}(\gamma_0) - \left(\widehat{R}^*_{fi}(\theta_0) - \widehat{R}^*_{gi}(\gamma_0)\right)\right] \xrightarrow{d} N(0, \sigma_U^2)$, where $\sigma_U^2 = \lim_{n\to\infty} \frac{1}{2n} \sum_i \left[\frac{E\left[v_{fi,t}^2\right]}{E\left[v_{fi,t}^\alpha\right]} - \frac{E\left[v_{gi,t}^2\right]}{E\left[v_{gi,t}^\alpha\right]}\right]^2$.*

Proof. From Proposition 5, we deduce that

$$\frac{1}{\sqrt{n}} \sum_i \left[R_{fi}(\theta_0) - \widehat{R}^*_{fi}(\theta_0)\right] \xrightarrow{d} N(0, \sigma_{U_f}^2)$$
$$\frac{1}{\sqrt{n}} \sum_i \left[R_{gi}(\gamma_0) - \widehat{R}^*_{gi}(\gamma_0)\right] \xrightarrow{d} N(0, \sigma_{U_g}^2)$$

We denote

$$\frac{1}{\sqrt{n}} \sum_i \left[R_{fi}(\theta_0) - \widehat{R}^*_{fi}(\theta_0)\right] = \sum_i V_{fi,T} \tag{144}$$
$$\frac{1}{\sqrt{n}} \sum_i \left[R_{gi}(\gamma_0) - \widehat{R}^*_{gi}(\gamma_0)\right] = \sum_i V_{gi,T} \tag{145}$$
$$\sigma_U^2 = \sigma_{U_f}^2 + \sigma_{U_g}^2 - 2Cov(\sum_i V_{fi,T}, \sum_i V_{gi,T}) \tag{146}$$

where

$$Cov(\sum_i V_{fi,T}, \sum_i V_{gi,T}) = \frac{1}{nT^2} \sum_i E\left[\frac{\sum_{t=2}^T v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right) \sum_{t=2}^T v_{gi,t}\left(\sum_{t'=1}^{t-1} v_{gi,t'}\right)}{E\left[v_{fi,t}^\alpha\right] E\left[v_{gi,t}^\alpha\right]}\right] \tag{147}$$

$$= \frac{T-1}{2nT} \sum_i \frac{\left(E\left[v_{fi,t}^2\right] E\left[v_{gi,t}^2\right]\right)}{E\left[v_{fi,t}^\alpha\right] E\left[v_{gi,t}^\alpha\right]} \tag{148}$$

the first equality holds because $E\left[v_{fi,t}v_{gi,t'}v_{fi,t'}v_{gi,t''}\right]=0$ and $E\left[v_{fi,t}v_{gi,t'}v_{fi,t''}v_{gi,t'''}\right]=0$ for any $t\neq t'\neq t''\neq t'''\neq t$, the second equality follows from $\sum_{t=1}^{T-1}(T-t)=\frac{T(T-1)}{2}$. We conclude that

$$\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}\left(\theta_0\right)-R_{gi}\left(\gamma_0\right)-\left(\widehat{R}^*_{fi}(\theta_0)-\widehat{R}^*_{gi}(\gamma_0)\right)\right]\xrightarrow{d}N(0,\sigma^2_U)$$

■

Proposition 7

$$\frac{1}{\sqrt{n}}\sum_i\left\{\left[R_{fi}(\theta_0)-\sum_i\widehat{R}^*_{fi}(\theta_0)\right]-\left[R_{1gi}(\theta_0)-\sum_i\widehat{R}^*_{1gi}(\theta_0)\right]-\left[R_{2gi}(\theta_0)-\sum_i\widehat{R}^*_{2gi}(\theta_0)\right]\right\}$$
$$\xrightarrow{d}N(0,\sigma^2_{U^{nested}_{t,n}})$$

where

$$\sigma^2_{U^{nested}_{t,n}}\equiv\sigma^2_{U_{1g}}=\lim_{n\to\infty}\frac{1}{2n}\sum_i\frac{\left(E\left[v^2_{1gi,t}\right]\right)^2}{\left(E\left[v^\alpha_{1gi,t}\right]\right)^2}$$

Proof. We denote $P_{fi}=\frac{1}{\sqrt{2J}}\sum_{t=1}^{t=2J}v_{fi,t}$, $P_{1gi}=\frac{1}{\sqrt{J}}\sum_{t=1}^{t=J}v_{1gi,t}$, $P_{2gi}=\frac{1}{\sqrt{J}}\sum_{t=J+1}^{t=2J}v_{2gi,t}$, $G_{fi}=E\left(\frac{1}{\sqrt{2J}}\sum_{t=1}^{t=2J}v_{fi,t}\right)^2$, $G_{1gi}=E(\frac{1}{\sqrt{J}}\sum_{t=1}^{t=J}v_{1gi,t})^2$, $G_{2gi}=E\left(\frac{1}{\sqrt{J}}\sum_{t=J+1}^{t=2J}v_{2gi,t}\right)^2$, under null hypothesis, $\alpha_{i1}=\alpha_{i2}=\alpha_i$, $E\left[v^\alpha_{fi,t}\right]=E\left[v^\alpha_{1gi,t}\right]=E\left[v^\alpha_{2gi,t}\right]$, $E\left(v^2_{2gi,t}\right)=E\left(v^2_{1gi,t}\right)=E\left(v^2_{fi,t}\right)$, $P_{fi}=\frac{1}{\sqrt{2}}\left(P_{1gi}+P_{2gi}\right)$, and that

$$\begin{aligned}
&R_{fi}\left(\theta_0\right)-R_{1gi}\left(\theta_0\right)-R_{2gi}\left(\theta_0\right)-\left[\widehat{R}^*_{fi}(\theta_0)-\widehat{R}^*_{1gi}(\theta_0)-\widehat{R}^*_{2gi}(\theta_0)\right]\\
=&-\frac{1}{2}\frac{\left[\frac{1}{\sqrt{2}}\left(P_{1gi}+P_{2gi}\right)\right]^2-P^2_{1gi}-P^2_{2gi}}{E\left[v^\alpha_{fi,t}\right]}-\left[-\frac{1}{2}\frac{G_{fi}-G_{1gi}-G_{2gi}}{E\left[v^\alpha_{fi,t}\right]}\right]\\
=&\frac{\frac{1}{2}P^2_{1gi}+\frac{1}{2}P^2_{2gi}-P_{1gi}P_{2gi}-G_{fi}}{2E\left[v^\alpha_{fi,t}\right]}\\
=&\frac{\frac{1}{2}\left(P^2_{1gi}-G_{1gi}\right)}{2E\left[v^\alpha_{fi,t}\right]}+\frac{\frac{1}{2}\left(P^2_{2gi}-G_{2gi}\right)}{2E\left[v^\alpha_{fi,t}\right]}-\frac{P_{1gi}P_{2gi}}{2E\left[v^\alpha_{fi,t}\right]}
\end{aligned}$$

From Proposition 5, we deduce that

$$\frac{1}{\sqrt{n}}\sum_i\left[R_{1gi}\left(\theta_0\right)-\widehat{R}^*_{1gi}(\theta_0)\right]=\frac{1}{\sqrt{n}}\sum_i\left[-\frac{1}{2}\frac{\left(P^2_{1gi}-G_{1gi}\right)}{E\left[v^\alpha_{1gi,t}\right]}\right]\xrightarrow{d}N(0,\sigma^2_{U_{1g}})$$

$$\frac{1}{\sqrt{n}}\sum_{i}\left[R_{2gi}\left(\theta_{0}\right)-\widehat{R}_{2gi}^{*}(\theta_{0})\right]=\frac{1}{\sqrt{n}}\sum_{i}\left[-\frac{1}{2}\frac{\left(P_{2gi}^{2}-G_{2gi}\right)}{E\left[v_{2gi,t}^{\alpha}\right]}\right]\xrightarrow{d}N(0,\sigma_{U_{2g}}^{2})$$

Since $E\left[P_{1gi}P_{2gi}\right]=0$,the mean becomes:

$$E[\frac{1}{\sqrt{n}}\sum_{i}\left\{\left[R_{fi}(\theta_{0})-\sum_{i}\widehat{R}_{fi}^{*}(\theta_{0})\right]-\left[R_{1gi}(\theta_{0})-\sum_{i}\widehat{R}_{1gi}^{*}(\theta_{0})\right]-\left[R_{2gi}(\theta_{0})-\sum_{i}\widehat{R}_{2gi}^{*}(\theta_{0})\right]\right\}]=0$$

and that

$$\begin{aligned}&Var\left[\frac{1}{\sqrt{n}}\sum_{i}(R_{fi}\left(\theta_{0}\right)-R_{1gi}\left(\theta_{0}\right)-R_{2gi}\left(\theta_{0}\right))\right]\\&=\frac{1}{2}\sigma_{U_{1g}}^{2}+E\left[\frac{1}{\sqrt{n}}\sum_{i}\frac{P_{1gi}P_{2gi}}{2E\left[v_{fi,t}^{\alpha}\right]}\right]^{2}=\sigma_{U_{1g}}^{2}\end{aligned}\tag{149}$$

where the first equality holds by $\sigma_{U_{1g}}^{2}=\sigma_{U_{2g}}^{2}$ under the null hypothesis and $E\left[P_{1gi}^{3}P_{2gi}\right]=0$, $E\left[P_{1gi}P_{2gi}^{3}\right]=0$. The second equality holds by the definition of P_{1gi},P_{1gi}, and $E\left[P_{1gi}^{2}\right]=E\left[P_{2gi}^{2}\right]$ under the null hypothesis. ■

Proposition 8 *under $H_{0}:E\left[LR_{nT}\left(\theta_{0},\gamma_{0}\right)\right]=0$:*

$$\frac{1}{\sqrt{nT}}\sum_{i}\sum_{t}\left\{\left(\psi_{f}\left(y_{it},\theta_{0},\alpha_{i}\right)-\psi_{g}\left(y_{it},\gamma_{0},\lambda_{i}\right)\right)\right\}\xrightarrow{d}N(0,\omega^{2})\tag{150}$$

where $\omega^{2}=\frac{1}{nT}\sum_{i}\cdot\sum_{t}\left[\left(\psi_{f}\left(y_{it},\theta_{0},\alpha_{i}\right)-\psi_{g}\left(y_{it},\gamma_{0},\lambda_{i}\right)\right)\right]^{2}$.

Proof. *Similar to Proposition 6, we denote*

$$L_{i,T}=\frac{1}{\sqrt{nT}}\sum_{t}\left(\psi_{f}\left(y_{it},\theta_{0},\alpha_{i}\right)-\psi_{g}\left(y_{it},\gamma_{0},\lambda_{i}\right)\right)\tag{151}$$

it is sufficient to verify the following Lyapunov condition: as $n\to\infty$,

$$\omega^{-3}\sum_{i}E\left(\left|L_{i,T}\right|^{3}\right)\to 0\tag{152}$$

which which is satisfied since as $n\to\infty$, let $P_{t,i}=\psi_{f}\left(y_{it},\theta_{0},\alpha_{i}\right)-\psi_{g}\left(y_{it},\gamma_{0},\lambda_{i}\right)$. Then,

$$\begin{aligned}&\omega^{-3}\sum_{i}E\left(\left|L_{i,T}\right|^{3}\right)\\=&n^{-\frac{3}{2}}T^{-\frac{3}{2}}\omega^{-3}\sum_{i}E\left[\left|\sum_{t}\left(\psi_{f}\left(y_{it},\theta_{0},\alpha_{i}\right)-\psi_{g}\left(y_{it},\gamma_{0},\lambda_{i}\right)\right)\right|^{3}\right]\end{aligned}$$

$$\le n^{-\frac{3}{2}} T^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[\sum_t |P_{t,i}|^3\right] + n^{\frac{3}{2}} T^{\frac{3}{2}} \omega^{3} \sum_i E\left[\sum_{t\neq t'} |P_{t,i}|\, P_{t',i}^2\right]$$
$$+ n^{-\frac{3}{2}} T^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[\sum_{t\neq t'\neq t''\neq t} \left|P_{t,i} P_{t',i} P_{t'',i}\right|\right]$$

Since

$$n^{-\frac{3}{2}} T^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[\sum_t |P_{t,i}|^3\right] \le n^{-\frac{3}{2}} T^{-\frac{1}{2}} \omega^{-3} \sum_i \sup_i \left(E\left[M\left(y_{i,t}\right)\right]\right)^3 = o_p\left(1\right) \tag{153}$$

According to Lemma 20, we deduce that

$$n^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[T^{-\frac{3}{2}} \sum_{t\neq t'} |P_{t,i}|\, P_{t',i}^2\right]$$
$$\le n^{-\frac{3}{2}} \omega^{-3} \sum_i \left(E\left[T^{-\frac{3}{2}} \sum_{t\neq t'} |P_{t,i}|\, P_{t',i}^2\right]^2\right)^{\frac{1}{2}} = o_p\left(1\right)$$

and that

$$n^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[T^{-\frac{3}{2}} \sum_{t\neq t'\neq t''\neq t} \left|P_{t,i} P_{t',i} P_{t'',i}\right|\right]$$
$$\le n^{-\frac{3}{2}} \omega^{-3} \sum_i \left(E\left[T^{-\frac{3}{2}} \sum_{t\neq t'\neq t''\neq t} \left|P_{t,i} P_{t',i} P_{t'',i}\right|\right]^2\right)^{\frac{1}{2}}$$
$$= o_p\left(1\right)$$

Therefore, we conclude that

$$\frac{1}{\sqrt{nT}} \sum_i \sum_t \left\{\left(\psi_f\left(y_{it}, \theta_0, \alpha_i\right) - \psi_g\left(y_{it}, \gamma_0, \lambda_i\right)\right)\right\} = \sum_i L_{i,T} \xrightarrow{d} N(0, \omega^2) \tag{154}$$

■

Proposition 9 *$\frac{1}{\sqrt{n}} LR_{nT}\left(\theta_0, \gamma_0\right) + \frac{1}{\sqrt{n}} \sum_i \left[R_{fi}\left(\theta_0\right) - R_{gi}\left(\gamma_0\right) - \left(R_{fi}^* - R_{gi}^*\right)\right] \xrightarrow{d} N(0, \sigma_W^2)$, where $\sigma_W^2 = T\omega^2 + \sigma_U^2$.*

Proof. We denote

$$W_{i,T} = V_{i,T} + L_{i,T} \tag{155}$$

$$V_{i,T} = -\frac{1}{\sqrt{n}}\frac{1}{T}\frac{\sum_{t=2}^{T} v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]} + \frac{1}{\sqrt{n}}\frac{1}{T}\frac{\sum_{t=2}^{T} v_{gi,t}\left(\sum_{t'=1}^{t-1} v_{gi,t'}\right)}{E\left[v_{gi,t}^{\alpha}\right]} \tag{156}$$

$$L_{i,T} = \frac{1}{\sqrt{n}}\left[\sum_t \left(\psi_f\left(y_{it},\theta_0,\alpha_i\right) - \psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right) - E\left[\sum_t \left(\psi_f\left(y_{it},\theta_0,\alpha_i\right) - \psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right)\right]\right] \tag{157}$$

Therefore we can write

$$\frac{1}{\sqrt{n}} LR_{nT}\left(\theta_0,\gamma_0\right) + \frac{1}{\sqrt{n}}\sum_i \left[R_{fi}\left(\theta_0\right) - R_{gi}\left(\gamma_0\right) - \left(R_{fi}^* - R_{gi}^*\right)\right] \tag{158}$$

$$= \sum_i V_{i,T} + \sum_i L_{i,T} = \sum_i W_{i,T} \tag{159}$$

From Proposition 6 and 8, we know that

$$\frac{1}{\sqrt{n}} LR_{nT}\left(\theta_0,\gamma_0\right) = \sum_i L_{i,T} \xrightarrow{d} N(0, T\omega^2)$$

$$\frac{1}{\sqrt{n}}\sum_i \left[R_{fi}\left(\theta_0\right) - R_{gi}\left(\gamma_0\right) - \left(\widehat{R}_{fi}^*(\theta_0) - \widehat{R}_{gi}^*(\gamma_0)\right)\right] = \sum_i V_{i,T} \xrightarrow{d} N(0,\sigma^2)$$

it is sufficient to verify the following Lyapunov condition: as $n \to \infty$,

$$\sigma_W^{-3}\sum_i E\left(\left|W_{i,T}\right|^3\right) \to 0 \tag{160}$$

which is satisfied since as $n \to \infty$,

$$\begin{aligned}
\sigma_W^{-3}\sum_i E\left(\left|W_{i,T}\right|^3\right) \leq & \sigma_W^{-3}\sum_i E\left(\left|V_{i,T}\right|^3 + \left|L_{i,T}\right|^3\right) \\
= & \frac{\sigma_U^3}{\sigma_W^3}\sigma_U^{-3}\sum_i E\left(\left|V_{i,T}\right|^3\right) + \frac{\left(\sqrt{T}\omega\right)^3}{\sigma_W^3}\left(\sqrt{T}\omega\right)^{-3}\sum_i E\left(\left|L_{i,T}\right|^3\right) \\
\to & 0
\end{aligned}$$

The first inequality holds by the convexity of the function $f\left(x\right) = \left|x\right|^3$. The first equality follows from Proposition 6 and 8. We denote

$$\begin{aligned}
V_{n,t} &= \left(v_{f1,t}, v_{f2,t}, \ldots v_{fn,t}, -v_{g1,t}, -v_{g2,t}, \ldots - v_{gn,t}\right)' \\
V_{n,t}^{\alpha} &= diag\left(v_{f1,t}^{\alpha}, v_{f2,t}^{\alpha}, \ldots v_{fn,t}^{\alpha}, -v_{g1,t}^{\alpha}, -v_{g2,t}^{\alpha}, \ldots - v_{gn,t}^{\alpha}\right) \\
H_n &= E\left[V_{n,t}^{\alpha}\right]
\end{aligned}$$

$$D_n = E\left[V_{n,t}V_{n,t}'\right]$$

and further consider

$$\begin{aligned}
&\sum_i E\left[L_{i,T}V_{i,T}\right] \\
=&-\frac{1}{nT}\sum_i E\left[\left[\sum_t \left(\psi_f\left(y_{it},\theta_0,\alpha_i\right)-\psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right)\right]\right. \\
&\left.*\left[\frac{\sum_{t=2}^T v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]}-\frac{1}{\sqrt{n}}\frac{1}{T}\frac{\sum_{t=2}^T v_{gi,t}\left(\sum_{t'=1}^{t-1} v_{gi,t'}\right)}{E\left[v_{gi,t}^{\alpha}\right]}\right]\right] \\
=&-\frac{1}{nT}E\left[\sum_{t=2}^{T}\sum_{t'=1}^{t-1} LP\left(\theta_0,\gamma_0\right)V_{n,t}'\left[H_n\right]^{-1}V_{n,t'}\right] \\
=&-\frac{1}{nT}E\left[\sum_{t=2}^{T}\sum_{t'=1}^{t-1} \varrho_{n,t}'D_n^{1/2}\left[H_n\right]^{-1}V_{n,t'}\right]
\end{aligned}$$

where $LP\left(\theta_0,\gamma_0\right)=\sum_i\left[\left(\psi_f\left(y_{it},\theta_0,\alpha_i\right)-\psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right)\right]$, $\varrho_{n,t}=\left(D_n^{1/2}\right)^{+}Cov\left[LP\left(\theta_0,\gamma_0\right),V_{n,t}\right]$, and $\left(D_n^{1/2}\right)^{+}$ denotes the Moore-Penrose inverse of $D_n^{1/2}$. Note that

$$\left(\left(D_n^{1/2}\right)^{+}Cov\left[LP\left(\theta_0,\gamma_0\right),V_{n,t}\right]\right)'D_n^{1/2}=Cov\left[LP\left(\theta_0,\gamma_0\right),V_{n,t}\right]'$$

Then,

$$\sum_i E\left[L_{i,T}V_{i,T}\right]=-\frac{1}{n}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\varrho_{n,t}'D_n^{1/2}\left[H_n\right]^{-1}V_{n,t'}=o_p\left(1\right)$$

follows from the fact that:

$$\begin{aligned}
&E\left[\sum_i E\left[L_{i,T}V_{i,T}\right]\right]^2 \\
=&\frac{1}{n^2}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)^2\varrho_{n,t}'D_n^{1/2}\left[H_n\right]^{-1}D_n\left[H_n\right]^{-1}D_n^{1/2}\varrho_{n,t} \\
\leq&\frac{\kappa_{\max}^2}{n}=o_p\left(1\right)
\end{aligned}$$

where $\kappa_{\max}^2$ is the maximum eigenvalue of matrix $\left(D_n\left[H_n\right]^{-1}\right)^2$. According to Lemma E.1 in Liao and Shi 2020, we have $\varrho_{n,t}'A\varrho_{n,t}\leq\kappa_{\max}\left(A\right)$ for any positive semi-definite matrix A.

$\kappa_{\max}(A)$ is the maximum eigenvalue of matrix A. The inequality holds by $\left(1-\frac{t}{T}\right)^2 \leq 1$. Therefore, we conclude that

$$\frac{1}{\sqrt{n}} LR_{nT}(\theta_0, \gamma_0) + \frac{1}{\sqrt{n}} \sum_i \left[R_{fi}(\theta_0) - R_{gi}(\gamma_0) - \left(R^*_{fi} - R^*_{gi}\right)\right] \xrightarrow{d} N(0, \sigma^2_W)$$

■

7.4 Theorems

Proof of Theorem 1. From equation 44, we have:

$$\begin{aligned} &\frac{1}{\sqrt{nT}}[LR_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right) - E\left[LR_{nT}(\theta_0, \gamma_0)\right]] - \frac{1}{\sqrt{nT}} \sum_i \left[\widehat{R}_{fi}(\tilde{\theta}) - \widehat{R}_{gi}(\tilde{\gamma})\right] \\ =&\frac{1}{\sqrt{nT}}[LR_{nT}(\theta_0, \gamma_0) - E\left[LR_{nT}(\theta_0, \gamma_0)\right]] + o_p(1) \end{aligned} \tag{T.1.1}$$

under $H_0 : E\left[LR_{nT}(\theta_0, \gamma_0)\right] = 0$:

$$\frac{1}{\sqrt{nT}} LR_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right) - \frac{1}{\sqrt{nT}} \sum_i \left[\widehat{R}_{fi}(\tilde{\theta}) - \widehat{R}_{gi}(\tilde{\gamma})\right] = \frac{1}{\sqrt{nT}} LR_{nT}(\theta_0, \gamma_0) + o_p(1) \tag{161}$$

from Proposition 8, we know that

$$\frac{1}{\sqrt{nT}} LR_{nT}(\theta_0, \gamma_0) \xrightarrow{d} N(0, \omega^2) \tag{162}$$

we denote $\widehat{\omega}^2_n$ as an estimator for ω^2, and $\widehat{\omega}^2_n = \frac{1}{nT}\sum_i\sum_t [\psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g(\tilde{\gamma}, \widehat{\alpha}_i(\tilde{\gamma}))]^2$, $\omega^2 = \lim_{n,T\to\infty} \frac{1}{nT}\sum_i\sum_t \left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \alpha_i(\gamma_0))\right]^2$. The consistency results for $\widehat{\omega}^2_n$ can be similarly proved as previous part, more specifically, let

$$\omega^2_n = \frac{1}{nT}\sum_i\sum_t \left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \alpha_i(\gamma_0))\right]^2$$

let $\frac{\partial u\left(y_{i,t}; \tilde{\tilde{\theta}}, \widehat{\alpha}_i(\tilde{\tilde{\theta}})\right)}{\partial \theta} = u^{\tilde{\tilde{\theta}}}_{fi,t}$, then we can prove that:

$$\begin{aligned} \widehat{\omega}^2_n - \omega^2 =&\frac{1}{nT}\sum_i\sum_t 2\left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \alpha_i(\gamma_0))\right] \\ &* \left\{\left[\psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g(\tilde{\gamma}, \widehat{\alpha}_i(\tilde{\gamma}))\right] - \left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \alpha_i(\gamma_0))\right]\right\} \\ &+ \frac{1}{nT}\sum_i\sum_t 2\left\{\left[\psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g(\tilde{\gamma}, \widehat{\alpha}_i(\tilde{\gamma}))\right] - \left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \alpha_i(\gamma_0))\right]\right\}^2 \end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{nT}\sum_i\sum_t 2[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\alpha_i(\gamma_0)\right)]\\
&*\{u_{fi,t}^{\tilde{\tilde{\theta}}}(\tilde{\tilde{\theta}}-\widehat{\theta})\left(\tilde{\theta}-\theta_0\right)+v_{fi,t}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]+\frac{1}{2}v_{fi,t}^{\alpha}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]^2\\
&+\frac{1}{6}\widetilde{v}_{fi,t}^{\alpha\alpha}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]^3-u_{gi,t}^{\tilde{\tilde{\gamma}}}(\tilde{\tilde{\gamma}}-\widehat{\gamma})\left(\tilde{\gamma}-\gamma_0\right)-v_{gi,t}\left[\widehat{\alpha}_i(\gamma_0)-\alpha_i(\gamma_0)\right]\\
&-\frac{1}{2}v_{gi,t}^{\alpha}\left[\widehat{\alpha}_i(\gamma_0)-\alpha_i(\gamma_0)\right]^2-\frac{1}{6}v_{gi,t}^{\alpha\alpha}\left[\widehat{\alpha}_i(\gamma_0)-\alpha_i(\gamma_0)\right]^3\}\\
&+\frac{1}{nT}\sum_i\sum_t 2*\left\{u_{fi,t}^{\tilde{\tilde{\theta}}}(\tilde{\tilde{\theta}}-\widehat{\theta})\left(\tilde{\theta}-\theta_0\right)+v_{fi,t}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]+\frac{1}{2}v_{fi,t}^{\alpha}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]^2\right.\\
&+\frac{1}{6}\widetilde{v}_{fi,t}^{\alpha\alpha}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]^3-u_{fi,t}^{\tilde{\tilde{\gamma}}}(\tilde{\tilde{\gamma}}-\widehat{\gamma})\left(\tilde{\gamma}-\gamma_0\right)-v_{gi,t}\left[\widehat{\alpha}_i(\gamma_0)-\alpha_i(\gamma_0)\right]\\
&-\frac{1}{2}v_{gi,t}^{\alpha}\left[\widehat{\alpha}_i(\gamma_0)-\alpha_i(\gamma_0)\right]^2-\frac{1}{6}v_{gi,t}^{\alpha\alpha}\left[\widehat{\alpha}_i(\gamma_0)-\alpha_i(\gamma_0)\right]^3\}^2\\
=&o_p(1)
\end{aligned}
$$

■

Proof of Theorem 2. Since the asymptotic result is established, we only need to prove the consistency of $\widehat{\sigma}^2_{Unested}$. As

$$
\widehat{\sigma}^2_{Unested}-\sigma^2_{Unested}=\frac{J-1}{2nJ}\sum_i\left[\frac{\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^2\right)^2}{\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^{\alpha}\right)^2}-\frac{\left(E\left[v_{1gi,t}^2\right]\right)^2}{\left(E\left[v_{1gi,t}^{\alpha}\right]\right)^2}\right]
$$

, it suffices to show that

$$
\frac{1}{n}\sum_i\left[\frac{\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^2\right)^2}{\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^{\alpha}\right)^2}-\frac{\left(E\left[v_{1gi,t}^2\right]\right)^2}{\left(E\left[v_{1gi,t}^{\alpha}\right]\right)^2}\right] \tag{163}
$$

$$
=\frac{1}{n}\sum_i\left\{\left[\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^2\right)^2-\left(E\left[v_{1gi,t}^2\right]\right)^2\right]\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^{\alpha}\right)^{-2}\right\} \tag{164}
$$

$$
+\frac{1}{n}\sum_i\left\{\left(E\left[v_{1gi,t}^2\right]\right)^2\left[\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^{\alpha}\right)^{-2}-\left(E\left[v_{1gi,t}^{\alpha}\right]\right)^{-2}\right]\right\} \tag{165}
$$

$$
=o_p(1) \tag{166}
$$

Since we can prove that

$$
\widehat{v}_{1gi,t}\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-v_{1gi,t}\left(\theta_0,\alpha_i(\theta_0)\right)=\widehat{v}_{1gi,t}\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-v_{1gi,t}\left(\theta_0,\widehat{\alpha}_i(\theta_0)\right)
$$

$$
\begin{aligned}
&+ v_{1gi,t}\left(\theta_0, \widehat{\alpha}_i(\theta_0)\right) - v_{1gi,t}\left(\theta_0, \alpha_i(\theta_0)\right) \\
=&\underbrace{\frac{\partial v_{1gi,t}(\widetilde{\widetilde{\widetilde{\theta}}}, \widehat{\alpha}_i(\widetilde{\widetilde{\widetilde{\theta}}}))}{\partial \theta}\left(\widetilde{\widetilde{\theta}} - \widehat{\theta}\right)\left(\widetilde{\theta} - \theta_0\right)}_{O_p\left(\frac{1}{nT}\right)} + \underbrace{v^{\alpha}_{1gi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]}_{O_p\left(\frac{1}{\sqrt{T}}\right)} \\
&+ \underbrace{\frac{1}{2}v^{\alpha\alpha}_{1gi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^2}_{O_p\left(\frac{1}{T}\right)} + \underbrace{\frac{1}{6}\widetilde{v}^{\alpha\alpha\alpha}_{1gi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^3}_{O_p\left(\frac{1}{T\sqrt{T}}\right)}
\end{aligned}
$$

where $\widetilde{v}^{\alpha\alpha\alpha}_{1gi,t} = v^{\alpha\alpha\alpha}_{1gi,t}\left(\theta_0, \widetilde{\alpha}_i(\theta_0)\right)$, $\widetilde{\alpha}_i(\theta_0)$ is between $\widehat{\alpha}_i(\theta_0)$ and $\alpha_i(\theta_0)$. Therefore it is shown that

$$
\begin{aligned}
&\frac{1}{J}\sum_t \widehat{v}^2_{1gi,t}\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \frac{1}{J}\sum_t v^2_{1gi,t}(\theta_0, \alpha_i(\theta_0)) \\
&= \frac{1}{J}\sum_t 2v_{1gi,t}\left(\widehat{v}_{1gi,t} - v_{1gi,t}\right) + \frac{1}{J}\sum_t 2\left(\widehat{v}_{1gi,t} - v_{1gi,t}\right)^2 \\
&= \underbrace{\left[\frac{1}{J}\sum_t 2v_{1gi,t}\frac{\partial v_{1gi,t}(\widetilde{\widetilde{\widetilde{\theta}}}, \widehat{\alpha}_i(\widetilde{\widetilde{\widetilde{\theta}}}))}{\partial \theta}\right]\left(\widetilde{\widetilde{\theta}} - \widehat{\theta}\right)\left(\widetilde{\theta} - \theta_0\right)}_{O_p\left(\frac{1}{nT}\right)} + \underbrace{[\frac{1}{J}\sum_t 2v_{1gi,t}v^{\alpha}_{1gi,t}]\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]}_{O_p\left(\frac{1}{\sqrt{T}}\right)} \\
&+ \underbrace{[\frac{1}{J}\sum_t v_{1gi,t}v^{\alpha\alpha}_{1gi,t}]\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^2}_{O_p\left(\frac{1}{T}\right)} + \underbrace{[\frac{1}{J}\sum_t \frac{1}{3}v_{1gi,t}\widetilde{v}^{\alpha\alpha\alpha}_{1gi,t}]\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^3}_{O_p\left(\frac{1}{T\sqrt{T}}\right)} + O_p\left(\frac{1}{T}\right) \\
&= o_p(1)
\end{aligned}
$$

and

$$
\left(E\left[v^{\alpha}_{1gi,t}\right]\right)^{-2}\left[\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}^2_{1gi,t}\right)^2 - \left(E\left[v^2_{1gi,t}\right]\right)^2\right] = o_p(1) \tag{T.1.4}
$$

Applying similar analysis to $\frac{1}{J}\sum_t \widehat{v}^{\alpha}_{1gi,t}$, we have

$$
\begin{aligned}
\frac{1}{J}\sum_t \widehat{v}^{\alpha}_{1gi,t} - \frac{1}{J}\sum_t v^{\alpha}_{1gi,t} =& \frac{1}{J}\sum_t \left[\widehat{v}^{\alpha}_{1gi,t}\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - v^{\alpha}_{1gi,t}\left(\theta_0, \widehat{\alpha}_i(\theta_0)\right)\right] \\
&+ \frac{1}{J}\sum_t \left[v^{\alpha}_{1gi,t}\left(\theta_0, \widehat{\alpha}_i(\theta_0)\right) - v^{\alpha}_{1gi,t}(\theta_0, \alpha_i(\theta_0))\right] \\
=& \frac{1}{J}\sum_t \frac{\partial v^{\alpha}_{1gi,t}(\widetilde{\widetilde{\widetilde{\theta}}}, \widehat{\alpha}_i(\widetilde{\widetilde{\widetilde{\theta}}}))}{\partial \theta}\left(\widetilde{\widetilde{\theta}} - \widehat{\theta}\right)\left(\widetilde{\theta} - \theta_0\right) + \left(\frac{1}{J}\sum_t v^{\alpha\alpha}_{1gi,t}\right)\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]
\end{aligned}
$$

$$+\frac{1}{2}\left(\frac{1}{J}\sum_t v_{1gi,t}^{\alpha\alpha\alpha}\right)\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]^2+\frac{1}{6}\left(\frac{1}{J}\sum_t v_{1gi,t}^{\alpha\alpha\alpha\alpha}\right)\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]^3$$

$$=O_p(1)O_p\left(\frac{1}{nT}\right)+O_p(1)O_p\left(\frac{1}{\sqrt{T}}\right)=o_p(1) \tag{T.1.5}$$

Based on law of large numbers,

$$\frac{1}{J}\sum_t v_{1gi,t}^{\alpha}-E\left[v_{1gi,t}^{\alpha}\right]=o_p(1) \tag{T.1.6}$$

$$\frac{1}{J}\sum_t v_{1gi,t}^{2}-E\left[v_{1gi,t}^{2}\right]=o_p(1) \tag{T.1.7}$$

Combining equation (T.1.4), (T.1.5), (T.1.6) and (T.1.7), we can prove that (T.1.2) is $o_p(1)$.

■

Proof of Theorem 3. Since

$$\begin{bmatrix}\sqrt{nT}(\widetilde{\theta}-\theta_0)\\ \sqrt{nT}(\widetilde{\gamma}-\gamma_0)\end{bmatrix}\xrightarrow{d}N(0,\Sigma_{\theta\gamma})$$

$$-LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)=-\frac{NT}{2}(\widetilde{\theta}-\theta_0)'\mathcal{I}_f(\widetilde{\theta}-\theta_0)+\frac{NT}{2}(\widetilde{\gamma}-\gamma_0)'\mathcal{I}_g(\widetilde{\gamma}-\gamma_0)+o_p(1)$$

$-2LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)$ can be considered as a quadratic form of a vector of normally distributed random variables:

$$\begin{bmatrix}\sqrt{nT}(\widetilde{\theta}-\theta_0)\\ \sqrt{nT}(\widetilde{\gamma}-\gamma_0)\end{bmatrix}' Q_{\theta\gamma}\begin{bmatrix}\sqrt{nT}(\widetilde{\theta}-\theta_0)\\ \sqrt{nT}\ (\widetilde{\gamma}-\gamma_0)\end{bmatrix}$$

where

$$\Sigma_{\theta\gamma}=\begin{bmatrix}\Sigma_\theta; & Cov_{\theta\gamma}\\ Cov_{\theta\gamma}; & \Sigma_\gamma\end{bmatrix};Q_{\theta\gamma}=\begin{bmatrix}-\mathcal{I}_f & 0\\ 0 & \mathcal{I}_g\end{bmatrix}$$

According to Lemma 11, $-2LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)$ is asymptotically distributed as a central chi-square if and only if

$$\Sigma_{\theta\gamma}Q_{\theta\gamma}\Sigma_{\theta\gamma}Q_{\theta\gamma}\Sigma_{\theta\gamma}=\Sigma_{\theta\gamma}Q_{\theta\gamma}\Sigma_{\theta\gamma} \tag{167}$$

If the information matrix identity holds:

$$\mathcal{I}_f=E\left[U_{fit}\left(\theta_0,\alpha_0\right)U'_{fit}\left(\theta_0,\alpha_0\right)\right]$$

$$\mathcal{I}_g = E\left[U_{git}\left(\theta_0,\alpha_0\right)U'_{git}\left(\theta_0,\alpha_0\right)\right]$$

then equation (167) implies that

$$\mathcal{I}_f - \left(E\left[U_{fit}\left(\theta_0,\alpha_0\right)U'_{git}\left(\theta_0,\alpha_0\right)\right]\right)\mathcal{I}_g^{-1}\left(E\left[U_{git}\left(\theta_0,\alpha_0\right)U'_{fit}\left(\theta_0,\alpha_0\right)\right]\right) = 0 \tag{168}$$

which holds under the null hypothesis, it is also known that the degree of freedom is:

$$tr\left(Q_{\theta\gamma}\Sigma_{\theta\gamma}\right) = \dim(\gamma) - \dim(\theta)$$

Lee and Phillips 2015 consider the following profile likelihood information criterion:

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\log f\left(z_{i,t};\tilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right) - \frac{1}{nT}\sum_{i=1}^{n}\widehat{R}^{*}_{fi}(\widetilde{\theta}) + \frac{1}{nT}\operatorname{tr}\left\{J_f(\widehat{G})^{-1}I_f(\widehat{G})\right\}$$

where

$$J_f(\widehat{G}) = -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial^2\log f\left(z_{i,t};\widetilde{\theta},\alpha_i(\widetilde{\theta})\right)}{\partial\theta\partial\theta'}$$

$$I_f(\widehat{G}) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial\log f\left(z_{i,t};\widetilde{\theta},\alpha_i(\widetilde{\theta})\right)}{\partial\theta}\frac{\partial\log f\left(z_{i,t};\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)}{\partial\theta'}$$

Under H_0, the information matrix identity holds and gives

$$\operatorname{tr}\left\{J_f(\widehat{G})^{-1}I_f(\widehat{G})\right\} = \frac{\dim(\theta)}{nT}$$

it means that chi-square distribution is also achieved, which is equivalent to our results. ■

Proof of Theorem 4. Since under H_0, the estimator of variance of $LR_{nT}\left(\theta_0,\gamma_0\right)$ is:

$$\begin{aligned}
&\frac{1}{nT}\sum_i\sum_t[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma},\widehat{\lambda}_i(\tilde{\gamma})\right)]^2\\
=&\frac{1}{nT}\sum_i\sum_t[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right) - \psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right) + \psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_f\left(\theta_0,\widehat{\alpha}_i(\theta_0)\right) + \psi_f\left(\theta_0,\widehat{\alpha}_i(\theta_0)\right)\\
&- \psi_f\left(\theta_0,\alpha_i(\theta_0)\right) - \psi_g\left(\widetilde{\gamma},\widehat{\lambda}_i(\widetilde{\gamma})\right) + \psi_g\left(\gamma_0,\widehat{\lambda}_i(\gamma_0)\right) - \psi_g\left(\gamma_0,\widehat{\lambda}_i(\gamma_0)\right) + \psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right)]^2\\
=&\frac{1}{nT}\sum_i\sum_t[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right) - \psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right) + \psi_f\left(\theta_0,\widehat{\alpha}_i(\theta_0)\right) - \psi_f\left(\theta_0,\alpha_i(\theta_0)\right)\\
&+ \psi_g\left(\gamma_0,\widehat{\lambda}_i(\gamma_0)\right) - \psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right)]^2 + o_p(1)
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{nT}\sum_i\sum_t\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right)+v_{f,it}\left\{\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right\}-v_{g,it}\left\{\widehat{\lambda}_i(\gamma_0)-\lambda_i(\gamma_0)\right\}\right]^2+o_p(1)\\
=&\frac{1}{nT}\sum_i\sum_t\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right)+v_{f,it}\left\{\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right\}-v_{g,it}\left\{\widehat{\lambda}_i(\gamma_0)-\lambda_i(\gamma_0)\right\}\right]^2+o_p(1)\\
=&\frac{1}{nT}\sum_i\sum_t\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right)-v_{f,it}\frac{\frac{1}{T}\sum_t v_{fi,t}}{E\left[v_{fi,t}^{\alpha}\right]}+v_{g,it}\frac{\frac{1}{T}\sum_t v_{gi,t}}{E\left[v_{gi,t}^{\alpha}\right]}\right]^2+o_p(1)\\
=&\frac{1}{nT}\sum_i\sum_t\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right)\right]^2+\frac{1}{nT}\sum_i\sum_t\left\{v_{f,it}\frac{\frac{1}{T}\sum_t v_{fi,t}}{E\left[v_{fi,t}^{\alpha}\right]}-v_{g,it}\frac{\frac{1}{T}\sum_t v_{gi,t}}{E\left[v_{gi,t}^{\alpha}\right]}\right\}^2\\
&-\frac{2}{nT}\sum_i\sum_t\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\lambda_i(\gamma_0)\right)\right]\left\{v_{f,it}\frac{\frac{1}{T}\sum_t v_{fi,t}}{E\left[v_{fi,t}^{\alpha}\right]}-v_{g,it}\frac{\frac{1}{T}\sum_t v_{gi,t}}{E\left[v_{gi,t}^{\alpha}\right]}\right\}+o_p(1)\\
=&\omega^2+\frac{1}{nT}\sum_i\frac{\left(E\left[v_{fi,t}^2\right]\right)^2}{\left(E\left[v_{fi,t}^{\alpha}\right]\right)^2}+\frac{1}{nT}\sum_i\frac{\left(E\left[v_{gi,t}^2\right]\right)^2}{\left(E\left[v_{gi,t}^{\alpha}\right]\right)^2}-2\frac{1}{nT}\sum_i\frac{\left(E\left[v_{fi,t}^2v_{gi,t}^2\right]\right)}{E\left[v_{fi,t}^{\alpha}\right]E\left[v_{gi,t}^{\alpha}\right]}+o_p(1)\\
=&\frac{1}{T}\sigma_W^2-\frac{1}{T}\sigma_U^2+\frac{2}{T}\sigma_U^2+o_p(1)=\frac{1}{T}\sigma_W^2+\frac{1}{T}\sigma_U^2+o_p(1)
\end{aligned}
$$

Thus we have the estimator of variance of $\frac{1}{\sqrt{n}}LR_{nT}\left(\theta_0,\gamma_0\right)$ is:

$$
\frac{1}{n}\sum_i\sum_t[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\psi_g\left(\tilde{\gamma},\widehat{\lambda}_i(\tilde{\gamma})\right)]^2=\sigma_W^2+\sigma_U^2+o_p(1)
$$

It means that $\frac{1}{n}\sum_i\sum_t\left[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\psi_g\left(\tilde{\gamma},\widehat{\lambda}_i(\tilde{\gamma})\right)\right]^2$ is a biased estimator of σ_W^2, we propose the bias-corrected variance term as follows:

$$
\widehat{\sigma}_W^2=\frac{1}{n}\sum_i\sum_t[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\psi_g\left(\tilde{\gamma},\widehat{\lambda}_i(\tilde{\gamma})\right)]^2-\widehat{\sigma}_U^2
$$

■