Breusch and Pagan's (1980) Test Revisited

Jinyong Hahn* Ruoyao Shi†

UCLA UC Riverside

June 22, 2021

Abstract

We consider the local asymptotic power of Breusch and Pagan's (1980) test for the general nonlinear models. The test is motivated by the random effects, but we consider the fixed effects for the alternative hypothesis, derive the local power, and show that the test has a power to detect the fixed effects. We also examine how the estimation noise of the maximum likelihood estimator changes the asymptotic distribution of the test under the null, and show that such a noise may be ignored in a large n large T situation, which may have a convenient implication in the possible application of the test to network models.

Keywords: Lagrange multiplier test, fixed effects, local power, error component model.

1 Introduction

Panel data analysis is most often concerned with the best way to overcome the presence of individual fixed effects. Although the fixed effects can be differenced away in linear models, such convenience is often unavailable in nonlinear models, and many papers were published to overcome the incidental parameters problems that arise as a consequence.

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

†Department of Economics, UC Riverside, Riverside, CA 92521. Email: ruoyao.shi@ucr.edu.

Such a concern would be a non-issue if the unobserved individual heterogeneity were not present in the data set to begin with. Without the fixed effects in the data, the statistical analysis does not have to involve any special technique, and most of the issues would disappear. Therefore, one may be interested in testing if there is any unobserved heterogeneity in the data set.

The most well-known test for detecting unobserved heterogeneity is due to Breusch and Pagan (1980, BP test hereafter). Although the BP test was originally developed to deal with both individual and time effects, the version of the BP test to detect only the individual effects seems to have received the most attention. Such a version of the BP test can be interpreted to be a test of overdispersion (Cox, 1983), and it is related to White's (1982) information matrix test, as was pointed out by Chesher (1984). See also Lancaster (1984) for the asymptotic distribution of the test under the null. From our point of view, the most convenient feature of the BP test is that it is a Lagrange Multiplier (LM) test, which requires calculation of the parameter estimates only under the null, i.e., under the null of no unobserved heterogeneity. This feature makes it pragmatically very attractive due to the simplification of computation; there is no unobserved heterogeneity under the null, and hence, the computational problem disappears. See also Engel (1984).

The first contribution of our paper is to address the question of power[1] of the BP test. To our knowledge, Honda (1985) is the only one who analyzed the asymptotic power of the BP test against *random* effects and he did so in the *linear* models. We make a progress over Honda (1985) in two dimensions. First, we derive the asymptotic power of the BP test against random effects in general *nonlinear* models.[2] Second, which is more important, we also consider the power against the alternative of *fixed* effects. By the fixed effects, we mean the type of general unobserved variables that may have arbitrary dependence structure with the observed explanatory variables. Chamberlain (1984) called such a variable the correlated random effects. The BP test was specifically designed to

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

[2]While Honda (1985) considered both individual and time effects, we focus only on individual effects. The presence of both effects presents an additional technical challenge in nonlinear models, and we leave it as a topic of future research.

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.[3] Our paper fills this gap in the literature and analyzes the local power of the BP test against the general alternative of fixed effects. Modifying Newey (1985), we obtain the asymptotic results, based on which we argue that the BP test in general has a power against the fixed effects.[4] We also find that the linear model has a peculiar feature and the power of the BP test against fixed effects may be lower than against random effects. As a by-product of our analysis of the linear model, we propose an intuitive LM-like test for detecting further neglected heterogeneity in the linear model with fixed effects.

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 panel data analysis. This is a generic problem for which we are unable to offer a solution. Yet at the same time, we see an interesting twist in the panel context. Nonlinear panel data analysis is often unable to eliminate the incidental parameters problem and ends up mitigating (reducing) the bias at best. In other words, if the BP test rejects the null of no unobserved heterogeneity, and a researcher adopts the usual bias correction for the panel data analysis, we should confront the fact that the bias is not completely removed anyway. This is in contrast with the cross sectional case where the estimators of the (typically more general) model under the alternative is associated with (asymptotically) unbiased estimation. Therefore, the pre-test problem in panel models may be less severe than in cross sectional models, from a pragmatic perspective. Another complication is the problem that the bias correc-

[3]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, Moon, and Snider (2017). Any test of conditional moment restrictions test is equivalent to a test of infinitely many unconditional moments, and hence, it may not be as appealing when compared to the simplicity of the BP test, except that the BP test was motivated to deal with the alternative of random effects.

[4]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.

tion technique in the two-way fixed effects models is not as well developed as in one-way models. To our knowledge, Fernández-Val and Weidner's (2016) paper is the only work in the literature that addresses the issue, and their result is predicated on a certain concavity assumption (Assumption 4.1.v), which limits its applicability. The severity of the pre-test bias is a topic that we leave as a future research topic.

The BP test was originally proposed to detect unobserved heterogeneities in two-way error component models, i.e., the panel models with both individual and time effects. In order to address such a structure, Honda (1985) adopted an asymptotic framework where both the cross sectional dimension (n) and the time series dimension (T) grow to infinity. In contrast, Chesher's (1984) and Lancaster's (1984) analyses are predicated on the IID assumption on observations, which would be sensible only with fixed n or T; otherwise, the dimension of observations would have to grow as a function of the sample size, invalidating the identical distribution assumption. Honda's (1985) technical analysis, while limited to the linear models, hints that the feasible version of the BP test based on the MLE does not need to address the noise of estimation in MLE in two-way models (for characterization of the asymptotic distribution under the null hypothesis). We analyze his argument carefully, and derive the mathematical implication that the noise in estimation of MLE does not affect the asymptotic distribution even for the nonlinear models if n and T are both large. This means that Lancaster's (1984) adjustment is unnecessary for two-way models,[5] which may have a convenient implication in the possible application of the test to network models.

2 Review of the LM Test

In this section, we consider the panel model with possible unobserved individual heterogeneity and present the LM test to detect neglected heterogeneity. It is largely a review of the BP test as well as Chesher (1984), albeit with some mild modification to make it easier to accommodate both fixed effects and random effects later. Throughout most of the

[5]We also find that the proof of Honda's (1985) main result is incorrect. See Footnote 15.

paper, we adopt the framework of the one-way error component model. Some interesting complications in two-way error component models are discussed later in Section 6.

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 $X_i = (X'_{i1}, \ldots, X'_{iT})'$. We will assume that the conditional density of Y_i given X_i is given by the function $f(y|\, x, \theta)$, 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 of the form

$$Y_{it} = x'_{it}\beta + \alpha_i + v_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T,$$

where α_i denotes the unobserved individual heterogeneity. Suppose that x_{it} does not include any intercept term, $x_i \equiv (x'_{i1}, \ldots, x'_{iT})'$, and let X_{it} denote x_{it} as well as the intercept term. Suppose that $v_{it} \sim N(0, \sigma_v^2)$ is independent of (X'_i, α_i) and is independent over i and t. We can then understand $\theta = (\alpha_i, \beta, \sigma_v^2)'$. 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.[6]

The heterogeneity of the first component θ_1 of θ can be modeled as $\theta_{0,1}$ plus a random variable. 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 $X_i = x$ is $k(\cdot)$. The conditional density of Y_i given $X_i = x$ and $\varepsilon_i = e$ is then equal to $f(y|\, x, \theta_0 + \eta e\iota) = f(y|\, x, (\theta_{0,1} + \eta\varepsilon, \theta_{0,2}, \ldots, \theta_{0,q}))$, where the vector ι is such that the first component is 1

[6] There is no reason that the LM test should be confined to the scalar effects, as is evident from Chesher's (1984) 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).

and the rest are all zeros. It follows that the conditional density of Y_i given $X_i = x$ is

$$h\left(y|\,x,\theta_0,\eta\right) \equiv \int f\left(y|\,x,\theta_0+\eta e\iota\right) k\left(e\right) de,$$

and note that $h\left(y|\,x,\theta,0\right) = f\left(y|\,x,\theta\right)$.

We consider the second order Taylor series expansion of $h\left(y|\,x,\theta,\eta\right)$ with respect to η around $(\theta,\eta) = (\theta_0, 0)$. Under the assumption that we can exchange differentiation and integration, we obtain

$$\begin{aligned}
\left.\frac{\partial h\left(y|\,x,\theta_0,\eta\right)}{\partial\eta}\right|_{\eta=0} &= \int \frac{\partial f\left(y|\,x,\theta_0\right)}{\partial\theta_1} e k\left(e\right) de = \frac{\partial f\left(y|\,x,\theta_0\right)}{\partial\theta_1}\int e k\left(e\right) de \\
&= \frac{\partial f\left(y|\,x,\theta_0\right)}{\partial\theta_1} E\left[\varepsilon_i\right] = 0, \\
\left.\frac{\partial^2 h\left(y|\,x,\theta_0,\eta\right)}{\partial\eta^2}\right|_{\eta=0} &= \int \frac{\partial^2 f\left(y|\,x,\theta_0\right)}{\partial\theta_1^2} e^2 k\left(e\right) de = \frac{\partial^2 f\left(y|\,x,\theta_0\right)}{\partial\theta_1^2}\int e^2 k\left(e\right) de \\
&= \frac{\partial^2 f\left(y|\,x,\theta_0\right)}{\partial\theta_1^2} E\left[\varepsilon_i^2\right].
\end{aligned}$$

Therefore, we have

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

where $\sigma_\varepsilon^2 = E\left[\varepsilon_i^2\right]$. Given the form of the expansion, it would make sense to consider the parameterization $h\left(y|\,x,\theta_0,\sqrt{\eta}\right)$ instead (i.e., $f\left(y|\,x,\theta_0+\sqrt{\eta}e\iota\right)$), which delivers the expansion[7]

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

This implies that

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

It follows that the LM test can be based on the score

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

[7] 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., $E\left[\varepsilon_i\right] = 0$.

or equivalently based on

$$\frac{\partial^2 f\left(y|\,x,\theta_0\right)/\,\partial\theta_1^2}{f\left(y|\,x,\theta_0\right)}=\frac{\partial^2\ln f\left(y|\,x,\theta_0\right)}{\partial\theta_1^2}+\left(\frac{\partial\ln f\left(y|\,x,\theta_0\right)}{\partial\theta_1}\right)^2, \tag{2}$$

which provides the basis of the information matrix test interpretation.

Note that the derivation above was predicated on the random effects assumption. If we allow arbitrary conditional density $k\left(\cdot|\,x\right)$ of ε_i given $X_i=x$, which is appropriate under the fixed effects specification, we would get

$$\begin{aligned}\left.\frac{\partial h\left(y|\,x,\theta_0,\eta\right)}{\partial\eta}\right|_{\eta=0}&=\int\frac{\partial f\left(y|\,x,\theta_0\right)}{\partial\theta_1}ek\left(e|\,x\right)de=\frac{\partial f\left(y|\,x,\theta_0\right)}{\partial\theta_1}\int ek\left(e|\,x\right)de\\&=\frac{\partial f\left(y|\,x,\theta_0\right)}{\partial\theta_1}\mu\left(x\right).\end{aligned} \tag{3}$$

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

$$E\left[\frac{\partial f\left(Y_i|\,X_i,\theta_0\right)/\,\partial\theta_1}{f\left(Y_i|\,X_i,\theta_0\right)}\mu\left(X_i\right)\right]=0 \tag{4}$$

holds for all $\mu\left(X_i\right)$. This prompted Hahn, Moon, and Snider (2017) to conclude that any test of the conditional moment restriction[8]

$$E\left[\left.\frac{\partial f\left(Y_i|\,X_i,\theta_0\right)/\,\partial\theta_1}{f\left(Y_i|\,X_i,\theta_0\right)}\right|X_i\right]=0 \tag{5}$$

can be a possible test of fixed effects. See Section 5 for other variants of this idea for linear models.

3 Power of the BP Test Against Random Effects

In the previous section, we summarized the current status of the literature. First, if we are to use the LM motivation to detect the fixed effects, an appropriate test would be to

[8]Hahn, Moon, and Snider (2017) recognized that the test of the conditional moment restrictions (5) would be a test of fixed effects, and they did not develop a separate test. Instead, they recommended using any test of conditional moment restrictions in the literature, such as Newey (1985), Bierens (1990), and Donald, Imbens, and Newey (2003).

test the conditional moment restriction (5). Because it is equivalent to an infinite number of unconditional moment restrictions (4), a practitioner needs to confront the resultant statistical complications. Second, the BP test (2) was developed to detect the random effects. Because it is an LM test, it suffices to estimate the parameters under the null hypothesis of no neglected heterogeneity, which may be very convenient computationally. Therefore, one may ask a question whether the BP test can actually detect the fixed effects, even though the fixed effects were not the initial target. Although some power may be sacrificed relative to a test of the conditional moment restrictions (5), perhaps such a cost may be justified from the pragmatic perspective of avoiding a potentially complicated statistical procedure. We are not aware of any paper in the literature that posed such a question, which our paper proposes to tackle.

In this section, we begin by examining the (local) power of the BP test against 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 models, an obvious gap in the literature if our library research is correct. Using (2), we can see that our BP statistic can be written as $m_n\left(\bar{\theta}_n\right)$, where

$$m_n\left(\theta\right) \equiv n^{-1}\sum_{i=1}^{n} m\left(Z_i,\theta\right), \quad m\left(z,\theta\right) \equiv \frac{\partial^2 f\left(y|\, x,\theta\right)/\,\partial\theta_1^2}{f\left(y|\, x,\theta\right)}, \tag{6}$$

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's (1985) analysis is almost tailor-made for our purpose, which we adopt as the main tool of analysis. 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 alternative of the form $\theta_{1,i} = \theta_{0,1} + n^{-1/4}\varepsilon_i$.[9] As a result, we will consider the local alternative of random

[9] 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.

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 $E[\varepsilon_i^2] = \sigma_\varepsilon^2$.

Below is the main result when we consider the alternative of random effects:

Theorem 1 *Under Assumptions 1 - 7 detailed in Appendix A, we get*

$$\sqrt{n} m_n\left(\bar{\theta}_n\right) \xrightarrow{d.} N\left(\frac{\sigma_\varepsilon^2}{2}\left(\kappa_1 - \kappa_2' \mathcal{I}^{-1}\kappa_2\right), \kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2\right),$$

where

$$\begin{aligned} s(z,\theta_0) &\equiv \frac{\partial f(y|\, x,\theta_0)/\partial\theta}{f(y|\, x,\theta_0)}, \\ \mathcal{I} &\equiv E\left[s(Z_i,\theta_0)\, s(Z_i,\theta_0)'\right] = -E\left[\frac{\partial s(Z_i,\theta_0)}{\partial\theta'}\right], \\ \kappa_1 &\equiv E\left[\left(\frac{\partial^2 f(Y_i|\, X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|\, X_i,\theta_0)}\right)^2\right], \\ \kappa_2 &\equiv E\left[\frac{\partial^2 f(Y_i|\, X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|\, X_i,\theta_0)} s(Z_i,\theta_0)\right]. \end{aligned}$$

Proof. In Appendix B. ■

The above theorem implies that (i) the test statistic would take the form

$$\frac{\left(\sqrt{n} m_n\left(\bar{\theta}_n\right)\right)^2}{\hat{\kappa}_1 - \hat{\kappa}_2' \hat{\mathcal{I}}^{-1}\hat{\kappa}_2}, \tag{7}$$

where $\hat{\kappa}_1, \hat{\kappa}_2, \hat{\mathcal{I}}$ are consistent estimators of $\kappa_1, \kappa_2, \mathcal{I}$; and (ii) the asymptotic distribution under $\theta_{1,i} = \theta_{0,1} + n^{-1/4}\varepsilon_i$ is non-central χ_1^2 distribution with noncentrality parameter equal to $\left(\frac{\sigma_\varepsilon^2}{2}\right)^2 (\kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2)$. Note that $\kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2$ can be interpreted to be the variance of the residual when $m(Z_i,\theta_0)$ is regressed on $s(Z_i,\theta_0)$. 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 pathological anomaly as $\kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2 = 0$ should not be expected in practice.

Remark 1 *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 explains the test statistic (7). See also Lancaster (1984).*

Remark 2 *Note that* κ_1 *denotes 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 Section 5.*

Remark 3 *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 6 in Appendix B. In other words, if the test 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 alternative 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).*[10] *If such a diagnostic test is desired, one can test the null hypothesis* $\kappa_2 = 0$ *by evaluating the test statistic based on*

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

Using standard arguments, it can be shown that the above expression is equal to

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

where[11]

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

It follows that

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

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

[11] 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 an alternative form of the asymptotic variance.

where $\kappa_3 \equiv E\left[\left(\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right)^2 s\left(Y_i|\, X_i, \theta_0\right) s\left(Y_i|\, X_i, \theta_0\right)'\right]$*, and the 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 static under the null* $\kappa_2 = 0$ *is* χ_q^2*.*

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 LM test statistic in equation (7) 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*.*

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

$$Y_{it} = \mathbb{I}\{x_{it}'\beta_0 + \alpha_{n,i} + v_{it} \geq 0\}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T. \tag{8}$$

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

$$\ln f\left(Y_i|\, X_i, \theta_0\right) = \sum_{t=1}^{T}\left(Y_{it}\ln\frac{\exp(X_{it}'\theta_0)}{1+\exp(X_{it}'\theta_0)} + (1-Y_{it})\ln\frac{1}{1+\exp(X_{it}'\theta_0)}\right).$$

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

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

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

and we can see that our test statistic is based on

$$m(Z_i, \theta_0) \equiv \frac{\partial^2 \ln f(Y_i|X_i, \theta_0)}{\partial\theta_1^2} + \left(\frac{\partial \ln f(Y_i|X_i, \theta_0)}{\partial\theta_1}\right)^2$$

$$= -\sum_{t=1}^{T} \Lambda_{it}(1-\Lambda_{it}) + \left(\sum_{t=1}^{T}(y_{it}-\Lambda_{it})\right)^2,$$

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

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

4 Power of the BP Test Against Fixed Effects

The discussion in Section 2 indicates that the parameterization $h\left(y|\, x, \theta_0, \sqrt{\eta}\right)$ is appropriate for local power analysis when $E\left[\varepsilon_i|\, X_i\right] = 0$ 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 parametrization captures the appropriate second order effects, as is evident in the derivation of (1). Therefore, a useful synthesis is to consider the local parameterization of the form

$$f\left(y\,\middle|\, x, \left(\theta_{0,1} + \frac{\mu(x)}{n^{1/2}} + \frac{\varepsilon^*}{n^{1/4}}, \theta_{0,2}, \ldots, \theta_{0,q}\right)\right), \tag{9}$$

where $E\left[\varepsilon^*|\, x\right] = 0$.

Below is the main result when we consider the alternative of fixed effects of the form (9):

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

$$\sqrt{n} m_n\left(\bar{\theta}_n\right) \xrightarrow{d.} N\left(\left[I_1, -\kappa_2'\mathcal{I}^{-1}\right]\left(K_F + K_R^*\right), \kappa_1 - \kappa_2'\mathcal{I}^{-1}\kappa_2\right),$$

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

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

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

Proof. In Appendix C. ■

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

$$Y_{it} = x_{it}'\beta_0 + \alpha_{n,i} + v_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T,$$

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[12]

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

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

$$\begin{aligned}\kappa_2&=E\left[s(Y_i|X_i,\theta_0)\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right]\\&=E\left[\left(\frac{1}{\sigma_v^2}\sum_{t=1}^T X_{it}(y_{it}-X_{it}'\theta_0)\right)\left(-\frac{T}{\sigma_v^2}+\left(\frac{1}{\sigma_v^2}\sum_{t=1}^T(Y_{it}-X_{it}'\theta_0)\right)^2\right)\right]\\&=0,\end{aligned}$$

leading to the implication

$$\left[I_1,-\kappa_2'\mathcal{I}^{-1}\right]K_F=0. \tag{10}$$

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

[12] We have

$$\ln f(Y_i|X_i,\theta_0)=C-\frac{1}{2\sigma_v^2}\sum_{t=1}^T(Y_{it}-X_{it}'\theta)^2,$$

so

$$s(Y_i|X_i,\theta_0)=\frac{1}{\sigma_v^2}\sum_{t=1}^T X_{it}(Y_{it}-X_{it}'\theta),\quad\text{and}\quad\frac{\partial^2\ln f(Y_i|X_i,\theta_0)}{\partial\theta\partial\theta'}=-\frac{1}{\sigma_v^2}\sum_{t=1}^T X_{it}X_{it}'.$$

In particular, we have

$$\frac{\partial\ln f(Y_i|X_i,\theta_0)}{\partial\theta_1}=\frac{1}{\sigma_v^2}\sum_{t=1}^T(Y_{it}-X_{it}'\theta)\quad\text{and}\quad\frac{\partial^2\ln f(Y_i|X_i,\theta_0)}{\partial\theta_1^2}=-\frac{T}{\sigma_v^2}.$$

We therefore see that

$$\begin{aligned}\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(y|x,\theta_0)}&=\frac{\partial^2\ln f(Y_i|X_i,\theta_0)}{\partial\theta_1^2}+\left(\frac{\partial\ln f(Y_i|X_i,\theta_0)}{\partial\theta_1}\right)^2\\&=-\frac{T}{\sigma_v^2}+\left(\frac{1}{\sigma_v^2}\sum_{t=1}^T(Y_{it}-X_{it}'\theta)\right)^2.\end{aligned}$$

Therefore, we should have

$$\begin{aligned}&E\left[s(Y_i|X_i,\theta_0)\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right]\\&=E\left[\left(\frac{1}{\sigma_v^2}\sum_{t=1}^T X_{it}(Y_{it}-X_{it}'\theta_0)\right)\left(-\frac{T}{\sigma_v^2}+\left(\frac{1}{\sigma_v^2}\sum_{t=1}^T(Y_{it}-X_{it}'\theta_0)\right)^2\right)\right].\end{aligned}$$

out that the issue is a little subtle, and the BP test does have a power to detect $\mu(X_i)$ as long as it is in the $n^{-1/4}$-neighborhood, not $n^{-1/2}$. We note that the components of $[I_1, -\kappa_2' \mathcal{I}^{-1}] K_F$ measures 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's (1984) derivation. We provide such an analysis in the next section.

5 In Depth Analysis of the Linear Model

In this section, we present two results. First, we show that the BP test is able to detect the fixed effects as long as they are in the $O\left(n^{-1/4}\right)$ neighborhood. This does not contradict our analysis in the previous section because we considered the case where the fixed effects are in the $O\left(n^{-1/2}\right)$ neighborhood. Second, we present a pragmatic version of the test of the conditional moment restriction (5) for the purpose of detecting the model considered by Bonhomme and Manresa (2015). The test can be argued to be a variant of the generic test of over-identification considered by Chamberlain (1984, Section 4.2), but geared for the particular variant of the panel model in Bonhomme and Manresa (2015).

We first present a more detailed analysis tailored to exploit the particulars of the linear model

$$Y_{it} = X_{it}'\theta + \frac{\mu(X_i) + \varepsilon_i^*}{n^{1/4}} + v_{it}, \quad t = 1, \ldots, T \tag{11}$$

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

$$\frac{\partial^2 f(Y_i | X_i, \theta_0) / \partial\theta_1^2}{f(Y_i | X_i, \theta_0)} = -\frac{T}{\sigma_v^2} + \left(\frac{1}{\sigma_v^2} \sum_{t=1}^{T} (Y_{it} - X_{it}'\theta)\right)^2,$$

so the counterpart of m is equal to

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\left(\sum_{t=1}^{T} \hat{v}_{it}\right)^2 - T\hat{\sigma}_v^2}{\left(\hat{\sigma}_v^2\right)^2} = \frac{1}{n} \frac{\hat{v}'[I_n \otimes (e_T e_T' - I_T)]\hat{v}}{\left(\hat{\sigma}_v^2\right)^2} \tag{12}$$

where the $\hat{v}_{it}$ denotes the OLS residual, e_T is a $T \times 1$ vector of ones, and $\hat{\sigma}_v^2 \equiv (nT)^{-1}$

$\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{v}_{it}^{2}$. We will assume that $T^{-1}E\left[\sum_{t=1}^{T} X_{it}X_{it}'\right]$ is positive definite with finite eigenvalues. We also assume that $E\left[\varepsilon_i^* v_{it}\right] = 0$ and $E\left[\varepsilon_i^* \mu\left(X_i\right)\right] = 0$.

In order to analyze the power of the BP test under the alternative (11), 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 (12). In Lemma 1, we show how Honda's (1985) Lemma 1 should be changed under the alternative of fixed effects:

Lemma 1 *Under (11), we have*

$$\begin{aligned}\frac{\hat{v}'[I_n \otimes (e_T e_T' - I_T)]\hat{v}}{n^{1/2}} &= \frac{1}{n^{1/2}} \sum_{i=1}^{n} \sum_{l \neq m}^{T} v_{il} v_{im} + T\left(T-1\right) E\left[\xi_i^2\right] \\ &\quad - 2T\lambda' Q^{-1} \lambda + \left(Q^{-1}\lambda\right)' S \left(Q^{-1}\lambda\right) + O_p\left(n^{-1/4}\right),\end{aligned}$$

where $Q \equiv T^{-1}E\left[\sum_{t=1}^{T} X_{it}X_{it}'\right]$, $\bar{X}_i \equiv T^{-1}\sum_{t=1}^{T} X_{it}$, $\lambda \equiv E\left[\bar{X}_i \mu\left(X_i\right)\right]$, $\xi_i \equiv \mu\left(X_i\right) + \varepsilon_i^*$ *and* $S \equiv E\left[\sum_{l \neq m}^{T} X_{il}X_{im}'\right]$.

Proof. In Appendix D. ■

In Lemma 1, 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 in the previous section, where the BP test was unable to detect $\mu\left(X_i\right)$ in the linear model. The difference is that the previous section considered the parameterization where $\mu\left(X_i\right)$ was too close to zero in the $O\left(n^{-1/2}\right)$ neighborhood, while we show in the current section 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.

In Section 2, a version of the test of conditional moment restrictions (5) was mentioned as a way of detecting neglected heterogeneity. We can easily see how this idea can be generalized to test the specification of Bonhomme and Manresa (2015), who considered estimating models of the form $Y_{it} = x_{it}'\beta + \alpha_{g_i t} + v_{it}$, where g_i denotes the group that the ith observation belongs to. This can be understood to be a generalization of the vanilla model $Y_{it} = x_{it}'\beta + \alpha_i + v_{it}$ (possibly with the time effects as well) against the alternative that there is some neglected heterogeneity in α_i possibly "correlated" with x_i ($\mu(x) \neq 0$), even under the presence of the individual fixed effects. Because in the linear model the BP test does not have a good power when the neglected heterogeneity may be correlated with x_i, we may consider adopting a version of the conditional moment test. For this purpose, it is useful to see that the LM test can be conducted through the moments[13]

$$E\left[\Delta x_{it}\left(\Delta Y_{it} - \Delta x_{it}'\beta_W\right)\right] = 0, \quad t = 2, \ldots, T, \tag{13}$$

where Δ denotes the first differencing operator, while β_W solves the moments

$$E\left[\sum_{t=1}^{T} \tilde{x}_{it}\left(\tilde{Y}_{it} - \tilde{x}_{it}'\beta_W\right)\right] = 0, \tag{14}$$

where $\tilde{x}_{it} = x_{it} - \bar{x}_i$, and $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$. If the test rejects, then we should suspect that there is neglected heterogeneity not captured by the usual individual fixed effects, and as such, there is a reason to consider the specification of Bonhomme and Manresa (2015). We applied the test to Bonhomme and Manresa's (2015) balanced panel, i.e., the data used by Acemoglu et al. (2008), where $n = 90$, $T = 7$ and $\dim(x) = 2$. We first run the OLS of ΔY_{it} on an intercept (to capture time dummies) and Δx_{it} for $t = 2, ..., 7$ and tabulate the results in Table 1. The estimates of the slope coefficients appear to be unstable over time, suggesting the presence of some neglected heterogeneity that lead to violation of (13).[14] For testing, we included the time dummy in the basic model, and adjusted for the estimation of the within estimator (14) when characterizing the asymptotic variance

[13] We do not consider $t = 1$ because of redundancy. If there is a prior suspicion that the fixed effects may have changed towards the end of the sample, one can test the subset of the moments in (13).

[14] Note the difference between the pooled first difference estimator and the two-way fixed effects estimator, presented in the last two columns of Table 1. (The two-way fixed effects estimator correspond to

of the empirical moment of (13) evaluated at the within estimator, allowing for possible heteroskedasticity. The p-value for all periods pooled is 0.0079, and the p-values for each period separately ($t = 2, ..., 7$) are: 0.0037, 0.1030, 0.0193, 0.1316, 0.0211, 0.0027. These results imply that the usual fixed effects specification is not flexible enough to explain the data, which may give credence to Bonhomme and Manresa's (2015) specification.

On the other hand, when Bonhomme and Manresa's (2015) specification is adopted (the number of groups $G = 4$), the LM test can be conducted through the moments

$$E\left[\sum_{g=1}^{G} x_{it}(Y_{it} - \alpha_{g,t} - x_{it}'\beta_{BM})I\{g_i = g\}\right] = 0, \quad t = 1, \ldots, T, \tag{15}$$

where β_{BM} solves the moments

$$\begin{aligned} E\left[\sum_{t=1}^{G}(Y_{it} - \alpha_{g,t} - x_{it}'\beta_{BM})I\{g_i = g\}\right] &= 0, \quad t = 1, \ldots, T, \\ E\left[\sum_{g=1}^{G}\sum_{t=1}^{T} x_{it}(Y_{it} - \alpha_{g,t} - x_{it}'\beta_{BM})I\{g_i = g\}\right] &= 0. \end{aligned} \tag{16}$$

The p-value for all periods pooled is 0.9990, and the p-values for each period separately ($t = 1, \ldots, 7$) are: 0.8507, 0.9905, 0.9748, 0.9421, 0.4979, 0.9993, 0.7634. These results are compatible with the claim that Bonhomme and Manresa's (2015) specification with four groups capture the time-varying heterogeneity in the data.

6 Discussion: Two-Way Error Component Model

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, we explained that this seeming contradiction can be understood by noticing

Table S10, Column 4 in Bonhomme and Manresa (2015).) Note that both estimators are consistent if the two-way fixed effects specification is correct. Therefore, this difference can be taken as another evidence of incorrect specification.

Table 1: OLS of ΔY_{it} on Δx_{it} (Standard Errors in Parentheses)

t	2	3	4	5	6	7	All Periods Pooled	Two-way FE
β_1	-0.3152	-0.2353	-0.4912	-0.0993	0.0124	-0.2543	-0.2623	0.2835
	(0.0798)	(0.1384)	(0.1465)	(0.0679)	(0.1712)	(0.0743)	(0.0495)	(0.0573)
β_2	0.1432	0.0576	0.1154	-0.0964	-0.4165	0.0393	-0.0300	-0.0313
	(0.1267)	(0.1880)	(0.0916)	(0.1121)	(0.1623)	(0.1129)	(0.0498)	(0.0490)

that Lancaster's (1984) adjustment is unnecessary for *linear* models. See Remark 2. In this section, we go one step further to show that Lancaster's (1984) adjustment is unnecessary for general two-way error component models. Lancaster (1984) implicitly adopted a "large n, fixed T" asymptotics, which is natural for one-way 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.[15] Variants of the two way error component model are adopted in the recent literature on networks, where the presence of unobserved individual effects is a challenge. See Graham (2017, 2020). It is therefore of interest to test whether these unobserved effects are present in the data or not. Our result in this section has a convenient pragmatic implication that the BP test can be used without the need to reflect the noise of estimating the MLE on

[15] Honda's (1985) Lemma 2 consists of several steps. He first establishes (correctly) that two normalized sums ($\sum_{i=1}^{N} p_i \big/ \sqrt{N}$ and $\sum_{t=1}^{T} w_t \big/ \sqrt{T}$) are asymptotically normal. He then establishes that $E\left[\left(\sum_{i=1}^{N} p_i \big/ \sqrt{N}\right)\left(\sum_{t=1}^{T} w_t \big/ \sqrt{T}\right)\right] = 0$, based on which he concludes that $\sum_{i=1}^{N} p_i \big/ \sqrt{N}$ and $\sum_{t=1}^{T} w_t \big/ \sqrt{T}$ are asymptotically independent. The last step is incorrect because the lack of correlation between two random normal variables does not guarantee independence, unless they are jointly normal to begin with. Therefore, the proof of his main result is invalid. It is not clear to us whether or how it is possible to fix his Lemma 2; the p_i and w_t are dependent on each other (if not correlated), which makes the asymptotic analysis non-trivial to us.

the asymptotic distribution.[16]

The two way error component model of interest includes as a special case the linear model with individual and time effects $Y_{it} = x_{it}'\beta + \alpha_i + \gamma_t + v_{it}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$. As in Honda (1985), we will assume that $n, T \to \infty$, which necessitates that we are a little more specific about the specification of the likelihood. We will assume that conditional on all the xs, αs, and γs, the joint likelihood of the Ys is given by $\prod_{i=1}^n \prod_{t=1}^T f\left(Y_{it} | x_{it}, \alpha_i, \gamma_t, \beta\right)$. This implies that the LM test would be based on

$$\frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\partial^2 \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1^2} + \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \left(\sum_{t=1}^{T} \frac{\partial \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1} \right)^2, \tag{17}$$

and

$$\frac{1}{n\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{\partial^2 \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1^2} + \frac{1}{n\sqrt{T}} \sum_{t=1}^{T} \left(\sum_{i=1}^{n} \frac{\partial \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta_1} \right)^2. \tag{18}$$

See Appendix E for justification of the normalization by $\sqrt{n}T$ and $n\sqrt{T}$.

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 model with $n, T \to \infty$ asymptotics, the noise does not affect the asymptotic distribution. For this purpose, it suffices to examine the distribution of (17) evaluated at the MLE

$$\frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\partial^2 \ln f\left(Y_{it} | X_{it}, \bar{\theta}_n\right)}{\partial \theta_1^2} + \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \left(\sum_{t=1}^{T} \frac{\partial \ln f\left(Y_{it} | X_{it}, \bar{\theta}_n\right)}{\partial \theta_1} \right)^2, \tag{19}$$

where we recognize that the MLE has the influence function proportional to $\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \frac{\partial \ln f\left(Y_{it} | X_{it}, \theta_0\right)}{\partial \theta}$. Most importantly for our purpose, we should have $\sqrt{nT}\left(\bar{\theta}_n - \theta_0\right) = O_p(1)$. It can be shown that under mild regularity conditions, (19) has the same distribution as (17) under the null.[17] By similar argument, we can conclude that (18) has the same

[16]Unlike the classical two-way error component model, where two types of errors (individual and time effects) are included, the network model has only one type of error but it is included similar to the two-way error component model. For example, $Y_{ij} = 1\left(g\left(X_i, X_j\right)' \beta_0 + U_i + U_j - V_{ij} \geq 0\right)$, where Us correspond to the random effects and V_{ij} has standard logistic distribution. This implies that the counterparts of (17) and (18) are identical, and we do not need to establish their joint asymptotic distribution.

[17]See Appendix E.

asymptotic distribution as its feasible counterpart, where the two θ_0 in (18) are replaced by $\bar{\theta}_n$.[18]

7 Summary

We revisited the BP test and derived several interesting results. We showed that the test has a power against fixed effects, even though it was developed to detect random effects. Because of the simplicity of the BP test as well as the complexity of nonlinear panel data analysis, this has a convenient implication for analysis of nonlinear panel models with fixed 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 one-way linear models or general two-way models. Given the similarity between the two-way models and some network models, this result has a convenient pragmatic implication for analysis of networks.

[18] Our conclusion only requires that (17) and (18) are both unaffected by the noise of the estimation of θ_0. Hence, the joint asymptotic distribution of the random vector consisting of (17) and (18), if it is correctly established, is unaffected by such a noise. Of course the problem in Honda's (1985) Lemma 2 needs to be fixed, which we are unable to do yet.

Appendices

A 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.*

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

Let $\gamma \equiv (\theta', \eta)'$ and $\gamma \in \Gamma$. Let θ_j $(j = 1, \ldots, q)$ denote the jth element of θ.

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* $h(y|x, \gamma) \equiv f(y|x, (\theta_1 + \eta\varepsilon, \theta_2, \ldots, \theta_q)')$*.*

Assumption 3′ *Let* ε_i^* *be a random variable with a conditional probability density function* $k(\cdot)$ *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* $h(y|x, \gamma) \equiv f(y|x, (\theta_1 + \eta^2\mu(x) + \eta\varepsilon^*, \theta_2, \ldots, \theta_q)')$*.*

Remark 4 *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^**), satisfy 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 $|A| = \max_{i,j} |a_{ij}|$.

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*.*

Let $f_X(x)$ denote the marginal probability density function of X_i.

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* $|\partial^2 \ln f(y|x,\theta)/\partial\theta_j\partial\theta_k|$ $(j,k = 1,\ldots,q)$ *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* $\{z : h(y|x,\theta) > 0\}$ *is independent of* θ*.*

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

We write $g = (m', s')'$, where $m(z,\theta)$ is defined in equation (6) and $s\left(y|\, x,\theta\right) \equiv \partial \ln f\left(y|\, x,\theta\right)/\,\partial\theta$ denotes the score. Define

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

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 $g_n(\theta) \equiv n^{-1}\sum_{i=1}^n g(Z_i,\theta)$, and when the expectation exists, $\phi(\theta,\gamma) \equiv \int g(z,\theta)h(z,\gamma)dz$. 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)[19] and is helpful for the proof of our theorems.

Lemma A.1 *Suppose that,* Θ *is compact; for each* θ *in* Θ*,* $g(z,\theta)$ *is a measurable function of* z*; and for almost all* z *in* $\mathcal{Z}$*,* $g(z,\theta)$ *is a continuous function of* θ*. Suppose that,* Γ *is compact; for each* γ *in* Γ*,* $h(z,\gamma)$ *is a measurable probability density on* $\mathcal{Z}$*; for almost all* z

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

in $\mathcal{Z}$, $h(z,\gamma)$ is a continuous function of γ. Suppose that there exists 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 $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{\Theta} |g_n(\theta) - \phi(\theta,\gamma_0)| = 0. \tag{20}$$

Proof. See Appendix of Newey (1985). ■

B Proof of Theorem 1

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

Step 1 Let

$$\phi_\eta(\theta) \equiv \int g(z,\theta)\, h(y|\, x, \theta_0, \sqrt{\eta})\, f_X(x)\, dz, \tag{21}$$

$$\bar{V}_\eta \equiv \int g(z,\theta_0)\, g(z,\theta_0)'\, h(y|\, x, \theta_0, \sqrt{\eta})\, f_X(x)\, dz - \phi_\eta(\theta_0)\phi_\eta(\theta_0)'. \tag{22}$$

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 hypotheses of Lemma A.1, implying that $\phi_\eta(\theta)$ exists and is continuous on Γ. Then by Assumption 3, we have

$$\lim_{\eta\to 0} \phi_\eta(\theta_0) = \phi_0(\theta_0). \tag{23}$$

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

$$\phi_0(\theta_0) = E\left[g(Z_i,\theta_0)\right] = \int g\left(z,\theta_0\right) f\left(y|\, x,\theta_0\right) f_X\left(x\right) dz = 0. \tag{24}$$

By Assumption 5, functions $[m(z,\theta)]^2$ and $s(y|x,\theta)s(y|x,\theta)'$ satisfy the hypotheses of Lemma A.1, and so does $s(y|x,\theta)m(z,\theta)$ by the Cauchy-Schwarz inequality. Applying Lemma A.1 to $\bar{V}_\eta$ and we get

$$\lim_{\eta\to 0} \bar{V}_\eta = V, \tag{25}$$

and note that

$$\begin{aligned}
V &= E_0\left[g\left(Z_i,\theta_0\right) g\left(Z_i,\theta_0\right)'\right] \\
&= \begin{bmatrix} E\left[\left(\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right)^2\right] & E\left[\frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)} s\left(Y_i|\, X_i,\theta_0\right)'\right] \\ E\left[s\left(Y_i|\, X_i,\theta_0\right) \frac{\partial^2 f(Y_i|X_i,\theta_0)/\partial\theta_1^2}{f(Y_i|X_i,\theta_0)}\right] & E\left[s\left(Y_i|\, X_i,\theta_0\right) s\left(Y_i|\, X_i,\theta_0\right)'\right] \end{bmatrix} \\
&= \begin{bmatrix} \kappa_1 & \kappa_2' \\ \kappa_2 & \mathcal{I} \end{bmatrix}.
\end{aligned}$$

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 $W_\eta(z) \equiv \lambda'\left[g(z,\theta_0) - \phi_\eta(\theta_0)\right]$, and let $W_{\eta,i} \equiv W_\eta(Z_i)$ 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 equations (21) and (22), 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 equation (25). For any $\delta > 0$, define the set $A_{\delta,\eta} \equiv \left\{z : |\lambda'\left[g(z,\theta_0) - \phi_\eta(\theta_0)\right]| > \delta\left(n\lambda'\bar{V}_\eta\lambda\right)^{1/2}\right\}$.[20] Note that Z_i $(i = 1,\ldots,n)$ are identically distributed, so for any $\epsilon > 0$, we have

$$\left(n\lambda'\bar{V}_\eta\lambda\right)^{-1} \sum_{i=1}^{n} \int_{|W_{\eta,i}|\geq\delta\left(n\lambda'\bar{V}_\eta\lambda\right)^{1/2}} |W_{\eta,i}|^2 h(Y_i|X_i,\theta_0,\sqrt{\eta}) f_X(X_i) dZ_i$$

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

$$= \left(\lambda'\bar{V}_\eta\lambda\right)^{-1}\int_{A_{\delta,\eta}}|W_\eta(z)|^2 h(y|x,\theta_0,\sqrt{\eta})f_X(x)dz$$

$$\leq \left(\lambda'\bar{V}_\eta\lambda\right)^{-1}2(q+1)|\lambda|^2\left(|\phi_\eta(\theta_0)|^2\int_{A_{\delta,\eta}}a(z)dz+\int_{A_{\delta,\eta}}b(z)a(z)dz\right), \tag{26}$$

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 equations (23) and (24), $\lim_{\eta\to 0}\phi_\eta(\theta_0)=0$. By equation (25), 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, equation (26) implies that the Lindberg condition is satisfied, and by the Lindberg-Feller CLT (e.g., p. 128 of Rao, 1971), we have $\left(n\lambda'\bar{V}_\eta\lambda\right)^{-1/2}\sum_{i=1}^n W_{\eta,i}\xrightarrow{d.}N(0,1)$, implying in turn that $n^{-1/2}\sum_{i=1}^n W_{\eta,i}\xrightarrow{d.}N(0,\lambda'V\lambda)$. This, together with the Cramér-Wold device, implies that for arbitrary sequence of DGP's with $\eta_n\to 0$ as $n\to\infty$,

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

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

$$n^{-1/2}\sum_{i=1}^n\left(g\left(Z_i,\theta_0\right)-\phi_{n^{-1/2}}\left(\theta_0\right)\right)\xrightarrow{d.}N(0,V). \tag{27}$$

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

$$K_R\equiv\frac{\partial}{\partial\eta}\left(\int g\left(z,\theta_0\right)f_X\left(x\right)\int f\left(y|\,x,\theta_0+\sqrt{\eta}e\iota\right)k\left(e\right)dedz\right)\Bigg|_{\eta=0} \tag{28}$$

$$=\left.\frac{\int\int g\left(z,\theta_0\right)f_X\left(x\right)\frac{\partial f\left(y|x,\theta_0+\sqrt{\eta}e\iota\right)}{\partial\theta_1}ek\left(e\right)dedz}{2\sqrt{\eta}}\right|_{\eta=0}$$

$$=\lim_{\eta\to 0}\frac{\frac{\sqrt{\eta}}{2}\int\int g\left(z,\theta_0\right)f_X\left(x\right)\frac{\partial f\left(y|x,\theta_0+\sqrt{\eta}e\iota\right)}{\partial\theta_1}ek\left(e\right)dedz}{\eta}.$$

Using the L'Hopital's rule, we write

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

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

from which we obtain

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

where we recall $\sigma_\varepsilon^2 = \int e^2 k(e^2) \, de$. Recalling that $g = (m', s')'$ helps us simplify as follows

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

Recall the definition of $\phi_\eta(\theta)$ in equation (21) and apply the mean-value theorem to equation (28) with $\eta_n = n^{-1/2}$, we get

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

Combined with the CLT in equation (27), we see that this implies that $\sqrt{n} g_n(\theta_0) \xrightarrow{d.} N(K_R, V)$.

Remark 5 *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*

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

with corresponding changes in the next section.

Step 4 We now show that $n^{1/2}\left(\bar{\theta}_n-\theta_0\right)=-\left(DH\right)^{-1}Dn^{1/2}g_n\left(\theta_0\right)+o_p\left(1\right)$, where $D\equiv[0,I_q]$ and $H\equiv E\left[\partial g\left(Z_i,\theta_0\right)/\partial\theta'\right]$. By the mean value theorem, we get $n^{1/2}g_n\left(\bar{\theta}_n\right)=n^{1/2}g_n\left(\theta_0\right)+\left[\partial g_n\left(\dot{\theta}_n\right)\Big/\partial\theta'\right]n^{1/2}\left(\bar{\theta}_n-\theta_0\right)$ for some $\dot{\theta}_n$ in 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 hypotheses of Lemma A.1, implying that $\partial g_n(\dot{\theta}_n)\Big/\partial\theta'\xrightarrow{p.}E\left[\partial g\left(Z_i,\theta_0\right)/\partial\theta'\right]$. 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 $n^{1/2}\left(\bar{\theta}_n-\theta_0\right)=-\left(DH\right)^{-1}Dn^{1/2}g_n\left(\theta_0\right)+o_p\left(1\right)$.

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

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

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

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

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

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

or

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

We therefore conclude that

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

and hence

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

Remark 6 *Note that* $-(DH)^{-1} = \mathcal{I}^{-1}$. *Combined with* $n^{1/2} g_n(\theta_0) \xrightarrow{d.} N(K_R, V)$, *which implies that* $Dn^{1/2} g_n(\theta_0) \xrightarrow{d.} N\left(\frac{\sigma_\varepsilon^2}{2}\kappa_2, \mathcal{I}\right)$, *we can see that* $n^{1/2}\left(\bar{\theta}_n - \theta_0\right) \xrightarrow{d.} N\left(\frac{\sigma_\varepsilon^2}{2}\mathcal{I}^{-1}\kappa_2, \mathcal{I}^{-1}\right)$. *Therefore, if* $\kappa_2 = 0$, *the MLE is asymptotically unbiased even under the alternative of random effects.*

Step 5 We now establish the 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 $L \equiv [I_1, 0]$. We also saw in the previous step that $n^{1/2} g_n\left(\bar{\theta}_n\right) = n^{1/2} g_n(\theta_0) + Hn^{1/2}\left(\bar{\theta}_n - \theta_0\right) + o_p(1)$, and $n^{1/2}\left(\bar{\theta}_n - \theta_0\right) = -(DH)^{-1} Dn^{1/2} g_n(\theta_0) + o_p(1)$. Therefore, we see that

$$
\begin{aligned}
\sqrt{n} m_n\left(\bar{\theta}_n\right) &= L\sqrt{n} g_n\left(\bar{\theta}_n\right) \\
&= L\sqrt{n} g_n(\theta_0) - LH(DH)^{-1} D\sqrt{n} g_n(\theta_0) + o_p(1) \\
&= L\left(I_{q+1} - H(DH)^{-1} D\right)\sqrt{n} g_n(\theta_0) + o_p(1) \\
&= [I_1, 0]\sqrt{n} g_n(\theta_0) - [I_1, 0] H(-\mathcal{I})^{-1}\sqrt{n} s_n(\theta_0) + o_p(1) \\
&= \sqrt{n} m_n(\theta_0) - \kappa_2' \mathcal{I}^{-1}\sqrt{n} s_n(\theta_0) + o_p(1) \\
&= \left[I_1, -\kappa_2' \mathcal{I}^{-1}\right]\sqrt{n} g_n(\theta_0) + o_p(1)
\end{aligned}
$$

while $\sqrt{n} g_n(\theta_0) \xrightarrow{d.} N(K_R, V)$. 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

$$
\begin{aligned}
\sqrt{n} m_n\left(\bar{\theta}_n\right) = L\sqrt{n} g_n\left(\bar{\theta}_n\right) &\xrightarrow{d.} N\left(\left[I, -\kappa_2' \mathcal{I}^{-1}\right] K_R, \kappa_1 - \kappa_2' \mathcal{I}^{-1}\kappa_2\right) \\
&\overset{d.}{=} N\left(\frac{\sigma_\varepsilon^2}{2}\left(\kappa_1 - \kappa_2' \mathcal{I}^{-1}\kappa_2\right), \kappa_1 - \kappa_2' \mathcal{I}^{-1}\kappa_2\right)
\end{aligned} \tag{29}
$$

in general.

C Proof of Theorem 2

The proof is essentially identical to the proof of Theorem 1, 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

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

where $\mu(x)$ denotes a function of x. Note that there is no other random variable. We now would like to calculate the counterpart of K_R. Due to Assumptions $3'$, 4, 5 and the dominated convergence theorem, we have

$$\begin{aligned}
K_F &\equiv \frac{\partial}{\partial \eta}\left(\int g(z,\theta_0) f_X(x) f(y|\, x, \theta_0 + \eta\mu(x)\iota)\, dz\right)\Big|_{\eta=0} \\
&= \int g(z,\theta_0) f_X(x) \frac{\partial f(y|\, x,\theta_0)}{\partial \theta_1}\mu(x)\, dz \\
&= \int g(z,\theta_0) s_1(y|\, x,\theta_0) f(y|\, x,\theta_0) f_X(x)\mu(x)\, dz \\
&= E\left[g(Z_i,\theta_0) s_1(Y_i|\, X_i,\theta_0)\mu(X_i)\right],
\end{aligned}$$

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

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

The rest of Newey's (1985) analysis still applies. We therefore have (in his symbols) $\sqrt{T} L_T g_T\left(\bar\theta_T\right) = LP\sqrt{T} g_T(\theta_0) + o_p(1)$ (2.6) and that $\sqrt{T} g_T(\theta_0) \xrightarrow{d.} N(K\delta, V)$ (2.7). Here, $L = [I_1, 0]$ and $D = [0, I_q]$. Because, $DH = -\mathcal{I}$, $Dg(z,\theta_0) = s(z,\theta_0)$, we still have (in our notation)

$$\begin{aligned}
L\sqrt{n} g_n\left(\bar\theta_n\right) &= LP\sqrt{n} g(\theta_0) + o_p(1) \\
&= L\left(I_{q+1} - H(DH)^{-1} D\right)\sqrt{n} g_n(\theta_0) + o_p(1) \\
&= [I_1, 0]\sqrt{n} g_n(\theta_0) - [I_1, 0] H(-\mathcal{I})^{-1}\sqrt{n} s_n(\theta_0) + o_p(1) \\
&= \sqrt{n} m_n(\theta_0) - \kappa_2' \mathcal{I}^{-1}\sqrt{n} s_n(\theta_0) + o_p(1) \\
&= \left[I_1, -\kappa_2'\mathcal{I}^{-1}\right]\sqrt{n} g_n(\theta_0) + o_p(1),
\end{aligned}$$

while

$$\sqrt{n} g_n(\theta_0) \xrightarrow{d.} N(K_F, V).$$

Because we still have

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

it follows that

$$\sqrt{n} m_n\left(\bar{\theta}_n\right) = L \sqrt{n} g_n\left(\bar{\theta}_n\right) \xrightarrow{d} N\left(\left[I_1, -\kappa_2' \mathcal{I}^{-1}\right] K_F, \kappa_1 - \kappa_2' \mathcal{I}^{-1} \kappa_2\right). \tag{30}$$

We now consider the fixed effects (9). After all, the whole calculation was based on the derivative of the form $f\left(y|\, x, \left(\theta_{0,1} + \eta\mu\left(x\right) + \sqrt{\eta}\varepsilon, \theta_{0,2}, \ldots, \theta_{0,q}\right)\right)$ and note that the derivative should be the sum of the derivatives of $f\left(y|\, x, \left(\theta_{0,1} + \eta\mu\left(x\right), \theta_{0,2}, \ldots, \theta_{0,q}\right)\right)$ 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 (29) and (30):

$$\sqrt{n} m_n\left(\bar{\theta}_n\right) = L \sqrt{n} g_n\left(\bar{\theta}_n\right) \xrightarrow{d.} N\left(\left[I_1, -\kappa_2' \mathcal{I}^{-1}\right]\left(K_F + K_R^*\right), \kappa_1 - \kappa_2' \mathcal{I}^{-1} \kappa_2\right),$$

where

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

is a heteroskedasticity-robust version of K_R based on ε_i^*. (See Remark 5.)

D Proof of Lemma 1

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

$$\begin{aligned}
\hat{\theta} &= \left(\sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} X_{it}'\right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} X_{it}\left(\frac{\mu\left(X_i\right)}{n^{1/4}} + \frac{\varepsilon_i^*}{n^{1/4}} + X_{it}'\theta + v_{it}\right)\right) \\
&= \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} X_{it}'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \bar{X}_i \mu\left(X_i\right)\right) / n^{1/4} \\
&+ \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} X_{it}'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \bar{X}_i \varepsilon_i^*\right) / n^{1/4}
\end{aligned}$$

$$+\theta+\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}X_{it}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}v_{it}\right).$$

We see that

$$\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i\mu\left(X_i\right)=\lambda+O_p\left(n^{-1/2}\right),$$
$$\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i\varepsilon_i^*=O_p\left(n^{-1/2}\right),$$
$$\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}v_{it}=O_p\left(n^{-1/2}\right),$$

so

$$n^{1/4}\left(\hat{\theta}-\theta\right)=Q^{-1}\lambda+O_p\left(n^{-1/2}\right)+O_p\left(n^{-1/4}\right)=Q^{-1}\lambda+O_p\left(n^{-1/4}\right).$$

Now we consider

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

Since the OLS residual $\hat{v}_{il}=\xi_i/n^{1/4}+v_{il}-X_{il}'\left(\hat{\theta}-\theta\right)$, we have

$$\begin{aligned}\frac{1}{n^{1/2}}\sum_{i=1}^{n}\sum_{l\neq m}^{T}\hat{v}_{il}\hat{v}_{im}&=\frac{1}{n^{1/2}}\sum_{i=1}^{n}\sum_{l\neq m}^{T}\left(\frac{\xi_i}{n^{1/4}}+v_{il}\right)\left(\frac{\xi_i}{n^{1/4}}+v_{im}\right)\\&-\frac{1}{n^{1/2}}\sum_{i=1}^{n}\sum_{l\neq m}^{T}\left(\frac{\xi_i}{n^{1/4}}+v_{il}\right)X_{im}'\left(\hat{\theta}-\theta\right)\\&-\frac{1}{n^{1/2}}\sum_{i=1}^{n}\sum_{l\neq m}^{T}\left(\frac{\xi_i}{n^{1/4}}+v_{im}\right)X_{il}'\left(\hat{\theta}-\theta\right)\\&+\frac{1}{n^{1/2}}\sum_{i=1}^{n}\sum_{l\neq m}^{T}\left(\hat{\theta}-\theta\right)'X_{il}X_{im}'\left(\hat{\theta}-\theta\right).\end{aligned}\tag{31}$$

Note that the first term can be rewritten

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

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

where we used the CLT for the second equality.

Now we show that the second term in equation (31) is $O_p(n^{1/4})$.

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} \sum_{l \neq m}^{T} \left(\frac{\xi_i}{n^{1/4}} + v_{il}\right) X_{im}' \left(\hat{\theta} - \theta\right)$$
$$= \frac{1}{n^{1/2}} \sum_{i=1}^{n} \sum_{l \neq m}^{T} v_{il} X_{im}' \left(\hat{\theta} - \theta\right) + \frac{1}{n^{3/4}} \sum_{i=1}^{n} \sum_{l \neq m}^{T} \xi_i X_{im}' \left(\hat{\theta} - \theta\right)$$
$$= \left(\hat{\theta} - \theta\right)' \left(n^{-1/2} \sum_{i=1}^{n} \sum_{l \neq m}^{T} v_{il} X_{im}\right) + n^{1/4} \left(\hat{\theta} - \theta\right) \left(n^{-1} \sum_{i=1}^{n} \xi_i \sum_{m=1}^{T} X_{im}\right).$$

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

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

we conclude that the second term on the far RHS is

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

Therefore, we should have the second term in equation (31) equal to $-T\lambda' Q^{-1}\lambda + O_p\left(n^{-1/4}\right)$. By the same argument, the third term is equation (31) is also $-T\lambda' Q^{-1}\lambda + O_p\left(n^{-1/4}\right)$ as the indices l and m are symmetric in this respect.

Finally, the fourth term in equation (31) can be written as

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} \sum_{l \neq m}^{T} \left(\hat{\theta} - \theta\right)' X_{il} X_{im}' \left(\hat{\theta} - \theta\right) = n^{1/4}\left(\hat{\theta} - \theta\right)' \left(n^{-1} \sum_{i=1}^{n} \sum_{l \neq m}^{T} X_{il} X_{im}'\right) n^{1/4}\left(\hat{\theta} - \theta\right)$$
$$= \left(Q^{-1}\lambda\right)' S \left(Q^{-1}\lambda\right) + O_p\left(n^{-1/4}\right).$$

E Technical Details of Section 6

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* $M(y,x)$ *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 (17) is $O_p(1)$ as $n,T \to \infty$. The same argument shows that (18) is also $O_p(1)$. First, we rewrite(17) as

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial^2 f(Y_{it}|X_{it},\theta_0)/\partial\theta_1^2}{f(Y_{it}|X_{it},\theta_0)} + \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t\neq t'}\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\frac{\partial \ln f(Y_{it'}|X_{it'},\theta_0)}{\partial\theta_1}. \tag{32}$$

Note that the first term in (32) 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 (32) is $O_p\left(\frac{1}{\sqrt{T}}\right)$.

As for the second term in (32), we see that $\sum_{t\neq t'}\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\frac{\partial \ln f(Y_{it'}|X_{it'},\theta_0)}{\partial\theta_1}$ has mean equal to zero and its variance is equal to

$$4\sum_{t\neq t'}E\left[\left(\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)^2\right]E\left[\left(\frac{\partial \ln f(Y_{it'}|X_{it'},\theta_0)}{\partial\theta_1}\right)^2\right],$$

and therefore, the second term in (32) has mean equal to zero and variance equal to

$$\frac{4}{nT^2}\sum_{i=1}^{n}\sum_{t\neq t'}E\left[\left(\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)^2\right]E\left[\left(\frac{\partial \ln f(Y_{it'}|X_{it'},\theta_0)}{\partial\theta_1}\right)^2\right] = O_p(1).$$

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 (19), and obtain

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

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

$$+ \begin{pmatrix} \frac{1}{nT\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial^3 \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1^2\partial\theta'} \\ +\frac{2}{nT\sqrt{T}}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)\left(\sum_{t=1}^{T}\frac{\partial^2 \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1\partial\theta'}\right) \end{pmatrix}\sqrt{nT}\left(\bar{\theta}_n-\theta_0\right)$$

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

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

Note that the last term above can be bounded above by

$$\begin{pmatrix} \frac{1}{2n\sqrt{n}T^2}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}M\left(Y_{it},X_{it}\right)\right) \\ +\frac{2}{n\sqrt{n}T^2}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}M\left(Y_{it},X_{it}\right)\right)\left(\sum_{t=1}^{T}M\left(Y_{it},X_{it}\right)\right) \end{pmatrix}\left\|\sqrt{nT}\left(\bar{\theta}_n-\theta_0\right)\right\|^2$$

$$= \left(O_p\left(\frac{nT}{n\sqrt{n}T^2}\right)+O_p\left(\frac{nT^2}{n\sqrt{n}T^2}\right)\right)O_p\left(1\right) = o_p\left(1\right),$$

so we obtain

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

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

$$+ \begin{pmatrix} \frac{1}{nT\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial^3 \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1^2\partial\theta'} \\ +\frac{2}{nT\sqrt{T}}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\frac{\partial \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1}\right)\left(\sum_{t=1}^{T}\frac{\partial^2 \ln f(Y_{it}|X_{it},\theta_0)}{\partial\theta_1\partial\theta'}\right) \end{pmatrix}\sqrt{nT}\left(\bar{\theta}_n-\theta_0\right)+o_p\left(1\right). \tag{33}$$

We now show that the third term in (33) is $o_p\left(1\right)$. 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

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

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

which we further write as

$$\frac{2}{n\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1}\left(\frac{1}{T}\sum_{t=1}^{T}E\left[\frac{\partial^2 \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)$$
$$+ \frac{2}{nT\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1}\left(\frac{\partial^2 \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'} - E\left[\frac{\partial^2 \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)$$
$$+ \frac{2}{nT\sqrt{T}}\sum_{i=1}^{n}\sum_{t\neq t'}\frac{\partial \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1}\left(\frac{\partial^2 \ln f\left(\left.Y_{it'}\right|X_{it'},\theta_0\right)}{\partial\theta_1\partial\theta'} - E\left[\frac{\partial^2 \ln f\left(\left.Y_{it'}\right|X_{it'},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right). \tag{34}$$

The first term has mean zero and variance equal to

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

so it should be $o_p(1)$. As for the second term of (34), we have

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

so it should be $o_p(1)$. As for the third term of (34), we note that conditional on Xs, it can be viewed as a sum of $\sum_{t\neq t'}\frac{\partial \ln f\left(\left.Y_{it}\right|X_{it},\theta_0\right)}{\partial\theta_1}\left(\frac{\partial^2 \ln f\left(\left.Y_{it'}\right|X_{it'},\theta_0\right)}{\partial\theta_1\partial\theta'} - E\left[\frac{\partial^2 \ln f\left(\left.Y_{it'}\right|X_{it'},\theta_0\right)}{\partial\theta_1\partial\theta'}\right]\right)$, which are independent over i. Therefore, it has mean zero and variance equal to

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

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

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

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

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

is of order T^3.

This implies that the third term of (34) has mean zero and variance of order

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

so it should be $o_p(1)$.

References

[1] Acemoglu, D., S. Johnson, J. Robinson, and P. Yared (2008): "Income and Democracy," American Economic Review 98, 808–842.

[2] Arellano, M. and J. Hahn (2007): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," in Blundell, R., W. K. Newey and T. Persson eds. Advances in Economics and Econometrics, Cambridge: Cambridge University Press.

[3] Bartle, R. G. (1966), "Elements of Integration." New York: John Wiley & Sons.

[4] Bierens, H. J. (1990), "A Consistent Conditional Moment Test of Functional Form," Econometrica 58, 1443–1458.

[5] Bonhomme, Stéphane, B., and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," Econometrica 83, 1147-1184.

[6] Breusch, T.S., and A.R. Pagan (1980): "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics," Review of Economic Studies 47, 239-253.

[7] Chamberlain, G. (1984): "Panel Data," in Griliches, Z., and M.D. Intriligator eds *Handbook of Econometrics Volume II*, Elsevier.

[8] Chesher, A. (1984): "Testing for Neglected Heterogeneity," Econometrica 52, 865-872.

[9] Cox, D.R. (1983): "Some Remarks on Overdispersion," Biometrica 70, 269-274.

[10] Donald, S., G. Imbens and W.K. Newey (2003), "Empirical Likelihood Estimation and Consistent Tests With Conditional Moment Restrictions," Journal of Econometrics 117, 55–93.

[11] Engle, R.F. (1984): "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in Griliches, Z., and M.D. Intriligator eds *Handbook of Econometrics Volume II*, Elsevier.

[12] Fernández-Val, I. and M. Weidner (2016): "Individual and Time Effects in Nonlinear Panel Models with Large N, T, " Journal of Econometrics 192, 291-312.

[13] Hahn, J., H. Moon, and C. Snider (2017): "LM Test of Neglected Correlated Random Effects and Its Application,"Journal of Business & Economic Statistics 35, 359-370.

[14] Hahn, J., and W.K. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Econometrica 72, 1295-1319.

[15] Hausman, J.A., and W.E. Taylor (1981): "Panel Data and Unobservable Individual Effects," Econometrica 49, 1377-1398.

[16] Honda, Y. (1985): "Testing the Error Components Model with Non-Normal Disturbances," Review of Economic Studies 52, 681-690.

[17] Lancaster, T. (1984): "The Covariance Matrix of the Information Matrix Test," Econometrica 52, 1051-1053.

[18] Newey, W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," Econometrica 53, 1047-1070.

[19] Rao, C. R. (1971), "Linear Statistical Inference and its Applications." New York: John Wiley & Sons.

[20] White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," Econometrica 50, 1-25.