Test of Neglected Heterogeneity in Dyadic Models

Jinyong Hahn* Hyungsik Roger Moon† Ruoyao Shi‡

UCLA USC & Yonsei UC Riverside

February, 2024

Abstract

We develop a Lagrange Multiplier (LM) test of neglected heterogeneity in dyadic models. The test statistic is derived by modifying Breusch and Pagan (1980)'s test. We establish the asymptotic distribution of the test statistic under the null using a novel martingale construction. We also consider the power of the LM test in generic panel models. Even though the test is motivated by random effects, we show that it has a power for detecting fixed effects as well. Finally, we examine how the estimation noise of the maximum likelihood estimator affects the asymptotic distribution of the test under the null, and show that such a noise may be ignored in large samples.

Keywords: Lagrange Multiplier test, dyadic regression model, error component panel regression model, fixed effects, local power.

JEL Classification: C12, C23.

Number of Pages: 74.

*Professor, Department of Economics, University of California Los Angeles, Los Angeles, CA 90095, USA. Email: hahn@econ.ucla.edu.

†Professor, Department of Economics, University of Southern California, Los Angeles, CA 90089, USA and School of Economics, Yonsei University, Seoul, Korea. Email: moonr@usc.edu.

‡Assistant Professor, Department of Economics, University of California Riverside, Riverside, CA 92521, USA. Email: ruoyao.shi@ucr.edu.

1 Introduction

Econometric analysis of dyadic data can be complicated under the presence of unobserved individual effects. If such effects are neglected, it would in general lead to inconsistency of standard estimators such as MLE. This is due to the fact that neglecting unobserved effects generally leads to model misspecification. It is also partly related to the well-known incidental parameter problem in panel data analysis. There do exist versions of dyadic models where standard estimators remain consistent even when unobserved effects are neglected. However, the presence of unobserved effects in the data affects the rate of convergence of standard estimators in a negative way, and the rate of convergence may be slower than what is expected for standard estimators.[1] As such, standard errors based on the presumption that the unobserved effects are present are too conservative if the unobserved effects are in fact absent.

For these reasons, it would be pragmatically useful to examine whether there are any neglected individual effects in dyadic regression models. This paper makes a contribution in this regard by developing a convenient statistical test of neglected heterogenous effects (even after controlling for observed dyadic specific explanatory variables), which is done by modifying Breusch and Pagan (1980)'s test (BP test hereafter). The BP test is a Lagrange Multiplier (LM) test originally developed for panel data analysis to deal with both individual and time effects, but the version of the BP test for detecting only the individual effects seems to have received the most attention.[2] From the viewpoint of dyadic regression models, the most convenient feature of the BP test is perhaps that it is an LM test, and as such, it requires calculation of the parameter estimates only under the null of no unobserved heterogeneity.[3] This feature simplifies the computation and makes it pragmatically very attractive, as the computational problem disappears due to the absence of unobserved heterogeneity under the null.

[1] Examples include Graham (2020) and Menzel (2017).

[2] Such a version of the BP test can be interpreted to be a test of overdispersion Cox (1983), and it is related to White (1982)'s information matrix test, as was pointed out by Chesher (1984). See also Lancaster (1984) for the asymptotic distribution of the test under the null.

[3] See also Engle (1984).

The LM test in dyadic models turns out to be very similar to the sum of the two BP test statistics in panel regressions, one for detection of individual effects and the other for detection of time effects. Therefore, the asymptotic properties of our modified BP test can be characterized by deriving the joint distribution of the two BP test statistics. Honda (1985) analyzed the asymptotic size of the two BP test statistics for *linear* panel regression models. Therefore, in principle, the asymptotic analysis of our test statistics would simply require nonlinear generalization of Honda (1985)'s argument. Unfortunately, we found that some of the arguments in Honda (1985)'s proof is incorrect. It turns out to be quite challenging to correct Honda (1985)'s proof. We overcome the problem by approximating the test statistics with a novel martingale that we construct in the appendix, which we consider to be an important contribution of the paper.

Our endeavor produced a few interesting by-products that are of independent interest in themselves. In addition to characterizing the asymptotic null distribution of the modified BP test in both dyadic regressions and panel regressions, we address the question of power of the BP test in generic panel models.[4] To our knowledge, Honda (1985) is the only one who analyzed the asymptotic power of the BP test, and he did so against *random* effects in *linear* models.

In this paper, we make significant progress over Honda (1985) in two dimensions. First, we derive the asymptotic power of the BP test against random effects in general *nonlinear* models. Second, which is more important, we also consider the power of the BP test against the alternative of *fixed* effects. By fixed effects, we mean the type of general unobserved variables that may have arbitrary dependence structure with the observed explanatory variables.[5] The BP test was specifically designed to detect the alternative of random effects, as is clear in the derivation by Breusch and Pagan (1980) or Chesher (1984). The random effects are by assumption independent of all the observable explanatory variables, so such an alternative may be argued to be restrictive.[6] Our paper fills this gap in the

[4]Throughout this paper, the power is defined to be the *local* power.

[5]Chamberlain (1984) called such a variable the correlated random effects.

[6]If the more general alternative of fixed effects is to be considered, one may adopt a version of the conditional moment restrictions test, as discussed in Hahn et al. (2017).

literature and analyzes the local power of the BP test against the general alternative of fixed effects. Modifying Newey (1985)'s argument,[7] we obtain the asymptotic results, based on which we argue that the BP test in general has a power against the fixed effects.[8]

We also derive the convenient implication that it is unnecessary to adjust for noise in estimation of MLE in two-way models (for characterization of the asymptotic distribution under the null hypothesis) under the asymptotics where both the cross sectional dimension (N) and the time series dimension (T) grow to infinity. This has a convenient implication in the application to dyadic models, but may superficially appear to contradict Lancaster (1984)'s result, which implies that such an adjustment is necessary. We explain that the difference can be explained by the fact that Lancaster (1984)'s analysis was on fixed T and $N \to \infty$ asymptotics.

We recognize that a specification test of the type analyzed in the paper is often associated with the pre-test bias in the usual cross sectional analysis, and we expect the same issue with uniformity in the application to the dyadic/panel data analysis. This is a generic problem for which we are unable to offer a general solution.

The paper is organized as follows. Our results for dyadic models are in Section 2. Section 3 presents results for local power of the BP test in nonlinear panel models. Section 4 presents an argument for why the asymptotic distribution under the null does not need to adjust for noise of estimation of MLE. Finite sample sizes and powers of the LM test in a nonlinear dyadic regression model are in Section 5. Section 6 concludes.

[7] The regularity conditions in Newey (1985) make it easy to obtain results along the line of Le Cam's Third Lemma.

[8] To be more precise, we show that the probability of rejection is higher under the alternative than under the null, i.e., we show that the BP test is locally unbiased.

2 Test of Neglected Heterogeneity in Dyadic Regression Models

In this section, we formulate the null hypothesis of no neglected heterogeneity in dyadic regression frameworks for either directed or undirected dyadic observations, and derive the limiting distributions of the test statistics under the null hypothesis. We also make a connection to the classical BP test against one-way and two-way error component panel data models.

Suppose that $\mathcal{N} := \{1, ..., N\}$ is the set of sample agents. A pair of two different agents constitute a dyad, $(i, j) \in \mathcal{N} \times \mathcal{N}$ with $i \neq j$. Let (Y_{ij}, X_{ij}) denote dyadic observations.

Undirected Dyadic Regression. Suppose that dyadic observations (Y_{ij}, X_{ij}) are undirected, that is, $(Y_{ij}, X_{ij}) = (Y_{ji}, X_{ji})$ for $i, j \in \mathcal{N}$ with $i \neq j$. The random effects likelihood of typical undirected dyadic regression models has a generic representation,

$$\int \cdots \int \prod_{i=1}^{N-1} \prod_{j>i}^{N} f\left(y_{ij} \middle| x_{ij}, \theta_0 + e_i \iota + e_j \iota\right) k\left(e_1\right) \cdots k\left(e_N\right) de_1 \ldots de_N. \tag{1}$$

Here, $f\left(y_{ij} \middle| x_{ij}, \theta_0 + \varepsilon_i \iota + \varepsilon_j \iota\right)$ denotes the marginal likelihood of y_{ij} given the observed explanatory variables x_{ij} as well as unobserved heterogenous effects ε_i and ε_j. The ι is a vector of the same dimension as the parameter of interest θ_0, where the first coordinate is equal to 1 and the rest are 0. Equation (1) captures the gist of the linear model of the form $Y_{ij} = X_{ij}' \theta_0 + \varepsilon_i + \varepsilon_j + v_{ij}$ (but not limited to the linear model), where we assume that the first component of X_{ij} is 1 (i.e., the intercept term), and we understand εs as representing the heterogeneity of the first component of θ_0. (More detailed discussion on the modeling is presented in the one-way error component model.) The density $f\left(\cdot \middle| \cdot\right)$ is then derived from the density of v_{ij}. Finally, the εs are assumed to be independent and identically distributed with density $k\left(\cdot\right)$, which gives rise to the above joint density.

In order to simplify the notation a bit, we rewrite the joint density in (1) as

$$E_\varepsilon \left[\prod_{i=1}^{N-1} \prod_{j>i}^{N} f(y_{ij} | \, x_{ij}, \theta_0 + \varepsilon_i \iota + \varepsilon_j \iota) \right],$$

where $E_\varepsilon[\cdot]$ denotes the expectation with respect to the εs, fixing everything else constant.

The LM test of overdispersion can be understood to be a test against the alternative where the density $k(\cdot)$ is very close to zero, loosely speaking. This is given a more rigorous meaning by considering the alternative parameterization indexed by a scale parameter η,

$$h\left(y|\, x, \theta_0, \sqrt{\eta}\right) := E_\varepsilon\left[\prod_{i=1}^{N-1}\prod_{j>i}^{N} f\left(y_{ij}|\, x_{ij}, \theta_0 + \sqrt{\eta}\varepsilon_i\iota + \sqrt{\eta}\varepsilon_j\iota\right)\right],$$

where the density of εs is fixed, and we test the null hypothesis that the scale parameter is zero, i.e.,

$$\mathbb{H}_0 : \eta = 0. \tag{2}$$

Directed Dyadic Regression If the dyadic observations (Y_{ij}, X_{ij}) are directed so that $(Y_{ij}, X_{ij}) \neq (Y_{ji}, X_{ji})$ for $i, j \in \mathcal{N}$ with $i \neq j$ in general, the likelihood of typical directed dyadic regression models has a generic representation,

$$\begin{aligned}
&\int \cdots \int \prod_{i=1}^{N}\prod_{j\neq i}^{N} f\left(y_{ij}|\, x_{ij}, \theta_0 + e_i\iota + e_j\iota\right) \times k_1(e_1)\cdots k_1(e_N)\, de_1 \ldots de_N \\
&:= E_\varepsilon\left[\prod_{i=1}^{N}\prod_{j\neq i}^{N} f(y_{ij}|\, x_{ij}, \theta_0 + \varepsilon_i\iota + \varepsilon_j\iota)\right], \qquad (3)
\end{aligned}$$

where ε_i and ε_j are unobserved effects of "out" agent i and "in" agent j, respectively.

Like in the undirected case, we reparameterize the likelihood as

$$h\left(y|\, x, \theta_0, \sqrt{\eta}\right) := E_\varepsilon\left[\prod_{i=1}^{N}\prod_{j\neq i}^{N} f\left(y_{ij}|\, x_{ij}, \theta_0 + \sqrt{\eta}\varepsilon_i\iota + \sqrt{\eta}\varepsilon_j\iota\right)\right]$$

and test

$$\mathbb{H}_0 : \eta = 0. \tag{4}$$

2.1 LM Test of Neglected Heterogeneity in Directed Dyadic Regression Models

The LM test of the null hypothesis (4) is based on the score function with the parameter estimate under the null restriction. To derive the LM test statistic in the directed dyadic regression, first we consider the first order derivative of the likelihood function with respect to the parameter η. Define

$$\bar{h}\left(y|\, x, \theta_0, \sqrt{\eta_1}, \sqrt{\eta_2}\right) := E_\varepsilon \left[\prod_{i=1}^{N} \prod_{j\neq i}^{N} f\left(y_{ij}|\, x_{ij}, \theta_0 + \sqrt{\eta_1}\varepsilon_i \iota + \sqrt{\eta_2}\varepsilon_j \iota\right)\right].$$

Then, by definition, we have

$$\left.\frac{\partial h\left(y|\, x, \theta_0, \sqrt{\eta}\right)}{\partial \eta}\right|_{\eta=0} = \left.\frac{\bar{h}\left(y|\, x, \theta_0, \sqrt{\eta_1}, 0\right)}{\partial \eta_1}\right|_{\eta_1=0} + \left.\frac{\partial \bar{h}\left(y|\, x, \theta_0, 0, \sqrt{\eta_2}\right)}{\partial \eta_2}\right|_{\eta_2=0}. \tag{5}$$

To proceed with (5), we take a detour and relate the two terms in (5) to the BP test statistic for overdispersion in one-way error component panel models.

2.1.1 Review of the BP Test in One-Way Error Component Panel Regression Models

We will consider the panel model with possible unobserved individual heterogeneity and present the LM test to detect neglected heterogeneity. For a one-way error component panel model, this is largely a review of Breusch and Pagan (1980) as well as Chesher (1984). Assume that we observe a random sample (Y_i, X_i), $i = 1, \ldots, N$. Y_i and X_i can be vectors. In the panel data analysis where each individual is observed over T time periods, we will have $Y_i = (Y_{i1}, \ldots, Y_{iT})$ and $X_i = (X'_{i1}, \ldots, X'_{iT})'$. Note that the first component of X_{it} be 1 (the intercept). We let X^*_{it} denotes X_{it} excluding the intercept and define X^*_i accordingly. We assume that the conditional density of Y_i given X_i is given by the function $f\left(y|\, x, \theta\right)$, where θ is a q-dimensional parameter that characterizes the density. Under the null hypothesis, the first component θ_1 of θ is fixed at $\theta_{0,1}$, but under the alternative hypothesis, it may be a random variable indexed by i. This is motivated by the linear model with one-way error component

$$Y_{it} = X^{*\prime}_{it}\beta + \alpha_i + v_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where α_i denotes the unobserved individual heterogeneity. Suppose that $v_{it} \sim N\left(0, \sigma_v^2\right)$ is independent of $(X^{*\prime}_i, \alpha_i)$, and is independent over i and t. We can then understand that the parameter $\theta = (\alpha_i, \beta, \sigma_v^2)'$ may be different across different individuals. Note that we assume that only the first component θ_1 of θ is allowed to be different across i, i.e., scalar random (or fixed) effects.[9] The heterogeneity of the first component θ_1 of θ can be modeled as $\theta_{0,1}$ plus a random variable ε_i. Under the random effects specification, the heterogeneity is independent of X_i, and therefore the conditional density of the heterogeneity given $X_i = x$ is equal to the marginal density. Under the random effects approach, it is also common to assume that the expectation of the heterogeneity is zero. In order to accentuate the local nature of the alternative, we may choose to write $\theta_{1,i} = \theta_{0,1} + \eta\varepsilon_i$, where $E[\varepsilon_i] = 0$ and $\eta \geq 0$ is a "small" number and the conditional density of ε_i given $X_i = x_i$ is $k(\cdot)$. The conditional density of Y_i given $X_i = x_i$ and $\varepsilon_i = e_i$ is then equal to $f(y_i|\, x_i, \theta_0 + \eta e_i \iota) = f(y_i|\, x_i, (\theta_{0,1} + \eta e_i, \theta_{0,2}, \ldots, \theta_{0,q}))$. It follows that the conditional density of Y_i given $X_i = x_i$ is

$$h(y_i|\, x_i, \theta_0, \eta) := E_\varepsilon \left[f(y_i|\, x_i, \theta_0 + \eta\varepsilon_i \iota) \right],$$

where the expectation is taken with respect to the distribution of ε. Note that $h(y_i|\, x_i, \theta, 0) = f(y_i|\, x_i, \theta)$. We consider the second order Taylor series expansion of $h(y_i|\, x_i, \theta, \eta)$ with respect to η around $(\theta, \eta) = (\theta_0, 0)$.

Under the assumption that we can exchange differentiation and integration, we obtain

$$\begin{aligned} \left.\frac{\partial h(y_i|\, x_i, \theta_0, \eta)}{\partial \eta}\right|_{\eta=0} &= E_\varepsilon \left[\frac{\partial f(y_i|\, x_i, \theta_0)}{\partial \theta_1} \varepsilon_i \right] = \frac{\partial f(y_i|\, x_i, \theta_0)}{\partial \theta_1} E[\varepsilon_i] = 0, \\ \left.\frac{\partial^2 h(y_i|\, x_i, \theta_0, \eta)}{\partial \eta^2}\right|_{\eta=0} &= E_\varepsilon \left[\frac{\partial^2 f(y_i|\, x_i, \theta_0)}{\partial \theta_1^2} \varepsilon_i^2 \right] = \frac{\partial^2 f(y_i|\, x_i, \theta_0)}{\partial \theta_1^2} E\left[\varepsilon_i^2\right]. \end{aligned} \tag{6}$$

Therefore, we have

$$h(y_i|\, x_i, \theta_0, \eta) = h(y_i|\, x_i, \theta_0, 0) + \frac{\eta^2}{2} \frac{\partial^2 f(y_i|\, x_i, \theta_0)}{\partial \theta_1^2} \sigma_\varepsilon^2 + o\left(\eta^2\right),$$

where $\sigma_\varepsilon^2 = E[\varepsilon_i^2]$. Given the form of the expansion, it would make sense to consider the parameterization $h\left(y_i|\, x_i, \theta_0, \sqrt{\eta}\right)$ instead (i.e., $E_\varepsilon\left[f\left(y_i|\, x_i, \theta_0 + \sqrt{\eta}\varepsilon_i\iota\right)\right]$), which delivers

[9] There is no reason that the LM test should be confined to the scalar effects, as is evident from Chesher (1984)'s derivation. On the other hand, the scalar effects are a common feature in many panel data analysis, and were the basis of the LM test as was presented in Breusch and Pagan (1980).

the expansion[10]

$$h\left(y_i|\,x_i,\theta_0,\sqrt{\eta}\right)=h\left(y_i|\,x_i,\theta_0,0\right)+\frac{\eta}{2}\frac{\partial^2 f\left(y_i|\,x_i,\theta_0\right)}{\partial\theta_1^2}\sigma_\varepsilon^2+o\left(\eta\right).$$

This implies that

$$\left.\frac{\partial h\left(y_i|\,x_i,\theta_0,\sqrt{\eta}\right)}{\partial\eta}\right|_{\eta=0}=\lim_{\eta\to 0}\frac{h\left(y_i|\,x_i,\theta_0,\sqrt{\eta}\right)-h\left(y_i|\,x_i,\theta_0,0\right)}{\eta}=\frac{1}{2}\frac{\partial^2 f\left(y_i|\,x_i,\theta_0\right)}{\partial\theta_1^2}\sigma_\varepsilon^2. \quad (7)$$

Now, we consider the joint conditional density of the entire data

$$h\left(y|\,x,\theta_0,\eta\right):=E_\varepsilon\left[\prod_{i=1}^{N}f\left(y_i|\,x_i,\theta_0+\eta\varepsilon_i\iota\right)\right],$$

and consider the form of the LM test statistic. It is straightforward to show that

$$\left.\frac{\partial h\left(y|\,x,\theta_0,\eta\right)}{\partial\eta}\right|_{\eta=0}=0,\quad \left.\frac{\partial^2 h\left(y|\,x,\theta_0,\eta\right)}{\partial\eta^2}\right|_{\eta=0}=\prod_{j=1}^{N}f\left(y_j|\,x_j,\theta_0\right)\left(\sum_{i=1}^{N}\frac{\partial^2 f\left(y_i|\,x_i,\theta_0\right)/\partial\theta_1^2}{f\left(y_i|\,x_i,\theta_0\right)}\right)\sigma_\varepsilon^2.$$

Therefore, it follows from a similar Taylor expansion argument that in the random effect one-way error component panel model (not necessarily linear), the LM test can be based on the score

$$\left.\frac{\partial h\left(y|\,x,\theta_0,\sqrt{\eta}\right)/\partial\eta}{h\left(y|\,x,\theta_0,0\right)}\right|_{\eta=0}=\frac{1}{2}\left(\sum_{i=1}^{N}\frac{\partial^2 f\left(y_i|\,x_i,\theta_0\right)/\partial\theta_1^2}{f\left(y_i|\,x_i,\theta_0\right)}\right)\sigma_\varepsilon^2,$$

or equivalently based on [11]

$$BP_{1way}:=\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\right)^2. \quad (8)$$

BP_{1way} is the same as the statistic in (12) below, and Appendix C gives the justification of the normalization by $\sqrt{N}T$.

[10]Chesher (1984) directly worked with $h\left(y|\,x,\theta_0,\sqrt{\eta}\right)$ and applied L'Hôpital's rule. The Taylor expansion adopted here makes it easier to understand the role of the zero mean assumption, i.e., $E\left[\varepsilon_i\right]=0$.

[11]Because $\sum_{t=1}^{T}\frac{\partial^2\ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1^2}+\left(\sum_{t=1}^{T}\frac{\partial\ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)^2$ has a zero expectation under correct specification, the LM statistic in (8) has the information matrix test interpretation. See Chesher (1984).

2.1.2 Back to the LM Test of Neglected Heterogeneity in Directed Dyadic Regression Models

We now get back to (5),

$$\left.\frac{\partial h\left(y|\,x,\theta_0,\sqrt{\eta}\right)}{\partial\eta}\right|_{\eta=0}=\left.\frac{\bar{h}\left(y|\,x,\theta_0,\sqrt{\eta_1},0\right)}{\partial\eta_1}\right|_{\eta_1=0}+\left.\frac{\partial\bar{h}\left(y|\,x,\theta_0,0,\sqrt{\eta_2}\right)}{\partial\eta_2}\right|_{\eta_2=0}$$
$$=I+II,\ \text{say.}$$

Notice that the terms I and II correspond to the scores of the one-way error component panel regression model. Therefore, in view of (8), we deduce that the LM test statistic of the null hypothesis (4) in the directed dyadic regression model is the sum of

$$\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\partial^2\ln f\left(Y_{ij}|\,X_{ij},\theta_0\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\left(\sum_{j\neq i}^{N}\frac{\partial\ln f\left(Y_{ij}|\,X_{ij},\theta_0\right)}{\partial\theta_1}\right)^2 \tag{9}$$

and

$$\frac{1}{\sqrt{N}N}\sum_{j=1}^{N}\sum_{i\neq j}^{N}\frac{\partial^2\ln f\left(Y_{ij}|\,X_{ij},\theta_0\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{N}N}\sum_{j=1}^{N}\left(\sum_{i\neq j}^{N}\frac{\partial\ln f\left(Y_{ij}|\,X_{ij},\theta_0\right)}{\partial\theta_1}\right)^2; \tag{10}$$

that is,

$$LM_d:=(9)+(10). \tag{11}$$

2.1.3 Comparison with the BP test in Two-Way Error Component Panel Regression Models

Notice that the directed dyadic regression model (3) and a two-way error component panel model are similar. A widely applied two-way error component panel regression model includes both individual and time effects as

$$Y_{it}=X_{it}^{*\prime}\beta+\alpha_i+\gamma_t+v_{it},\quad i=1,\ldots,N,\ t=1,\ldots,T.$$

Compared to (3), the two-way panel model pairs cross-section indexed by i and time series indexed by t. The dimensions of cross section (N) and time series (T) are different in general, and if the panel is balanced, we observe all the pairs (Y_{it},X_{it}). Similar to the one-way panel model, we can understand that $\theta = (\alpha_i + \gamma_t, \beta, \sigma_v^2)$. The heterogeneity of the first component θ_1 of θ can then be modeled as $\theta_{0,1}$ plus a random variable ε_{1i} that differs across i, as well as another random variable ε_{2t} that varies over t. If we assume that the joint conditional density of all Y_{it} given all X_{it}, ε_{1i} and ε_{2t} is

$$\bar{h}\left(y\right|x, \theta_0, \eta_1, \eta_2) := E_\varepsilon\left[\prod_{i=1}^{N}\prod_{t=1}^{T} f\left(y_{it}\right| x_{it}, \theta_0 + \eta_1\varepsilon_{1i}\iota + \eta_2\varepsilon_{2t}\iota)\right],$$

then we recognize that the BP test statistic can be obtained by separately differentiating with respect to η_1 and η_2. This implies that the BP test statistic for two-way error components is a two-dimensional vector, with the first component being BP_{1way} defined in (8)

$$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2 \ln f\left(Y_{it}\right| X_{it}, \theta_0)}{\partial\theta_1^2} + \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial \ln f\left(Y_{it}\right| X_{it}, \theta_0)}{\partial\theta_1}\right)^2 \tag{12}$$

and the second component being the counterpart of BP_{1way} when the alternative one-way model under consideration contains only time effects

$$\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial^2 \ln f\left(Y_{it}\right| X_{it}, \theta_0)}{\partial\theta_1^2} + \frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\frac{\partial \ln f\left(Y_{it}\right| X_{it}, \theta_0)}{\partial\theta_1}\right)^2. \tag{13}$$

That is,

$$BP_{2way} := ((12), (13))'. \tag{14}$$

See Appendix C for justification of the normalization by $\sqrt{N}T$ and $N\sqrt{T}$. Therefore, our test statistic LM_d for the dyadic model in (11) is almost equal to the sum of the two components of BP_{2way} for the special case $N = T$.[12]

2.1.4 Asymptotic Distribution of the LM test statistic LM_d in (11)

In this subsection, we discuss how to establish the asymptotic theory of the LM test statistic LM_d in (11). Comparing LM_d in the directed dyadic regression model and the

[12]By the same token, a directed dyadic model where each individual has two fixed effects (e.g., each country has an exporter effect ε_{i1} and an importer effect ε_{i2}) is mathematically identical to the panel model with time and individual fixed effects, with $N = T$ and the $i = t$ observations excluded. Alternatively, one can think of a particular form of heterogeneity $\varepsilon_i = \mathbb{I}\{exporter\}_i * \varepsilon_{i1} + \mathbb{I}\{importer\}_i * \varepsilon_{i2}$, and our analysis in this paper remains unchanged.

sum of the two components of the traditional BP test statistic BP_{2way} in (14) in the two-way error component panel model, we find that the two are almost equal to each other, except that LM_d only considers the $i \neq j$ terms (i.e., the $i \neq t$ terms in the panel model) with $N = T$.

The traditional BP test in the two-way error component linear panel regression model was first proposed by Breusch and Pagan (1980), and then subsequently Honda (1985) derived the asymptotic distribution of the test statistic. Given these existing studies and the closeness between the two test statistics mentioned above, one may think that the asymptotic null distribution of the LM test statistic (11) can be easily derived using the result in Honda (1985)'s. Unfortunately, it is not the case.

The problem with Honda (1985)'s analysis is that in proving the joint asymptotic distribution of (12) and (13), or more precisely the counterparts of (12) and (13) in linear models, Honda (1985, Lemma 2) asserts that if two uncorrelated sequences of random variables both converge to normal distributions, then they jointly converge to a bivariate normal distribution. This argument is incorrect.[13] Fortunately, his result can be saved, although it turned out to be quite challenging. In the rest of this subsection, we will present a correct argument for the asymptotic independence by using a novel martingale construction and applying a martingale central limit theorem.

To make the discussion above concrete, it is convenient to rewrite the statistic in (12) as

$$\begin{aligned} &\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(\frac{\partial^2 \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1^2}+\left(\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\right)\right) \\ &+\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\frac{\partial \ln f\left(Y_{is}|\, X_{is},\theta_0\right)}{\partial\theta_1}. \end{aligned} \tag{15}$$

Under the null, because $\frac{\partial^2 \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1^2}+\left(\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)^2$ is independent over i and t with mean zero, the first term in (15) has mean zero and variance of order $\frac{1}{NT^2}O\left(NT\right) = O\left(T^{-1}\right)$. Therefore, the statistic in (12) is asymptotically equivalent to the second term $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\frac{\partial \ln f(Y_{is}|X_{is},\theta_0)}{\partial\theta_1}$ in (15) under the null. Likewise, the statis-

[13]See Kuersteiner and Prucha (2013)'s Example 1 for related discussion.

tic in (13) is asymptotically equivalent to $\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\sum_{i,j=1,i\neq j}^{N}\frac{\partial\ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\frac{\partial\ln f(Y_{jt}|X_{jt},\theta_0)}{\partial\theta_1}$ under the null.

For simplicity of notation, let U_{it} denote $\frac{\partial\ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}$ for $i=1,\ldots,N$ and $t=1,\ldots,T$. So, our objective is to analyze the joint asymptotic distribution of $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}$ and $\frac{1}{N\sqrt{T}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt}$. Honda (1985) argued that these two statistics are uncorrelated, and drew the conclusion that they are asymptotically independent, which is the mistake that we will fix by constructing a novel martingale structure.[14]

We use the Cramér-Wold theorem and define

$$A_{NT}:=\varrho\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}+\frac{1}{N\sqrt{N}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt}. \qquad (16)$$

Without loss of generality, we will assume that $|\varrho|<\infty$. The $|\varrho|=\infty$ case is where we are only interested in the marginal distribution of $\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}$, which can be established using symmetry by considering the $\varrho=0$ case.[15]

Lemma 1 *Suppose that U_{it} are iid*[16] *with variance σ_U^2 and a finite fourth moment across i,t. Suppose that $N,T\to\infty$ with $\frac{N}{T}\to\kappa$ and $0<\kappa<\infty$. We then have*

$$A_{NT}\Rightarrow N\left(0,2\left(\frac{1}{\kappa}+\frac{\varrho^2}{\kappa^2}\right)\sigma_U^4\right).$$

Proof. In Appendix A.2. ■

Lemma 2 *Suppose that U_{it} are iid with variance σ_U^2 and a finite fourth moment across i,t. Suppose that $N,T\to\infty$ with $\frac{N}{T}\to\kappa$ and $0<\kappa<\infty$. We then have*

$$\begin{pmatrix}\frac{1}{N\sqrt{T}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt}\\ \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}\end{pmatrix}\Rightarrow N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}2\sigma_U^4 & 0\\ 0 & 2\sigma_U^4\end{bmatrix}\right).$$

[14] The construction is technical in nature, and can be found in Appendix A.1.

[15] The two cases $\varrho=0$ and $|\varrho|=\infty$ correspond to the first half of Lemma 2 in Honda (1985), which was correctly proven.

[16] The iid assumption will be satisfied if $(X_{i,t},Y_{i,t})$ are iid. The iid assumption is sufficient but not necessary for validity of our martingale structure, which is a crucial component in our proof. As long as Xs are strictly exogenous, the martingale will continue to be valid, so we do not need that $(X_{i,t},Y_{i,t})$ are iid. On the other hand, the iid assumption facilitates other technical analysis, which will be complicated when the iid assumption is violated.

Proof. In Appendix A.3. ■

Comparison between the two statistics in Lemma 2 (for the directed dyadic model if $N = T$) and the two statistics in (12) and (13) for the panel model suggests that the only difference between them is that the test statistics for the directed dyadic model only considers the $i \neq j$ terms (i.e., the $i \neq t$ terms in the panel model). It turns out to be the case that the deletion of the (i, i) observations has an asymptotically negligible effect. This is formally summarized in the following theorem, which is our main result for the directed dyadic model.

Theorem 1 *Suppose that U_{ij} are iid with variance σ_U^2 and a finite fourth moment across i, j. We then have*

$$LM_d \Rightarrow N\left(0, 4\sigma_U^4\right).$$

Proof. In Appendix A.4. ■

Remark 1 *The analysis in Section 4 implies that the result in Theorem 1 carries over to the feasible version of LM_d where θ_0 is estimated. To be more precise, let $\hat{\sigma}_U^2$ and $\bar{\theta}_N$ denote a consistent estimator of σ_U^2 and the MLE of θ_0, then we have $\left(l_N(\bar{\theta}_N)\right)^2 / 4\hat{\sigma}_U^4 \Rightarrow \chi_1^2$ under the null hypothesis of no unobserved heterogeneity, where*

$$l_{1N}(\bar{\theta}_N) := \frac{1}{\sqrt{N}N} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{\partial^2 \ln f\left(Y_{ij} | X_{ij}, \bar{\theta}_N\right)}{\partial \theta_1^2} + \frac{1}{\sqrt{N}N} \sum_{i=1}^{N} \left(\sum_{j \neq i}^{N} \frac{\partial \ln f\left(Y_{ij} | X_{ij}, \bar{\theta}_N\right)}{\partial \theta_1} \right)^2,$$

$$l_{2N}(\bar{\theta}_N) := \frac{1}{\sqrt{N}N} \sum_{j=1}^{N} \sum_{i \neq j}^{N} \frac{\partial^2 \ln f\left(Y_{ij} | X_{ij}, \bar{\theta}_N\right)}{\partial \theta_1^2} + \frac{1}{\sqrt{N}N} \sum_{j=1}^{N} \left(\sum_{i \neq j}^{N} \frac{\partial \ln f\left(Y_{ij} | X_{ij}, \bar{\theta}_N\right)}{\partial \theta_1} \right)^2.$$

and $l_N(\bar{\theta}_N) := l_{1N}(\bar{\theta}_N) + l_{2N}(\bar{\theta}_N)$.

2.2 LM Test of Neglected Heterogeneity in Undirected Dyadic Regression Models

For undirected dyadic regression, $Y_{ij} = Y_{ji}$ and $X_{ij} = X_{ji}$, and we can write the joint conditional density of all Y_{ij} given all X_{ij} and ε_i as

$$\bar{h}\left(y | x, \theta_0, \sqrt{\eta_1}, \ldots, \sqrt{\eta_N}\right) := E_{\varepsilon}\left[\prod_{i=1}^{N-1} \prod_{j>i}^{N} f\left(y_{i,j} | x_{i,j}, \theta_0 + \sqrt{\eta_i}\varepsilon_i \iota + \sqrt{\eta_j}\varepsilon_j \iota\right)\right].$$

Similar to the derivation of (5) and (11), we recognize that the LM test statistic can be obtained by taking separate first order differentiation with respect to $\eta_1 \dots \eta_N$ and then adding them together. In particular, we see that our LM test statistic would be

$$LM_{ud} := \frac{1}{N\sqrt{N}} \sum_{i=1}^{N} \left(\sum_{j \neq i}^{N} \frac{\partial^2 \ln f\left(Y_{ij} \middle| X_{ij}, \theta_0\right)}{\partial \theta_1^2} + \left(\sum_{j \neq i}^{N} \frac{\partial \ln f\left(Y_{ij} \middle| X_{ij}, \theta_0\right)}{\partial \theta_1} \right)^2 \right), \qquad (17)$$

which is almost identical to the BP test statistic in the one-way error component panel model, as shown in (8), except that we are using j instead of t and excluding the terms where $i = j$. Because $Y_{ij} = Y_{ji}$, the martingale needs to be constructed more carefully. For this purpose, we rewrite the statistic in (17) as

$$\begin{aligned} LM_{ud} = {} & \frac{2}{N\sqrt{N}} \sum_{i=1}^{N} \sum_{j>i}^{N} \left(\frac{\partial^2 \ln f\left(Y_{ij} \middle| X_{ij}, \theta_0\right)}{\partial \theta_1^2} + \left(\frac{\partial \ln f\left(Y_{ij} \middle| X_{ij}, \theta_0\right)}{\partial \theta_1} \right)^2 \right) \\ & + \frac{1}{N\sqrt{N}} \sum_{i=1}^{N} \left(\sum_{j \neq i}^{N} \frac{\partial \ln f\left(Y_{ij} \middle| X_{ij}, \theta_0\right)}{\partial \theta_1} \left(\sum_{j' \neq i, j' \neq j}^{N} \frac{\partial \ln f\left(Y_{ij'} \middle| X_{ij'}, \theta_0\right)}{\partial \theta_1} \right) \right) \end{aligned}$$

The first term on the right is a sum of $N(N-1)/2$ iid random variables with zero mean under the null, so its variance is of order $\left(\frac{1}{N\sqrt{N}}\right)^2 O\left(N(N-1)\right) = o(1)$. Therefore, the test statistic is asymptotically equivalent to the second term under the null; that is

$$LM_{ud} = \frac{1}{N\sqrt{N}} \sum_{i=1}^{N} \left(\sum_{j \neq i}^{N} \frac{\partial \ln f\left(Y_{ij} \middle| X_{ij}, \theta_0\right)}{\partial \theta_1} \left(\sum_{j' \neq i, j' \neq j}^{N} \frac{\partial \ln f\left(Y_{ij'} \middle| X_{ij'}, \theta_0\right)}{\partial \theta_1} \right) \right) + o_p(1). \qquad (18)$$

We let $U_{ij} = \frac{\partial \ln f\left(Y_{ij} | X_{ij}, \theta_0\right)}{\partial \theta_1}$ as before if $i \neq j$, and define $U_{ij} = 0$ if $i = j$. We then write the main term in (18) as

$$A_N = \frac{1}{N\sqrt{N}} \sum_{i=1}^{N} \sum_{j, j'=1, j \neq j'}^{N} U_{ij} U_{ij'},$$

so it superficially resembles A_{NT}, the term that we analyzed in Lemma 1, except that $U_{ij} = U_{ji}$ in the current setup.

Lemma 3 *Suppose U_{ij} are iid with variance σ_U^2 and a finite fourth moment across i, j ($i \neq j$). We then have*

$$A_N \Rightarrow N\left(0, 2\sigma_U^4\right).$$

Proof. In Appendix A.5. ■

Then, from (18), we have the following theorem.

Theorem 2 *Suppose U_{ij} are iid with variance σ_U^2 and a finite fourth moment across i, j ($i \neq j$). We then have*

$$LM_{ud} \Rightarrow N\left(0, 2\sigma_U^4\right).$$

3 Power of the BP Test

We now analyze the power aspect of the LM test. Given that the LM test statistic in dyadic models is asymptotically equivalent to the sum of two components (12) and (13) of the traditional BP test statistic in panel models with two-way error components and that (12) equals BP_{1way} defined in (8), it suffices to consider the power properties of BP_{1way}, the BP test statistic in panel models with one-way error components.[17] So, in this section, we focus on the analysis of the power of the BP test in generic one-way error component panel models. In order to understand the power properties better, we adopt the fixed T approach. After all, with large T, the power cannot decrease. We consider two kinds of alternatives. First, we consider the random effects, where the individual effects are assumed to be independent of the explanatory variable X. The derivation in (6) shows that the LM test was motivated by the random effects assumption, since the density of ε there does not depend on X. Second, we consider the fixed effects, where the conditional distribution of the individual effects on the explanatory variable X may depend on the realization of X. If we allow arbitrary conditional density $k(\cdot|\, x)$ of ε_i given $X_i = x_i$, which is appropriate under the fixed effects specification, we would change the derivation

[17] Lemma 2 shows that (12) and (13), the two components of BP_{2way}, are asymptotically independent. The two components of LM_d, (9) and (10), therefore, are also asymptotically independent as shown in the proof of Theorem 1. Since their asymptotic distributions are contiguous (verified but not shown), it suffices to focus on BP_{1way}. (The asymptotic bias of LM_d, the source of the power, is simply two times the asymptotic bias of BP_{1way}.)

in (6) to

$$\left.\frac{\partial h\left(y_i|\, x_i,\theta_0,\eta\right)}{\partial\eta}\right|_{\eta=0} = \int \frac{\partial f\left(y_i|\, x_i,\theta_0\right)}{\partial\theta_1} ek\left(e|\, x_i\right) de = \frac{\partial f\left(y_i|\, x_i,\theta_0\right)}{\partial\theta_1}\int ek\left(e|\, x_i\right) de$$
$$= \frac{\partial f\left(y_i|\, x_i,\theta_0\right)}{\partial\theta_1}\mu\left(x_i\right),$$

where we define $\mu\left(x\right) := E\left[\varepsilon_i|\, X_i = x\right]$. Because one can consider arbitrary specification of $\mu\left(x\right)$, the score test that is against all possible specifications of the fixed effects would test whether the equality

$$E\left[\frac{\partial f\left(Y_i|\, X_i,\theta_0\right)/\,\partial\theta_1}{f\left(Y_i|\, X_i,\theta_0\right)}\mu\left(X_i\right)\right] = 0 \tag{19}$$

holds for all $\mu\left(X_i\right)$.[18] Because the score test is equivalent to an infinite number of unconditional moment restrictions (19), a practitioner needs to confront the resultant statistical complications.[19] Our objective is not to develop different tests for different alternatives. Rather, we would like to examine the power properties of the BP test against general alternatives including fixed effects, even though BP_{1way} in (8) was initially developed to detect the random effects. This is an interesting question for practice because the BP test is relatively simple to implement; it is an LM test, and therefore, it suffices to estimate the parameters under the null hypothesis of no neglected heterogeneity, which may be very convenient computationally. Therefore, one may ask the question of whether the BP test can actually detect the fixed effects, even though the fixed effects were not the initial target and therefore, the BP test is not expected to be as powerful as the test of infinite number of moment restrictions when $\mu(\cdot) \neq 0$. We show that the BP test has a power even against the fixed effects, although its power may be weak in linear models.

3.1 Power of the BP Test Against Random Effects

We begin with the random effects. Even though it seems to be such an elementary question, we have not found a literature that deals with the power of the BP test in general nonlinear

[18] This prompted Hahn et al. (2017) to conclude that any test of the conditional moment restriction $E\left[\left.\frac{\partial f(Y_i|X_i,\theta_0)/\partial\theta_1}{f(Y_i|X_i,\theta_0)}\right| X_i\right] = 0$ can be a possible test of fixed effects.

[19] A pragmatic choice would be to work with $\mu(X_i) = X_i$, an approach similar to the generic test of over-identification considered by Chamberlain (1984, Section 4).

models, an obvious gap in the literature if our library research is correct. Using (8), we can see that our BP statistic can be written as $m_N\left(\bar{\theta}_N\right)$, where

$$m_N(\theta) := N^{-1}\sum_{i=1}^{N} m(Z_i,\theta), \quad m(z,\theta) := \frac{\partial^2 f(y|\, x,\theta)/\,\partial\theta_1^2}{f(y|\, x,\theta)}, \tag{20}$$

z denotes the observed data vector, and $\bar{\theta}_N$ denotes the MLE of θ under the null hypothesis of no unobserved heterogeneity. The local power can be analyzed by deriving the asymptotic distribution under the appropriate sequence of DGP's under the alternative of random effects. Newey (1985)'s analysis is almost tailor-made for our purpose, which we adopt as the main tool of analysis.[20] Minor differences do exist. For example, the discussion in the previous section suggests that the local power analysis should be conducted by examining the $N^{1/4} = \sqrt{N^{1/2}}$-neighborhood, i.e., by examining the local alternatives of the form $\theta_{1,i} = \theta_{0,1} + N^{-1/4}\varepsilon_i$.[21] As a result, we will consider the local alternatives of random effects where (i) $\theta_{1,i} = \theta_{0,1} + N^{-1/4}\varepsilon_i$, (ii) ε_i is independent of X_i; (iii) $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = \sigma_\varepsilon^2$.

Next theorem gives the local power property of the BP test against the alternatives of random effects.

Theorem 3 *Under Assumptions 1 - 7 detailed in Appendix B.1, we get*

$$\sqrt{N} m_N\left(\bar{\theta}_N\right) \Rightarrow N\left(\frac{\sigma_\varepsilon^2}{2}\left(\kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2\right), \kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2\right),$$

where

$$\begin{aligned} s(z,\theta_0) &:= \frac{\partial f(y|\, x,\theta_0)/\,\partial\theta}{f(y|\, x,\theta_0)}, \\ \mathcal{I} &:= E\left[s(Z_i,\theta_0)\, s(Z_i,\theta_0)'\right] = -E\left[\frac{\partial s(Z_i,\theta_0)}{\partial\theta'}\right], \\ \kappa_1 &:= E\left[\left(\frac{\partial^2 f(Y_i|\, X_i,\theta_0)/\,\partial\theta_1^2}{f(Y_i|\, X_i,\theta_0)}\right)^2\right], \end{aligned}$$

[20] In this section, we also adopt Newey (1985)'s notation wherever is possible.

[21] Honda (1985) worked with a case that includes both the individual and time effects, and assumed that $N, T \to \infty$ at the same rate. Because we are working with models without time effects and with fixed T, it is a little difficult to make a direct comparison.

$$\kappa_2 := E\left[\frac{\partial^2 f\left(Y_i|\, X_i, \theta_0\right)/\,\partial \theta_1^2}{f\left(Y_i|\, X_i, \theta_0\right)} s\left(Z_i, \theta_0\right)\right].$$

Proof. In Appendix B.2. ■

Theorem 3 implies that (i) we could use a squared standardized BP test statistic

$$\frac{\left(\sqrt{N} m_N\left(\bar{\theta}_N\right)\right)^2}{\hat{\kappa}_1 - \hat{\kappa}_2' \hat{\mathcal{I}}^{-1} \hat{\kappa}_2}, \tag{21}$$

where $\hat{\kappa}_1, \hat{\kappa}_2, \hat{\mathcal{I}}$ are consistent estimators of $\kappa_1, \kappa_2, \mathcal{I}$; and (ii) its asymptotic distribution under $\theta_{1,i} = \theta_{0,1} + N^{-1/4}\varepsilon_i$ is a non-central χ_1^2 distribution with noncentrality parameter $\left(\frac{\sigma_\varepsilon^2}{2}\right)^2 (\kappa_1 - \kappa_2' \mathcal{I}^{-1}\kappa_2)$. Note that $\kappa_1 - \kappa_2' \mathcal{I}^{-1}\kappa_2$ can be interpreted to be the variance of the residual when $m\left(Z_i, \theta_0\right)$ is regressed on $s\left(Z_i, \theta_0\right)$. Unless such residual variance is equal to zero, we should expect that the BP test would have a power against the random effects, i.e., the probability of rejection is higher under the alternative than under the null. Given that $\kappa_1 - \kappa_2' \mathcal{I}^{-1}\kappa_2$ is equal to the asymptotic variance of $\sqrt{N} m_N\left(\bar{\theta}_N\right)$, we can conclude that such a pathological anomaly as $\kappa_1 - \kappa_2' \mathcal{I}^{-1}\kappa_2 = 0$ should not be expected in practice.

As an example, let's consider a panel logit model where

$$Y_{it} = \mathbb{I}\{X_{it}^{*\prime}\beta_0 + \alpha_{N,i} + v_{it} \geq 0\}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{22}$$

where $\alpha_{N,i} = \alpha_0$ under the null and $\alpha_{N,i} = \alpha_0 + N^{-1/4}\varepsilon_i$ under the alternative. Let $\theta = (\alpha_i, \beta')'$, and assume that v_{it} are errors such that the log conditional density of Y_i given X_i and θ is characterized by

$$\ln f\left(Y_i|\, X_i, \theta\right) = \sum_{t=1}^{T}\left(Y_{it} \ln \frac{\exp(X_{it}'\theta)}{1 + \exp(X_{it}'\theta)} + (1 - Y_{it}) \ln \frac{1}{1 + \exp(X_{it}'\theta)}\right).$$

Let $\Lambda_{it}\left(\theta\right) := \exp\left(X_{it}'\theta\right)/\left(1 + \exp\left(X_{it}'\theta\right)\right)$ and note that $\partial \Lambda_{it}/\,\partial\theta = \Lambda_{it}(1 - \Lambda_{it})X_{it}$. Then we have

$$s\left(Y_i|\, X_i, \theta_0\right) = \sum_{t=1}^{T}(Y_{it} - \Lambda_{it})X_{it},$$

$$\frac{\partial^2 \ln f\left(Y_i|\, X_i, \theta_0\right)}{\partial\theta\partial\theta'} = -\sum_{t=1}^{T}\Lambda_{it}(1 - \Lambda_{it})X_{it}X_{it}'.$$

We can see that the BP test statistic is based on

$$
\begin{aligned}
m(Z_i, \theta_0) &:= \frac{\partial^2 \ln f(Y_i|X_i, \theta_0)}{\partial \theta_1^2} + \left(\frac{\partial \ln f(Y_i|X_i, \theta_0)}{\partial \theta_1}\right)^2 \\
&= -\sum_{t=1}^{T} \Lambda_{it}(1-\Lambda_{it}) + \left(\sum_{t=1}^{T}(Y_{it} - \Lambda_{it})\right)^2,
\end{aligned}
$$

and if we assume that Y_{it} and Y_{is} $(t \neq s)$ are independent given X_i, then we have

$$
\begin{aligned}
\mathcal{I} &= E\left[\sum_{t=1}^{T} \Lambda_{it}(1-\Lambda_{it}) X_{it} X_{it}'\right], \\
\kappa_1 &= E\left[\sum_{t=1}^{T} \Lambda_{it}(1-\Lambda_{it})(1-2\Lambda_{it})^2 + 2\sum_{t,s,=1,t\neq s}^{T} \Lambda_{it}(1-\Lambda_{it})\Lambda_{is}(1-\Lambda_{is})\right], \\
\kappa_2 &= E\left[\sum_{t=1}^{T} \Lambda_{it}(1-\Lambda_{it})(1-2\Lambda_{it}) X_{it}\right].
\end{aligned}
$$

3.1.1 Discussion

Remark 2 *Under the null, we can take* $\sigma_\varepsilon^2 = 0$*, so the asymptotic null distribution is* $N(0, \kappa_1 - \kappa_2' \mathcal{I}^{-1} \kappa_2)$*, which justifies the test statistic (21). See also Lancaster (1984).*

Remark 3 *Note that* κ_1 *is equal to the variance of* $m(z, \theta_0)$ *under the null. Therefore, the component* $-\kappa_2' \mathcal{I}^{-1} \kappa_2$ *represents the noise of estimating the MLE* $\bar{\theta}_n$ *as part of the test statistic. It turns out that the linear panel model is a special case where* $\kappa_2 = 0$*, and the test statistic does not need to be adjusted for the noise of estimating the MLE/OLS. See Appendix B.4.*

Remark 4 *The* κ_2 *has yet another interpretation. If* $\kappa_2 = 0$*, the MLE is asymptotically unbiased even under the alternative of random effects, as discussed in Remark 10 in Appendix B.2. In other words, if the statistic* $\frac{\partial^2 f(Y_i|X_i,\theta_0) / \partial \theta_1^2}{f(Y_i|X_i,\theta_0)}$ *is uncorrelated with the score* $s(Y_i|X_i, \theta_0)$*, the MLE is not affected under the alternatives of random effects. Note that* κ_2 *is identical to the numerator of the bias formula in panel data analysis as discussed in Hahn and Newey (2004, p.1315).*[22] *If such a diagnostic test is desired, one can test the*

[22]See the bias formula involving V_{2it} in the second to last displayed equation. The V_{2it} there is equivalent to our test statistic. See also Arellano and Hahn (2007, Section 3.1) for a similar expression.

null hypothesis $\kappa_2 = 0$ by evaluating the test statistic based on

$$N^{1/2}\hat{\kappa}_2 := N^{-1/2}\sum_{i=1}^{N}\frac{\partial^2 f\left(Y_i|\, X_i, \bar{\theta}_N\right)/\partial\theta_1^2}{f\left(Y_i|\, X_i, \bar{\theta}_N\right)} s\left(Y_i|\, X_i, \bar{\theta}_N\right).$$

Using standard arguments, it can be shown that this statistic is equal to

$$N^{-1/2}\sum_{i=1}^{N}\frac{\partial^2 f\left(Y_i|\, X_i, \theta_0\right)/\partial\theta_1^2}{f\left(Y_i|\, X_i, \theta_0\right)} s\left(Y_i|\, X_i, \theta_0\right) + \kappa_4'\mathcal{I}^{-1}N^{-1/2}\sum_{i=1}^{N} s\left(Y_i|\, X_i, \theta_0\right) + o_p\left(1\right),$$

where[23]

$$\kappa_4 := E\left[\frac{\partial}{\partial\theta}\left(\frac{\partial^2 f\left(Y_i|\, X_i, \theta_0\right)/\partial\theta_1^2}{f\left(Y_i|\, X_i, \theta_0\right)}\right) s\left(Y_i|\, X_i, \theta_0\right)\right] + E\left[\frac{\partial^2 f\left(Y_i|\, X_i, \theta_0\right)/\partial\theta_1^2}{f\left(Y_i|\, X_i, \theta_0\right)}\frac{\partial s\left(Y_i|\, X_i, \theta_0\right)}{\partial\theta}\right].$$

It follows that

$$N^{-1/2}\sum_{i=1}^{N}\frac{\partial^2 f\left(Y_i|\, X_i, \bar{\theta}_N\right)/\partial\theta_1^2}{f\left(Y_i|\, X_i, \bar{\theta}_N\right)} s\left(Y_i|\, X_i, \bar{\theta}_N\right) \to N\left(0, \kappa_3 - \kappa_4'\mathcal{I}^{-1}\kappa_4\right),$$

where $\kappa_3 := E\left[\left(\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right)^2 s\left(Y_i|\, X_i, \theta_0\right) s\left(Y_i|\, X_i, \theta_0\right)'\right]$, and the squared standardized test statistic takes the form $N\hat{\kappa}_2'\left(\hat{\kappa}_3 - \hat{\kappa}_4'\hat{\mathcal{I}}^{-1}\hat{\kappa}_4\right)^{-1}\hat{\kappa}_2$, where $\hat{\kappa}_3$ and $\hat{\kappa}_4$ are straightforward sample analogs of κ_3 and κ_4. Obviously, the distribution of the test statistic under the null of $\kappa_2 = 0$ is χ_q^2 (where q is the dimension of θ).

A sequential test procedure can therefore be used in practice. First, test whether there is neglected heterogeneity in the random effects form, i.e., whether $E\left[m(Z_i, \theta_0)\right] = 0$, by comparing the BP test statistic in (21) with $\chi^2_{1,1-\alpha}$, the upper α level critical value from the χ_1^2 distribution. If this test rejects the null, then proceed to test whether $\kappa_2 = 0$ by comparing $N\hat{\kappa}_2'\left(\hat{\kappa}_3 - \hat{\kappa}_4'\hat{\mathcal{I}}^{-1}\hat{\kappa}_4\right)^{-1}\hat{\kappa}_2$ with $\chi^2_{q,1-\alpha}$, the upper α level critical value from the χ_q^2 distribution. If the null is not rejected, then the neglected heterogeneity does not significantly affect the inference based on the MLE which does not take it into account. This sequential procedure has an overall false rejection probability (weakly) smaller than α.

[23]Note that

$$\kappa_4 = -E\left[\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)} s(Y_i|X_i,\theta_0)s(Y_i|X_i,\theta_0)'\right]$$

if $\kappa_2 = 0$, which may provide a basis for an alternative form of the asymptotic variance.

3.2 Power of the BP Test Against Fixed Effects

The discussion leading up to (19) indicates that the parameterization $h\left(y|\, x, \theta_0, \sqrt{\eta}\right)$ is appropriate for local power analysis when $\mu(X_i) = E\left[\varepsilon_i|\, X_i\right] = 0$, while the parameterization $h\left(y|\, x, \theta_0, \eta\right)$ is appropriate for local power analysis when $E\left[\varepsilon_i|\, X_i\right] \neq 0$. The former parameterization captures the appropriate second order effects, as is evident in the derivation of (7), and the latter captures the first order effects. Therefore, a useful synthesis is to combine the two and to consider the local parameterization of the form

$$f\left(y_i \left| x_i, \left(\theta_{0,1} + \frac{\mu\left(x_i\right)}{N^{1/2}} + \frac{\varepsilon_i^*}{N^{1/4}}, \theta_{0,2}, \ldots, \theta_{0,q}\right)\right.\right), \tag{23}$$

where $E\left[\varepsilon_i^*|\, x_i\right] = 0$.

Next theorem gives the local power property of the BP test against the alternatives of fixed effects of the form (23).[24]

Theorem 4 *Under Assumptions 1 - 2, 3′ and 4 - 7 detailed in Appendix B.1, we get*

$$\sqrt{N} m_N\left(\bar{\theta}_N\right) \Rightarrow N\left(\left[I_1, -\kappa_2' \mathcal{I}^{-1}\right]\left(K_F + K_R^*\right), \kappa_1 - \kappa_2' \mathcal{I}^{-1} \kappa_2\right),$$

where I_k is the $k \times k$ identity matrix ($k = 1$ here),

$$K_F := \begin{bmatrix} E\left\{E\left[\left.\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)} s_1\left(Y_i|\, X_i, \theta_0\right)\right| X_i\right] \mu\left(X_i\right)\right\} \\ E\left\{E\left[s\left(Y_i|\, X_i, \theta_0\right) s_1\left(Y_i|\, X_i, \theta_0\right)|\, X_i\right] \mu\left(X_i\right)\right\} \end{bmatrix},$$

$$K_R^* := \frac{1}{2}\begin{bmatrix} E\left\{E\left[\left(\varepsilon_i^*\right)^2\big|\, X_i\right]\left(\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right)^2\right\} \\ E\left\{E\left[\left(\varepsilon_i^*\right)^2\big|\, X_i\right] s\left(Y_i|\, X_i, \theta_0\right) \frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right\} \end{bmatrix},$$

and $s_1\left(Y_i|\, X_i, \theta_0\right)$ denotes the first coordinate of the score $s\left(Y_i|\, X_i, \theta_0\right)$.

Proof. In Appendix B.3. ■

Similar to the random effect case, Theorem 4 implies that (i) we could use the same squared standardized BP test statistic in (21) in the fixed effect case; and (ii) its asymptotic distribution under $\theta_{0,1} + \frac{\mu(x)}{N^{1/2}} + \frac{\varepsilon^*}{N^{1/4}}$ is a non-central χ_1^2 distribution with noncentrality parameter $\frac{\left(\left[I_1, -\kappa_2'\mathcal{I}^{-1}\right]\left(K_F + K_R^*\right)\right)^2}{\kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2}$.

[24] Theorem 4 makes a marginal contribution over Hahn et al. (2017)'s Proposition 2 by explicitly accounting for the noise of estimating $\bar{\theta}_N$.

4 A Pragmatic Aspect of the LM Test

In practice, we would have to confront the fact that θ_0 is estimated and examine how the noise of estimating θ_0 by the MLE $\bar{\theta}_n$ affects the distribution of the test statistic under the null. We will argue that for the two-way error component panel models with $N, T \to \infty$ asymptotics, the noise does not affect the asymptotic distribution. For this purpose, it suffices to examine the distribution of (12) evaluated at the MLE

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\partial^2 \ln f\left(Y_{it} \middle| X_{it}, \bar{\theta}_n\right)}{\partial \theta_1^2} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \frac{\partial \ln f\left(Y_{it} \middle| X_{it}, \bar{\theta}_n\right)}{\partial \theta_1} \right)^2, \tag{24}$$

where we recognize that the MLE is such that $\sqrt{NT}\left(\bar{\theta}_n - \theta_0\right) = O_p(1)$ under the null, with the $N, T \to \infty$ asymptotics. Under such asymptotics, it can be shown that (24) has the same distribution as (12) under the null.[25] By a similar argument, we can conclude that (13) has the same asymptotic distribution as its feasible counterpart, where the two θ_0 in (13) are replaced by $\bar{\theta}_n$.[26]

Honda (1985) analyzed the asymptotic properties of the BP test for the linear model, and his analysis indicates that the feasible test statistic evaluated at the MLE does not need to reflect the noise of estimating the MLE. On the other hand, Lancaster (1984) showed that it is in general necessary to adjust for such noise in general nonlinear models. In Section 3.1.1, we explained that this seeming contradiction can be understood by noticing that Lancaster (1984)'s adjustment is unnecessary for *linear* models.[27] In this section, we went one step further to show that Lancaster (1984)'s adjustment is unnecessary for general two-way error component panel models. Lancaster (1984) implicitly adopted "large N, fixed T" asymptotics, which is natural for one-way error component panel models. In contrast, the two-way models make it necessary to adopt a "large N, large T" asymptotic framework. Given that the natural asymptotic frameworks are different, there is no logical contradiction.

[25]See Appendix C.

[26]Our conclusion only requires that (12) and (13) are both unaffected by the noise of the estimation of θ_0. Hence, the joint asymptotic distribution of the random vector consisting of (12) and (13), if it is correctly established, is unaffected by such a noise.

[27]See Remark 3.

Our result in this section has convenient pragmatic implications beyond two-way error component panel models. First, due to the equality between (8) and (12), Lancaster (1984)'s adjustment is unnecessary for one-way error component panel models if a "large N, large T" asymptotic framework is adopted. Second, and more importantly, due to the asymptotic equivalence between LM_d in (11) and the sum of (12) and (13) when $N = T$, Lancaster (1984)'s adjustment is also unnecessary for directed dyadic regression models.

5 Monte Carlo Experiments

We explore finite sample properties of the LM test in two directed dyadic link formation models with unobserved individual heterogeneity. Model 1 is

$$\begin{aligned} y_{ij} &= \mathbb{I}\{x_{ij}\theta_0 + \varepsilon_i + \varepsilon_j + v_{ij} > 0\}, \\ x_{ij} &= w_i * w_j, \\ w_i &\sim iid\ U\left(-\frac{1}{2}, \frac{1}{2}\right), \\ \varepsilon_i &\sim iid\ N(0, \sigma_\varepsilon^2), \\ v_{ij} &\sim iid\ N(0, 1), \end{aligned}$$

where $\theta_0 = 1$, $i \neq j$, and $i, j \in \{1, \ldots, N\}$. We run $R = 5000$ Monte Carlo repetitions, and for each one we vary $N \in \{23, 32\}$ and $\sigma_\varepsilon^2 \in \{1/2, 1/8, 1/32, 0\}$.[28] ε_is in this model are random effects since their distribution does not depend on x_{ij}s.

Model 2 is the same as Model 1, except it contains fixed effects $\varepsilon_i \sim N(\sigma_\varepsilon x_i^{max}, \sigma_\varepsilon^2(1 + x_i^{max}))$, and ε_is are mutually independent, where $x_i^{max} \equiv \max_{j \neq i} x_{ij}$.

Finite sample sizes and powers of the LM test for these models are reported in Panel A of Table 1. The results are consistent with our theoretical analysis.

To understand the effect of misspecification of the conditional likelihood $f(Y_{ij}|X_{ij}, \theta)$, we also let $v_{ij} \sim iid\ logistic(0, \frac{\sqrt{3}}{\pi})$ and run 5000 Monte Carlo repetitions of our test assuming the standard normal distribution of v_{ij}. The results are reported in Panel B of

[28]We choose these N values such that the effective sample sizes, $N(N-1)$, are roughly 500 and 1000, respectively.

Table 1. The rejection rates increase due to the misspecification, but not by much. More importantly, the sizes are still well controlled by the nominal level, so our test appears to be reasonably robust to mild misspecification of the conditional likelihood.

Table 1: Rejection Rates of Level 5% LM Test

	Panel A: correct specification, i.e., $v_{ij} \sim iid\ N(0,1)$							
	Model 1 (Random Effects)				Model 2 (Fixed Effects)			
σ_ε^2	1/2	1/8	1/32	0	1/2	1/8	1/32	0
$N = 23$	1.000	0.947	0.280	0.027	1.000	0.971	0.343	0.027
$N = 32$	1.000	0.999	0.653	0.033	1.000	1.000	0.754	0.033
	Panel B: misspecification, i.e., $v_{ij} \sim iid\ logistic(0, \sqrt{3}/\pi)$							
	Model 1 (Random Effects)				Model 2 (Fixed Effects)			
σ_ε^2	1/2	1/8	1/32	0	1/2	1/8	1/32	0
$N = 23$	1.000	0.981	0.386	0.028	1.000	0.990	0.466	0.028
$N = 32$	1.000	1.000	0.778	0.035	1.000	1.000	0.861	0.035

6 Summary

We developed a test of neglected individual effects in dyadic regression models by modifying the BP test. The test statistic is almost identical to the the sum of two components of the BP test statistic in the panel data analysis for testing the presence of two-way error components. Asymptotic distribution of the test statistic is carefully derived based on a novel martingale argument.

We also derived several interesting results about the BP test in generic panel data analysis. We showed that the test has a power against fixed effects, even though it was developed to detect random effects. We also analyzed the nature of the distortion to the asymptotic distribution induced by the noise of estimating the MLE, and found that the noise need not be accounted for with one-way linear models or general two-way models.

References

Arellano, Manuel and Jinyong Hahn, *Understanding bias in nonlinear panel models: Some recent developments*, Vol. 43 of Advances in Economics and Econometrics, Cambridge: Cambridge Univerisity Press,

Bartle, Robert Gardner, *The Elements Of Integration*, New York: John Wiley & Sons, 1966.

Breusch, Trevor S and Adrian R Pagan, "The Lagrange multiplier test and its applications to model specification in econometrics," *The Review of Economic Studies*, 1980, *47* (1), 239–253.

Chamberlain, Gary, *Panel data*, Vol. II of Handbook of Econometrics, Elsevier,

Chesher, Andrew, "Testing for neglected heterogeneity," *Econometrica*, 1984, *52*, 865–872.

Cox, David R, "Some remarks on overdispersion," *Biometrika*, 1983, *70* (1), 269–274.

Engle, Robert F, *Wald, likelihood ratio, and Lagrange multiplier tests in econometrics*, Vol. II of Handbook of Econometrics, Elsevier,

Graham, Bryan S, "Dyadic regression," *The Econometric Analysis of Network Data*, 2020, pp. 23–40.

Hahn, Jinyong and Whitney Newey, "Jackknife and analytical bias reduction for nonlinear panel models," *Econometrica*, 2004, *72* (4), 1295–1319.

_ , Hyungsik Roger Moon, and Connan Snider, "LM test of neglected correlated random effects and its application," *Journal of Business & Economic Statistics*, 2017, *35* (3), 359–370.

Hall, Peter and Christopher C Heyde, *Martingale limit theory and its application*, Academic press, New York, 1980.

Hausman, Jerry A and William E Taylor, "Panel data and unobservable individual effects," *Econometrica*, 1981, *49*, 1377–1398.

Honda, Yuzo, "Testing the error components model with non-normal disturbances," *The Review of Economic Studies*, 1985, *52* (4), 681–690.

Kuersteiner, Guido M and Ingmar R Prucha, "Limit theory for panel data models with cross sectional dependence and sequential exogeneity," *Journal of Econometrics*, 2013, *174* (2), 107–126.

Lancaster, Tony, "The covariance matrix of the information matrix test," *Econometrica*, 1984, *52*, 1051–1053.

Menzel, Konrad, "Bootstrap with clustering in two or more dimensions," *arXiv preprint arXiv:1703.03043*, 2017.

Newey, Whitney K, "Maximum likelihood specification testing and conditional moment tests," *Econometrica*, 1985, *53*, 1047–1070.

Rao, Calyampudi Radhakrishna, *Linear statistical inference and its applications*, New York: John Wiley & Sons, 1971.

White, Halbert, "Maximum likelihood estimation of misspecified models," *Econometrica*, 1982, *50*, 1–25.

Online Appendices

for

"Test of Neglected Heterogeneity in Dyadic Models"

by

Jinyong Hahn, Hyungsik Roger Moon and Ruoyao Shi

A Proof of the Theorems in Section 2

A.1 Construction of a Martingale Difference Sequence Array

We assume $T := T(N)$, where $T(N) \to \infty$ is an increasing function of N with $\frac{N}{T} \to \kappa$ and $0 < \kappa < \infty$. For simplicity, we will often use notation T instead of $T(N)$.

Our proof of Theorem 1 relies on a novel construction of a martingale difference sequence, which we explicitly discuss here. We can rewrite $\sqrt{N} A_{NT}$ as

$$\begin{aligned}
\sqrt{N} A_{NT} &= \varrho \frac{1}{N} \sum_{i=1}^{N} \sum_{t,s=1,t\neq s}^{T} U_{it} U_{is} + \frac{1}{N} \sum_{i,j=1,i\neq j}^{N} \sum_{t=1}^{T} U_{it} U_{jt} \\
&= \varrho \frac{2}{N} \sum_{i=2}^{N} \sum_{t=2}^{T} \left(\sum_{s=1}^{t-1} U_{is} \right) U_{it} + \frac{2}{N} \sum_{i=2}^{N} \sum_{t=2}^{T} \left(\sum_{j=1}^{i-1} U_{jt} \right) U_{it} \\
&\quad + \varrho \frac{2}{N} \sum_{t=2}^{T} \left(\sum_{s=1}^{t-1} U_{1s} \right) U_{1t} + \frac{2}{N} \sum_{i=2}^{N} \left(\sum_{j=1}^{i-1} U_{j1} \right) U_{i1} \\
&= \frac{2}{N} \left[\sum_{i=2}^{N} \sum_{t=2}^{T} \left\{ \left(\sum_{j=1}^{i-1} U_{jt} \right) + \varrho \left(\sum_{s=1}^{t-1} U_{is} \right) \right\} U_{it} \right] + O_p(1) .
\end{aligned} \tag{A.1}$$

Remark 5 *The $O_p(1)$ term in* (A.1) *is equal to*

$$\varrho \frac{2}{N} \sum_{t=2}^{T} \left(\sum_{s=1}^{t-1} U_{1s} \right) U_{1t} + \frac{2}{N} \sum_{i=2}^{N} \left(\sum_{j=1}^{i-1} U_{j1} \right) U_{i1} = \frac{2\varrho}{N} \sum_{1\leq s<t\leq T} U_{1s} U_{1t} + \frac{2}{N} \sum_{1\leq j<i\leq N} U_{j1} U_{i1},$$

which is indeed of stochastic order of $O_p(1)$ because

$$E\left[\left(\frac{2\varrho}{N} \sum_{1\leq s<t\leq T} U_{1s} U_{1t} + \frac{2}{N} \sum_{1\leq j<i\leq N} U_{j1} U_{i1} \right)^2 \right]$$

$$
\begin{aligned}
&\leq 2\varrho^2 E\left[\left(\frac{2}{N}\sum_{1\leq s<t\leq T} U_{1s}U_{1t}\right)^2\right] + 2E\left[\left(\frac{2}{N}\sum_{1\leq j<i\leq N} U_{j1}U_{i1}\right)^2\right] \\
&= \frac{8\varrho^2}{N^2}\sum_{1\leq s<t\leq T} E\left[(U_{1s}U_{1t})^2\right] + \frac{8}{N^2}\sum_{1\leq j<i\leq N} E\left[(U_{j1}U_{i1})^2\right] \\
&= \frac{8\varrho^2}{N^2}\frac{T\,(T-1)}{2}\sigma_U^4 + \frac{8}{N^2}\frac{N\,(N-1)}{2}\sigma_U^4 \\
&= O(1)
\end{aligned}
$$

as $N, T \to \infty$ *with* $\frac{N}{T} \to \kappa$ *and* $0 < \kappa < \infty$.

Define

$$
Q_{it} := \frac{2}{N}\left\{\left(\sum_{j=1}^{i-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{is}\right)\right\} U_{it}
$$

for $2 \leq i \leq N$ and $2 \leq t \leq T$, and we can rewrite the leading term of (A.1)

$$
\frac{2}{N}\sum_{i=2}^{N}\sum_{t=2}^{T}\left\{\left(\sum_{j=1}^{i-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{is}\right)\right\} U_{it} = \sum_{i=2}^{N}\sum_{t=2}^{T} Q_{it}.
$$

Assume $N \leq T$ in the following, and the case where $N > T$ can be handled symmetrically. Define

$$
Z_{N,2} := Q_{22} = \frac{2}{N}\left(U_{12} + \varrho U_{21}\right) U_{22}.
$$

Next, for $3 \leq n \leq N$, define

$$
\begin{aligned}
Z_{N,n} &:= \sum_{t=2}^{n-1} Q_{nt} + \sum_{i=2}^{n-1} Q_{in} + Q_{nn} \\
&= \frac{2}{N}\sum_{t=2}^{n-1}\left\{\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right\} U_{nt} + \frac{2}{N}\sum_{i=2}^{n-1}\left\{\left(\sum_{j=1}^{i-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{is}\right)\right\} U_{in} \\
&+ \frac{2}{N}\left\{\left(\sum_{j=1}^{n-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{ns}\right)\right\} U_{nn}.
\end{aligned}
$$

Also, if $N < T$, then for $N + 1 \leq n \leq T$, define

$$
Z_{N,n} := \sum_{i=2}^{N} Q_{in} = \frac{2}{N}\sum_{i=2}^{N}\left\{\left(\sum_{j=1}^{i-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{is}\right)\right\} U_{in}.
$$

Therefore, the leading term of (A.1) can be further rewritten as

$$\frac{2}{N}\sum_{i=2}^{N}\sum_{t=2}^{T}\left\{\left(\sum_{j=1}^{i-1}U_{jt}\right)+\varrho\left(\sum_{s=1}^{t-1}U_{is}\right)\right\}U_{it}=\sum_{i=2}^{N}\sum_{t=2}^{T}Q_{it}=\sum_{n=2}^{T}Z_{N,n}.$$

In sum, we see that

$$\sqrt{N}A_{NT}=\sum_{n=2}^{T}Z_{N,n}+O_p\left(1\right), \tag{A.2}$$

and we will analyze $\sum_{n=2}^{T}Z_{N,n}$ using the martingale described below.

For $1\leq n\leq N$, define $\mathcal{F}_{N,n}$ to be the sigma field generated by $\{U_{it}\ :\ 1\leq i,t\leq n\}$. For $N+1\leq n\leq T$, define $\mathcal{F}_{N,n}$ to be the sigma field generated by $\{U_{it}\ :\ 1\leq i\leq N,\ 1\leq t\leq n\}$.

By definition, we have $\mathcal{F}_{N,1}\subset\mathcal{F}_{N,2}\subset\cdots\subset\mathcal{F}_{N,T}$, and $Z_{N,n}$ is measurable with respect to $\mathcal{F}_{N,n}$ for $2\leq n\leq T$. Also, we have

$$E\left(Z_{N,n}|\mathcal{F}_{N,n-1}\right)=0 \text{ and } E\left(|Z_{N,n}|\right)<\infty,$$

for $2\leq n\leq T$. Therefore, $\{Z_{N,n},\mathcal{F}_{N,n}\ ,2\leq n\leq T\}$ is a martingale difference sequence array.

A.2 Proof of Lemma 1

With the assumption $N<T$, we define

$$B_{NN}:=\sum_{n=2}^{N}Z_{N,n}=\frac{2}{N}\left[\sum_{i=2}^{N}\sum_{t=2}^{N}\left\{\left(\sum_{j=1}^{i-1}U_{jt}\right)+\varrho\left(\sum_{s=1}^{t-1}U_{is}\right)\right\}U_{it}\right],$$

$$C_{NT}:=\sum_{n=N+1}^{T}Z_{N,n}=\frac{2}{N}\left[\sum_{i=2}^{N}\sum_{t=N+1}^{T}\left\{\left(\sum_{j=1}^{i-1}U_{jt}\right)+\varrho\left(\sum_{s=1}^{t-1}U_{is}\right)\right\}U_{it}\right].$$

If $N=T$, then let $C_{NT}=0$. We can rewrite

$$\sqrt{N}A_{NT}=B_{NN}+C_{NT}+O_p\left(1\right).$$

We will work with Hall and Heyde (1980)'s Corollary 3.3 to obtain the CLT. Because of complexity of notation, we will take care of B_{NN} and C_{NT} separately and obtain intermediate results in the next two subsections. All the intermediate results assume the conditions in Lemma 1, and we skip the conditions in stating the intermediate results. The intermediate results can be combined to show the result of Lemma 1 as follows.

Lemma 4 $E\left[\left|\sum_{n=2}^{T} E\left[\left.\frac{Z_{N,n}^{2}}{N}\right|\mathcal{F}_{N,n-1}\right]-2\left(\frac{1}{\kappa}+\frac{\varrho^{2}}{\kappa^{2}}\right)\sigma_{U}^{4}\right|^{2}\right]=o\left(1\right).$

Proof. From Lemmas 9 and 13, we have

$$\begin{aligned}\sum_{n=2}^{T} E\left[\frac{Z_{N,n}^{2}}{N}\right] &= \sum_{n=2}^{T} E\left(E\left[\left.\frac{Z_{N,n}^{2}}{N}\right|\mathcal{F}_{N,n-1}\right]\right)\\ &= 2\left(\varrho^{2}+1\right)\sigma_{U}^{4}+2\left(T-N\right)\frac{N+N\varrho^{2}+T\varrho^{2}}{N^{2}}\sigma_{U}^{4}+o\left(1\right)\\ &= \frac{2}{N^{2}}T\left(T\varrho^{2}+N\right)\sigma_{U}^{4}+o\left(1\right),\end{aligned}$$

from which we conclude that

$$\sum_{n=2}^{T} E\left[\frac{Z_{N,n}^{2}}{N}\right]=2\left(\frac{1}{\kappa}+\frac{\varrho^{2}}{\kappa^{2}}\right)\sigma_{U}^{4}+o\left(1\right),$$

where we recall that $\kappa=\lim\frac{N}{T}$. Combined with

$$E\left[\left|\sum_{n=2}^{T}\left(E\left[\left.\frac{Z_{N,n}^{2}}{N}\right|\mathcal{F}_{N,n-1}\right]-E\left[\frac{Z_{N,n}^{2}}{N}\right]\right)\right|^{2}\right]=o\left(1\right),$$

which follows from Lemmas 8 and 12, we obtain the conclusion. ■

Lemma 5 $\sum_{n=2}^{T} E\left[\frac{Z_{N,n}^{4}}{N^{2}}\right]=o\left(1\right).$

Proof. It is a consequence of Lemmas 10 and 14. ■

By Lemmas 4 and 5, we can see that Hall and Heyde (1980)'s (3.38) and (3.40) are satisfied with $p=2$. Therefore, we can conclude by Hall and Heyde (1980)'s Corollary 3.3 that

$$\sum_{n=2}^{T}\frac{Z_{N,n}}{\sqrt{N}}\Rightarrow N\left(0,2\left(\frac{1}{\kappa}+\frac{\varrho^{2}}{\kappa^{2}}\right)\sigma_{U}^{4}\right),$$

which, combined with (A.2), implies the result of Lemma 1.

Remark 6 *Lemma 5 implies* $\max_{2\leq n\leq T} E\left[\left.\frac{Z_{N,n}^{2}}{N}\right|\mathcal{F}_{N,n-1}\right]=o_{p}\left(1\right)$, *which is a counterpart of Hall and Heyde (1980)'s (3.34). This can be seen by letting* $W_{n}:=E\left[\left.\frac{Z_{N,n}^{2}}{N}\right|\mathcal{F}_{N,n-1}\right]\geq 0$, *and noting that we have for any* $\epsilon>0$,

$$P\left(\max_{2\leq n\leq T} W_{n}\geq\epsilon\right)\leq P\left(\sum_{n=2}^{T} W_{n}^{2}\geq\epsilon^{2}\right)$$

$$
\begin{aligned}
&\leq \frac{1}{\epsilon^2}\sum_{n=2}^{T} E(W_n^2) \\
&= \frac{1}{\epsilon^2}\sum_{n=2}^{T} E\left[\left(E\left[\left.\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right]\right)^2\right] \\
&\leq \frac{1}{\epsilon^2}\sum_{n=2}^{T} E\left[E\left[\left.\left(\frac{Z_{N,n}^2}{N}\right)^2\right|\mathcal{F}_{N,n-1}\right]\right] \\
&= \frac{1}{\epsilon^2}\sum_{n=2}^{T} E\left[\frac{Z_{N,n}^4}{N^2}\right].
\end{aligned}
$$

Remark 7 *Lemma 5 also implies that Hall and Heyde (1980)'s (3.36), i.e., the Lindeberg-Feller condition, is satisfied. For this purpose, note that*

$$
\sum_{n=2}^{T} E\left[\frac{Z_{N,n}^2}{N}\mathbb{I}\left\{\left|\frac{Z_{N,n}}{\sqrt{N}}\right| \geq \epsilon\right\}\right] \leq \frac{1}{\epsilon^2}\sum_{n=2}^{T} E\left[\left(\frac{Z_{N,n}}{\sqrt{N}}\right)^4\right].
$$

A.2.1 Analysis of B_{NN}

We will assume that $n \leq N$ below. Recall that

$$
Z_{N,2} = \frac{2}{N}\left(U_{12} + \varrho U_{21}\right) U_{22},
$$

and for $3 \leq n \leq N$,

$$
\begin{aligned}
Z_{N,n} := &\frac{2}{N}\sum_{t=2}^{n-1}\left\{\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right\} U_{nt} + \frac{2}{N}\sum_{i=2}^{n-1}\left\{\left(\sum_{j=1}^{i-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{is}\right)\right\} U_{in} \\
&+ \frac{2}{N}\left\{\left(\sum_{j=1}^{n-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{ns}\right)\right\} U_{nn}.
\end{aligned}
$$

Notice that

$$
E\left[\left.\frac{Z_{N,2}^2}{N}\right|\mathcal{F}_{N,1}\right] = \frac{4}{N^3}(1+\varrho^2)\sigma_U^4. \tag{A.3}
$$

Lemma 6 *For $3 \leq n \leq N$, we have*

$$
E\left[\left.\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right] = \frac{4\sigma_U^4}{N^3}\sum_{t=2}^{n-1}\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2 + \frac{4\varrho^2\sigma_U^4}{N^3}\sum_{i=2}^{n-1}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2 + \frac{2\left(1+\varrho^2\right)n\left(n-1\right)}{N^3}\sigma_U^4.
$$

Proof. We have

$$\begin{aligned}
E\left[\left. Z_{N,n}^2 \right| \mathcal{F}_{N,n-1}\right] &= \frac{4}{N^2} \sum_{t=2}^{n-1} \left\{ \left(\sum_{j=1}^{n-1} \frac{U_{jt}}{\sigma_U} \right)^2 \sigma_U^2 + \varrho^2 (t-1) \sigma_U^2 \right\} \sigma_U^2 \\
&\quad + \frac{4}{N^2} \sum_{i=2}^{n-1} \left\{ (i-1) \sigma_U^2 + \varrho^2 \left(\sum_{s=1}^{n-1} \frac{U_{is}}{\sigma_U} \right)^2 \sigma_U^2 \right\} \sigma_U^2 \\
&\quad + \frac{4}{N^2} \left\{ (n-1) \sigma_U^2 + \varrho^2 (n-1) \sigma_U^2 \right\} \sigma_U^2, \qquad \text{(A.4)}
\end{aligned}$$

from which we derive the implication. ■

Lemma 7 *Suppose that the kurtosis of* $\frac{U_{jt}}{\sigma_U}$ *is* $K+1$*. We then have*

$$\operatorname{Var}\left(\left(\sum_{j=1}^{n} \frac{U_{jt}}{\sigma_U} \right)^2 \right) = 2n^2 + (K-2)\, n, \text{ and } \operatorname{Var}\left(\left(\sum_{s=1}^{n} \frac{U_{is}}{\sigma_U} \right)^2 \right) = 2n^2 + (K-2)\, n.$$

Proof. We have

$$\begin{aligned}
E\left[\left(\sum_{j=1}^{n} \frac{U_{jt}}{\sigma_U} \right)^4 \right] &= \sum_{j=1}^{n} E\left[\left(\frac{U_{jt}}{\sigma_U} \right)^4 \right] + 3 \sum_{j \neq j'} E\left[\left(\frac{U_{jt}}{\sigma_U} \right)^2 \right] E\left[\left(\frac{U_{j't}}{\sigma_U} \right)^2 \right] \\
&= n (K+1) + 3n (n-1), \qquad \text{(A.5)}
\end{aligned}$$

so

$$\begin{aligned}
\operatorname{Var}\left(\left(\sum_{j=1}^{n} \frac{U_{jt}}{\sigma_U} \right)^2 \right) &= E\left[\left(\sum_{j=1}^{n} \frac{U_{jt}}{\sigma_U} \right)^4 \right] - \left(E\left[\left(\sum_{j=1}^{n} \frac{U_{jt}}{\sigma_U} \right)^2 \right] \right)^2 \\
&= n (K+1) + 3n (n-1) - n^2 \\
&= 2n^2 + (K-2)\, n.
\end{aligned}$$

The second statement holds by symmetry. ■

Lemma 8 $E\left[\left| \sum_{n=2}^{N} \left(E\left[\left. \frac{Z_{N,n}^2}{N} \right| \mathcal{F}_{N,n-1} \right] - E\left[\frac{Z_{N,n}^2}{N} \right] \right) \right|^2 \right] = o(1)$.

Proof. Since $\left(\frac{1}{N} \sum_{n=1}^{N} a_n \right)^2 \leq \frac{1}{N} \sum_{n=1}^{N} a_n^2$, the desired result of the lemma follows if we show

$$E\left[\frac{1}{N} \sum_{n=2}^{N} \left(E\left[\left. Z_{N,n}^2 \right| \mathcal{F}_{N,n-1} \right] - E\left[Z_{N,n}^2 \right] \right)^2 \right] = o(1).$$

Notice that by the independence among U_{it} and the definition of $\mathcal{F}_{N,1}$, we have $E\left[\left. Z_{N,2}^2 \right| \mathcal{F}_{N,1}\right] - E\left[Z_{N,2}^2\right] = 0$. Then,

$$\begin{aligned}
&\frac{1}{\sigma_U^8} E\left[\frac{1}{N}\sum_{n=2}^{N}\left(E\left[\left. Z_{N,n}^2 \right| \mathcal{F}_{N,n-1}\right] - E\left[Z_{N,n}^2\right]\right)^2\right] \\
&= \frac{1}{\sigma_U^8} E\left[\frac{1}{N}\sum_{n=3}^{N}\left(E\left[\left. Z_{N,n}^2 \right| \mathcal{F}_{N,n-1}\right] - E\left[Z_{N,n}^2\right]\right)^2\right] \\
&= \frac{1}{\sigma_U^8}\frac{1}{N}\sum_{n=3}^{N}\mathrm{Var}\left(E\left[\left. Z_{N,n}^2 \right| \mathcal{F}_{N,n-1}\right]\right) \\
&= \frac{1}{\sigma_U^8}\frac{1}{N}\sum_{n=3}^{N}\mathrm{Var}\left[\left(\frac{4}{N^2}\sum_{t=2}^{n-1}\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2 + \frac{4\varrho^2}{N^2}\sum_{i=2}^{n-1}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right)\sigma_U^4\right],
\end{aligned}$$

where we used Lemma 6 for the last equality. Therefore, we have

$$\begin{aligned}
&\frac{1}{\sigma_U^8} E\left[\frac{1}{N}\sum_{n=2}^{N}\left(E\left[\left. Z_{N,n}^2 \right| \mathcal{F}_{N,n-1}\right] - E\left[Z_{N,n}^2\right]\right)^2\right] \\
&= \frac{1}{N}\sum_{n=3}^{N}\mathrm{Var}\left(\frac{4}{N^2}\sum_{t=2}^{n-1}\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2 + \frac{4\varrho^2}{N^2}\sum_{i=2}^{n-1}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right) \\
&\leq \frac{1}{N}\sum_{n=3}^{N}\left[2\,\mathrm{Var}\left(\frac{4}{N^2}\sum_{t=2}^{n-1}\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2\right) + 2\,\mathrm{Var}\left(\frac{4\varrho^2}{N^2}\sum_{i=2}^{n-1}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right)\right] \\
&= \frac{32}{N^5}\sum_{n=3}^{N}\left(\mathrm{Var}\left(\sum_{t=2}^{n-1}\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2\right) + \varrho^4\,\mathrm{Var}\left(\sum_{i=2}^{n-1}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right)\right) \\
&= \frac{32}{N^5}\sum_{n=3}^{N}\left[\sum_{t=2}^{n-1}\mathrm{Var}\left(\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2\right) + \varrho^4\sum_{i=2}^{n-1}\mathrm{Var}\left(\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right)\right],
\end{aligned}$$

where in the last equality, we used the fact that (i) $\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2$ are independent over t; and (ii) $\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2$ are independent over i. Using Lemma 7, we obtain

$$\begin{aligned}
\mathrm{Var}\left(\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2\right) &= 2\left(n-1\right)^2 + \left(K-2\right)\left(n-1\right), \\
\mathrm{Var}\left(\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right) &= 2\left(n-1\right)^2 + \left(K-2\right)\left(n-1\right),
\end{aligned}$$

from which we obtain

$$\frac{1}{\sigma_U^8}E\left[\frac{1}{N}\sum_{n=2}^{N}\left(E\left[Z_{N,n}^2\right|\mathcal{F}_{N,n-1}\right]-E\left[Z_{N,n}^2\right]\right)^2\right]\le\frac{32\left(1+\varrho^4\right)}{N^5}\sum_{n=3}^{N}\left[\sum_{t=2}^{n-1}2\left(n-1\right)^2+\sum_{t=2}^{n-1}(K-2)\left(n-1\right)\right]$$
$$=O(N^{-1})=o(1),$$

as required. ■

Lemma 9 $\sum_{n=2}^{N}E\left(E\left[\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right]\right)=2\left(\varrho^2+1\right)\sigma_U^4+o\left(1\right).$

Proof. For $3\le n$, from (A.4), we obtain

$$\begin{aligned}E\left(E\left[Z_{N,n}^2\right|\mathcal{F}_{N,n-1}\right]\right)&=\frac{4}{N^2}\sum_{t=2}^{n-1}\left\{E\left[\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2\right]\sigma_U^2+\varrho^2\left(t-1\right)\sigma_U^2\right\}\sigma_U^2\\&+\frac{4}{N^2}\sum_{i=2}^{n-1}\left\{\left(i-1\right)\sigma_U^2+\varrho^2E\left[\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right]\sigma_U^2\right\}\sigma_U^2\\&+\frac{4}{N^2}\left\{\left(n-1\right)\sigma_U^2+\varrho^2\left(n-1\right)\sigma_U^2\right\}\sigma_U^2,\end{aligned}$$

from which together with (A.3) we obtain

$$\begin{aligned}&\frac{1}{\sigma_U^4}\sum_{n=2}^{N}E\left(E\left[\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right]\right)\\&=\frac{4}{N^3}(1+\varrho^2)+\sum_{n=3}^{N}\left(\frac{4}{N^3}\sum_{t=2}^{n-1}\left(\left(n-1\right)+\varrho^2\left(t-1\right)\right)\right)\\&+\sum_{n=3}^{N}\left(\frac{4}{N^3}\sum_{i=2}^{n-1}\left(\left(i-1\right)+\varrho^2\left(n-1\right)\right)\right)+\frac{4\left(1+\varrho^2\right)}{N^3}\sum_{n=3}^{N}\left(n-1\right)\\&\to2\left(\varrho^2+1\right).\end{aligned}$$

■

Lemma 10 $\sum_{n=2}^{N}E\left[\frac{Z_{N,n}^4}{N^2}\right]=o\left(1\right).$

Proof. It is straightforward to show that

$$E\left[\frac{Z_{N,2}^4}{N^2}\right]=o\left(1\right).$$

In what follows, we show

$$\sum_{n=3}^{N} E\left[\frac{Z_{N,n}^4}{N^2}\right] = o\left(1\right).$$

Recall that for $3 \leq n \leq N$, we have

$$\begin{aligned}
Z_{N,n} := & \frac{2}{N}\sum_{t=2}^{n-1}\left\{\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right\} U_{nt} \\
& + \frac{2}{N}\sum_{i=2}^{n-1}\left\{\left(\sum_{j=1}^{i-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{is}\right)\right\} U_{in} \\
& + \frac{2}{N}\left\{\left(\sum_{j=1}^{n-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{ns}\right)\right\} U_{nn},
\end{aligned}$$

so we see that $Z_{N,n}$ is the sum of $n-2+n-2+1$ terms of $\frac{2}{N}\left\{\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right\} U_{nt}$, $\frac{2}{N}\left\{\left(\sum_{j=1}^{i-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{is}\right)\right\} U_{in}$, and $\frac{2}{N}\left\{\left(\sum_{j=1}^{n-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{ns}\right)\right\} U_{nn}$, where $i, t = 2, ..., n-1$. For notational simplicity, here let's call each term W_j with $j = 1, \ldots, 2n-3$. We first note that W_j and $W_{j'}$ are independent conditional on $\mathcal{F}_{N,n-1}$ due to the independence among U_{nt}, U_{in} and U_{nn}. Then, we have

$$E\left[\left(\sum_{j=1}^{2n-3} W_j\right)^4 \middle| \mathcal{F}_{N,n-1}\right] = \sum_{j=1}^{2n-3} E\left[\left. W_j^4 \right| \mathcal{F}_{N,n-1}\right] + 3\sum_{j\neq j'} E\left[\left. W_j^2 W_{j'}^2 \right| \mathcal{F}_{N,n-1}\right].$$

Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
E\left[\left(\sum_{j=1}^{2n-3} W_j\right)^4 \middle| \mathcal{F}_{N,n-1}\right] & \leq \sum_{j=1}^{2n-3} E\left[\left. W_j^4 \right| \mathcal{F}_{N,n-1}\right] + 3\sum_{j\neq j'} \sqrt{E\left[\left. W_j^4 \right| \mathcal{F}_{N,n-1}\right]}\sqrt{E\left[\left. W_{j'}^4 \right| \mathcal{F}_{N,n-1}\right]} \\
& \leq \sum_{j=1}^{2n-3} 3E\left[\left. W_j^4 \right| \mathcal{F}_{n-1}\right] + 3\sum_{j\neq j'} \sqrt{E\left[\left. W_j^4 \right| \mathcal{F}_{N,n-1}\right]}\sqrt{E\left[\left. W_{j'}^4 \right| \mathcal{F}_{N,n-1}\right]} \\
& = 3\left(\sum_{j=1}^{2n-3} \sqrt{E\left[\left. W_j^4 \right| \mathcal{F}_{N,n-1}\right]}\right)^2.
\end{aligned}$$

In what follows, we derive an upper bound of $\left(\sum_{j=1}^{2n-3} \sqrt{E\left[\left. W_j^4 \right| \mathcal{F}_{N,n-1}\right]}\right)^2$.

Note that

$$E\left[\left(\frac{2}{N}\left\{\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right\} U_{nt}\right)^4 \middle| \mathcal{F}_{N,n-1}\right]$$

$$= \left(\frac{2}{N}\right)^4 E\left[\left(\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right)^4 \middle| \mathcal{F}_{N,n-1}\right] E\left[U_{nt}^4\right]$$
$$= \left(\frac{2}{N}\right)^4 E\left[\left(\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right)^4 \middle| \mathcal{F}_{N,n-1}\right] (K+1)\,\sigma_U^4.$$

From (A.5) and symmetry, we have

$$E\left[\left(\sum_{s=1}^{t-1} \frac{U_{ns}}{\sigma_U}\right)^4\right] = (t-1)(K+1) + 3(t-1)(t-2)$$
$$= 3\,(t-1)^2 + (K-2)\,(t-1)\,.$$

Then,

$$E\left[\left(\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right)^4 \middle| \mathcal{F}_{N,n-1}\right]$$
$$= \left(\sum_{j=1}^{n-1} U_{jt}\right)^4 + 6\varrho^2\left(\sum_{j=1}^{n-1} U_{jt}\right)^2 (t-1)\,\sigma_U^2$$
$$+ 4\varrho^3\left(\sum_{j=1}^{n-1} U_{jt}\right) E\left[\left(\sum_{s=1}^{t-1} U_{ns}\right)^3\right] + \varrho^4\left(3\,(t-1)^2 + (K-2)\,(t-1)\right)\sigma_U^4.$$

Using that

$$E\left[\left(\sum_{s=1}^{t-1} U_{ns}\right)^3\right] = E\left[\sum_{s=1}^{t-1} U_{ns}^3\right] = (t-1)\,S\sigma_U^3,$$

where S denotes the skewness of U_{it}/σ_U, we further conclude that

$$E\left[\left(\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho\left(\sum_{s=1}^{t-1} U_{ns}\right)\right)^4 \middle| \mathcal{F}_{N,n-1}\right]$$
$$= \left(\sum_{j=1}^{n-1} U_{jt}\right)^4 + 6\varrho^2\left(\sum_{j=1}^{n-1} U_{jt}\right)^2 (t-1)\,\sigma_U^2$$
$$+ 4\varrho^3 S\left(\sum_{j=1}^{n-1} U_{jt}\right)(t-1)\,\sigma_U^3 + \varrho^4\left(3\,(t-1)^2 + (K-2)\,(t-1)\right)\sigma_U^4. \tag{A.6}$$

Notice that if $\varrho = 0$,

$$(A.6) = \left(\sum_{j=1}^{n-1} U_{jt}\right)^4.$$

On the other hand, if $\varrho \neq 0$,

$$(A.6) \leq \left(\sum_{j=1}^{n-1} U_{jt}\right)^4 + 8\varrho^2 \left(\sum_{j=1}^{n-1} U_{jt}\right)^2 (t-1)\,\sigma_U^2$$
$$+ \varrho^4 \left(3\,(t-1)^2 + (K-2)\,(t-1)\right) \sigma_U^4 + 2(t-1)\varrho^4 S^2 \sigma_U^4,$$

where we used the fact that

$$6\varrho^2 z^2\,(t-1)\,\sigma_U^2 + 4\varrho^3 S z\,(t-1)\,\sigma_U^3 \leq 8\varrho^2 z^2\,(t-1)\,\sigma_U^2 + 2(t-1)\varrho^4 S^2 \sigma_U^4$$

for all z on the real line and $t \geq 2$. Then, we can find a finite constant M (that only depend on $K, S, \sigma_U^2, \varrho$) sufficiently large such that

$$\varrho^4 \left(3\,(t-1)^2 + (K-2)\,(t-1)\right) \sigma_U^4 + 2(t-1)\varrho^4 S^2 \sigma_U^4 \leq M^2\,(t-1)^2\,,$$
$$1 \leq 2M(t-1),$$
$$8\varrho^2\,(t-1)\,\sigma_U^2 \leq 2M\,(t-1)\,,$$

for all $t = 2, 3, ...$ For example, it is possible if

$$M \geq \max\left\{ \left(3\varrho^4\sigma_U^4 + |K-2|\varrho^4\sigma_U^4 + 2\varrho^4 S^2 \sigma_U^4\right)^{1/2}, \frac{1}{2}, 4\varrho^2\sigma_U^2 \right\}.$$

Then, it follows that

$$E\left[\left(\frac{2}{N} \left\{ \left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho \left(\sum_{s=1}^{t-1} U_{ns}\right) \right\} U_{nt} \right)^4 \middle| \mathcal{F}_{N,n-1} \right]$$
$$= \left(\frac{2}{N}\right)^4 E\left[\left(\left(\sum_{j=1}^{n-1} U_{jt}\right) + \varrho \left(\sum_{s=1}^{t-1} U_{ns}\right) \right)^4 \middle| \mathcal{F}_{N,n-1} \right] (K+1)\,\sigma_U^4$$
$$\leq \left(\frac{2}{N}\right)^4 \left(\left(\sum_{j=1}^{n-1} U_{jt}\right)^4 + 2\,(t-1)\,M \left(\sum_{j=1}^{n-1} U_{jt}\right)^2 + M^2\,(t-1)^2 \right) (K+1)\,\sigma_U^4$$
$$\leq \frac{C}{N^4} \left(\left(\sum_{j=1}^{n-1} U_{jt}\right)^2 + M\,(t-1) \right)^2,$$

and here and in what follows C denotes a generic constant. Likewise, we get

$$E\left[\left(\frac{2}{N} \left\{ \left(\sum_{j=1}^{i-1} U_{jn}\right) + \varrho \left(\sum_{s=1}^{n-1} U_{is}\right) \right\} U_{in} \right)^4 \middle| \mathcal{F}_{N,n-1} \right] \leq \frac{C}{N^4} \left(\left(\sum_{s=1}^{n-1} U_{is}\right)^2 + M\,(i-1) \right)^2.$$

Using (A.5) and symmetry, we also get

$$E\left[\left(\frac{2}{N}\left\{\left(\sum_{j=1}^{n-1}U_{jn}\right)+\varrho\left(\sum_{s=1}^{n-1}U_{ns}\right)\right\}U_{nn}\right)^4 \middle| \mathcal{F}_{N,n-1}\right]\leq\frac{C}{N^4}\left(n-1\right)^2.$$

Then we derive the following upper bound of $\sum_{j=1}^{2n-3}\sqrt{E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]}$ as

$$\begin{aligned}\sum_{j=1}^{2n-3}\sqrt{E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]}\leq&\sum_{t=2}^{n-1}\sqrt{\frac{C}{N^4}\left(\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2+M\left(t-1\right)\right)^2}\\&+\sum_{i=2}^{n-1}\sqrt{\frac{C}{N^4}\left(\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2+M\left(i-1\right)\right)^2}+\sqrt{\frac{C}{N^4}\left(n-1\right)^2}.\end{aligned}$$

Using the above bound, we obtain

$$\begin{aligned}&E\left[\left.\frac{Z_{N,n}^4}{N^2}\right|\mathcal{F}_{N,n-1}\right]=E\left[\frac{1}{N^2}\left(\sum_{j=1}^{2n-3}W_j\right)^4\middle|\mathcal{F}_{N,n-1}\right]\leq\frac{3}{N^2}\left(\sum_{j=1}^{2n-3}\sqrt{E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]}\right)^2\\&\leq\frac{C}{N^6}\left(\sum_{t=2}^{n-1}\left(\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2+M\left(t-1\right)\right)+\sum_{i=2}^{n-1}\left(\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2+M\left(i-1\right)\right)+\left(n-1\right)\right)^2\\&=\frac{C}{N^6}(D_{n,1}+D_{n,2}+D_{n,3}+D_{n,4}+D_{n,5})^2\\&\leq\frac{C}{N^6}(D_{n,1}^2+D_{n,2}^2+D_{n,3}^2+D_{n,4}^2+D_{n,5}^2),\end{aligned}$$

where $D_{n,1}:=\sum_{t=2}^{n-1}\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^2$, $D_{n,2}:=M\sum_{t=2}^{n-1}(t-1)$, $D_{n,3}:=\sum_{i=2}^{n-1}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2$, $D_{n,4}:=M\sum_{i=2}^{n-1}(i-1)$ and $D_{n,5}:=n-1$. By the Cauchy-Schwartz inequality and (A.5), we have

$$E(D_{n,1}^2)\leq n\sum_{t=2}^{n-1}E\left[\left(\sum_{j=1}^{n-1}\frac{U_{jt}}{\sigma_U}\right)^4\right]\leq n\sum_{t=2}^{n-1}\left((n-1)(K+1)+3(n-1)(n-2)\right)\leq Cn^4.$$

Similarly, we can show

$$E(D_{n,3}^2)\leq Cn^4.$$

Also, it is straight forward to see that

$$D_{n,2}^2+D_{n,4}^2+D_{n,5}^2\leq Cn^4$$

for $3 \leq n$. Then,

$$\sum_{n=3}^{N} E\left[\frac{Z_{N,n}^4}{N^2}\right] = \sum_{n=3}^{N} E\left[E\left[\left.\frac{Z_{N,n}^4}{N^2}\right|\mathcal{F}_{N,n-1}\right]\right] \leq \frac{C}{N^6}\sum_{n=3}^{N} n^4 = o(1),$$

as desired for the lemma. ■

A.2.2 Analysis of C_{NT}

We will assume that $n > N$ below. Recall that

$$Z_{N,n} := \frac{2}{N}\sum_{i=2}^{N}\left\{\left(\sum_{j=1}^{i-1} U_{jn}\right) + \varrho\left(\sum_{s=1}^{n-1} U_{is}\right)\right\} U_{in}$$

for $N+1 \leq n \leq T$.

Lemma 11 *For $N+1 \leq n \leq T$, we have*

$$E\left[\left.\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right] = \frac{2}{N^2}(N-1)\,\sigma_U^4 + \frac{4\varrho^2\sigma_U^2}{N^3}\sum_{i=2}^{N}\left(\sum_{s=1}^{n-1} U_{is}\right)^2.$$

Proof. It follows from

$$\begin{aligned} E\left[\left.Z_{N,n}^2\right|\mathcal{F}_{N,n-1}\right] &= \frac{4}{N^2}\sum_{i=2}^{N}\left\{(i-1)\,\sigma_U^2 + \varrho^2\left(\sum_{s=1}^{n-1} U_{is}\right)^2\right\}\sigma_U^2 \\ &= \frac{2}{N}(N-1)\,\sigma_U^4 + \frac{4\varrho^2\sigma_U^2}{N^2}\sum_{i=2}^{N}\left(\sum_{s=1}^{n-1} U_{is}\right)^2. \end{aligned}$$

■

Lemma 12 $E\left[\left|\sum_{n=N+1}^{T}\left(E\left[\left.\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right] - E\left[\frac{Z_{N,n}^2}{N}\right]\right)\right|^2\right] = o(1).$

Proof. Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} &\left|\sum_{n=N+1}^{T}\left(E\left[\left.\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right] - E\left[\frac{Z_{N,n}^2}{N}\right]\right)\right| \\ &= \left|\frac{1}{N}\sum_{n=N+1}^{T}\left(E\left[\left.Z_{N,n}^2\right|\mathcal{F}_{N,n-1}\right] - E\left[Z_{N,n}^2\right]\right)\right| \\ &\leq \frac{T-N}{N}\left(\frac{1}{T-N}\sum_{n=N+1}^{T}\left(E\left[\left.Z_{N,n}^2\right|\mathcal{F}_{N,n-1}\right] - E\left[Z_{N,n}^2\right]\right)^2\right)^{1/2} \end{aligned}$$

$$= \frac{\sqrt{T-N}}{\sqrt{N}} \left(\frac{1}{N} \sum_{n=N+1}^{T} \left(E\left[Z_{N,n}^2 \middle| \mathcal{F}_{N,n-1} \right] - E\left[Z_{N,n}^2 \right] \right)^2 \right)^{1/2}.$$

Since $\frac{T-N}{N} \to \frac{1}{\kappa} - 1$, the lemma follows if we show

$$E\left[\frac{1}{N} \sum_{n=N+1}^{T} \left(E\left[Z_{N,n}^2 \middle| \mathcal{F}_{N,n-1} \right] - E\left[Z_{N,n}^2 \right] \right)^2 \right] = o(1).$$

We have

$$\begin{aligned}
&\frac{1}{\sigma_U^8} E\left[\frac{1}{N} \sum_{n=N+1}^{T} \left(E\left[Z_{N,n}^2 \middle| \mathcal{F}_{N,n-1} \right] - E\left[Z_{N,n}^2 \right] \right)^2 \right] \\
&= \frac{1}{\sigma_U^8} \frac{1}{N} \sum_{n=N+1}^{T} \mathrm{Var}\left(E\left[Z_{N,n}^2 \middle| \mathcal{F}_{N,n-1} \right] \right) \\
&= \frac{1}{\sigma_U^8} \frac{1}{N} \sum_{n=N+1}^{T} \mathrm{Var}\left(\frac{4\varrho^2 \sigma_U^4}{N^2} \sum_{i=2}^{N} \left(\sum_{s=1}^{n-1} \frac{U_{is}}{\sigma_U} \right)^2 \right) \\
&= \frac{16\varrho^4}{N^5} \sum_{n=N+1}^{T} \mathrm{Var}\left(\sum_{i=2}^{N} \left(\sum_{s=1}^{n-1} \frac{U_{is}}{\sigma_U} \right)^2 \right) \\
&= \frac{16\varrho^4}{N^5} \sum_{n=N+1}^{T} \sum_{i=2}^{N} \mathrm{Var}\left(\left(\sum_{s=1}^{n-1} \frac{U_{is}}{\sigma_U} \right)^2 \right),
\end{aligned}$$

where we used Lemma 11 in the second equality, and in the last equality, we used the fact that $\left(\sum_{s=1}^{n-1} \frac{U_{is}}{\sigma_U} \right)^2$ are independent over i.

Using Lemma 7, we get

$$\mathrm{Var}\left(\left(\sum_{s=1}^{n-1} \frac{U_{is}}{\sigma_U} \right)^2 \right) = 2\left(n-1\right)^2 + \left(K-2\right)\left(n-1\right).$$

It follows that

$$\begin{aligned}
&\frac{1}{\sigma_U^8} E\left[\frac{1}{N} \sum_{n=N+1}^{T} \left(E\left[Z_{N,n}^2 \middle| \mathcal{F}_{N,n-1} \right] - E\left[Z_{N,n}^2 \right] \right)^2 \right] \\
&= \frac{16\varrho^4}{N^5} \sum_{n=N+1}^{T} \sum_{i=2}^{N} \left(2\left(n-1\right)^2 + \left(K-2\right)\left(n-1\right) \right) \\
&= O(N^{-1}) = o(1),
\end{aligned}$$

from which we get the desired conclusion. ■

Lemma 13 $\sum_{n=N+1}^{T} E\left(E\left[\left.\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right]\right) = 2\left(T-N\right)\frac{N+N\varrho^2+T\varrho^2}{N^2}\sigma_U^4 + o\left(1\right).$

Proof. Using Lemma 11, we have

$$\begin{aligned}
\frac{1}{\sigma_U^4}\sum_{n=N+1}^{T} E\left(E\left[\left.\frac{Z_{N,n}^2}{N}\right|\mathcal{F}_{N,n-1}\right]\right) &= \sum_{n=N+1}^{T} E\left[\frac{2}{N^2}\left(N-1\right)+\frac{4\varrho^2}{N^3}\sum_{i=2}^{N}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right] \\
&= \sum_{n=N+1}^{T}\left(\frac{2}{N^2}\left(N-1\right)+\frac{4\varrho^2}{N^3}\sum_{i=2}^{N}\left(n-1\right)\right) \\
&= \frac{2}{N}T-\frac{2}{N^2}T+\frac{2}{N}-2+\frac{4}{N}\varrho^2-2\varrho^2 \\
&\quad -\frac{2}{N^2}T\varrho^2+\frac{2}{N^2}T^2\varrho^2-\frac{2}{N^3}T^2\varrho^2+\frac{2}{N^3}T\varrho^2-\frac{2}{N^2}\varrho^2 \\
&= -2\varrho^2+\frac{2}{N}T+\frac{2}{N^2}T^2\varrho^2-2+o\left(1\right).
\end{aligned}$$

■

Lemma 14 $\sum_{n=N+1}^{T} E\left[\frac{Z_{N,n}^4}{N^2}\right] = o(1).$

Proof. Recall that for $N+1 \le n \le T$, we have

$$Z_{N,n} := \frac{2}{N}\sum_{i=2}^{N}\left\{\left(\sum_{j=1}^{i-1}U_{jn}\right)+\varrho\left(\sum_{s=1}^{n-1}U_{is}\right)\right\}U_{in},$$

so we see that $Z_{N,n}$ consists of $N-1$ terms like $\frac{2}{N}\left\{\left(\sum_{j=1}^{i-1}U_{jn}\right)+\varrho\left(\sum_{s=1}^{n-1}U_{is}\right)\right\}U_{in}$. Let's call them W_j with $j=1,\ldots,N-1$. We first note that W_j and $W_{j'}$ are independent conditional on $\mathcal{F}_{N,n-1}$ due to the independence among U_{in} $(i=2,\ldots,N)$. Then, we have

$$E\left[\left.\left(\sum_{j=1}^{N-1}W_j\right)^4\right|\mathcal{F}_{N,n-1}\right] = \sum_{j=1}^{N-1}E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]+3\sum_{j\neq j'}E\left[\left.W_j^2W_{j'}^2\right|\mathcal{F}_{N,n-1}\right].$$

Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
E\left[\left.\left(\sum_{j=1}^{N-1}W_j\right)^4\right|\mathcal{F}_{N,n-1}\right] &\le \sum_{j=1}^{N-1}E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]+3\sum_{j\neq j'}\sqrt{E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]}\sqrt{E\left[\left.W_{j'}^4\right|\mathcal{F}_{N,n-1}\right]} \\
&\le \sum_{j=1}^{N-1}3E\left[\left.W_j^4\right|\mathcal{F}_{n-1}\right]+3\sum_{j\neq j'}\sqrt{E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]}\sqrt{E\left[\left.W_{j'}^4\right|\mathcal{F}_{N,n-1}\right]}
\end{aligned}$$

$$= 3\left(\sum_{j=1}^{N-1}\sqrt{E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]}\right)^2.$$

As in the proof of Lemma 10, we can bound

$$E\left[\left.\left(\frac{2}{N}\left\{\left(\sum_{j=1}^{i-1}U_{jn}\right)+\varrho\left(\sum_{s=1}^{n-1}U_{is}\right)\right\}U_{in}\right)^4\right|\mathcal{F}_{N,n-1}\right]\leq\frac{C}{N^4}\left(\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2+M\left(i-1\right)\right)^2$$

for some finite generic constant C. Therefore, we get the following bound

$$\begin{aligned}
E\left[\left.\frac{Z_{N,n}^4}{N^2}\right|\mathcal{F}_{N,n-1}\right] &= E\left[\left.\frac{1}{N^2}\left(\sum_{j=1}^{N-1}W_i\right)^4\right|\mathcal{F}_{N,n-1}\right]\leq\frac{3}{N^2}\left(\sum_{j=1}^{N-1}\sqrt{E\left[\left.W_j^4\right|\mathcal{F}_{N,n-1}\right]}\right)^2\\
&\leq\frac{C}{N^6}\left(\sum_{i=2}^{N}\left(M\left(i-1\right)+\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right)\right)^2\\
&=\frac{C}{N^6}\left(\frac{M}{2}N\left(N-1\right)+\sum_{i=2}^{N}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right)^2\\
&\leq\frac{C}{N^6}\left(2\left(\frac{M}{2}N\left(N-1\right)\right)^2+2\left(\sum_{i=2}^{N}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^2\right)^2\right)\\
&\leq\frac{C}{N^6}\left(N^2(N-1)^2+(N-1)\sum_{i=2}^{N}\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^4\right),
\end{aligned}$$

where we use the Cauchy-Schwarz inequality in the last line. By (A.5) and symmetry, we have

$$E\left(\left(\sum_{s=1}^{n-1}\frac{U_{is}}{\sigma_U}\right)^4\right)=(n-1)(K+1)+3(n-1)(n-2).$$

Therefore,

$$\begin{aligned}
\sum_{n=N+1}^{T}E\left[\frac{Z_{N,n}^4}{N^2}\right] &= \sum_{n=N+1}^{T}E\left[E\left[\left.\frac{Z_{N,n}^4}{N^2}\right|\mathcal{F}_{N,n-1}\right]\right]\\
&\leq C\left(\frac{(T-N)N^2(N-1)^2}{N^6}+\frac{(N-1)^2}{N^2}\frac{1}{N^4}\sum_{n=N+1}^{T}\left((n-1)(K+1)+3(n-1)(n-2)\right)\right)\\
&= o(1),
\end{aligned}$$

as required for the lemma. ■

A.3 Proof of Lemma 2

Note that the following equality holds if $\varrho = 0$:

$$\begin{aligned}
\frac{1}{N\sqrt{T}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt} &= \frac{\sqrt{N}}{\sqrt{T}}\left(\frac{1}{N\sqrt{N}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt} + \varrho\frac{1}{N\sqrt{N}}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}\right) \\
&= \frac{\sqrt{N}}{\sqrt{T}}A_{NT},
\end{aligned}$$

so its marginal asymptotic distribution can be deduced from Lemma 1 by taking $\varrho = 0$ and multiplying the asymptotic variance by $\left(\lim \frac{\sqrt{N}}{\sqrt{T}}\right)^2 = \kappa$. In other words, Lemma 1 implies that

$$\frac{1}{N\sqrt{T}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt} \Rightarrow N\left(0, 2\sigma_U^4\right).$$

By symmetry, we can deduce that

$$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is} \Rightarrow N\left(0, 2\sigma_U^4\right).$$

Lemma 1 implies that in addition to these marginal asymptotic results, their joint asymptotic distribution is also normal. Letting *cov* denote the asymptotic covariance, we see that the asymptotic variance of $\frac{1}{N\sqrt{T}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt}+\xi\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}$ would be

$$2\sigma_U^4 + \xi^2\cdot 2\sigma_U^4 + 2\xi cov = 2\sigma_U^4 + \varrho^2\frac{1}{\kappa}\cdot 2\sigma_U^4 + 2\xi cov = 2\left(1+\frac{\varrho^2}{\kappa}\right)\sigma_U^4 + 2\xi cov.$$

On the other hand, the joint distribution can be taken care of by analyzing

$$\begin{aligned}
&\frac{1}{N\sqrt{T}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt} + \xi\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is} \\
&= \frac{\sqrt{N}}{\sqrt{T}}\left(\frac{1}{N\sqrt{N}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt} + \xi\sqrt{\frac{N}{T}}\frac{1}{N\sqrt{N}}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}\right) \\
&= \frac{\sqrt{N}}{\sqrt{T}}\left(\frac{1}{N\sqrt{N}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt} + \varrho\frac{1}{N\sqrt{N}}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}\right) \\
&= \sqrt{\frac{N}{T}}A_{NT}
\end{aligned}$$

for $\varrho = \xi\sqrt{\kappa}$, where we assume for convenience that N/T is fixed. According to Lemma 1, we obtain that the asymptotic variance of $\frac{1}{N\sqrt{T}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt}+\xi\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}$ is equal to $\frac{N}{T}\cdot 2\left(\frac{1}{\kappa}+\frac{\varrho^2}{\kappa^2}\right)\sigma_U^4 = 2\left(1+\frac{\varrho^2}{\kappa}\right)\sigma_U^4$ regardless of the value of ξ. So, we conclude that the asymptotic covariance *cov* should be equal to zero. To conclude, Lemma 1 implies that

$$\left(\frac{1}{N\sqrt{T}}\sum_{i,j=1,i\neq j}^{N}\sum_{t=1}^{T}U_{it}U_{jt},\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t,s=1,t\neq s}^{T}U_{it}U_{is}\right)' \Rightarrow N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}2\sigma_U^4 & 0\\ 0 & 2\sigma_U^4\end{bmatrix}\right).$$

A.4 Proof of Theorem 1

We assume that $N = T$ and examine the contribution of the term with $i = t$. In other words, for (12), we are looking at

$$\begin{aligned}
&\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\sum_{t=1}^{N}\frac{\partial^2 \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\left(\sum_{t=1}^{N}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\\
&-\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\sum_{t\neq i}^{N}\frac{\partial^2 \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1^2}-\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\left(\sum_{t\neq i}^{N}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\\
&=\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\frac{\partial^2 \ln f\left(Y_{ii}|\, X_{ii},\theta_0\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\left(\frac{\partial \ln f\left(Y_{ii}|\, X_{ii},\theta_0\right)}{\partial\theta_1}\right)^2\\
&+\frac{2}{\sqrt{N}N}\sum_{i=1}^{N}\left(\frac{\partial \ln f\left(Y_{ii}|\, X_{ii},\theta_0\right)}{\partial\theta_1}\right)\left(\sum_{t\neq i}^{N}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right)\\
&=\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\left(\frac{\partial^2 \ln f\left(Y_{ii}|\, X_{ii},\theta_0\right)}{\partial\theta_1^2}+\left(\frac{\partial \ln f\left(Y_{ii}|\, X_{ii},\theta_0\right)}{\partial\theta_1}\right)^2\right) &\text{(A.7)}\\
&+\frac{2}{N}\sum_{i=1}^{N}\left(\frac{\partial \ln f\left(Y_{ii}|\, X_{ii},\theta_0\right)}{\partial\theta_1}\right)\left(\frac{1}{\sqrt{N}}\sum_{t\neq i}^{N}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right). &\text{(A.8)}
\end{aligned}$$

Because $\frac{\partial^2 \ln f(Y_{ii}|X_{ii},\theta_0)}{\partial\theta_1^2}+\left(\frac{\partial \ln f(Y_{ii}|X_{ii},\theta_0)}{\partial\theta_1}\right)^2$ is iid over i with mean zero, we see that the term in (A.7) is of order $O\left(N^{-1}\right)$. As for the term in (A.8), we see that

$$E\left[\left(\frac{\partial \ln f\left(Y_{ii}|\, X_{ii},\theta_0\right)}{\partial\theta_1}\right)\left(\frac{1}{\sqrt{N}}\sum_{t\neq i}^{N}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right)\right]=0,$$

and

$$
\begin{aligned}
&E\left[\left(\frac{\partial \ln f\left(\left.Y_{ii}\right|X_{ii},\theta_0\right)}{\partial\theta_1}\right)^2\left(\frac{1}{\sqrt{N}}\sum_{t\neq i}^{N}\frac{\partial \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\right]\\
&=E\left[\left(\frac{\partial \ln f\left(\left.Y_{ii}\right|X_{ii},\theta_0\right)}{\partial\theta_1}\right)^2\right]E\left[\left(\frac{1}{\sqrt{N}}\sum_{t\neq i}^{N}\frac{\partial \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\right]\\
&=O\left(1\right),
\end{aligned}
$$

so

$$
E\left[\frac{2}{N}\sum_{i=1}^{N}\left(\frac{\partial \ln f\left(\left.Y_{ii}\right|X_{ii},\theta_0\right)}{\partial\theta_1}\right)\left(\frac{1}{\sqrt{N}}\sum_{t\neq i}^{N}\frac{\partial \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1}\right)\right]=0,
$$

$$
E\left[\left(\frac{2}{N}\sum_{i=1}^{N}\left(\frac{\partial \ln f\left(\left.Y_{ii}\right|X_{ii},\theta_0\right)}{\partial\theta_1}\right)\left(\frac{1}{\sqrt{N}}\sum_{t\neq i}^{N}\frac{\partial \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1}\right)\right)^2\right]=O\left(N^{-1}\right).
$$

Therefore, the sum of (A.7) and (A.8) are of order $O\left(N^{-1/2}\right)$; that is, ignoring the $i=t$ terms in (12) does not add any complications to the asymptotic distribution analysis. The same conclusion can be drawn for (13), too. This shows that the difference between the two test statistics LM_d and BP_{two} is small, as argued at the beginning of Section 2.1.4.

In addition, recall that we have shown after (15) that the two statistics in Lemma 2 are asymptotically equivalent to (12) and (13), respectively. Then the result of this theorem follows from Lemma 2.

A.5 Proof of Lemma 3

Letting $Q_{ij}:=\frac{2}{N}\left(\sum_{j'=1}^{j-1}U_{ij'}\right)U_{ij}$ for $2\leq i,j,j'\leq N$, we can follow the argument in Remark 5 to rewrite

$$
\begin{aligned}
\sqrt{N}A_N&=\frac{2}{N}\sum_{i=2}^{N}\sum_{j=2}^{N}\left(\sum_{j'=1}^{j-1}U_{ij'}\right)U_{ij}+\frac{2}{N}\sum_{j=2}^{N}\left(\sum_{j'=1}^{j-1}U_{1j'}\right)U_{1j}\\
&=\sum_{i=2}^{N}\sum_{j=2}^{N}Q_{ij}+O_p(1).
\end{aligned}
$$

For $n\geq 3$, define

$$
Z_{N,n}^{*}:=\sum_{i=2}^{n-1}Q_{ni}+\sum_{i=2}^{n-1}Q_{in}+Q_{nn}
$$

$$= \frac{2}{N}\sum_{i=2}^{n-1}\left(\sum_{j=1}^{i-1}U_{nj}\right)U_{ni} + \frac{2}{N}\sum_{i=2}^{n-1}\left(\sum_{j=1}^{n-1}U_{ij}\right)U_{in} + \frac{2}{N}\left(\sum_{j=1}^{n-1}U_{nj}\right)U_{nn}$$
$$= \frac{2}{N}\sum_{i=2}^{n-1}\left(\sum_{j=1}^{i-1}U_{nj}\right)U_{ni} + \frac{2}{N}\sum_{i=2}^{n-1}\left(\sum_{j=1}^{n-1}U_{ij}\right)U_{in},$$

where the last equality holds because $U_{nn} = 0$. Then we get

$$\sqrt{N}A_N = \sum_{n=2}^{N} Z^*_{N,n} + O_p(1).$$

In order to overcome the difficulty related to the dependence, note that we can write the second component of $Z^*_{N,n}$ above as

$$\begin{aligned}
\sum_{i=2}^{n-1}\left(\sum_{j=1}^{n-1}U_{ij}\right)U_{in} &= \sum_{i=2}^{n-1}U_{i1}U_{in} + \sum_{i=2}^{n-1}\left(\sum_{j=2}^{n-1}U_{ij}\right)U_{in} \\
&= \sum_{i=2}^{n-1}U_{i1}U_{in} + \sum_{i,j=2,i\neq j}^{n-1}U_{ij}U_{in} \\
&= \sum_{i=2}^{n-1}U_{i1}U_{in} + \sum_{i=2,i<j}^{n-1}U_{ij}U_{in} + \sum_{i=2,i>j}^{n-1}U_{ij}U_{in} \\
&= \sum_{i=2}^{n-1}U_{i1}U_{in} + \sum_{i=2,i<j}^{n-1}U_{ij}U_{in} + \sum_{i=2,j>i}^{n-1}U_{ji}U_{jn} \\
&= \sum_{i=2}^{n-1}U_{i1}U_{in} + \sum_{i=2,i<j}^{n-1}\left(U_{ij}U_{in} + U_{ij}U_{jn}\right) \\
&= \sum_{i=2}^{n-1}U_{i1}U_{ni} + \sum_{i=2,i<j}^{n-1}U_{ij}\left(U_{ni} + U_{nj}\right),
\end{aligned}$$

where we used $U_{ii} = 0$ in the second equality, simply interchanged the indices i and j in the last term of the fourth equality, and the fifth and the sixth equalities hold because $U_{ji} = U_{ij}$. It follows that

$$Z^*_{N,n} = \frac{2}{N}\sum_{i=2}^{n-1}U_{n1}U_{ni} + \frac{2}{N}\sum_{i=2,i<j}^{n-1}U_{ni}U_{nj} + \frac{2}{N}\sum_{i=2}^{n-1}U_{i1}U_{ni} + \frac{2}{N}\sum_{i=2,i<j}^{n-1}U_{ij}\left(U_{ni} + U_{nj}\right).$$

Note that

$$E\left[\frac{1}{N}\sum_{n=2}^{N}\sum_{i=2}^{n-1}U_{n1}U_{ni}\right] = 0,$$

$$E\left[\left(\frac{1}{N}\sum_{n=2}^{N}\sum_{i=2}^{n-1}U_{n1}U_{ni}\right)^2\right]=\frac{1}{N^2}\sum_{n=2}^{N}\sum_{i=2}^{n-1}\sigma_U^4=O\left(1\right),$$

and

$$E\left[\frac{1}{N}\sum_{n=2}^{N}\sum_{i=2}^{n-1}U_{i1}U_{ni}\right]=0,$$

$$E\left[\left(\frac{1}{N}\sum_{n=2}^{N}\sum_{i=2}^{n-1}U_{i1}U_{ni}\right)^2\right]=\frac{1}{N^2}\sum_{n=2}^{N}\sum_{i=2}^{n-1}\sigma_U^4=O\left(1\right),$$

so we can further write

$$\sqrt{N}A_N=\frac{2}{N}\sum_{n=2}^{N}\sum_{i=2,i<j}^{n-1}U_{ni}U_{nj}+\frac{2}{N}\sum_{n=2}^{N}\sum_{i=2,i<j}^{n-1}U_{ij}\left(U_{ni}+U_{nj}\right)+O_p\left(1\right).$$

For $n\geq 4$, define

$$Z_{N,n}:=\frac{2}{N}\sum_{i=2,i<j}^{n-1}U_{ni}U_{nj}+\frac{2}{N}\sum_{i=2,i<j}^{n-1}U_{ij}\left(U_{ni}+U_{nj}\right),$$

then we have

$$\sqrt{N}A_N=\sum_{n=4}^{N}Z_{N,n}+O_p(1). \tag{A.9}$$

And because U_{ij} are iid, we have

$$E\left[\left.Z_{N,n}\right|\mathcal{F}_{N,n-1}\right]=0,$$

where $\mathcal{F}_{N,n}, 3\leq n\leq N$, is the filtration defined in page 33 with $U_{ii}=0$.

Again, we will work with Hall and Heyde (1980)'s Corollary 3.3 to obtain the CLT. By Lemmas 16, 17 and 18 below, we see that Hall and Heyde (1980)'s (3.38) and (3.40) are satisfied with $p=2$. Therefore, we can conclude by Hall and Heyde (1980)'s Corollary 3.3 that

$$\sum_{n=4}^{N}\frac{Z_{N,n}}{\sqrt{N}}\Rightarrow N\left(0,2\sigma_U^4\right),$$

which together with (A.9) implies the result of Lemma 3.

Lemma 15

$$E\left[\left.Z_{N,n}^2\right|\mathcal{F}_{N,n-1}\right]=\frac{2\left(n-2\right)\left(n-3\right)}{N^2}\sigma_U^4+\frac{8\sigma_U^4}{N^2}\sum_{i=2,i<j}^{n-1}\left(\frac{U_{ij}}{\sigma_U}\right)^2$$

$$+\frac{4\sigma_U^2}{N^2}\sum_{i=2}^{n-3}\sum_{j>i,j'>i,j'\neq j}^{n-1}U_{ij}U_{ij'}+\frac{4\sigma_U^2}{N^2}\sum_{j=4}^{n-1}\sum_{i<j,i'<j,i'\neq i}^{n-2}U_{ij}U_{i'j}$$
$$+\frac{4\sigma_U^2}{N^2}\sum_{2=i'<i<j}^{n-1}U_{ij}U_{i'i}+\frac{4\sigma_U^2}{N^2}\sum_{2=i<j<j'}^{n-1}U_{ij}U_{jj'}.$$

Proof. We have

$$\begin{aligned}
Z_{N,n}^2 &= \left(\frac{2}{N}\sum_{i=2,i<j}^{n-1}\left(U_{ni}U_{nj}+U_{ij}\left(U_{ni}+U_{nj}\right)\right)\right)^2\\
&= \frac{4}{N^2}\sum_{i=2,i<j}^{n-1}\left(U_{ni}U_{nj}+U_{ij}\left(U_{ni}+U_{nj}\right)\right)^2\\
&+\frac{4}{N^2}\sum_{i=2,i<j,i'=2,i'<j',i'\neq i\text{ or }j'\neq j}^{n-1}\left(U_{ni}U_{nj}+U_{ij}\left(U_{ni}+U_{nj}\right)\right)\left(U_{ni'}U_{nj'}+U_{i'j'}\left(U_{ni'}+U_{nj'}\right)\right).
\end{aligned}$$

We start with the analysis of the first term on the right, and note that

$$\begin{aligned}
E\left[\frac{4}{N^2}\sum_{i=2,i<j}^{n-1}\left(U_{ni}U_{nj}+U_{ij}\left(U_{ni}+U_{nj}\right)\right)^2\middle|\mathcal{F}_{N,n-1}\right] &= \frac{4}{N^2}\sum_{i=2,i<j}^{n-1}\left(\sigma_U^4+U_{ij}^2\left(2\sigma_U^2\right)\right)\\
&= \frac{2\left(n-2\right)\left(n-3\right)}{N^2}\sigma_U^4+\frac{8\sigma_U^2}{N^2}\sum_{i=2,i<j}^{n-1}U_{ij}^2.
\end{aligned}$$

As for the second term on the right, we need to consider the following four cases:

Case I: $i'=i, j'\neq j$, leading to

$$\frac{4}{N^2}\sum_{i=2,i<j,i'=2,i'<j',i'=i,j'\neq j}^{n-1}E\left[\left(U_{ni}U_{nj}+U_{ij}\left(U_{ni}+U_{nj}\right)\right)\left(U_{ni}U_{nj'}+U_{ij'}\left(U_{ni}+U_{nj'}\right)\right)\middle|\mathcal{F}_{N,n-1}\right]$$
$$=\frac{4}{N^2}\sum_{i=2}^{n-3}\sum_{j>i,j'>i,j'\neq j}^{n-1}U_{ij}U_{ij'}\sigma_U^2;$$

Case II: $i'\neq i, j'=j$, leading to

$$\frac{4}{N^2}\sum_{i=2,i<j,i'=2,i'<j',i'\neq i,j'=j}^{n-1}E\left[\left(U_{ni}U_{nj}+U_{ij}\left(U_{ni}+U_{nj}\right)\right)\left(U_{ni'}U_{nj}+U_{i'j}\left(U_{ni'}+U_{nj}\right)\right)\middle|\mathcal{F}_{N,n-1}\right]$$
$$=\frac{4}{N^2}\sum_{j=4}^{n-1}\sum_{i<j,i'<j,i'\neq i}^{n-2}U_{ij}U_{i'j}\sigma_U^2;$$

Case III: $i'<j'=i<j$, leading to

$$\frac{4}{N^2}\sum_{2=i'<j'=i<j}^{n-1}E\left[\left(U_{ni}U_{nj}+U_{ij}\left(U_{ni}+U_{nj}\right)\right)\left(U_{ni'}U_{ni}+U_{i'i}\left(U_{ni'}+U_{ni}\right)\right)\middle|\mathcal{F}_{N,n-1}\right]$$

$$= \frac{4}{N^2} \sum_{2=i'<i<j}^{n-1} U_{ij} U_{i'i} \sigma_U^2;$$

Case IV: $i < j = i' < j'$, leading to

$$\frac{4}{N^2} \sum_{2=i<j=i'<j'}^{n-1} E\left[\left(U_{ni} U_{nj} + U_{ij} \left(U_{ni} + U_{nj} \right) \right) \left(U_{nj} U_{nj'} + U_{jj'} \left(U_{nj} + U_{nj'} \right) \right) \middle| \mathcal{F}_{N,n-1} \right]$$
$$= \frac{4}{N^2} \sum_{2=i<j<j'}^{n-1} U_{ij} U_{jj'} \sigma_U^2.$$

Combining these four cases and the first term gives the result of the lemma. ■

Lemma 16 $E\left[\left| \sum_{n=4}^{N} \left(E\left[\frac{Z_{N,n}^2}{N} \middle| \mathcal{F}_{N,n-1} \right] - E\left[\frac{Z_{N,n}^2}{N} \right] \right) \right|^2 \right] = o\left(1\right).$

Proof. Since $\left(\frac{1}{N} \sum_{n=1}^{N} a_n \right)^2 \leq \frac{1}{N} \sum_{n=1}^{N} a_n^2$, the desired result of the lemma follows if we show

$$E\left[\frac{1}{N} \sum_{n=4}^{N} \left(E\left[Z_{N,n}^2 \middle| \mathcal{F}_{N,n-1} \right] - E\left[Z_{N,n}^2 \right] \right)^2 \right] = o(1).$$

We have

$$\frac{1}{\sigma_U^8} E\left[\frac{1}{N} \sum_{n=4}^{N} \left(E\left[Z_{N,n}^2 \middle| \mathcal{F}_{N,n-1} \right] - E\left[Z_{N,n}^2 \right] \right)^2 \right] = \frac{1}{\sigma_U^8} \frac{1}{N} \sum_{n=4}^{N} \mathrm{Var}\left(E\left[Z_{N,n}^2 \middle| \mathcal{F}_{N,n-1} \right] \right).$$

Using Lemma 15, it suffices to show that

$$\frac{1}{N} \sum_{n=4}^{N} \mathrm{Var}\left(\frac{8\sigma_U^4}{N^2} \sum_{i=2,i<j}^{n-1} \left(\frac{U_{ij}}{\sigma_U} \right)^2 \right) = o\left(1\right), \tag{A.10}$$

$$\frac{1}{N} \sum_{n=4}^{N} \mathrm{Var}\left(\frac{4\sigma_U^2}{N^2} \sum_{i=2}^{n-3} \sum_{j>i,j'>i,j'\neq j}^{n-1} U_{ij} U_{ij'} \right) = o\left(1\right), \tag{A.11}$$

$$\frac{1}{N} \sum_{n=4}^{N} \mathrm{Var}\left(\frac{4\sigma_U^2}{N^2} \sum_{j=4}^{n-1} \sum_{i<j,i'<j,i'\neq i}^{n-2} U_{ij} U_{i'j} \right) = o\left(1\right), \tag{A.12}$$

$$\frac{1}{N} \sum_{n=4}^{N} \mathrm{Var}\left(\frac{4\sigma_U^2}{N^2} \sum_{2=i'<i<j}^{n-1} U_{ij} U_{i'i} \right) = o\left(1\right), \tag{A.13}$$

$$\frac{1}{N} \sum_{n=4}^{N} \mathrm{Var}\left(\frac{4\sigma_U^2}{N^2} \sum_{2=i<j<j'}^{n-1} U_{ij} U_{jj'} \right) = o\left(1\right). \tag{A.14}$$

In order to prove (A.10), we note that by the iid of U_{ij},

$$
\begin{aligned}
\frac{1}{N}\sum_{n=4}^{N}\mathrm{Var}\left(\frac{8\sigma_U^4}{N^2}\sum_{i=2,i<j}^{n-1}\left(\frac{U_{ij}}{\sigma_U}\right)^2\right) &= \frac{C}{N^5}\sum_{n=4}^{N}\mathrm{Var}\left(\sum_{i=2,i<j}^{n-1}\left(\frac{U_{ij}}{\sigma_U}\right)^2\right) \\
&= \frac{C}{N^5}\sum_{n=4}^{N}\sum_{i=2,i<j}^{n-1}\mathrm{Var}\left(\left(\frac{U_{ij}}{\sigma_U}\right)^2\right) \\
&= O\left(N^{-2}\right) = o\left(1\right).
\end{aligned}
$$

In order to prove (A.11), we note that U_{ij} are iid, so

$$
\begin{aligned}
\frac{1}{N}\sum_{n=4}^{N}\mathrm{Var}\left(\frac{4\sigma_U^2}{N^2}\sum_{i=2}^{n-3}\sum_{j>i,j'>i,j'\neq j}^{n-1}U_{ij}U_{ij'}\right) &= \frac{C}{N^5}\sum_{n=4}^{N}E\left[\left(\sum_{i=2}^{n-3}\sum_{j>i,j'>i,j'\neq j}^{n-1}\frac{U_{ij}}{\sigma_U}\frac{U_{ij'}}{\sigma_U}\right)^2\right] \\
&= \frac{C}{N^5}\sum_{n=4}^{N}\sum_{i=2}^{n-3}\sum_{j>i,j'>i,j'\neq j}^{n-1}E\left[\left(\frac{U_{ij}}{\sigma_U}\right)^2\left(\frac{U_{ij'}}{\sigma_U}\right)^2\right] \\
&= \frac{C}{N^5}\sum_{n=4}^{N}\sum_{i=2}^{n-3}\sum_{j>i,j'>i,j'\neq j}^{n-1}\mathrm{Var}\left(\frac{U_{ij}}{\sigma_U}\right)\mathrm{Var}\left(\frac{U_{ij'}}{\sigma_U}\right) \\
&= O\left(N^{-1}\right) = o\left(1\right),
\end{aligned}
$$

where the first equality holds because $E\left(\sum_{i=2}^{n-3}\sum_{j>i,j'>i,j'\neq j}^{n-1}\frac{U_{ij}}{\sigma_U}\frac{U_{ij'}}{\sigma_U}\right) = 0$ due to $E(U_{ij}) = 0$ and the iid of U_{ij}, the second equality holds because we know that the cross product terms are zero due to the fact that $i < j$, $i < j'$ and $j \neq j'$, the third equality holds because $E\left(U_{ij}\right) = 0$ and U_{ij} are iid, and the fourth equality holds because we recognize that $\sum_{i=2}^{n-3}\sum_{j>i,j'>i,j'\neq j}^{n-1}\mathrm{Var}\left(\frac{U_{ij}}{\sigma_U}\right)\mathrm{Var}\left(\frac{U_{ij'}}{\sigma_U}\right) = O\left(n^3\right)$ uniformly. We can prove (A.12) - (A.14) similarly. ■

Lemma 17 $\sum_{n=4}^{N}E\left(E\left[\frac{Z_{N,n}^2}{N}\middle|\mathcal{F}_{N,n-1}\right]\right) = 2\sigma_U^4 + o\left(1\right)$.

Proof. Note that the last four terms in Lemma 15 have zero expectations due to the iid of U_{ij}. Therefore, Lemma 15 implies that

$$
\begin{aligned}
\frac{1}{\sigma_U^4}\sum_{n=4}^{N}E\left(E\left[\frac{Z_{N,n}^2}{N}\middle|\mathcal{F}_{N,n-1}\right]\right) &= \frac{1}{\sigma_U^4}\sum_{n=4}^{N}\frac{4\sigma_U^4}{N^3}\frac{\left(n-2\right)\left(n-3\right)}{2} + \frac{1}{\sigma_U^4}\sum_{n=4}^{N}\left(\frac{8\sigma_U^4}{N^3}\sum_{i=2,i<j}^{n-1}E\left[\left(\frac{U_{ij}}{\sigma_U}\right)^2\right]\right) \\
&= \frac{2}{3}\frac{\left(N-1\right)\left(N-2\right)\left(N-3\right)}{N^3} + \frac{8}{N^3}\sum_{n=4}^{N}\frac{\left(n-2\right)\left(n-3\right)}{2}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2}{3}\frac{(N-1)(N-2)(N-3)}{N^3} + \frac{4}{3}\frac{(N-1)(N-2)(N-3)}{N^3} \\
&\to 2,
\end{aligned}
$$

where in the second equality we used the definition of σ_U^2, and in the third equality we used the fact that $\sum_{n=4}^{N}(n-2)(n-3) = (N-1)(N-2)(N-3)/3$. ■

Lemma 18 $\sum_{n=4}^{N} E\left[\frac{Z_{N,n}^4}{N^2}\right] = o(1)$.

Proof. We rewrite

$$
\begin{aligned}
Z_{N,n} &= \frac{2}{N}\sum_{i=2,i<j}^{n-1} U_{ni}U_{nj} + \frac{2}{N}\sum_{i=2,i<j}^{n-1} U_{ij}\left(U_{ni}+U_{nj}\right) \\
&:= \frac{2}{N}\sum_{j=3}^{n-1} V_{jn,1} + \frac{2}{N}\sum_{j=2}^{n-2} V_{jn,2} + \frac{2}{N}\sum_{j=3}^{n-1} V_{jn,3},
\end{aligned}
$$

where we use the symmetry $U_{ij} = U_{ji}$, and define

$$
\begin{aligned}
V_{jn,1} &:= \left(\sum_{i=2}^{j-1} U_{in}\right) U_{jn} := W_{jn,1}U_{jn}, \\
V_{jn,2} &:= \left(\sum_{i=j+1}^{n-1} U_{ij}\right) U_{jn} := W_{jn,2}U_{jn}, \\
V_{jn,3} &:= \left(\sum_{i=2}^{j-1} U_{ji}\right) U_{jn} := W_{jn,3}U_{jn}.
\end{aligned}
$$

The desired result $\sum_{n=4}^{N} E\left[\frac{Z_{N,n}^4}{N^2}\right] = o(1)$ follows if we show $\frac{1}{N^6}\sum_{n=4}^{N} E\left[\left(\sum_{j=2}^{n-2} V_{jn,2}\right)^4\right] = o(1)$ and $\frac{1}{N^6}\sum_{n=4}^{N} E\left[\left(\sum_{j=3}^{n-1} V_{jn,m}\right)^4\right] = o(1)$ with $m = 1,3$, which follow if we show

$$
E\left[\left(\sum_{j=2}^{n-2} V_{jn,2}\right)^4\right] = O(n^4), \text{ and } E\left[\left(\sum_{j=3}^{n-1} V_{jn,m}\right)^4\right] = O(n^4), \tag{A.15}
$$

where $m = 1,3$.

First note that there is a generic constant C such that

$$
E\left[(W_{jn,1})^4\right] \le Cj^2, \;\; E\left[(W_{jn,2})^4\right] \le C(n-j)^2, \;\; E\left[(W_{jn,3})^4\right] \le Cj^2,
$$

because each $W_{jn,m}$ $(m = 1, 2, 3)$ is a sum of $(j-2)$ or $(n-j-1)$ iid random variables with mean zero and finite fourth moment. Our proof of (A.15) consists of three parts.

Part (a): We show $E\left[\left(\sum_{j=3}^{n-1} V_{jn,1}\right)^4\right] = O\left(n^4\right)$.

Notice that

$$E\left[V_{j_1n,1}V_{j_2n,1}V_{j_3n,1}V_{j_4n,1}\right] = 0$$

unless one of the following cases happens. $(a1)$: $j_1 = j_2 = j_3 = j_4$; $(a2)$: three indices are equal to each other and larger than the remaining one index (i.e., $(j_1 = j_3 = j_4) > j_2$); $(a3)$: two indices are equal to each other, the remaining two indices are equal to each other, and the pairs take different values (i.e., $(j_1 = j_3) \neq (j_2 = j_4)$); or $(a4)$: two indices are equal to each other, remaining two indices are different from each other and from the first two identical indices (i.e., $(j_1 = j_4) > j_2 > j_3$). Recall that the expectation is zero if $j_3 > (j_1 = j_4) > j_2$, $j_3 > j_2 > (j_1 = j_4)$, or $j_1 > j_2 > j_3 > j_4$.

Case $(a1)$: We have

$$\sum_{j=3}^{n-1} E\left[W_{jn,1}^4 U_{jn}^4\right] = \sum_{j=3}^{n-1} E\left[W_{jn,1}^4\right] E\left[U_{jn}^4\right] = O\left(\sum_{j=3}^{n-1} j^2\right) = O(n^3),$$

as desired.

Case $(a2)$: By the Hölder's inequality, we have

$$\begin{aligned}
&\sum_{j_1=3}^{n-1} \sum_{j_2=3,<j_1}^{n-1} E[(W_{j_1n,1}U_{j_1n})^3(W_{j_2n,1}U_{j_2n})] \\
\leq\ & \sum_{j_1=3}^{n-1} \sum_{j_2=3,<j_1}^{n-1} \left(E[(W_{j_1n,1}U_{j_1n})^3]^{4/3}\right)^{3/4} \left(E\left[(W_{j_2n,1}U_{j_2n})^4\right]\right)^{1/4} \\
\leq\ & \left[\sum_{j_1=3}^{n-1} \left(E[(W_{j_1n,1}U_{j_1n})^4]\right)^{3/4}\right] \left[\sum_{j_2=3}^{n-1} \left(E\left[(W_{j_2n,1}U_{j_2n})^4\right]\right)^{1/4}\right] \\
=\ & \left[\sum_{j=3}^{n-1} \left(E\left[W_{jn,1}^4\right]\right)^{3/4} \left(E\left[U_{jn}^4\right]\right)^{3/4}\right] \left[\sum_{j=3}^{n-1} \left(E\left[W_{jn,1}^4\right]\right)^{1/4} \left(E\left[U_{jn}^4\right]\right)^{1/4}\right] \\
=\ & O\left(\sum_{j=3}^{n-1} j^{3/2}\right) O\left(\sum_{j=3}^{n-1} j^{1/2}\right) = O\left(n^{5/2}\frac{1}{n}\sum_{j=3}^{n-1}\left(\frac{j}{n}\right)^{3/2}\right) O\left(n^{3/2}\frac{1}{n}\sum_{j=3}^{n-1}\left(\frac{j}{n}\right)^{1/2}\right) \\
=\ & n^4 O\left(\int_0^1 x^{3/2}dx\right) O\left(\int_0^1 x^{1/2}dx\right) = n^4 O(1)O(1) = O(n^4),
\end{aligned}$$

as desired.

Case $(a3)$: By using a similar argument in the proof of Case $(a2)$ with the Cauchy-Schwarz's inequality instead of the Hölder's inequality, we can show that

$$\sum_{j_1=3}^{n-1} \sum_{j_2=3,\neq j_1}^{n-1} E[(W_{j_1n,1}U_{j_1n})^2(W_{j_2n,1}U_{j_2n})^2] \leq O(n^4).$$

Case $(a4)$: We then have to deal with

$$\begin{aligned}
&\sum_{j_1=3}^{n-1} \sum_{j_2=3,<j_1}^{n-1} \sum_{j_3=3,<j_1,<j_2}^{n-1} E\left[V_{j_1n,1}^2 V_{j_2n,1} V_{j_3n,1}\right] \\
&= \sum_{j_3<j_2<j_1}^{n-1} \sum_{i_1=2}^{j_1-1} \sum_{i_1'=2}^{j_1-1} \sum_{i_2=2}^{j_2-1} \sum_{i_3=2}^{j_3-1} E\left[U_{i_1n}U_{i_1'n}U_{i_2n}U_{i_3n}U_{j_1n}^2U_{j_2n}U_{j_3n}\right] \\
&= \sum_{j_3<j_2<j_1}^{n-1} \sum_{i_1=2}^{j_1-1} \sum_{i_1'=2}^{j_1-1} \sum_{i_2=2}^{j_2-1} \sum_{i_3=2}^{j_3-1} E\left[U_{i_1n}U_{i_1'n}U_{i_2n}U_{i_3n}U_{j_2n}U_{j_3n}\right] E\left[U_{j_1n}^2\right].
\end{aligned}$$

We further consider several cases in case $(a4)$:

- Case I: Let's first assume that $i_2 = i_3$. If this is the case, we have $i_2 = i_3 < j_3 < j_2$, so we can have $E\left[U_{i_1n}U_{i_1'n}U_{i_2n}U_{i_3n}U_{j_2n}U_{j_3n}\right] \neq 0$ only if $(i_1, i_1') = (j_2, j_3)$ or $(i_1', i_1) = (j_2, j_3)$. This contributes

  $$\sum_{j_3<j_2<j_1}^{n-1} \sum_{i_3=2}^{j_3-1} E\left[U_{i_3n}^2U_{j_2n}^2U_{j_3n}^2\right] E\left(U_{j_1n}^2\right) = O\left(\sum_{j_1=4}^{n-1}\sum_{j_2=3}^{j_1-1}\sum_{j_3=2}^{j_2-1}(j_3-2)\right) = O\left(n^4\right)$$

  to the sum.

- Case II: Let's assume that $i_2 < i_3$. If this is the case, we have $i_2 < i_3 < j_3 < j_2$, so we have $E\left[U_{i_1n}U_{i_1'n}U_{i_2n}U_{i_3n}U_{j_2n}U_{j_3n}\right] = 0$ regardless of the value of (i_1, i_1').

- Case III: Let's assume that $i_2 > i_3$.

  - Case III (1): If we have $i_2 < j_3$, we have $i_3 < i_2 < j_3 < j_2$, $E\left[U_{i_1n}U_{i_1'n}U_{i_2n}U_{i_3n}U_{j_2n}U_{j_3n}\right] = 0$ regardless of the value of (i_1, i_1').

  - Case III (2): Let's assume that $i_2 > i_3$, and $j_3 \leq i_2$, which means that $i_3 < j_3 \leq i_2 < j_2$.

* Case III (2-a): Let's assume that $i_2 > i_3$, and $j_3 < i_2$, which means that $i_3 < j_3 < i_2 < j_2$. If this is the case, we would have $E\left[U_{i_1 n} U_{i_1' n} U_{i_2 n} U_{i_3 n} U_{j_2 n} U_{j_3 n}\right] = 0$ regardless of the value of (i_1, i_1').

* Case III (2-b): Let's assume that $i_2 > i_3$, and $j_3 = i_2$, which means that $i_3 < j_3 = i_2 < j_2$. If this is the case, we would have $E\left[U_{i_1 n} U_{i_1' n} U_{i_2 n} U_{i_3 n} U_{j_2 n} U_{j_3 n}\right] \neq 0$ only if $(i_1, i_1') = (i_3, j_2)$ or $(i_1, i_1') = (j_2, i_3)$. This contributes

$$\sum_{j_3 < j_2 < j_1}^{n-1} \sum_{i_1=2}^{j_3-1} E\left[U_{i_1 n}^2 U_{j_2 n}^2 U_{j_3 n}^2\right] E\left(U_{j_1 n}^2\right) + \sum_{j_3 < j_2 < j_1}^{n-1} \sum_{i_3=2}^{j_3-1} E\left[U_{i_3 n}^2 U_{j_2 n}^2 U_{j_3 n}^2\right] E\left(U_{j_1 n}^2\right)$$
$$= O\left(\sum_{j_1=4}^{n-1} \sum_{j_2=3}^{j_1-1} \sum_{j_3=2}^{j_2-1} (j_3 - 2)\right) + O\left(\sum_{j_1=4}^{n-1} \sum_{j_2=3}^{j_1-1} \sum_{j_3=2}^{j_2-1} (j_3 - 2)\right) = O\left(n^4\right).$$

Part (b): We show $E\left[\left(\sum_{j=3}^{n-1} V_{jn,2}\right)^4\right] = O(n^4)$.

Notice that

$$E\left[V_{j_1 n,2} V_{j_2 n,2} V_{j_3 n,2} V_{j_4 n,2}\right] = 0$$

unless one of the following cases happens. $(b1)$: $j_1 = j_2 = j_3 = j_4$; or $(b2)$: two indices are equal to each other, the remaining two indices are equal to each other, and the pairs take different values (i.e., $(j_1 = j_3) \neq (j_2 = j_4)$). The reason is as follows. Suppose that one index, say j_1, is different from the other three indices, say j_2, j_3, j_4. Then, $U_{j_1,n}$ is independent of the rest, and so we have

$$E\left[V_{j_1 n,2} V_{j_2 n,2} V_{j_3 n,2} V_{j_4 n,2}\right]$$
$$= E\left[W_{n-1 j_1,2} W_{n-1 j_2,2} U_{j_2 n} W_{n-1 j_3,2} U_{j_3,n} W_{n-1 j_4,2} U_{j_4,n}\right] E\left[U_{j_1,n}\right] = 0.$$

The desired results in the cases $(b1)$ and $(b2)$ follow by the similar arguments used in cases $(a1)$ and $(a3)$, respectively.

Part (c): The desired result $E\left[\left(\sum_{j=3}^{n-1} V_{jn,3}\right)^4\right] = O(n^4)$ follows by the same argument used in Part (b). ■

B Detailed Calculations for Section 3

B.1 Regularity Conditions

Assumption 1 *The observed data Z_i ($i = 1, \ldots, N$) are independently and identically distributed. Z_i belongs to a measure space $\mathcal{Z}$ and consists of two subvectors X_i and Y_i such that $Z_i = (Y_i', X_i')'$, and the conditional probability density function of Y_i given X_i is $h(y|x, \theta_0, \eta)$, where θ_0 is a q-dimensional parameter and η is a scalar parameter.*

Let $\theta \in \Theta \subset \mathbb{R}^q$. Let $\gamma := (\theta', \eta)'$ and $\gamma \in \Gamma$. Let θ_j ($j = 1, \ldots, q$) denote the jth element of θ.

Assumption 2 *For all θ in Θ and almost all z in $\mathcal{Z}$, $h(y|x, \theta, 0) = f(y|x, \theta)$.*

Assumption 3 *Let ε_i be a random variable that is independent of X_i and has a probability density function $k(\cdot)$ such that $\int ek(e)de = 0$ and $\int e^2 k(e)de = \sigma_\varepsilon^2$. Define $h(y|x, \gamma) := f(y|x, (\theta_1 + \eta\varepsilon, \theta_2, \ldots, \theta_q)')$.*

Assumption 3′ *Let ε_i^* be a random variable with a conditional probability density function $k(\cdot|x)$ such that $\int ek(e|x)de = 0$ for all x in the support $\mathcal{X}$ of X and $\sup_{x\in\mathcal{X}} \int e^2 k(e|x)de < \infty$. Let $\mu(x)$ denote a function of x and define $h(y|x, \gamma) := f(y|x, (\theta_1 + \eta^2\mu(x) + \eta\varepsilon^*, \theta_2, \ldots, \theta_q)')$.*

Remark 8 *Although $h(y|x, \gamma)$ in Assumptions 3 and 3′ is conceptually different from that in Assumptions 1 and 2, the former, when integrating out ε (or ε^*), satisfies the conditions in Assumptions 1 and 2. For this reason, we will slightly abuse the notation and use $h(y|x, \gamma)$ to denote both.*

For a matrix $A = [a_{ij}]$, let $|A| = \max_{i,j} |a_{ij}|$. Let $f_X(x)$ denote the marginal probability density function of X_i.

Assumption 4 *For almost all z in $\mathcal{Z}$, $\ln f(y|x, \theta)$ is twice continuously differentiable with respect to θ_1, and $f(y|x, \theta)$, $\ln f(y|x, \theta)$, $\partial \ln f(y|x, \theta)/\partial\theta_1$ and $\partial^2 \ln f(y|x, \theta)/\partial\theta_1^2$ are all measurable functions of z for each θ in Θ, where Θ is a compact subsets of $\mathbb{R}^q$. For* *almost all z in $\mathcal{Z}$, $\ln f(y|x,\theta)$, $\partial \ln f(y|x,\theta)/\partial\theta_1$ and $\partial^2 \ln f(y|x,\theta)/\partial\theta_1^2$ are all continuously differentiable with respect to θ. $\partial \ln f(y|x,\theta)/\partial\theta$, $\partial\left(\partial \ln f(y|x,\theta)/\partial\theta_1\right)/\partial\theta$ and $\partial\left(\partial^2 \ln f(y|x,\theta)/\partial\theta_1^2\right)/\partial\theta$ are all measurable functions of z for each θ in Θ. $f_X(x)$ is a measurable function of z for each θ in Θ. θ_0 is an element of the interior of Θ.*

Assumption 5 *There exist measurable functions $a(z)$ and $b(z)$ such that $|f(y|x,\theta)f_X(x)| \leq a(z)$ and $|\partial f(y|x,\theta)/\partial\theta|$, $|\partial^2 f(y|x,\theta)/\partial\theta_1^2|$, $|\partial^2 \ln f(y|x,\theta)/\partial\theta_1^2|^2$, $|\partial \ln f(y|x,\theta)/\partial\theta_1|^4$, $|\partial^3 \ln f(y|x,\theta)/\partial\theta_1^2\partial\theta'|$, $|\partial \ln f(y|x,\theta)/\partial\theta|$ and $|\partial^2 \ln f(y|x,\theta)/\partial\theta\partial\theta'|$ are each less than $b(z)$. Further, it is the case that $\int a(z)dz < +\infty$ and $\int b(z)a(z)dz < +\infty$, and that the set $\{z : f(y|x,\theta) > 0\}$ is independent of θ.*

Assumption 6 *If $\theta \neq \theta_0$, then $A := \{z : f(y|x,\theta) \neq f(y|x,\theta_0)\}$ satisfies $\int_A f(y|x,\theta_0)dy > 0$.*

We write $g(z,\theta) = (m(z,\theta)', s(z,\theta)')'$, where $m(z,\theta)$ is defined in equation (20) and $s(z,\theta) = s\left(y|\, x,\theta\right) := \partial \ln f\left(y|\, x,\theta\right)/\,\partial\theta$ denotes the score. Define

$$V := \int g\left(z,\theta_0\right) g\left(z,\theta_0\right)' f\left(y|\, x,\theta_0\right) f_X\left(x\right) dz.$$

Assumption 7 *The matrix V is nonsingular.*

Now we define some general notation. Suppose that $g(z,\theta)$ is a scalar function, $h(z,\gamma)$ is a density of Z with parameter γ, and Z_i $(i = 1,\ldots,N)$ is a sequence of observations from $h(z,\gamma)$, where an extra subscript γ on Z_i is suppressed for notational convenience. Define $g_N(\theta) := N^{-1}\sum_{i=1}^N g(Z_i,\theta)$, and when the expectation exists, $\phi(\theta,\gamma) := \int g(z,\theta)h(z,\gamma)dz$. This notation does not refer to the specific functions defined elsewhere in this paper, and it will be used only in the following lemma, which is a restatement of Lemma A.1 in Newey (1985)[29] and is helpful for the proof of Theorems 3 and 4.

Lemma B.1 *Suppose that, for each θ in Θ, $g(z,\theta)$ is a measurable function of z, for almost all z in $\mathcal{Z}$ a continuous function of θ, and Θ is compact. Suppose that, for each γ*

[29]We slightly modify the notation in Newey (1985) to suit our paper.

in Γ, $h(z,\gamma)$ *is a measurable probability density on* $\mathcal{Z}$, *for almost all* z *in* $\mathcal{Z}$ *a continuous function of* γ, *and* Γ *is compact. Suppose that there exist measurable functions* $a(z)$ *and* $b(z)$ *such that* $h(z,\gamma) \leq a(z)$ *and* $|g(z,\theta)| \leq b(z)$ *with*

$$\int b(z)a(z)dz < +\infty, \quad \int a(z)dz < +\infty.$$

Then $\phi(\theta,\gamma)$ *exists and is continuous on* $\Theta \times \Gamma$. *Suppose, in addition, that* $Z_1, \ldots, Z_N$ *are independent observations with density* $h(z,\gamma_N)$ *where* $\lim_{N\to\infty} \gamma_N = \gamma_0$. *Then for all* $\varepsilon > 0$,

$$\lim_{N\to\infty} P\left(\sup_{\Theta} |g_N(\theta) - \phi(\theta,\gamma_0)| \geq \varepsilon\right) = 0. \tag{B.1}$$

Proof. See Appendix of Newey (1985). ■

B.2 Proof of Theorem 3

For ease of reading, we will follow Newey (1985, Proof of Lemma 2.1) as closely as possible.

Step 1 Let

$$\phi_\eta(\theta) := \int g(z,\theta)\, h(y|\, x, \theta_0, \sqrt{\eta})\, f_X(x)\, dz, \tag{B.2}$$

$$\bar{V}_\eta := \int g(z,\theta_0)\, g(z,\theta_0)'\, h(y|\, x, \theta_0, \sqrt{\eta})\, f_X(x)\, dz - \phi_\eta(\theta_0)\phi_\eta(\theta_0)'. \tag{B.3}$$

By Assumptions 4 and 5, the elements of $g(z,\theta)$ and the density $h(y|x,\theta,\sqrt{\eta})f_X(x)$ satisfy the conditions of Lemma B.1, implying that $\phi_\eta(\theta)$ exists and is continuous on Γ. Then by Assumption 3, we have

$$\lim_{\eta\to 0} \phi_\eta(\theta_0) = \phi_0(\theta_0), \text{ and } \lim_{\eta\to 0} \bar{V}_\eta = V. \tag{B.4}$$

Due to Assumptions 3 - 5, the dominated convergence theorem (e.g., Bartle, 1966, Corollary 5.9) allows one to differentiate the integrand function in the identity $\int f(y|x,\theta)f_X(x)$ $dz = 1$, which yields the following identities for θ in the interior of Θ:

$$E\left[s(Y_i|X_i,\theta)\right] = \int \frac{\partial f(y|x,\theta)}{\partial\theta} f_X(x)dz = \int s(y|x,\theta)f(y|x,\theta)f_X(x)dz = 0,$$

and

$$E\left[m(Z_i,\theta)\right] = \int \frac{\partial^2 f(y|x,\theta)/\partial\theta_1^2}{f(y|x,\theta)} f(y|x,\theta) f_X(x) dz = 0.$$

In light of Assumption 2, these identities evaluating at θ_0 immediately imply that

$$\phi_0(\theta_0) = E\left[g(Z_i,\theta_0)\right] = \int g\left(z,\theta_0\right) f\left(y|\, x,\theta_0\right) f_X\left(x\right) dz = 0. \tag{B.5}$$

By Assumption 5, functions $[m(z,\theta)]^2$ and $s(y|x,\theta)s(y|x,\theta)'$ satisfy the conditions of Lemma B.1, and so does $s(y|x,\theta)m(z,\theta)$ by the Cauchy-Schwarz inequality.

Step 2 In this step, we will first establish a central limit theorem (CLT) for $N^{-1/2}\sum_{i=1}^{N}$ $(g(Z_i,\theta_0) - \phi_{\eta_N}(\theta_0))$ under arbitrary sequence of DGP's with $\eta_N \to 0$ as $N \to \infty$. Define a function $W_\eta(z) := \lambda'\left[g(z,\theta_0) - \phi_\eta(\theta_0)\right]$, and let $W_{\eta,i} := W_\eta(Z_i)$ for $i = 1,\ldots,N$, where λ is a $(q+1)$-dimensional non-zero vector. By the definitions of $\phi_\eta(\theta)$ and $\bar{V}_\eta$ in (B.2) and (B.3), we know that $W_{\eta,i}$ has mean zero and variance $\lambda'\bar{V}_\eta\lambda$, which is positive for small η by Assumption 7 and (B.4). For any $\delta > 0$, define the set $A_{\delta,\eta} := \left\{z : |W_\eta(z)| \geq \delta\left(N\lambda'\bar{V}_\eta\lambda\right)^{1/2}\right\}$.[30] Note that Z_i $(i = 1,\ldots,N)$ are identically distributed, so for any $\delta > 0$, we have

$$\begin{aligned}
&\left(N\lambda'\bar{V}_\eta\lambda\right)^{-1}\sum_{i=1}^{N}\int_{|W_{\eta,i}|\geq\delta\left(N\lambda'\bar{V}_\eta\lambda\right)^{1/2}} |W_{\eta,i}|^2 h(Y_i|X_i,\theta_0,\sqrt{\eta}) f_X(X_i) dZ_i \\
=\ &\left(\lambda'\bar{V}_\eta\lambda\right)^{-1}\int_{A_{\delta,\eta}} |W_\eta(z)|^2 h(y|x,\theta_0,\sqrt{\eta}) f_X(x) dz \\
\leq\ &\left(\lambda'\bar{V}_\eta\lambda\right)^{-1} 2(q+1)|\lambda|^2\left(|\phi_\eta(\theta_0)|^2\int_{A_{\delta,\eta}} a(z)dz + \int_{A_{\delta,\eta}} b(z)a(z)dz\right),
\end{aligned} \tag{B.6}$$

where the last inequality holds by Assumption 5 and the simple inequality that $(a+b)^2 \leq 2(a^2+b^2)$ for any $a,b \in \mathbb{R}$. By (B.4) and (B.5), $\lim_{\eta\to 0}\phi_\eta(\theta_0) = 0$. By (B.4), we have $\lim_{\eta\to 0}\lambda'\bar{V}_\eta\lambda = \lambda' V\lambda > 0$, so $A_{\delta,\eta}$ converges to an empty set as $N \to \infty$, implying that $\lim_{N\to\infty}\int_{A_{\delta,\eta}} a(z)dz = 0$ and $\lim_{N\to\infty}\int_{A_{\delta,\eta}} b(z)a(z)dz = 0$. Therefore, (B.6) implies that the Lindberg condition is satisfied, and by the Lindberg-Feller CLT (e.g., p.128 of Rao, 1971), we have $\left(N\lambda'\bar{V}_\eta\lambda\right)^{-1/2}\sum_{i=1}^{N} W_{\eta,i} \Rightarrow N(0,1)$, implying in turn that

[30] Let $A_{\delta,\eta} = \emptyset$ if $\lambda'\bar{V}_\eta\lambda < 0$.

$N^{-1/2}\sum_{i=1}^{N} W_{\eta,i} \Rightarrow N(0,\lambda' V\lambda)$. This, together with the Cramér-Wold device, implies that for arbitrary sequence of DGP's with $\eta_N \to 0$ as $N \to \infty$,

$$N^{-1/2}\sum_{i=1}^{N}\left(g\left(Z_i,\theta_0\right)-\phi_{\eta_N}\left(\theta_0\right)\right) \Rightarrow N(0,V).$$

Then, we apply this CLT to a particular sequence $\eta_N = N^{-1/2}$ and get

$$N^{-1/2}\sum_{i=1}^{N}\left(g\left(Z_i,\theta_0\right)-\phi_{N^{-1/2}}\left(\theta_0\right)\right) \Rightarrow N(0,V). \tag{B.7}$$

Step 3 Due to Assumptions 3 - 5 and the dominated convergence theorem, we calculate the derivative of $\int g\left(z,\theta_0\right) h\left(y|\, x,\theta_0,\eta\right) f_X\left(x\right) k\left(e\right) dedz$ with respect to η as follows

$$\begin{aligned}
K_R &:= \frac{\partial}{\partial\eta}\left(\int g\left(z,\theta_0\right) f_X\left(x\right)\int f\left(y|\, x,\theta_0+\sqrt{\eta}e\iota\right) k\left(e\right) dedz\right)\Bigg|_{\eta=0} \qquad\qquad (B.8)\\
&= \left.\frac{\int\int g\left(z,\theta_0\right) f_X\left(x\right)\frac{\partial f\left(y|x,\theta_0+\sqrt{\eta}e\iota\right)}{\partial\theta_1} ek\left(e\right) dedz}{2\sqrt{\eta}}\right|_{\eta=0}\\
&= \lim_{\eta\to 0}\frac{\frac{\sqrt{\eta}}{2}\int\int g\left(z,\theta_0\right) f_X\left(x\right)\frac{\partial f\left(y|x,\theta_0+\sqrt{\eta}e\iota\right)}{\partial\theta_1} ek\left(e\right) dedz}{\eta}.
\end{aligned}$$

Using the L'Hôpital's rule, we get

$$\begin{aligned}
K_R &= \lim_{\eta\to 0}\frac{\begin{pmatrix}\frac{1}{4\sqrt{\eta}}\int\int g\left(z,\theta_0\right) f_X\left(x\right)\frac{\partial f\left(y|x,\theta_0+\sqrt{\eta}e\iota\right)}{\partial\theta_1} ek\left(e\right) dedz\\ +\frac{\sqrt{\eta}}{2}\frac{1}{2\sqrt{\eta}}\int\int g\left(z,\theta_0\right) f_X\left(x\right)\frac{\partial^2 f\left(y|x,\theta_0+\sqrt{\eta}e\iota\right)}{\partial\theta_1^2} e^2 k\left(e\right) dedz\end{pmatrix}}{1}\\
&= \frac{K_R}{2}+\frac{1}{4}\lim_{\eta\to 0}\int\int g\left(z,\theta_0\right) f_X\left(x\right)\frac{\partial^2 f\left(y|\, x,\theta_0+\sqrt{\eta}e\iota\right)}{\partial\theta_1^2} e^2 k\left(e\right) dedz,
\end{aligned}$$

from which we obtain

$$\begin{aligned}
K_R &= \frac{1}{2}\int\int g\left(z,\theta_0\right) f_X\left(x\right)\frac{\partial^2 f\left(y|\, x,\theta_0\right)}{\partial\theta_1^2} e^2 k\left(e\right) dedz\\
&= \frac{\sigma_\varepsilon^2}{2}\int g\left(z,\theta_0\right) f_X\left(x\right)\frac{\partial^2 f\left(y|\, x,\theta_0\right)}{\partial\theta_1^2} dz\\
&= \frac{\sigma_\varepsilon^2}{2}\int g\left(z,\theta_0\right)\frac{\partial^2 f\left(y|\, x,\theta_0\right)/\,\partial\theta_1^2}{f\left(y|\, x,\theta_0\right)} f\left(y|\, x,\theta_0\right) f_X\left(x\right) dz\\
&= \frac{\sigma_\varepsilon^2}{2}E\left[g\left(Z_i,\theta_0\right)\frac{\partial^2 f\left(Y_i|\, X_i,\theta_0\right)/\,\partial\theta_1^2}{f\left(Y_i|\, X_i,\theta_0\right)}\right],
\end{aligned}$$

where we recall $\sigma_\varepsilon^2 = \int e^2 k(e)\, de$. Recalling that $g(z,\theta) = (m(z,\theta)', s(z,\theta)')'$ helps us obtain the following:

$$K_R = \frac{\sigma_\varepsilon^2}{2} \begin{bmatrix} E\left[\left(\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right)^2\right] \\ E\left[s\left(Y_i|\, X_i,\theta_0\right)\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right] \end{bmatrix} = \frac{\sigma_\varepsilon^2}{2}\begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix}.$$

Recall the definition of $\phi_\eta(\theta)$ in (B.2) and apply the mean value theorem to (B.8) with $\eta_N = N^{-1/2}$, we get

$$\begin{aligned} K_R &= \lim_{N\to\infty} N^{1/2}\left(\int g(z,\theta_0)\, f_X(x)\int f\left(y|\, x,\theta_0+\sqrt{N^{-1/2}}e\iota\right)k(e)\, dedz - \int g(z,\theta_0)\, f(z|\,\theta_0)\, dz\right) \\ &= \lim_{N\to\infty} N^{1/2}\left(\int g(z,\theta_0)\, h\left(y|\, x,\theta_0,\sqrt{N^{-1/2}}\right) f_X(x)dz - \int g(z,\theta_0)\, h(y|\, x,\theta_0,0)\, f_X(x)dz\right) \\ &= \lim_{N\to\infty} N^{1/2}\left(\phi_{N^{-1/2}}(\theta_0) - \phi_0(\theta_0)\right). \end{aligned}$$

Combined with the CLT in (B.7), we see that this implies that $\sqrt{N}g_N(\theta_0) \xrightarrow{d.} N(K_R, V)$, where we define $g_N(\theta) := N^{-1}\sum_{i=1}^N g(Z_i,\theta)$.[31]

Remark 9 *We can in principle address heteroscedasticity as long as $E[\varepsilon|\, x] = 0$ is satisfied. This of course implies that K_R should be redefined as*

$$\frac{1}{2}\begin{bmatrix} E\left\{E\left[\varepsilon_i^2|\, X_i\right]\left(\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right)^2\right\} \\ E\left\{E\left[\varepsilon_i^2|\, X_i\right] s\left(Y_i|\, X_i,\theta_0\right)\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right\} \end{bmatrix},$$

with corresponding changes in the following steps.

Step 4 We now show that $N^{1/2}\left(\bar\theta_N - \theta_0\right) = -(DH)^{-1}DN^{1/2}g_N(\theta_0) + o_p(1)$, where $D := [0, I_q]$ and $H := E\left[\partial g(Z_i,\theta_0)/\partial\theta'\right]$. By the mean value theorem, we get $N^{1/2}g_N\left(\bar\theta_N\right) = N^{1/2}g_N(\theta_0) + \left[\partial g_N\left(\dot\theta_N\right)\big/\partial\theta'\right]N^{1/2}\left(\bar\theta_N - \theta_0\right)$ for some $\dot\theta_N$ on the line segment connecting $\bar\theta_N$ and θ_0. By Assumptions 4 and 5, we know that $h(y|x,\gamma)f_X(x)$ and the constituent elements of $\partial g_N(\dot\theta_N)/\partial\theta'$ satisfy the conditions of Lemma B.1, implying that

[31] Note that $g_N(\theta)$ and $g(z,\theta)$ here are different from those introduced before Lemma B.1. The former are based on the specific functions $m(z,\theta)$ and $s(z,\theta)$ used discussed in this paper, while the latter are generic notation used only to state Lemma B.1, i.e., Lemma A.1 in Newey (1985).

$\partial g_N(\dot{\theta}_N) \Big/ \partial\theta' \xrightarrow{p.} E\left[\partial g\left(Z_i,\theta_0\right)/\,\partial\theta'\right]$. This, combined with the standard $\sqrt{N}$-consistency of the MLE $\bar{\theta}_N$, implies that $N^{1/2}g_N\left(\bar{\theta}_N\right) = N^{1/2}g_N\left(\theta_0\right) + HN^{1/2}\left(\bar{\theta}_N - \theta_0\right) + o_p\left(1\right)$. Because the MLE satisfies $0 = Dg_N\left(\bar{\theta}_N\right)$ by definition, it follows that $0 = DN^{1/2}g_N\left(\theta_0\right) + DHN^{1/2}\left(\bar{\theta}_N - \theta_0\right) + o_p\left(1\right)$, from which we obtain $N^{1/2}\left(\bar{\theta}_N - \theta_0\right) = -\left(DH\right)^{-1}DN^{1/2}g_N\left(\theta_0\right) + o_p\left(1\right)$.

By the definition of $g(z,\theta)$ and H, we note that

$$H = \begin{bmatrix} E\left[\frac{\partial m(Z_i,\theta_0)}{\partial\theta'}\right] \\ E\left[\frac{\partial s(Y_i|X_i,\theta_0)}{\partial\theta'}\right] \end{bmatrix} = \begin{bmatrix} E\left[\frac{\partial m(Z_i,\theta_0)}{\partial\theta'}\right] \\ -\mathcal{I} \end{bmatrix}.$$

We now simplify the following coordinates of H,

$$E\left[\frac{\partial m\left(Z_i,\theta_0\right)}{\partial\theta}\right] = E\left[\frac{\partial}{\partial\theta}\frac{\partial^2 f\left(Y_i|\,X_i,\theta_0\right)/\,\partial\theta_1^2}{f\left(Y_i|\,X_i,\theta_0\right)}\right],$$

a bit. For this purpose, we start with the observation that

$$0 = \int \frac{\partial^2 f\left(y|\,x,\theta\right)/\,\partial\theta_1^2}{f\left(y|\,x,\theta\right)} f\left(y|\,x,\theta\right)dy$$

for all θ. Assumptions 3 - 5 and the dominated convergence theorem allow differentiating both sides with respect to θ and getting

$$0 = \int \frac{\partial}{\partial\theta}\frac{\partial^2 f\left(y|\,x,\theta\right)/\,\partial\theta_1^2}{f\left(y|\,x,\theta\right)} f\left(y|\,x,\theta\right)dy + \int \frac{\partial^2 f\left(y|\,x,\theta\right)/\,\partial\theta_1^2}{f\left(y|\,x,\theta\right)}\frac{\partial f\left(y|\,x,\theta\right)/\,\partial\theta}{f\left(y|\,x,\theta\right)} f\left(y|\,x,\theta\right)dy,$$

or

$$0 = E\left[\frac{\partial}{\partial\theta}\frac{\partial^2 f\left(Y_i|\,X_i,\theta_0\right)/\,\partial\theta_1^2}{f\left(Y_i|\,X_i,\theta_0\right)}\right] + E\left[\frac{\partial^2 f\left(Y_i|\,X_i,\theta_0\right)/\,\partial\theta_1^2}{f\left(Y_i|\,X_i,\theta_0\right)} s\left(Y_i|\,X_i,\theta_0\right)\right].$$

We therefore conclude that

$$E\left[\frac{\partial m\left(Z_i,\theta_0\right)}{\partial\theta}\right] = -E\left[\frac{\partial^2 f\left(Y_i|\,X_i,\theta_0\right)/\,\partial\theta_1^2}{f\left(Y_i|\,X_i,\theta_0\right)} s\left(Y_i|\,X_i,\theta_0\right)\right] = -\kappa_2,$$

and hence

$$H = \begin{bmatrix} -\kappa_2' \\ -\mathcal{I} \end{bmatrix}.$$

Remark 10 *Note that* $-\left(DH\right)^{-1} = \mathcal{I}^{-1}$. *Combined with* $N^{1/2}g_N\left(\theta_0\right) \Rightarrow N\left(K_R, V\right)$, *which implies that* $DN^{1/2}g_N\left(\theta_0\right) \Rightarrow N\left(\frac{\sigma_\varepsilon^2}{2}\kappa_2, \mathcal{I}\right)$, *we can see that* $N^{1/2}\left(\bar{\theta}_N - \theta_0\right) \Rightarrow N\left(\frac{\sigma_\varepsilon^2}{2}\mathcal{I}^{-1}\kappa_2, \mathcal{I}^{-1}\right)$. *Therefore, if* $\kappa_2 = 0$, *the MLE is asymptotically unbiased even under the alternative of random effects.*

Step 5 We now establish the asymptotic distribution of $m_N\left(\bar{\theta}_N\right)$. For this purpose, we note that $\sqrt{N}m_N\left(\bar{\theta}_N\right) = L\sqrt{N}g_N\left(\bar{\theta}_N\right)$ for $L := [I_1, 0]$. We also saw in the previous step that $N^{1/2}g_N\left(\bar{\theta}_N\right) = N^{1/2}g_N\left(\theta_0\right) + HN^{1/2}\left(\bar{\theta}_N - \theta_0\right) + o_p\left(1\right)$, and $N^{1/2}(\bar{\theta}_N - \theta_0) = -(DH)^{-1}DN^{1/2}g_N(\theta_0) + o_p(1)$. Let $s_N(\theta) := N^{-1}\sum_{i=1}^{N} s(Z_i, \theta)$. Therefore, we see that

$$
\begin{aligned}
\sqrt{N}m_N\left(\bar{\theta}_N\right) &= L\sqrt{N}g_N\left(\bar{\theta}_N\right) \\
&= L\sqrt{N}g_N\left(\theta_0\right) - LH\left(DH\right)^{-1}D\sqrt{N}g_N\left(\theta_0\right) + o_p\left(1\right) \\
&= L\left(I_{q+1} - H\left(DH\right)^{-1}D\right)\sqrt{N}g_N\left(\theta_0\right) + o_p\left(1\right) \\
&= [I_1, 0]\sqrt{N}g_N\left(\theta_0\right) - [I_1, 0]\,H\left(-\mathcal{I}\right)^{-1}\sqrt{N}s_N\left(\theta_0\right) + o_p\left(1\right) \\
&= \sqrt{N}m_N\left(\theta_0\right) - \kappa_2'\mathcal{I}^{-1}\sqrt{N}s_N\left(\theta_0\right) + o_p\left(1\right) \\
&= \left[I_1, -\kappa_2'\mathcal{I}^{-1}\right]\sqrt{N}g_N\left(\theta_0\right) + o_p\left(1\right),
\end{aligned}
$$

while in Step 3 we showed $\sqrt{N}g_N\left(\theta_0\right) \Rightarrow N\left(K_R, V\right)$. Because $[I_1, -\kappa_2'\mathcal{I}^{-1}]\,V\,[I_1, -\kappa_2'\mathcal{I}^{-1}]' = \kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2$, it follows that

$$
\sqrt{N}m_N\left(\bar{\theta}_N\right) = L\sqrt{N}g_N\left(\bar{\theta}_N\right) \Rightarrow N\left(\left[I_1, -\kappa_2'\mathcal{I}^{-1}\right]K_R, \kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2\right) \qquad \text{(B.9)}
$$

as desired, since $[I_1, -\kappa_2'\mathcal{I}^{-1}]\,K_R = \frac{\sigma_\varepsilon^2}{2}\left(\kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2\right)$.

B.3 Proof of Theorem 4

The proof is essentially identical to the proof of Theorem 3, except that we need to calculate the counterpart of K_R in Step 3. We begin by considering the special case where the neglected heterogeneity takes the form

$$
f\left(y\,\middle|\,x, \left(\theta_{0,1} + \frac{\mu\left(x\right)}{N^{1/2}}, \theta_{0,2}, \ldots, \theta_{0,q}\right)\right).
$$

Note that there is no other random variable yet (such as ε^*). We now would like to calculate the counterpart of K_R. Due to Assumptions 3′, 4, 5 and the dominated convergence theorem, we have

$$
K_F := \frac{\partial}{\partial\eta}\left(\int g\left(z, \theta_0\right) f_X\left(x\right) f\left(y|\, x, \theta_0 + \eta\mu\left(x\right)\iota\right) dz\right)\Bigg|_{\eta=0}
$$

$$
\begin{aligned}
&= \int g(z,\theta_0) f_X(x) \frac{\partial f(y|x,\theta_0)}{\partial \theta_1} \mu(x)\, dz \\
&= \int g(z,\theta_0) s_1(y|x,\theta_0) f(y|x,\theta_0) f_X(x) \mu(x)\, dz \\
&= E\left[g(Z_i,\theta_0) s_1(Y_i|X_i,\theta_0) \mu(X_i)\right],
\end{aligned}
$$

where $s_1(y|x,\theta_0)$ is the first coordinate of the score function. By applying the mean value theorem with $\eta_N = N^{-1/2}$ as before, we get $K_F = \lim_{N\to\infty} N^{1/2}\left(\phi_{N^{-1/2}}(\theta_0) - \phi_0(\theta_0)\right)$. Note that K_F can be written as

$$
\begin{aligned}
K_F &:= E\begin{bmatrix} \frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)} s_1(Y_i|X_i,\theta_0)\mu(X_i) \\ s(Y_i|X_i,\theta_0) s_1(Y_i|X_i,\theta_0)\mu(X_i) \end{bmatrix} \\
&= \begin{bmatrix} E\left\{ E\left[\left. \frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)} s_1(Y_i|X_i,\theta_0) \right| X_i \right] \mu(X_i) \right\} \\ E\left\{ E\left[s(Y_i|X_i,\theta_0) s_1(Y_i|X_i,\theta_0) | X_i\right] \mu(X_i)\right\} \end{bmatrix}.
\end{aligned}
$$

The rest of Newey (1985)'s analysis still applies, in particular, (2.6) and (2.7) in Newey (1985) hold.[32] We therefore have (in his notation) $L_T = L = [I_1, 0]$, $D = [0, I_q]$, $P = I_{q+1} - H(DH)^{-1} D$, $K = K_F$, and we consider $\sqrt{N}$ instead of $\sqrt{T}$. Because $DH = -\mathcal{I}$, $Dg(z,\theta_0) = s(z,\theta_0)$, we still have (in our notation)

$$
\begin{aligned}
L\sqrt{N} g_N\left(\bar{\theta}_N\right) &= LP\sqrt{N} g_N(\theta_0) + o_p(1) \\
&= L\left(I_{q+1} - H(DH)^{-1} D\right)\sqrt{N} g_N(\theta_0) + o_p(1) \\
&= [I_1, 0]\sqrt{N} g_N(\theta_0) - [I_1, 0] H(-\mathcal{I})^{-1}\sqrt{N} s_N(\theta_0) + o_p(1) \\
&= \sqrt{N} m_N(\theta_0) - \kappa_2' \mathcal{I}^{-1}\sqrt{N} s_N(\theta_0) + o_p(1) \\
&= \left[I_1, -\kappa_2'\mathcal{I}^{-1}\right]\sqrt{N} g_N(\theta_0) + o_p(1),
\end{aligned}
$$

while

$$
\sqrt{N} g_N(\theta_0) \Rightarrow N(K_F, V).
$$

Because we still have

$$
\left[I_1, -\kappa_2'\mathcal{I}^{-1}\right] V \left[I_1, -\kappa_2'\mathcal{I}^{-1}\right]' = \kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2,
$$

[32] Newey (1985)'s (2.6) is $\sqrt{T} L_T g_T\left(\bar{\theta}_T\right) = LP\sqrt{T} g_T(\theta_0) + o_p(1)$, and (2.7) is $\sqrt{T} g_T(\theta_0) \Rightarrow N(K\delta, V)$.

it follows that

$$\sqrt{N} m_N\left(\bar{\theta}_N\right)=L \sqrt{N} g_N\left(\bar{\theta}_N\right) \Rightarrow N\left(\left[I_1,-\kappa_2^{\prime} \mathcal{I}^{-1}\right] K_F, \kappa_1-\kappa_2^{\prime} \mathcal{I}^{-1} \kappa_2\right). \tag{B.10}$$

We now consider the fixed effects (23). After all, the whole calculation of K_R was based on the derivative of $f\left(y \mid x,\left(\theta_{0,1}+\eta \mu(x)+\sqrt{\eta} \varepsilon, \theta_{0,2}, \ldots, \theta_{0,q}\right)\right)$ and note that the derivative should be the sum of the derivatives of $f\left(y \mid x,\left(\theta_{0,1}+\eta \mu(x), \theta_{0,2}, \ldots, \theta_{0,q}\right)\right)$ and $f(y|x,(\theta_{0,1}+\sqrt{\eta}\varepsilon, \theta_{0,2}, \ldots, \theta_{0,q}))$. The asymptotic bias is then equal to the sum of two asymptotic biases in (B.9) and (B.10):

$$\sqrt{N} m_N\left(\bar{\theta}_N\right)=L \sqrt{N} g_N\left(\bar{\theta}_N\right) \Rightarrow N\left(\left[I_1,-\kappa_2^{\prime} \mathcal{I}^{-1}\right]\left(K_F+K_R^*\right), \kappa_1-\kappa_2^{\prime} \mathcal{I}^{-1} \kappa_2\right),$$

where

$$K_R^*:=\frac{1}{2}\left[\begin{array}{c} E\left\{E\left[\left.\left(\varepsilon_i^*\right)^2\right| X_i\right]\left(\frac{\partial^2 f\left(Y_i \mid X_i, \theta_0\right) / \partial \theta_1^2}{f\left(Y_i \mid X_i, \theta_0\right)}\right)^2\right\} \\ E\left\{E\left[\left.\left(\varepsilon_i^*\right)^2\right| X_i\right] s\left(Y_i \mid X_i, \theta_0\right) \frac{\partial^2 f\left(Y_i \mid X_i, \theta_0\right) / \partial \theta_1^2}{f\left(Y_i \mid X_i, \theta_0\right)}\right\}\end{array}\right]$$

is a heteroskedasticity-robust version of K_R based on ε_i^*. (See Remark 9.)

B.4 In-Depth Analysis of the Linear Model

Note that the fixed effects can be decomposed into two components, $\mu(X_i)$ and ε_i^*. Their distinct roles are best understood by considering a linear panel data model

$$Y_{it}=X_{it}^{*\prime} \beta_0+\alpha_{N,i}+v_{it}, \quad i=1, \ldots, N, \quad t=1, \ldots, T,$$

where $\alpha_{N,i}=\alpha_0$ under the null and $\alpha_{N,i}=\alpha_0+N^{-1/2} \mu(x_i)+N^{-1/4} \varepsilon_i^*$ under the alternative. It is clear that the correlation between x_{it} and $\mu(x_i)$ induces the bias in the OLS (i.e., MLE), while the presence of ε_i^* does not affect the unbiasedness property of the OLS. Although even ε_i^* induces the MLE to be biased in general nonlinear models, the distinctive roles are quite clear in linear models. It turns out that the BP test does not have any power against the presence of $\mu(x_i)$ in linear models. This is because in linear models, we have

$$E\left[\left.\frac{\partial^2 f\left(Y_i \mid X_i, \theta_0\right) / \partial \theta_1^2}{f\left(Y_i \mid X_i, \theta_0\right)} s_1\left(Y_i \mid X_i, \theta_0\right)\right| X_i\right]=0,$$

so that the first component of K_F is equal to 0, and it can be shown that

$$\begin{aligned}\kappa_2 &= E\left[s\left(Y_i|\, X_i,\theta_0\right)\frac{\partial^2 f\left(Y_i|\, X_i,\theta_0\right)/\,\partial\theta_1^2}{f\left(Y_i|\, X_i,\theta_0\right)}\right] \\ &= E\left[\left(\frac{1}{\sigma_v^2}\sum_{t=1}^{T}X_{it}\left(Y_{it}-\alpha_0-X_{it}^{*\prime}\beta_0\right)\right)\left(-\frac{T}{\sigma_v^2}+\left(\frac{1}{\sigma_v^2}\sum_{t=1}^{T}\left(Y_{it}-\alpha_0-X_{it}^{*\prime}\beta_0\right)\right)^2\right)\right]=0,\end{aligned} \tag{B.11}$$

leading to the implication

$$\left[I_1,-\kappa_2'\mathcal{I}^{-1}\right]K_F=0.$$

Remark 11 *Equation (B.11) holds because for linear models, we have*

$$\ln f\left(Y_i|\, X_i,\theta\right)=C-\frac{1}{2\sigma_v^2}\sum_{t=1}^{T}\left(Y_{it}-\alpha-X_{it}^{*\prime}\beta\right)^2,$$

where C is a constant, so

$$s\left(Y_i|\, X_i,\theta\right)=\frac{1}{\sigma_v^2}\sum_{t=1}^{T}X_{it}\left(Y_{it}-\alpha-X_{it}^{*\prime}\beta\right),\ \ \textit{and}\ \ \frac{\partial^2\ln f\left(Y_i|\, X_i,\theta\right)}{\partial\theta\partial\theta'}=-\frac{1}{\sigma_v^2}\sum_{t=1}^{T}X_{it}X_{it}'.$$

In particular, we have

$$s_1\left(Y_i|\, X_i,\theta\right)=\frac{\partial\ln f\left(Y_i|\, X_i,\theta\right)}{\partial\theta_1}=\frac{1}{\sigma_v^2}\sum_{t=1}^{T}\left(Y_{it}-\alpha-X_{it}^{*\prime}\beta\right),\ \ \textit{and}\ \ \frac{\partial^2\ln f\left(Y_i|\, X_i,\theta\right)}{\partial\theta_1^2}=-\frac{T}{\sigma_v^2}.$$

We therefore see that

$$\begin{aligned}\frac{\partial^2 f\left(Y_i|\, X_i,\theta\right)/\,\partial\theta_1^2}{f\left(y|\, x,\theta\right)} &= \frac{\partial^2\ln f\left(Y_i|\, X_i,\theta\right)}{\partial\theta_1^2}+\left(\frac{\partial\ln f\left(Y_i|\, X_i,\theta\right)}{\partial\theta_1}\right)^2 \\ &= -\frac{T}{\sigma_v^2}+\left(\frac{1}{\sigma_v^2}\sum_{t=1}^{T}\left(Y_{it}-\alpha-X_{it}^{*\prime}\beta\right)\right)^2.\end{aligned}$$

Therefore, we should have

$$\begin{aligned}&E\left[s\left(Y_i|\, X_i,\theta_0\right)\frac{\partial^2 f\left(Y_i|\, X_i,\theta_0\right)/\,\partial\theta_1^2}{f\left(Y_i|\, X_i,\theta_0\right)}\right] \\ &= E\left[\left(\frac{1}{\sigma_v^2}\sum_{t=1}^{T}X_{it}\left(Y_{it}-\alpha_0-X_{it}^{*\prime}\beta_0\right)\right)\left(-\frac{T}{\sigma_v^2}+\left(\frac{1}{\sigma_v^2}\sum_{t=1}^{T}\left(Y_{it}-\alpha_0-X_{it}^{*\prime}\beta_0\right)\right)^2\right)\right].\end{aligned}$$

This seems to indicate that, at least for linear models, the BP test has zero power against fixed effects in the sense that it is unable to detect the presence of $\mu(X_i)$. It turns out that the issue is a little subtle, and the BP test does have the power to detect $\mu(X_i)$ as long as it is in the $N^{-1/4}$-neighborhood, instead of the $N^{-1/2}$-neighborhood. We note that the components of $[I_1, -\kappa_2' \mathcal{I}^{-1}] K_F$ measure the first order effect of misspecification on the asymptotic mean of $m_N(\bar{\theta}_N)$. When $[I_1, -\kappa_2' \mathcal{I}^{-1}] K_F$ is zero, we can make a more refined analysis by going through the second order derivative, similar in spirit to Chesher (1984)'s derivation. To be more specific, we zoom in on the following linear model

$$Y_{it} = X_{it}^{*\prime}\beta_0 + \frac{\mu(X_i) + \varepsilon_i^*}{N^{1/4}} + v_{it}, \quad t = 1, \ldots, T, \tag{B.12}$$

with $v_{it} \sim N(0, \sigma_v^2)$. Note that the fixed effects, especially $\mu(X_i)$, are in the $O\left(N^{-1/4}\right)$ neighborhood. Also note that

$$\frac{\partial^2 f(Y_i|X_i, \theta) / \partial \theta_1^2}{f(Y_i|X_i, \theta)} = -\frac{T}{\sigma_v^2} + \left(\frac{1}{\sigma_v^2} \sum_{t=1}^{T} (Y_{it} - \alpha - X_{it}^{*\prime}\beta)\right)^2,$$

so the counterpart of $\sqrt{N} m_N(\bar{\theta}_N)$ is equal to

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\left(\sum_{t=1}^{T} \hat{v}_{it}\right)^2 - T\hat{\sigma}_v^2}{(\hat{\sigma}_v^2)^2} = \frac{N^{-1/2}\hat{v}'[I_N \otimes (e_T e_T' - I_T)]\hat{v}}{(\hat{\sigma}_v^2)^2} \tag{B.13}$$

where $\hat{v}_{it}$ denotes the OLS residual, e_T is a $T \times 1$ vector of ones, and $\hat{\sigma}_v^2 := (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{v}_{it}^2$. We will assume that $T^{-1} E\left[\sum_{t=1}^{T} X_{it}^* X_{it}^{*\prime}\right]$ is positive definite with finite eigenvalues. We also assume that $E[\varepsilon_i^* v_{it}] = 0$ and $E[\varepsilon_i^* \mu(X_i)] = 0$.

In order to analyze the power of the BP test under the alternative (B.12), it suffices to analyze the asymptotic mean of the numerator $N^{-1/2}\hat{v}'[I_N \otimes (e_T e_T' - I_T)]\hat{v}$ of (B.13). In Lemma 19, we show how Honda (1985)'s Lemma 1 should be modified under the alternative of fixed effects:

Lemma 19 *Under (B.12), we have*

$$\frac{\hat{v}'[I_N \otimes (e_T e_T' - I_T)]\hat{v}}{N^{1/2}} = \frac{1}{N^{1/2}} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} v_{il} v_{im} + T(T-1) E\left[\xi_i^2\right]$$

$$- 2T\left(T-1\right)\lambda' Q^{-1}\lambda + \left(Q^{-1}\lambda\right)' S\left(Q^{-1}\lambda\right) + O_p\left(N^{-1/4}\right),$$

where $Q := T^{-1}E\left[\sum_{t=1}^{T} X_{it}^{*} X_{it}^{*\prime}\right]$, $\bar{X}_i^{*} := T^{-1}\sum_{t=1}^{T} X_{it}^{*}$, $\lambda := E\left[\bar{X}_i^{*}\mu\left(X_i\right)\right]$, $\xi_i := \mu\left(X_i\right) + \varepsilon_i^{*}$ *and* $S := E\left[\sum_{l,m=1,l\neq m}^{T} X_{il}^{*} X_{im}^{*\prime}\right]$.

Proof. In Appendix B.5. ■

In Lemma 19, we can clearly see that the presence of $\mu\left(X_i\right)$ affects the asymptotic bias through λ and $E\left[\xi_i^2\right]$. This is in contrast to the argument earlier in this section, where the BP test was unable to detect $\mu\left(X_i\right)$ in the linear model. The difference is that in the parameterization (23), $\mu\left(X_i\right)$ was too close to zero in the $O\left(N^{-1/2}\right)$ neighborhood for the linear model, while we show here that $\mu\left(X_i\right)$ can be detected by the BP test if $\mu\left(X_i\right)$ is not too close to zero. There already exists a well-known test (Hausman and Taylor, 1981) for the linear model, which can be shown to have a power against the local misspecification in the $O\left(N^{-1/2}\right)$ neighborhood. The test by Hausman and Taylor (1981) does not have a counterpart in the nonlinear panel models, probably because the fixed effects estimator is not asymptotically unbiased for fixed T for nonlinear models (even after bias reduction). The BP test was shown to be able to detect fixed effects in nonlinear models, so it makes sense to examine whether the BP test has such a property for linear models. Our analysis in the current section leads to the practical conclusion that the BP test may be best suited for nonlinear models.

B.5 Proof of Lemma 19

First, we can see that the OLS $\hat{\beta}$ is not $\sqrt{N}$-consistent under the misspecification. In fact, we have

$$\begin{aligned}
\hat{\beta} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*} X_{it}^{*\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}\left(\frac{\mu\left(X_i\right)}{N^{1/4}} + \frac{\varepsilon_i^{*}}{N^{1/4}} + X_{it}^{*\prime}\beta_0 + v_{it}\right)\right) \\
&= \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*} X_{it}^{*\prime}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\bar{X}_i^{*}\mu\left(X_i\right)\right)/N^{1/4} \\
&+ \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*} X_{it}^{*\prime}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\bar{X}_i^{*}\varepsilon_i^{*}\right)/N^{1/4}
\end{aligned}$$

$$+\beta_0+\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}X_{it}^{*\prime}\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}v_{it}\right).$$

We see that for fixed T,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}X_{it}^{*\prime}=Q+O_p\left(N^{-1/2}\right),$$

$$\frac{1}{N}\sum_{i=1}^{N}\bar{X}_i^{*}\mu\left(X_i\right)=\lambda+O_p\left(N^{-1/2}\right),$$

$$\frac{1}{N}\sum_{i=1}^{N}\bar{X}_i^{*}\varepsilon_i^{*}=O_p\left(N^{-1/2}\right),$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}v_{it}=O_p\left(N^{-1/2}\right),$$

so

$$N^{1/4}\left(\hat{\beta}-\beta_0\right)=Q^{-1}\lambda+O_p\left(N^{-1/2}\right)+O_p\left(N^{-1/4}\right)=Q^{-1}\lambda+O_p\left(N^{-1/4}\right). \tag{B.14}$$

Now we consider

$$\hat{v}'[I_N\otimes(e_Te_T'-I_T)]\hat{v}=\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}\hat{v}_{il}\hat{v}_{im}.$$

Since the OLS residual $\hat{v}_{il}=\xi_i/N^{1/4}+v_{il}-X_{il}^{*\prime}\left(\hat{\beta}-\beta_0\right)$ with $\xi_i=\mu(X_i)+\varepsilon_i^{*}$, we have

$$\begin{aligned}\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}\hat{v}_{il}\hat{v}_{im}&=\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}\left(\frac{\xi_i}{N^{1/4}}+v_{il}\right)\left(\frac{\xi_i}{N^{1/4}}+v_{im}\right)\\&-\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}\left(\frac{\xi_i}{N^{1/4}}+v_{il}\right)X_{im}^{*\prime}\left(\hat{\beta}-\beta_0\right)\\&-\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}\left(\frac{\xi_i}{N^{1/4}}+v_{im}\right)X_{il}^{*\prime}\left(\hat{\beta}-\beta_0\right)\\&+\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}\left(\hat{\beta}-\beta_0\right)'X_{il}^{*}X_{im}^{*\prime}\left(\hat{\beta}-\beta_0\right).\end{aligned} \tag{B.15}$$

Note that the first term in (B.15) can be rewritten as

$$\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}\left(\frac{\xi_i}{N^{1/4}}+v_{il}\right)\left(\frac{\xi_i}{N^{1/4}}+v_{im}\right)$$

$$
\begin{aligned}
&= \frac{1}{N^{1/2}} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} v_{il} v_{im} + \frac{T(T-1)}{N} \sum_{i=1}^{N} \xi_i^2 + \frac{2(T-1)}{N^{3/4}} \sum_{i=1}^{N} \xi_i \sum_{l=1}^{T} v_{il} \\
&= \frac{1}{N^{1/2}} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} v_{il} v_{im} + T(T-1) \left(E\left[\xi_i^2\right] + O_p\left(N^{-1/2}\right)\right) + O_p(N^{-1/4}) \\
&= \frac{1}{N^{1/2}} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} v_{il} v_{im} + T\left(T-1\right) E\left[\xi_i^2\right] + O_p(N^{-1/4}),
\end{aligned}
$$

where we used the CLT for the second equality.

Now we show that the second term in (B.15) is $O_p(N^{1/4})$.

$$
\begin{aligned}
&\frac{1}{N^{1/2}} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} \left(\frac{\xi_i}{N^{1/4}} + v_{il} \right) X_{im}^{*\prime} \left(\hat{\beta} - \beta_0\right) \\
&= \frac{1}{N^{1/2}} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} v_{il} X_{im}^{*\prime} \left(\hat{\beta} - \beta_0\right) + \frac{1}{N^{3/4}} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} \xi_i X_{im}^{*\prime} \left(\hat{\beta} - \beta_0\right) \\
&= \left(\hat{\beta} - \beta_0\right)' \left(\frac{1}{N^{1/2}} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} v_{il} X_{im}^{*} \right) + N^{1/4} \left(\hat{\beta} - \beta_0\right)' \left(\frac{T-1}{N} \sum_{i=1}^{N} \xi_i \sum_{m=1}^{T} X_{im}^{*} \right).
\end{aligned}
$$

Because $E\left[\sum_{l,m=1,l\neq m}^{T} v_{il} X_{im}^{*}\right] = 0$, we can apply the CLT and conclude that $N^{-1/2} \sum_{i=1}^{N} \sum_{l,m=1,l\neq m}^{T} v_{il} X_{im}^{*} = O_p(1)$. Combined with the result that $N^{1/4}\left(\hat{\beta} - \beta_0\right) = Q^{-1}\lambda + O_p\left(N^{-1/4}\right)$, we conclude that the first term on the far right-hand side is $O_p\left(N^{-1/4}\right)$. Noting that

$$
\begin{aligned}
\frac{T-1}{N} \sum_{i=1}^{N} \xi_i \sum_{m=1}^{T} X_{im}^{*} &= (T-1) E\left[\xi_i \sum_{m=1}^{T} X_{im}^{*} \right] + O_p\left(N^{-1/2}\right) \\
&= (T-1) E\left[\left(\mu\left(X_i\right) + \varepsilon_i^{*}\right) \left(T \bar{X}_i^{*}\right) \right] + O_p\left(N^{-1/2}\right) \\
&= T(T-1) E\left[\bar{X}_i^{*} \mu\left(X_i\right) \right] + O_p\left(N^{-1/2}\right) = T(T-1)\lambda + O_p\left(N^{-1/2}\right),
\end{aligned}
$$

we conclude that the second term on the far right-hand side is

$$
N^{1/4}\left(\hat{\beta} - \beta_0\right)' \left(\frac{T-1}{N} \sum_{i=1}^{N} \xi_i \sum_{m=1}^{T} X_{im}^{*} \right) = T(T-1)\lambda' Q^{-1} \lambda + O_p\left(N^{-1/4}\right).
$$

Combined with (B.14), we should have that the second term in (B.15) is equal to $-T(T-1)\lambda' Q^{-1}\lambda + O_p\left(N^{-1/4}\right)$. By the same argument, the third term in (B.15) is also $-T(T-1)\lambda' Q^{-1}\lambda + O_p\left(N^{-1/4}\right)$ as the indices l and m are symmetric in this respect.

Finally, by (B.14), the fourth term in (B.15) can be written as

$$
\begin{aligned}
&\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}\left(\hat{\beta}-\beta_0\right)' X_{il}^{*}X_{im}^{*\prime}\left(\hat{\beta}-\beta_0\right)\\
&= N^{1/4}\left(\hat{\beta}-\beta_0\right)'\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{l,m=1,l\neq m}^{T}X_{il}^{*}X_{im}^{*\prime}\right)N^{1/4}\left(\hat{\beta}-\beta_0\right)\\
&= \left(Q^{-1}\lambda\right)' S\left(Q^{-1}\lambda\right)+O_p\left(N^{-1/4}\right).
\end{aligned}
$$

The result of the lemma follows by combining our analysis of each term of (B.15).

C Technical Details of Section 4

This section makes the following additional assumption.

Assumption 8 *(i)* $\sup_{i,t} E\left[\frac{\partial^2 f(Y_{it}|X_{it},\theta_0)/\partial\theta_1^2}{f(Y_{it}|X_{it},\theta_0)}\right] < \infty$*; (ii)* $\sup_{i,t} E\left[\left(\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)^2\right] < \infty$*; (iii) there exists some function* $M(y,x)$ *such that* $\left|\frac{\partial^4 \ln f(Y_{it}|X_{it},\theta)}{\partial\theta_1^2\partial\theta\partial\theta'}\right| \leq M(Y_{it},X_{it})$, $\left|\frac{\partial^3 \ln f(Y_{it}|X_{it},\theta)}{\partial\theta_1\partial\theta\partial\theta'}\right| \leq M(Y_{it},X_{it})$, $\left|\frac{\partial^2 \ln f(Y_{it}|X_{it},\theta)}{\partial\theta_1\partial\theta}\right|^2 \leq M(Y_{it},X_{it})$, $\left|\frac{\partial \ln f(Y_{it}|X_{it},\theta)}{\partial\theta_1}\right|^2 \leq M(Y_{it},X_{it})$, *and* $\sup_{i,t} E[M(Y_{it},X_{it})] < \infty$.

We first show that (12) is $O_p(1)$ as $N,T\to\infty$. The same argument shows that (13) is also $O_p(1)$. First, we rewrite (12) as

$$
\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2 f\left(Y_{it}|\,X_{it},\theta_0\right)/\,\partial\theta_1^2}{f\left(Y_{it}|\,X_{it},\theta_0\right)}+\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t,t'=1,t\neq t'}^{T}\frac{\partial \ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\frac{\partial \ln f\left(Y_{it'}|\,X_{it'},\theta_0\right)}{\partial\theta_1}. \tag{C.1}
$$

Note that the first term in (C.1) is a zero mean random variable, and conditional on the Xs, it is a sum of random variables independent over i and t. Therefore, the first term in (C.1) is $O_p\left(T^{-1/2}\right)$.

As for the second term in (C.1), we see that $\sum_{t,t'=1,t\neq t'}^{T}\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\frac{\partial \ln f(Y_{it'}|X_{it'},\theta_0)}{\partial\theta_1}$ has mean equal to zero and its variance is equal to

$$
\sum_{t,t'=1,t\neq t'}^{T}E\left[\left(\frac{\partial \ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\right]E\left[\left(\frac{\partial \ln f\left(Y_{it'}|\,X_{it'},\theta_0\right)}{\partial\theta_1}\right)^2\right],
$$

and therefore, the second term in (C.1) has mean equal to zero and variance equal to

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,t'=1,t\neq t'}^{T} E\left[\left(\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\right] E\left[\left(\frac{\partial \ln f\left(Y_{it'}|\, X_{it'},\theta_0\right)}{\partial\theta_1}\right)^2\right] = O_p\left(1\right).$$

In order to establish that the noise of estimating θ_0 does not affect the distribution of the test statistic under the null, we first apply the second order mean value theorem to (24), and obtain

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2 \ln f\left(Y_{it}|\, X_{it},\bar{\theta}_n\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\bar{\theta}_n\right)}{\partial\theta_1}\right)^2\\
&=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2 \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\\
&+\left(\begin{array}{c}\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^3 \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1^2\partial\theta'}\\ +\frac{2}{NT\sqrt{T}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)\left(\sum_{t=1}^{T}\frac{\partial^2 \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1\partial\theta'}\right)\end{array}\right)\sqrt{NT}\left(\bar{\theta}_n-\theta_0\right)\\
&+\left(\sqrt{NT}\left(\bar{\theta}_n-\theta_0\right)\right)'\left(\begin{array}{c}\frac{1}{2N\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^4 \ln f\left(Y_{it}|X_{it},\widetilde{\theta}\right)}{\partial\theta_1^2\partial\theta\partial\theta'}\\ +\frac{1}{N\sqrt{NT^2}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial^2 \ln f\left(Y_{it}|X_{it},\widetilde{\theta}\right)}{\partial\theta_1\partial\theta}\right)\left(\sum_{t=1}^{T}\frac{\partial^2 \ln f\left(Y_{it}|X_{it},\widetilde{\theta}\right)}{\partial\theta_1\partial\theta'}\right)\\ +\frac{1}{N\sqrt{NT^2}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial \ln f\left(Y_{it}|X_{it},\widetilde{\theta}\right)}{\partial\theta_1}\right)\left(\sum_{t=1}^{T}\frac{\partial^3 \ln f\left(Y_{it}|X_{it},\widetilde{\theta}\right)}{\partial\theta_1\partial\theta\partial\theta'}\right)\end{array}\right)\left(\sqrt{NT}\left(\bar{\theta}_n-\theta_0\right)\right)
\end{aligned}$$

for some $\widetilde{\theta}$ between θ_0 and $\bar{\theta}_n$.

Note that the last term above can be bounded from above by

$$\begin{aligned}
&\left(\begin{array}{c}\frac{1}{2N\sqrt{NT^2}}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}M\left(Y_{it},X_{it}\right)\right)\\ +\frac{2}{N\sqrt{NT^2}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}M\left(Y_{it},X_{it}\right)\right)\left(\sum_{t=1}^{T}M\left(Y_{it},X_{it}\right)\right)\end{array}\right)\left|\sqrt{NT}\left(\bar{\theta}_n-\theta_0\right)\right|^2\\
&=\left(O_p\left(\frac{NT}{N\sqrt{NT^2}}\right)+O_p\left(\frac{NT^2}{N\sqrt{NT^2}}\right)\right)O_p\left(1\right)=o_p\left(1\right),
\end{aligned}$$

so we obtain that (24) can be rewritten as

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2 \ln f\left(Y_{it}|\, X_{it},\bar{\theta}_n\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\bar{\theta}_n\right)}{\partial\theta_1}\right)^2\\
&=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2 \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1^2}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial \ln f\left(Y_{it}|\, X_{it},\theta_0\right)}{\partial\theta_1}\right)^2
\end{aligned}$$

$$+\left(\begin{array}{c}\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^3\ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1^2\partial\theta'}\\+\frac{2}{NT\sqrt{T}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial\ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)\left(\sum_{t=1}^{T}\frac{\partial^2\ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1\partial\theta'}\right)\end{array}\right)\sqrt{NT}\left(\bar{\theta}_n-\theta_0\right)+o_p\left(1\right). \tag{C.2}$$

We now show that the third term in (C.2) is $o_p(1)$. First, we have

$$\left|\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\partial^3\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1^2\partial\theta'}\right)\right|\leq\frac{1}{\sqrt{T}}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}M\left(Y_{it},X_{it}\right)\right)=o_p\left(1\right).$$

Second, we have

$$\begin{aligned}&\frac{2}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}\right)\\&=\frac{2}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}E\left[\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)\\&+\frac{2}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}-E\left[\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)\right),\end{aligned}$$

which we further write as

$$\begin{aligned}&\frac{2}{N\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\left(\frac{1}{T}\sum_{t'=1}^{T}E\left[\frac{\partial^2\ln f\left(Y_{it'}|\,X_{it'},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)\\&+\frac{2}{NT\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\left(\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}-E\left[\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)\\&+\frac{2}{NT\sqrt{T}}\sum_{i=1}^{N}\sum_{t,t'=1,t\neq t'}^{T}\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\left(\frac{\partial^2\ln f\left(Y_{it'}|\,X_{it'},\theta_0\right)}{\partial\theta_1\partial\theta'}-E\left[\frac{\partial^2\ln f\left(Y_{it'}|\,X_{it'},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right).\end{aligned} \tag{C.3}$$

The first term of (C.3) has mean zero and variance equal to

$$\begin{aligned}&\frac{4}{N^2T}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[\left(\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\right)^2\right]\left(\frac{1}{T}\sum_{t'=1}^{T}E\left[\frac{\partial^2\ln f\left(Y_{it'}|\,X_{it'},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)^2\\&\leq\frac{4}{N^2T}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[M\left(Y_{it},X_{it}\right)\right]\left(\frac{1}{T}\sum_{t=1}^{T}E\left[M\left(Y_{it},X_{it}\right)\right]\right)^2=O\left(N^{-1}\right),\end{aligned}$$

so it should be $o_p(1)$. As for the second term of (C.3), we have

$$E\left|\frac{2}{NT\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1}\left(\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}-E\left[\frac{\partial^2\ln f\left(Y_{it}|\,X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)\right|$$

$$\leq \frac{2}{\sqrt{T}} \sup_{i,t} E\left[\left| \frac{\partial \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1} \left(\frac{\partial^2 \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1 \partial \theta'} - E\left[\frac{\partial^2 \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1 \partial \theta'} \right] \right) \right|\right]$$

$$\leq \frac{2}{\sqrt{T}} \sup_{i,t} \sqrt{E\left[\left| \frac{\partial \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1} \right|^2\right]} \sqrt{E\left[\left| \left(\frac{\partial^2 \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1 \partial \theta'} - E\left[\frac{\partial^2 \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1 \partial \theta'} \right] \right) \right|^2\right]}$$

$$= O\left(T^{-1/2}\right),$$

so it should be $o_p(1)$. As for the third term of (C.3), we note that conditional on Xs, it can be viewed as a sum of $\sum_{t,t'=1,t\neq t'}^{T} \frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial \theta_1} \left(\frac{\partial^2 \ln f(Y_{it'}|X_{it'},\theta_0)}{\partial \theta_1 \partial \theta'} - E\left[\frac{\partial^2 \ln f(Y_{it'}|X_{it'},\theta_0)}{\partial \theta_1 \partial \theta'} \right] \right)$, which are independent over i. Therefore, it has mean zero and variance equal to

$$\frac{4}{N^2 T^3} \sum_{i=1}^{N} \mathrm{Var}\left(\sum_{t,t'=1,t\neq t'}^{T} \frac{\partial \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1} \left(\frac{\partial^2 \ln f\left(Y_{it'} | X_{it'}, \theta_0\right)}{\partial \theta_1 \partial \theta'} - E\left[\frac{\partial^2 \ln f\left(Y_{it'} | X_{it'}, \theta_0\right)}{\partial \theta_1 \partial \theta'} \right] \right) \right).$$

By a similar reasoning, we can see that for $t \neq t'$,

$$\frac{\partial \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1} \left(\frac{\partial^2 \ln f\left(Y_{it'} | X_{it'}, \theta_0\right)}{\partial \theta_1 \partial \theta'} - E\left[\frac{\partial^2 \ln f\left(Y_{it'} | X_{it'}, \theta_0\right)}{\partial \theta_1 \partial \theta'} \right] \right)$$

has mean equal to zero and the variance uniformly bounded over i and t. Therefore, we can conclude that the variance of

$$\sum_{t,t'=1,t\neq t'}^{T} \frac{\partial \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1} \left(\frac{\partial^2 \ln f\left(Y_{it'} | X_{it'}, \theta_0\right)}{\partial \theta_1 \partial \theta'} - E\left[\frac{\partial^2 \ln f\left(Y_{it'} | X_{it'}, \theta_0\right)}{\partial \theta_1 \partial \theta'} \right] \right)$$

is of order T^3.

This implies that the third term of (C.3) has mean zero and variance of order $O\left(N^{-1}\right)$, so it should be $o_p(1)$.