

Large-N and Large-T Properties of Panel Data Estimators and the Hausman Test*

Seung Chan Ahn^{†‡}
Department of Economics
Arizona State University

Hyungsik Roger Moon
Department of Economics
University of Southern California

This Version: August 2001

Abstract

This paper examines the asymptotic properties of the popular within, GLS estimators and the Hausman test for panel data models with both large numbers of cross-section (N) and time-series (T) observations. The model we consider includes the regressors with deterministic trends in mean as well as time invariant regressors. If a time-varying regressor is correlated with time invariant regressors, the time series of the time-varying regressor is not ergodic. Our asymptotic results are obtained considering the dependence of such non-ergodic time-varying regressors. We find that the within estimator is as efficient as the GLS estimator. Despite this asymptotic equivalence, however, the Hausman statistic, which is essentially a distance measure between the two estimators, is well defined and asymptotically χ^2 -distributed under the random effects assumption.

*We would like to thank Geert Ridder for helpful discussions. We also appreciate the comments of seminar participants at Arizona State University, the University of British Columbia, and the University of California, Davis.

[†]Corresponding Author: Seung C. Ahn, Department of Economics, Arizona State University, Tempe, AZ 85287; tel) 480-965-6574; email) miniahn@asu.edu.

[‡]The first author gratefully acknowledges the financial support of the College of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of the College of Business.

1 Introduction

Panel data refers to the data with repeated time-series observations (T) for a large number (N) of cross-sectional units (e.g., individuals, households, or firms). An important advantage of using these data is that they allow researchers to control for unobservable heterogeneity, that is, systematic differences across cross-sectional units (e.g., individuals, households, firms, countries). Error-components models have been widely used to control for these individual differences. These models assume that stochastic error terms have two components: an unobservable time-invariant individual effect which captures the unobservable individual heterogeneity and the usual random noise. The most popular estimation methods for panel data models are the within and the generalized least squares (GLS) estimators. For the panel data with large N and small T , the appropriate choice of estimators depends on whether or not regressors are correlated with the unobservable individual effect. An important advantage of using the within estimator (least squares on data transformed into deviations from individual means) is that it is consistent even if regressors are correlated with the individual effect. However, a serious defect of the estimator is its inability to estimate the impact of time-invariant regressors.¹ The GLS estimator is often used in the literature as a treatment of this problem, but it is not without its own defect: The consistency of the GLS crucially depends on a strong assumption that no regressor is correlated with the effect (random effects assumption). Use of the estimator thus requires a statistical test that can empirically validate this strong assumption. A Hausman statistic (1978) is commonly used for this purpose (e.g., Hausman and Taylor, 1981; Cornwell and Rupert, 1988; or Baltagi and Khanti-Akom, 1990).

In this paper, we study the asymptotic properties of the within, GLS estimators and the Hausman statistic for a general error-components model with both large numbers of cross-section and time-series observations. The GLS estimator has been known to be asymptotically equivalent to the within estimator for the cases with infinite N and T (see, for example, Hsiao, Chapter 3, 1986; Mátyás and Sevestre, Chapter 4, 1992; and Baltagi, Chapter 2, 1995). This asymptotic equivalence result has been obtained using a naive sequential limit method ($T \rightarrow \infty$ followed by $N \rightarrow \infty$) and some strong assumptions such as fixed regressors. This result naturally raises a couple of questions regarding the asymptotic properties of the Hausman test. Firstly, the Hausman statistic could be viewed as a distance measure between the within and GLS estimators. Then, does the equivalence result indicate that the Hausman statistic should have a degenerating or nonstandard asymptotic distribution under the random effects assumption? Secondly, does the equivalence result also imply that the Haus-

¹Estimation of the effect of a certain time invariant variable on a dependent variable could be an important task in a broad range of empirical research. Examples would be the labor studies about the effects of schooling or gender on individual workers' earnings, and the macroeconomic studies about the effect of a country's geographic location (e.g., whether the country is located in Europe or Asia) on its economic growth. The within estimator is inappropriate for such studies.

man test would have low power to detect any violation of the random effects assumption when T is large? This paper is concerned to answer these questions.

The analysis of panel data with both large N and T becomes increasingly important in the literature. While a variety of estimation and model specification testing techniques have been introduced and proposed in the panel data literature, most of these methods are limited to the analysis of data with large N and small T . An obvious reason for this limited approach is that most of the available panel data have only a short history. However, panel data with a large number of time-series observations have been increasingly more available in recent years in many economic fields such as international finance, finance, industrial organization, and economic growth. Furthermore, popular panel data, such as the Panel Study of Income Dynamics (PSID) and the National Longitudinal Surveys (NLS), contain increasingly more time-series observations as they are updated regularly over the years. Consistent with this trend, some recent studies have examined the large- N and large- T properties of the within and GLS estimators for error-component models.² For example, Phillips and Moon (1999) and Kao (1999) establish the asymptotic normality of the within estimator for the cases in which regressors follow unit root processes. Extending these studies, Choi (1998) considers a general random effects model which contains both unit-root and covariance-stationary regressors. For this model, he derives the asymptotic distributions of both the within and GLS estimators. These papers however do not consider the asymptotic properties of the Hausman test.

The general model we consider in this paper is different from the models considered by these studies in two ways. First, our model contains the time-varying regressors that are serially dependent and heteroskedastic over time with or without time trends. These variables could be cross-sectionally heteroskedastic or homogeneous. Second, our model contains time invariant regressors that are correlated or uncorrelated with other time-varying regressors. If a time-varying regressor is correlated with time-invariant regressors, the time series of it is non-ergodic because the influence of the time invariant random regressors is persistent in all time periods.³ However, under the assumption that such time-varying regressors satisfy mixing properties conditionally on time invariant random regressors, we can derive the limiting distributions of various forms of sample averages of panel data when both $N, T \rightarrow \infty$. These intermediate results are used to establish the asymptotic distributions of the panel data estimators and the Hausman test.

Most of the asymptotic results derived in the paper hold as $N, T \rightarrow \infty$ without any particular sequence. In addition, we do not make any assumption

²Some other studies have considered different panel data models with large N and large T . For example, Levin and Lin (1992, 1993), Quah (1994), Pesaran, Shin and Im (1997), and Higin and Zakrajsěk (1999) develop unit-root tests for data with large N and large T . Alvarez and Arellano (1998) and Hahn and Kuersteiner (2000) examine the large- N and large- T properties of generalized method of moments (GMM) and within estimators for stationary dynamic panel data models.

³The time series of the panel data considered in Phillips and Moon (1999) is also non-ergodic. However, in their paper the non-ergodicity arises due to stochastic trends generated by unit root processes.

about the relative sizes of N and T . Previous studies of nonstationary panel data typically assume that T is relatively greater than N . Differently from these studies, our approximation theories can apply to any panel data set with large N and large T regardless of their asymptotic ratio.

The main findings of this paper are as follows. Consistent with the previous studies, we find that the within and GLS estimators (of coefficients on time-varying regressors) are asymptotically equivalent. Nonetheless, the Hausman statistic is asymptotically χ^2 -distributed under the random effects assumption. This seemingly contradictory result can be explained by our finding that the differences between the within and GLS estimators converge in probability to zeros much faster than the two estimators converge in probability to the true values of coefficients (at the same speed). For example, for a simple model with a no-trend time-varying regressor, we find that the within and GLS estimators (of the coefficient on the only time-varying regressor) are both \sqrt{NT} -consistent and asymptotically normal. In addition, the two estimators are \sqrt{NT} -equivalent in the sense that the difference between the two estimators is $o_p(1/\sqrt{NT})$. However, we also find that the difference is $O_p(1/\sqrt{NT^2})$ or $O_p(1/\sqrt{NT^3})$ depending on whether data are cross-sectionally heteroskedastic or homoskedastic. This implies that the within and GLS estimators are not equivalent in the orders of $\sqrt{NT^2}$ or $\sqrt{NT^3}$. Furthermore, it is shown that under the random effects assumption, the differences between the within and GLS estimators are asymptotically normally distributed. From this, we show that the Hausman test statistic is asymptotically χ^2 -distributed. In addition, our analysis under a series of local alternative assumptions indicates that the Hausman test retains power to detect violations of the random effects assumption even if $T \rightarrow \infty$.

This paper is organized as follows. Section 2 introduces the panel model of interest here, and defines the within, GLS estimators and the Hausman test. For several simple illustrative models, we derive the asymptotic distributions of the within estimators, the GLS estimators, and the Hausman test statistic. We show that the convergence rates of the estimators and the Hausman test statistic are sensitive to the unknown data generating structure. Section 3 defines a conditional α -mixing coefficient. Using this, we propose a conditional α -mixing process and discuss its properties. In Section 4, we provide our general asymptotic results. Concluding remarks follow in section 5. All the technical derivations and proofs are presented in the Appendix.

2 Preliminaries

2.1 Estimation and Specification Test

The model under discussion here is given:

$$y_{it} = \beta' x_{it} + \gamma' z_i + \zeta + \varepsilon_{it} = \delta' w_{it} + \zeta + \varepsilon_{it}; \varepsilon_{it} = u_i + v_{it}, \quad (1)$$

where $i = 1, \dots, N$ denotes cross-sectional (individual) observations, $t = 1, \dots,$

T denotes time, $w_{it} = (x_{it}, z_i)'$, and $\delta = (\beta', \gamma')'$. In model (1), x_{it} is a $k \times 1$ vector of time-varying regressors, z_i is a $g \times 1$ vector of time-invariant regressors, ζ is an overall intercept term, and the error ε_{it} contains a time-invariant individual effect u_i and random noise v_{it} . We consider the case of both large numbers of individual and time series observations, so asymptotic properties of the estimators and statistics for model (1) apply as $N, T \rightarrow \infty$. The orders of convergence rates of some estimators depend on whether or not the model contains an overall intercept term. This problem will be addressed later.

We assume that data are distributed independently (but not necessarily identically) across different i , and that the v_{it} are independently and identically distributed (i.i.d.) with $\text{var}(v_{it}) = \sigma_v^2$. We further assume that $u_i, x_{i1}, \dots, x_{iT}$ and z_i are strictly exogenous with respect to v_{it} ; that is,

$$E(v_{it} \mid u_i, x_{i1}, \dots, x_{iT}) = 0,$$

for any i and t . This assumption rules out the cases in which the set of regressors includes lagged dependent variables or predetermined regressors. Detailed assumptions about the regressors $x_{i1}, \dots, x_{iT}, z_i$ will be introduced later.

For convenience, we adopt the following notational rule: For any $p \times 1$ vector a_{it} , we denote $\bar{a}_i = \frac{1}{T} \sum_t a_{it}$; $\tilde{a}_{it} = a_{it} - \bar{a}_i$; $\bar{a} = \frac{1}{N} \sum_i \bar{a}_i$; $\tilde{a}_i = \bar{a}_i - \bar{a}$. Thus, for example, for $w_{it} = (x'_{it}, z'_i)'$, we have $\bar{w}_i = (\bar{x}'_i, z'_i)'$; $\tilde{w}_{it} = (\tilde{x}'_{it}, 0_{1 \times g})'$; $\bar{w} = (\bar{x}', \bar{z})$; $\tilde{w}_i = ((\bar{x}_i - \bar{x})', (\bar{z}_i - \bar{z})')'$.

When the regressors are correlated with the individual effect, both of the OLS and GLS estimators of δ are biased and inconsistent. This problem has been traditionally addressed by the use of the within estimator (OLS on data transformed into deviations from individual means):

$$\hat{\beta}_w = (\sum_{i,t} \tilde{x}_{it} \tilde{x}'_{it})^{-1} \sum_{i,t} \tilde{x}_{it} \tilde{y}'_{it}.$$

Under our assumptions, the variance-covariance matrix of the within estimator is given:

$$\text{Var}(\hat{\beta}_w) = \sigma_\varepsilon^2 (\sum_{i,t} \tilde{x}_{it} \tilde{x}'_{it})^{-1}. \quad (2)$$

Although the within method provides a consistent estimate of β , a serious defect is its inability to identify γ , the impact of time-invariant regressors. A popular treatment of this problem is the random effects (RE) assumption under which the u_i are random and uncorrelated with the regressors:

$$E(u_i \mid x_{i1}, \dots, x_{iT}, z_i) = 0. \quad (3)$$

Under this assumption, all of the parameters in model (1) can be consistently estimated. For example, a simple but consistent estimator is the between estimator (OLS on data transformed into individual means):

$$\hat{\delta}_b = (\hat{\beta}'_b, \hat{\gamma}'_b)' = (\sum_i \tilde{w}_i \tilde{w}'_i)^{-1} \sum_i \tilde{w}_i \tilde{y}_i.$$

However, as Balestra and Nerlove (1966) suggest, under the RE assumption, an efficient estimator is the GLS estimator of the following form:

$$\begin{aligned} \hat{\delta}_g &= [\sum_{i,t} (\tilde{w}_{it} + \theta_T \tilde{w}_i) (\tilde{w}_{it} + \theta_T \tilde{w}_i)']^{-1} \sum_{i,t} (\tilde{w}_{it} + \theta_T \tilde{w}_i) (\tilde{y}_{it} + \theta_T \tilde{y}_i) \\ &= [\sum_{i,t} \tilde{w}_{it} \tilde{w}'_{it} + T\theta_T^2 \sum_i \tilde{w}_i \tilde{w}'_i]^{-1} [\sum_{i,t} \tilde{w}_{it} \tilde{y}_i + T\theta_T^2 \sum_i \tilde{w}_i \tilde{y}_i], \end{aligned}$$

where $\theta_T = \sqrt{\sigma_v^2 / (T\sigma_u^2 + \sigma_v^2)}$. The variance-covariance matrix of this estimator is given:

$$\text{Var}(\hat{\delta}_g) = \sigma_v^2 [\sum_{i,t} \tilde{w}_{it} \tilde{w}'_{it} + T\theta_T^2 \sum_i \tilde{w}_i \tilde{w}'_i]^{-1}. \quad (4)$$

For notational convenience, we assume that σ_u^2 and σ_v^2 are known, while in practice they must be estimated.⁴

An important advantage of the GLS estimator over the within estimator is that it allows researchers to estimate γ . In addition, the GLS estimator of β is more efficient than the within estimator of β , because $[\text{Var}(\hat{\beta}_w) - \text{Var}(\hat{\beta}_g)]$ is positive definite so long as $\theta_T > 0$.⁵ Despite these desirable properties, it is important to notice that the consistency of the GLS estimator crucially depends on the RE assumption (3). Accordingly, the legitimacy of the RE assumption should be tested to justify the use of GLS. In the literature, a Hausman test (1978) has been widely used for this purpose. The statistic used for this test is a distance measure between the within and GLS estimators of β :

$$\mathcal{HM}_{NT} \equiv (\hat{\beta}_w - \hat{\beta}_g)' [\text{Var}(\hat{\beta}_w) - \text{Var}(\hat{\beta}_g)]^{-1} (\hat{\beta}_w - \hat{\beta}_g). \quad (5)$$

For the cases in which T is fixed and $N \rightarrow \infty$, the RE assumption warrants that the Hausman statistic \mathcal{HM}_{NT} is asymptotically χ^2 -distributed with degrees of freedom equal to k . This result is a direct outcome of the fact that for fixed T , the GLS estimator $\hat{\beta}_g$ is asymptotically more efficient than the within estimator $\hat{\beta}_w$, and that the difference between the two estimators is asymptotically normal; specifically, as $N \rightarrow \infty$,

$$\sqrt{NT}(\hat{\beta}_w - \hat{\beta}_g) \implies N(0, \text{plim}_{N \rightarrow \infty} NT[\text{Var}(\hat{\beta}_w) - \text{Var}(\hat{\beta}_g)]), \quad (6)$$

where “ \implies ” means “converges in distribution.”

An important condition that guarantees (6) is that $\theta_T > 0$. If $\theta_T = 0$, then the within and GLS estimators become identical and the Hausman statistic is not defined. Observe now that $\theta_T \rightarrow 0$ as $T \rightarrow \infty$. This observation naturally raises several issues related with the asymptotic properties of the Hausman test as $T \rightarrow \infty$. In order to clarify the nature of the problem, consider model (1), but without the time-invariant regressors and the overall intercept term (ζ). Assume that x_{it} contains a single time-varying regressor which is independently and identically distributed over different i and t . For this simple model, we can easily show $\text{plim}_{N,T \rightarrow \infty} NT \text{Var}(\hat{\beta}_w) = \text{plim}_{N,T \rightarrow \infty} NT \text{Var}(\hat{\beta}_g)$, using the fact that $\theta_T \rightarrow 0$ as $T \rightarrow \infty$. This asymptotic equality immediately implies that the within and GLS estimators of β are asymptotically equivalent; that is, $\text{plim}_{N,T \rightarrow \infty} \sqrt{NT}(\hat{\beta}_w - \hat{\beta}_g) = 0_{k \times 1}$. This preliminary finding raises several questions. First, does this equivalence result hold for the general cases with

⁴The conventional estimates are given:

$$\hat{\sigma}_v^2 = \sum_{i,t} (\tilde{y}_{it} - \tilde{x}_{it} \hat{\beta}_w)^2 / [N(T-1)]; \hat{\sigma}_u^2 = \sum_i (\tilde{y}_i - \tilde{w}_i \hat{\delta}_b)^2 / N - \hat{\sigma}_v^2 / T.$$

⁵This efficiency gain of course results from the fact that the GLS estimator utilizes between variations in \tilde{x}_i .

time-varying regressors with arbitrary autocovariance structure? Second, what is the asymptotic distribution of the Hausman statistic when $N, T \rightarrow \infty$? Is the statistic \mathcal{HM}_{NT} χ^2 -distributed despite the equivalence result? Third, does the Hausman test have power to detect violations of the RE assumption when T is large? Our equivalence result implies that between variations in data become less informative for the GLS estimation of β as $T \rightarrow \infty$. Then, the GLS estimator of β may remain consistent even if the RE assumption is violated. If this is the case, the power of the Hausman test might be inversely related to the size of T . We will attempt to answer these questions in the following sections.

What makes it complex to investigate the asymptotic properties of the within, GLS estimators and the Hausman statistic is that their convergence rates crucially depend on data generating processes. The following subsection considers several simple cases to illustrate this point.

2.2 Preliminary Results

This section considers several simple examples demonstrating that the convergence rates of the within, GLS estimators and the Hausman statistic crucially depend on whether or not data are cross-sectionally heteroskedastic, and whether or not time-varying regressors contain time trends. For model (1), we can easily show that

$$\widehat{\beta}_w - \beta = A_{NT}^{-1} a_{NT}; \quad (7)$$

$$\widehat{\beta}_b - \beta = (B_{NT} - C_{NT} H_N^{-1} C'_{NT})^{-1} [b_{NT} - C_{NT} H_N^{-1} c_{NT}], \quad (8)$$

where,

$$A_{NT} = \sum_{i,t} \tilde{x}_{it} \tilde{x}'_{it}; B_{NT} = \sum_i \tilde{x}_i \tilde{x}'_i; C_{NT} = \sum_i \tilde{x}_i \tilde{z}'_i; H_N = \sum_i \tilde{z}_i \tilde{z}'_i;$$

$$a_{NT} = \sum_{i,t} \tilde{x}_{it} v_{it}; b_{NT} = \sum_i \tilde{x}_i (u_i + \bar{v}_i); c_{NT} = \sum_i \tilde{z}_i (u_i + \bar{v}_i).$$

Using (7) and (8), we can also show that the GLS estimator is a convex combination of the within and between estimators:

$$\begin{aligned} \widehat{\beta}_g - \beta &= [A_{NT} + T\theta_T^2 (B_{NT} - C_{NT} H_N^{-1} C'_{NT})]^{-1} \\ &\quad \times [A_{NT} (\widehat{\beta}_w - \beta) + T\theta_T^2 (B_{NT} - C_{NT} H_N^{-1} C'_{NT}) (\widehat{\beta}_b - \beta)]. \end{aligned} \quad (9)$$

Using (7), (8) and (9), we can also obtain

$$\begin{aligned} \widehat{\beta}_w - \widehat{\beta}_g &= [A_{NT} + T\theta_T^2 (B_{NT} - C_{NT} H_N^{-1} C'_{NT})]^{-1} \\ &\quad \times T\theta_T^2 (B_{NT} - C_{NT} H_N^{-1} C'_{NT}) [(\widehat{\beta}_w - \beta) - (\widehat{\beta}_b - \beta)]; \end{aligned} \quad (10)$$

$$Var(\widehat{\beta}_w) - Var(\widehat{\beta}_g) = A_{NT}^{-1} - [A_{NT} + T\theta_T^2 (B_{NT} - C_{NT} H_N^{-1} C'_{NT})]^{-1}. \quad (11)$$

Equation (10) provides some insight into the convergence rate of the Hausman test statistic. Note that $(\widehat{\beta}_w - \widehat{\beta}_g)$ depends on both $(\widehat{\beta}_w - \beta)$ and $(\widehat{\beta}_b - \beta)$. Apparently, the between estimator $\widehat{\beta}_b$ exploits only N between-individual variations,

while the within estimator $\widehat{\beta}_w$ is computed based on $N(T-1)$ within-individual variations. Accordingly, $(\widehat{\beta}_b - \beta)$ converges to a zero vector in probability much slower than $(\widehat{\beta}_w - \beta)$ does. Thus, we can conjecture that the convergence rate of $(\widehat{\beta}_w - \widehat{\beta}_g)$ will depend on that of $(\widehat{\beta}_b - \beta)$, not $(\widehat{\beta}_w - \beta)$. Indeed, we below justify this conjecture.

In this subsection, we only consider a simple model which has a single time-varying regressor (x_{it}) and a single time-invariant regressor (z_i). Accordingly, all of the terms defined in (7)-(11) are scalars. We consider asymptotics under the RE assumption (3). To save space, this section only considers the estimators of β and the Hausman test. The asymptotic distributions of the estimators of γ will be discussed in Section 4. Throughout the examples below, we assume that the z_i are *i.i.d.* over different i with $N(0, \sigma_z^2)$. In addition, we introduce a notation e_{it} to denote a white noise component in the time-varying regressor x_{it} . We assume that the e_{it} are *i.i.d.* over different i and t with $N(0, \sigma_e^2)$, and are uncorrelated with the z_i .

CASE 1: We here consider a case in which the time-varying regressor x_{it} contains a time trend of order m . Specifically, we assume:

$$x_{it} = \Theta_{1,i}t^m + e_{it}. \quad (12)$$

We assume the parameters $\Theta_{1,i}$ are fixed with finite $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \Theta_{1,i} = \lim_{N \rightarrow \infty} \bar{\Theta}_1 \equiv p_{1,1}$ and $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \Theta_{1,i}^2 \equiv p_{1,2}$. We can allow them to be random without changing our results, but at the cost of analytical complexity.⁶ We consider two possible cases: one in which the parameters $\Theta_{1,i}$ are heterogeneous, and the other in which they are constant over different individuals. Allowing the $\Theta_{1,i}$ to be different across different individuals, we allow the x_{it} be cross-sectionally heteroskedastic. In contrast, if the $\Theta_{1,i}$ are constant over different i , the x_{it} become cross-sectionally homogeneous. As we show below, the convergence rates of the between estimator and Hausman test statistic are different in the two cases. Furthermore, whether or not the model is estimated with an overall intercept could matter for convergence rates.

To be more specific, consider the three terms B_{NT} , C_{NT} , and b_{NT} defined below (8). A straightforward algebra reveals that with $r_t \equiv t/T$,

$$\begin{aligned} B_{NT} &= \sum_i (\Theta_{1,i} - \bar{\Theta}_1)^2 \left(T^m \frac{1}{T} \sum_t r_t^m \right)^2 \\ &\quad + 2 \sum_i (\Theta_{1,i} - \bar{\Theta}_1) \left(T^m \frac{1}{T} \sum_t r_t^m \right) (\bar{e}_i - \bar{e}) + \sum_i (\bar{e}_i - \bar{e})^2; \\ C_{NT} &= \sum_i (\Theta_{1,i} - \bar{\Theta}_1) \left(T^m \frac{1}{T} \sum_t r_t^m \right) (\bar{z}_i - \bar{z}) + \sum_i (\bar{e}_i - \bar{e})(\bar{z}_i - \bar{z}); \end{aligned}$$

⁶We can consider a more general case: for example, $x_{it} = a_i t^m + \Theta_i + \Theta_t + b_i z_i + e_{it}$. However, the same asymptotic results apply to this general model. This is so because the trend term (t^m) dominates asymptotics.

$$\begin{aligned}
b_{NT} &= \sum_i (\Theta_{1,i} - \bar{\Theta}_1) \left(T^m \frac{1}{T} \sum_t r_t^m \right) u_i + \sum_i (\bar{e}_i - \bar{e}) \hat{u}_i \\
&\quad + \sum_i (\Theta_{1,i} - \bar{\Theta}_1) \left(T^m \frac{1}{T} \sum_t r_t^m \right) \bar{v}_i + \sum_i (\bar{e}_i - \bar{e}) \bar{v}_i.
\end{aligned}$$

From these three equations, it is obvious that the terms including T^m will be the dominant factors determining the asymptotic properties of B_{NT} , C_{NT} , and b_{NT} . However, if the parameters $\Theta_{1,i}$ are constant over different individuals so that $\Theta_{1,i} - \bar{\Theta}_1 = 0$, none of B_{NT} , C_{NT} , and b_{NT} depends on T^m . For this case, the asymptotic properties of the three terms depend on $(e_i - \bar{e})$. This result indicates that the asymptotic distribution of the between estimator $\hat{\beta}_b$, which is a function of B_{NT} , C_{NT} , and b_{NT} , will depend on whether the parameters $\Theta_{1,i}$ are cross-sectionally heteroskedastic or homoskedastic. Somewhat interestingly, however, the distinction between these two cases becomes unimportant when the model has no intercept term ($\zeta = 0$) and is estimated with this restriction. For such a case, B_{NT} , C_{NT} and b_{NT} depend on \bar{x}_i instead of \tilde{x}_i . With \bar{x}_i , the terms $(\Theta_{1,i} - \bar{\Theta}_1)$ and $(\bar{e}_i - \bar{e})$ in B_{NT} , C_{NT} , and b_{NT} are replaced by $\Theta_{1,i}$ and \bar{e}_i , respectively. Then, it is clear that the trend term T^m remains as a dominating factor whether or not the $\Theta_{1,i}$ are heterogenous.

We now consider the asymptotic distributions of the within, between, GLS estimators and the Hausman statistic under the two alternative assumptions about the parameters $\Theta_{1,i}$.

CASE 1.1: Assume that the parameters $\Theta_{1,i}$ are heterogeneous over different individuals; that is, $p_{1,2} - p_{1,1}^2 \neq 0$. For this case, we can easily show:

$$\begin{aligned}
plim_{N,T \rightarrow \infty} \frac{1}{T^{2m}} \frac{1}{NT} A_{NT} &= p_{1,2} q_1; \\
plim_{N,T \rightarrow \infty} \frac{1}{T^{2m}} \frac{1}{N} B_{NT} &= \frac{p_{1,2} - p_{1,1}^2}{(m+1)^2}; \\
plim_{N,T \rightarrow \infty} \frac{1}{T^m} \frac{1}{N} C_{NT} &= 0; \\
plim_{N \rightarrow \infty} \frac{1}{N} H_N &= \sigma_z^2,
\end{aligned}$$

where $q_1 = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t [(t/T)^m - \frac{1}{T} \sum_t (t/T)^m]^2 = \int_0^1 [r^m - 1/(m+1)]^2 dr$.⁷ The first two equalities are obtained using the fact that $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_t (t/T)^m = \int_0^1 r^m dr = 1/(m+1)$. In addition, we can also show that as $(N, T \rightarrow \infty)$,

$$\begin{aligned}
\frac{1}{T^m} \frac{1}{\sqrt{NT}} a_{NT} &\implies N(0, p_{1,2} q_1 \sigma_v^2); \\
\frac{1}{T^m} \frac{1}{\sqrt{N}} b_{NT} &\implies N\left(0, \sigma_u^2 \frac{p_{1,2} - p_{1,1}^2}{(m+1)^2}\right);
\end{aligned}$$

⁷We can obtain these results using the fact that under given assumptions, $\frac{1}{\sqrt{T}} \sum_t e_{it}$ and $\frac{1}{\sqrt{T}} \sum_t v_{it}$ are i.i.d. over different i with $N(0, \sigma_e^2)$ and $N(0, \sigma_v^2)$, respectively, for any T .

$$\frac{1}{\sqrt{N}}c_{NT} \implies N(0, \sigma_z^2 \sigma_u^2).$$

Using these results and the fact that $\lim_{N \rightarrow \infty} T\theta_T^2 = \sigma_v^2/\sigma_u^2$, we can show that as $(N, T \rightarrow \infty)$,

$$T^m \sqrt{NT}(\hat{\beta}_w - \beta) \implies N\left(0, \frac{\sigma_v^2}{p_{1,2}q_1}\right); \quad (13)$$

$$T^m \sqrt{N}(\hat{\beta}_b - \beta) \implies N\left(0, \sigma_u^2 \frac{(m+1)^2}{p_{1,2} - p_{1,1}^2}\right); \quad (14)$$

$$\begin{aligned} T^m \sqrt{NT}(\hat{\beta}_g - \beta) &= T^m \sqrt{NT}(\hat{\beta}_w - \beta) \\ &\quad + \frac{1}{\sqrt{T}} \frac{\sigma_v^2}{\sigma_u^2} \frac{(p_{1,2} - p_{1,1}^2)}{p_{1,2}q_1(m+1)^2} T^m \sqrt{N}(\hat{\beta}_b - \beta) \\ &\quad + o_p(1/\sqrt{T}); \end{aligned} \quad (15)$$

$$\begin{aligned} T^m \sqrt{NT^2}(\hat{\beta}_w - \hat{\beta}_g) &= -\frac{\sigma_v^2}{\sigma_u^2} \frac{(p_{1,2} - p_{1,1}^2)}{p_{1,2}q_1(m+1)^2} T^m \sqrt{N}(\hat{\beta}_b - \beta) + o_p(1) \\ &\implies N\left(0, \frac{\sigma_v^4}{\sigma_u^2} \frac{(p_{1,2} - p_{1,1}^2)^2}{(p_{1,2}q_1)^2(m+1)^2}\right); \end{aligned} \quad (16)$$

$$plim_{N,T \rightarrow \infty} NT^{2m+2} [Var(\hat{\beta}_w) - Var(\hat{\beta}_g)] = \frac{\sigma_v^4}{\sigma_u^2} \frac{(p_{1,2} - p_{1,1}^2)^2}{(p_{1,2}q_1)^2(m+1)^2}. \quad (17)$$

Several remarks follow. First, not surprisingly, all of the within, between and GLS estimators are superconsistent when the time-varying regressor x_{it} contains a time trend. Second, from (15), we can see that the two estimators $\hat{\beta}_w$ and $\hat{\beta}_g$ are $T^m \sqrt{NT}$ -equivalent in the sense that $(\hat{\beta}_w - \hat{\beta}_g)$ is $o_p(1/T^m \sqrt{NT})$. This is so because the second term in the right-hand side of (15) is $O_p(1/\sqrt{T})$. Nonetheless, from (16), we can see that $(\hat{\beta}_w - \hat{\beta}_g)$ is $O_p(1/T^m \sqrt{NT^2})$ and asymptotically normal. These results indicate that the within and GLS estimators are equivalent to each other by the order of $T^m \sqrt{NT}$, but not by the order of $T^m \sqrt{NT^2}$. Third, from (16) and (17), we can see that the Hausman statistic is asymptotically χ^2 -distributed. Fourth, when the model is estimated without an intercept term because $\zeta = 0$, all of the results (14)-(17) are still valid with $p_{1,2}$ replacing $(p_{1,2} - p_{1,1}^2)$.

Finally, (16) provides some intuition about the power property of the Hausman test. Observe that the asymptotic distribution of $(\hat{\beta}_w - \hat{\beta}_g)$ depends on that of $(\hat{\beta}_b - \beta)$. From this, we can conjecture that the Hausman statistic is for testing consistency of the between estimator $\hat{\beta}_b$, not exactly for testing the RE assumption. In fact, the RE assumption (3) is not a necessary condition for the asymptotic unbiasedness of $\hat{\beta}_b$. For example, if the effect is correlated with z_i ,

but not with x_{it} , $\hat{\beta}_b$ could be asymptotically unbiased, as we find in section 4.⁸ Thus, the Hausman test does not have power to detect the violations of the RE assumption in the direction in which $\hat{\beta}_b$ remains asymptotically unbiased. This issue will be further explored later.

CASE 1.2: We now assume that $\Theta_{1,i} = \Theta_1$ for all i ; that is, $p_{1,2} - p_{1,1}^2 = 0$. As we have discussed above, the terms B_{NT} , C_{NT} , and b_{NT} do not depend on T^m . Furthermore, the asymptotic distribution of the between estimator $\hat{\beta}_b$ depends on the e_{it} instead of the $\Theta_{1,i}$. Specifically, we can easily show that

$$\begin{aligned} plim_{N,T \rightarrow \infty} \frac{T}{N} B_{NT} &= \sigma_e^2; \\ plim_{N,T \rightarrow \infty} \frac{\sqrt{T}}{N} C_{NT} &= 0; \\ \frac{\sqrt{T}}{\sqrt{N}} b_{NT} &\implies N(0, \sigma_u^2 \sigma_e^2), \end{aligned}$$

while other asymptotics are essentially the same as those obtained for CASE 1.1. The asymptotic distributions of the within and GLS estimators are the same under both CASE 1.1 and CASE 1.2, but those of the between and the Hausman statistic are different. For CASE 1.2,

$$\sqrt{\frac{N}{T}}(\hat{\beta}_b - \beta) \implies N\left(0, \frac{\sigma_u^2}{\sigma_e^2}\right); \quad (18)$$

$$\begin{aligned} T^{2m} \sqrt{NT^3}(\hat{\beta}_w - \hat{\beta}_g) &= -\frac{\sigma_v^2}{\sigma_u^2} \frac{\sigma_e^2}{\Theta_1^2 q_1} \sqrt{\frac{N}{T}}(\hat{\beta}_b - \beta) + o_p(1) \\ &\implies N\left(0, \frac{\sigma_v^4 \sigma_e^2}{\Theta_1^4 q_1^2 \sigma_u^2}\right); \end{aligned} \quad (19)$$

$$plim_{N,T \rightarrow \infty} NT^{4m+3} [Var(\hat{\beta}_w) - Var(\hat{\beta}_g)] = \frac{\sigma_v^4 \sigma_e^2}{\Theta_1^4 q_1^2 \sigma_u^2}. \quad (20)$$

Several comments follow. First, the between estimator is no longer superconsistent if the time trend in x_{it} is common to every individual (i.e., the parameters $\Theta_{1,i}$ are the same for i). An interesting result is obtained when $N/T \rightarrow c < \infty$. For this case, the between estimator is inconsistent, although it is still asymptotically unbiased. This implies that the between estimator is an inconsistent estimator for the analysis of cross-sectionally homogeneous panel data unless N is substantially larger than T . Second, the convergence rate of $(\hat{\beta}_w - \hat{\beta}_g)$ is quite different between CASE 1.1 and 1.2. Notice that the convergence rate of $(\hat{\beta}_w - \hat{\beta}_g)$ is $\sqrt{NT^{4m+3}}$ for CASE 1.2, while it is $\sqrt{NT^{2m+2}}$ for CASE 1.1.

⁸In contrast, the between estimator of γ , $\hat{\gamma}_b$, is inconsistent whenever the RE assumption is violated.

Thus, $(\widehat{\beta}_w - \widehat{\beta}_g)$ converges in probability to zero much faster in CASE 1.2 than in CASE 1.1. Nonetheless, the Hausman statistic is asymptotically χ^2 -distributed in both cases.

CASE 2: We now consider two simple examples in which the time-varying regressor x_{it} is stationary without trend. Assume:

$$x_{it} = \Theta_{2,i} + \Psi_{2,t} + e_{it}, \quad (21)$$

where $\Theta_{2,i}$ and $\Psi_{2,t}$ are fixed individual-specific and time-specific effects, respectively. Define

$$\begin{aligned} p_{2,1} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \Theta_{2,i}; p_{2,2} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \Theta_{2,i}^2; \\ q_{2,1} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t \Psi_{2,t}; q_{2,2} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t \Psi_{2,t}^2. \end{aligned}$$

Notice that if the $\Theta_{2,i}$ are allowed to vary across different i , the x_{it} become cross-sectionally heteroskedastic. Similarly to CASE 1, we will demonstrate that the convergence rates of the between estimator and the Hausman statistic depend on whether the x_{it} is cross-sectionally heteroskedastic or homogeneous.

CASE 2.1: Assume that the $\Theta_{2,i}$ vary across different i ; that is, $p_{2,1} - p_{2,2}^2 \neq 0$. With this assumption, we can easily show:

$$\begin{aligned} \text{plim}_{N,T \rightarrow \infty} \frac{1}{NT} A_{NT} &= q_{2,2} - q_{2,1}^2 + \sigma_e^2; \\ \text{plim}_{N,T \rightarrow \infty} \frac{1}{N} B_{NT} &= p_{2,2} - p_{2,1}^2; \\ \text{plim}_{N,T \rightarrow \infty} \frac{1}{N} C_{NT} &= 0; \\ \text{plim}_{N \rightarrow \infty} \frac{1}{N} H_N &= \sigma_z^2; \end{aligned}$$

$$\begin{aligned} \frac{1}{\sqrt{NT}} a_{NT} &\Rightarrow N(0, \sigma_v^2 (q_{2,2} - q_{2,1}^2 + \sigma_e^2)); \\ \frac{1}{\sqrt{N}} b_{NT} &\Rightarrow N(0, \sigma_u^2 (p_{2,2} - p_{2,1}^2)); \\ \frac{1}{\sqrt{N}} c_{NT} &\Rightarrow N(0, \sigma_u^2 \sigma_z^2). \end{aligned}$$

With these results, we can show

$$\sqrt{NT}(\widehat{\beta}_w - \beta) \Rightarrow N\left(0, \frac{\sigma_v^2}{(q_{2,2} - q_{2,1}^2 + \sigma_e^2)}\right); \quad (22)$$

$$\sqrt{N}(\widehat{\beta}_b - \beta) \Rightarrow N\left(0, \frac{\sigma_u^2}{(p_{2,2} - p_{2,1}^2)}\right); \quad (23)$$

$$\begin{aligned}
\sqrt{NT}(\hat{\beta}_g - \beta) &= \sqrt{NT}(\hat{\beta}_w - \beta) \\
&+ \frac{1}{\sqrt{T}} \frac{\sigma_v^2}{\sigma_u^2} \frac{(p_{2,2} - p_{2,1}^2)}{(q_{2,2} - q_{2,1}^2 + \sigma_e^2)} \sqrt{N}(\hat{\beta}_b - \beta) \\
&+ o_p\left(\frac{1}{\sqrt{T}}\right); \tag{24}
\end{aligned}$$

$$\begin{aligned}
\sqrt{NT^2}(\hat{\beta}_w - \hat{\beta}_g) &= -\frac{\sigma_v^2}{\sigma_u^2} \frac{(p_{2,2} - p_{2,1}^2)}{(q_{2,2} - q_{2,1}^2 + \sigma_e^2)} \sqrt{N}(\hat{\beta}_b - \beta) + o_p(1) \\
\implies &N\left(0, \frac{\sigma_v^4}{\sigma_u^2} \frac{(p_{2,2} - p_{2,1}^2)}{(q_{2,2} - q_{2,1}^2 + \sigma_e^2)^2}\right); \tag{25}
\end{aligned}$$

$$\text{plim}_{N,T \rightarrow \infty} NT^2 [\text{Var}(\hat{\beta}_w) - \text{Var}(\hat{\beta}_g)] = \frac{\sigma_v^4}{\sigma_u^2} \frac{(p_{2,2} - p_{2,1}^2)}{(q_{2,2} - q_{2,1}^2 + \sigma_e^2)^2}. \tag{26}$$

Note that (22) - (26) are essentially the same as (13) - (17) except that they do not include the time trend T^m .

CASE 2.2: Now, we consider the case in which the $\Theta_{2,i}$ are constant over different i ; that is, $p_{2,2} - p_{2,1}^2 = 0$. If the model contains no intercept term ($\zeta = 0$) and it is estimated with this restriction, all of the results (22) - (26) are still valid with $p_{2,2}$ replacing $(p_{2,2} - p_{2,1}^2)$. However, if the model contains an intercept term, the assumption of cross-sectional homoskedasticity affects the convergence rates of the between estimator and the Hausman statistic, while it does not to the within and GLS estimators. To see this, we assume that $\Theta_{2,i} = 0$, for all i , without loss of generality. Then, we can show

$$\sqrt{\frac{N}{T}}(\hat{\beta}_b - \beta) \implies N\left(0, \frac{\sigma_v^2}{\sigma_e^2}\right); \tag{27}$$

$$\begin{aligned}
\sqrt{NT^3}(\hat{\beta}_w - \hat{\beta}_g) &= -\frac{\sigma_v^2}{\sigma_u^2} \frac{\sigma_e^2}{(q_{22,2} - q_{22,1}^2 + \sigma_e^2)} \sqrt{\frac{N}{T}}(\hat{\beta}_b - \beta) + o_p(1) \\
\implies &N\left(0, \frac{\sigma_v^4}{\sigma_u^2} \frac{\sigma_e^2}{(q_{22,2} - q_{22,1}^2 + \sigma_e^2)^2}\right); \tag{28}
\end{aligned}$$

$$\text{plim}_{N,T \rightarrow \infty} NT^3 [\text{Var}(\hat{\beta}_w) - \text{Var}(\hat{\beta}_g)] = \frac{\sigma_v^4 \sigma_e^2}{\sigma_u^2 (q_{22,2} - q_{22,1}^2 + \sigma_e^2)^2}, \tag{29}$$

as $(N, T \rightarrow \infty)$. Observe that the between estimator is $\sqrt{\frac{N}{T}}$ -consistent as in CASE 1.2.

CASE 3: So far, we have considered the cases in which the time-varying regressor x_{it} and the time invariant regressor z_i are uncorrelated. We now

examine the cases in which this assumption is relaxed. The degree of the correlation between the x_{it} and z_i may vary over time. As we demonstrate below, the asymptotic properties of the panel data estimators and the Hausman test statistic depend on how the correlation varies over time. The basic model we consider here is given by

$$x_{it} = \Theta_{3,i} + \Psi_{3,t} + \Pi_i z_i / t^m + e_{it}, \quad (30)$$

where the $\Theta_{3,i}$ and Π_i are individual-specific fixed parameters, the $\Psi_{3,t}$ are the time-specific fixed effects, m is a non-negative real number. Observe that because of the presence of the $\Theta_{3,i}$ and Π_i , the x_{it} are not i.i.d. over different i . The correlation between x_{it} and z_i decreases over time if $m > 0$. In contrast, $m = 0$ implies that the correlation remains constant over time. We will not report our detailed asymptotic results for model (30) with heterogeneous $\Theta_{3,i}$, because they are essentially the same as those we obtain for CASE 2.1. This is so because the terms $\Theta_{3,i}$ dominate and the terms $\Pi_i z_i / t^m$ become irrelevant in asymptotics. Thus, we set $\Theta_{3,i} = 0$ for all i . In addition, we set $\Psi_{3,t} = 0$ for all t . We do so because presence of the time effects is irrelevant for convergence rates of panel data estimators and the Hausman statistic. For CASE 3, the within and GLS estimators are always \sqrt{NT} -consistent regardless of the size of m . Thus, we only report the asymptotic results for the between estimator and the Hausman statistic.

For the cases in which the parameters Π_i are the same for all i , it is easy to show that the between estimator $\hat{\beta}_b$ does not depend on $\Pi_i z_i / t^m$. For such cases, the terms $\Pi_i z_i / t^m$ do not play any important role in asymptotics. In fact, when the parameters Π_i are the same for all i , we obtain exactly the same asymptotic results as those for CASE 2.2. This result is due to the fact that the individual mean of the time-varying regressor \bar{x}_i becomes a linear function of the time invariant regressor z_i if the Π_i are the same for all i . This particular case does not seem to be of practical importance, because it assumes an overly restrictive covariance structure of regressors. Thus we only consider the cases in which the Π_i are heterogeneous over different i .

We examine three possible cases: $m \in (\frac{1}{2}, \infty]$, $m = \frac{1}{2}$, and $m \in [0, \frac{1}{2})$. We do so because, depending on the size of m , one (or both) of the two terms e_{it} and $\Pi_i z_i / t^m$ in x_{it} becomes a dominating factor in determining the convergence rates of the between estimator $\hat{\beta}_b$ and the Hausman statistic $\mathcal{H}\mathcal{M}_{TN}$.

CASE 3.1: Assume that $m \in (\frac{1}{2}, \infty]$. This is the case where the correlation between x_{it} and z_i fades away quickly over time. Thus, one could expect that the correlation between x_{it} and z_i (through the term $\Pi_i z_i / t^m$) would not play any important role in asymptotics. Indeed, a straightforward algebra, which is not reported here, justifies this conjecture: The term e_{it} in x_{it} dominates $\Pi_i z_i / t^m$ asymptotically, and thus, this is essentially the same case as CASE 2.2.⁹

⁹We can obtain this result using the fact that $\lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \sum_t t^{-m} = 0$, if $m > \frac{1}{2}$.

CASE 3.2: We now assume $m = 1/2$. For this case, define

$$p_{3,1} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \Pi_i; p_{3,2} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \Pi_i^2,$$

and $q_3 = \lim_{T \rightarrow \infty} \frac{1}{T^{1-m}} \int_0^1 r^{-m} dr = \frac{1}{1-m}$ for $m \leq \frac{1}{2}$. With this notation, a little algebra shows that as $(N, T \rightarrow \infty)$,

$$\sqrt{\frac{N}{T}}(\hat{\beta}_b - \beta) \Rightarrow N \left(0, \frac{\sigma_u^2}{(p_{3,2} - p_{3,1}^2)q_3^2\sigma_z^2 + \sigma_e^2} \right).$$

Observe that the asymptotic variance of the between estimator $\hat{\beta}_b$ depends on both the terms σ_e^2 and $(p_{3,2} - p_{3,1}^2)q_3^2\sigma_z^2$. That is, both the terms e_{it} and $\Pi_i z_i / t^m$ in x_{it} are important in the asymptotics of the between estimator $\hat{\beta}_b$. This implies that the correlation between the x_{it} and z_i , when it decreases reasonably slowly over time, matters for the asymptotic distribution of the between estimator $\hat{\beta}_b$. Nonetheless, the convergence rate of $\hat{\beta}_b$ is the same as that of $\hat{\beta}_w$ for CASES 2.2 and 3.1. We can also show

$$\begin{aligned} \sqrt{NT^3}(\hat{\beta}_w - \hat{\beta}_g) &= -\frac{\sigma_v^2}{\sigma_u^2} \frac{(p_{3,2} - p_{3,1}^2)q_3^2\sigma_z^2 + \sigma_e^2}{\sigma_e^2} \sqrt{\frac{N}{T}}(\hat{\beta}_b - \beta) + o_p(1) \\ &\Rightarrow N \left(0, \frac{\sigma_v^4}{\sigma_u^2} \frac{(p_{3,2} - p_{3,1}^2)q_3^2\sigma_z^2 + \sigma_e^2}{\sigma_e^4} \right); \end{aligned}$$

$$plim_{N,T \rightarrow \infty} NT^3 [Var(\hat{\beta}_w) - Var(\hat{\beta}_g)] = \frac{\sigma_v^4}{\sigma_u^2} \frac{(p_{3,2} - p_{3,1}^2)q_3^2\sigma_z^2 + \sigma_e^2}{\sigma_e^4},$$

both of which imply that the Hausman statistic is asymptotically χ^2 -distributed.

CASE 3.3: Finally, we consider the case in which $m \in [0, \frac{1}{2})$, where the correlation between x_{it} and z_i decays over time slowly. Note that the correlation remains constant over time if $m = 0$. We can show

$$\sqrt{\frac{N}{T^{2m}}}(\hat{\beta}_b - \beta) \Rightarrow N \left(0, \frac{\sigma_u^2}{(p_{3,2} - p_{3,1}^2)q_3^2\sigma_z^2} \right).$$

Observe that the asymptotic distribution of $\hat{\beta}_b$ no longer depends on σ_e^2 . This implies that the term e_{it} in x_{it} dominates $\Pi_i z_i / t^m$ in the asymptotics for $\hat{\beta}_b$. Furthermore, the convergence rate of $\hat{\beta}_b$ now depend on m . Specifically, so long as $m < \frac{1}{2}$, the convergence rate increases as m decreases. In particular, when the correlation between x_{it} and z_i remains constant over time ($m = 0$), the between estimator $\hat{\beta}_b$ is \sqrt{N} -consistent as in CASE 2.1. This is so because, in this case, the term $\Pi_i z_i$ takes the role of the Θ_i term in CASE 2.1. Finally, the following results indicate that the convergence rate of the Hausman statistic $\mathcal{H}\mathcal{M}_{NT}$ also depends on m :

$$\begin{aligned}
\sqrt{NT^{2m+2}}(\widehat{\beta}_w - \widehat{\beta}_g) &= -\frac{\sigma_v^2 (p_{3,2} - p_{3,1}^2) q_3^2 \sigma_z^2}{\sigma_u^2 \sigma_e^2} \sqrt{\frac{N}{T^{2m}}} (\widehat{\beta}_b - \beta) + o_p(1) \\
&\Rightarrow N \left(0, \frac{\sigma_v^4 (p_{3,2} - p_{3,1}^2) q_3^2 \sigma_z^2}{\sigma_u^4 \sigma_e^4} \right); \\
plim_{N,T \rightarrow \infty} NT^{2m+2} [Var(\widehat{\beta}_w) - Var(\widehat{\beta}_g)] &= \frac{\sigma_v^4 (p_{3,2} - p_{3,1}^2) q_3^2 \sigma_z^2}{\sigma_u^4 \sigma_e^4}.
\end{aligned}$$

So far, we have considered several simple cases to demonstrate how the convergence rates of the popular panel data estimators and the Hausman test are sensitive to data generating processes. For these simple cases, all of the relevant asymptotics can be obtained in a straightforward manner. In the following sections, we will show that the main results obtained from this section apply to more general cases in which regressors are serially dependent with arbitrary covariance structures.

3 Conditional α - Mixing

In the asymptotic analysis of the general model (1) with large T , some technical difficulties arise when some of the time varying regressors x_{it} are correlated with the time invariant regressors z_i . For such cases, the temporal dependence of the time-varying regressors may persist through their correlations with the time invariant regressors; that is, the time series of x_{it} may not be ergodic in time. Thus, for general asymptotic results, we need to study the probability limits of the random variables containing time-averages of such non-ergodic regressors (*i.e.*, B_{NT} and b_{NT} in Section 2.2). In this section, we discuss the assumptions that can facilitate derivations of the (joint) limits of such random variables as $(N, T \rightarrow \infty)$ simultaneously.

Consider CASE 3.3 with $m = 0$. Observe that the time series of x_{it} is not ergodic, because of the presence of the time invariant random component $\Pi_i z_i$ in x_{it} . In addition, $cov(x_{it}, x_{i,t+l}) = E(x_{it} - E(x_{it}))(x_{i,t+l} - E(x_{i,t+l})) = \Pi_i^2 \sigma_z^2 \not\rightarrow 0$ as $l \rightarrow \infty$. Thus, the temporal dependence of x_{it} does not decay. Despite these problems, we were able to obtain handy asymptotic results based on the two strong assumptions: $E(x_{it} | z_i) = \Pi_i z_i$, and the conditional terms $e_{it} = x_{it} - E(x_{it} | z_i)$ are i.i.d. over time. This example illustrates that under some certain conditions imposed on non-ergodic time-varying regressors, we can analyze the asymptotic properties of sample averages of panel data. In fact, our major findings from CASE 3 remain unaltered even if we alternatively assume that $E(x_{it} | z_i)$ is an arbitrary nonlinear function of z_i , and/or the e_{it} are autocorrelated, so long as the e_{it} satisfy the conditions we discuss in detail below. Formally, we consider a mixing model that is defined conditionally on the sigma field generated by time invariant regressors z_i , which we call a *conditional*

mixing model. For this model, we can establish the joint limits of the sample averages of panel data whose time series are non-ergodic, as we show in Section 4 and Appendix.

Suppose that (Ω, \mathcal{F}, P) is a basic probability space, and $\mathcal{G}, \mathcal{H}, \mathcal{Z}$ are sub-sigma fields of \mathcal{F} with $\mathcal{Z} \subset \mathcal{G}$ and $\mathcal{Z} \subset \mathcal{H}$. Then, a conditional α -mixing coefficient between two sub-sigma fields \mathcal{G} and \mathcal{H} on \mathcal{Z} is defined as

$$\alpha_{\mathcal{Z}}(\mathcal{G}, \mathcal{H}) = \sup_{G \in \mathcal{G}, H \in \mathcal{H}} |P_{\mathcal{Z}}(G \cap H) - P_{\mathcal{Z}}(G)P_{\mathcal{Z}}(H)|, \quad (31)$$

where $P_{\mathcal{Z}}(\cdot)$ denotes a conditional probability defined on the sigma field \mathcal{Z} .¹⁰ This conditional α -mixing coefficient is a straightforward extension of the usual α -mixing coefficient, except that it uses the conditional probability $P_{\mathcal{Z}}(\cdot)$ instead of the usual unconditional probability $P(\cdot)$.

The general definition of the conditional α -mixing coefficient can apply to our panel data model as follows. Suppressing the subscript i for convenience, assume that $\{x_t\}_t$ and z are scalar random variables, respectively defined in the probability space (Ω, \mathcal{F}, P) , where $\sup_t E|x_t|^{2q} < \infty$, for some $q > 1$.¹¹ Define

$$\mathcal{F}_{-\infty}^t = \sigma(\dots, x_{t-1}, x_t); \mathcal{F}_{t+d}^{\infty} = \sigma(x_{t+m}, x_{t+m+1}, \dots); \mathcal{Z} = \sigma(z),$$

where \mathcal{Z} is assumed to be a non-trivial sigma field, *i.e.*, in \mathcal{Z} there exists a subset A of Ω with $0 < P(A) < 1$. Define

$$\alpha_{\mathcal{Z}}(d) = \sup_t \sup_{G \in \mathcal{F}_{-\infty}^t, H \in \mathcal{F}_{t+d}^{\infty}} |P_{\mathcal{Z}}(G \cap H) - P_{\mathcal{Z}}(G)P_{\mathcal{Z}}(H)|. \quad (32)$$

With this definition, we will say that the sequence $\{x_t\}$ is conditionally α -mixing if and only if

$$\alpha_{\mathcal{Z}}(d) \rightarrow 0 \text{ a.s.}, \quad (33)$$

as $d \rightarrow \infty$, where the almost sure convergence of $\alpha_{\mathcal{Z}}(d)$ holds with respect to an outer probability measure P^* of the probability space (Ω, \mathcal{F}, P) .¹²

A technical problem in using the conditional α -mixing coefficient $\alpha_{\mathcal{Z}}(d)$ (as well as $\alpha_{\mathcal{Z}}(\mathcal{G}, \mathcal{H})$) is that it is not necessarily measurable with respect to the conditioning sigma field \mathcal{Z} . This problem raises some technical difficulties in deriving useful inequalities. For example, following the usual techniques related to (unconditional) α -mixing coefficients, one may expect that the following conditional versions of α -mixing inequalities hold:

$$|E_{\mathcal{Z}}(x_t x_{t+d}) - (E_{\mathcal{Z}} x_t)(E_{\mathcal{Z}} x_{t+d})| \leq 2 \frac{q}{q-1} \left(\alpha_{\mathcal{Z}}(d)^{\frac{q-1}{q}} \right) \left(\sup_t (E_{\mathcal{Z}} |x_t|^{2q}) \right)^{\frac{1}{q}}; \quad (34)$$

$$|E_{\mathcal{Z}}(x_t x_{t+d}) - (E_{\mathcal{Z}} x_t)(E_{\mathcal{Z}} x_{t+d})| \leq 8 \left(\alpha_{\mathcal{Z}}(d)^{\frac{q-1}{q}} \right) \left(\sup_t (E_{\mathcal{Z}} |x_t|^{2q}) \right)^{\frac{1}{q}}, \quad (35)$$

¹⁰We also could define similar conditional mixing coefficients of β -mixing and ϕ -mixing.

¹¹The x_t need not be strictly stationary.

¹²For the details of the outer probability measure P^* , readers may refer to Chapter 1.2 of van der Vaart and Wellner (1996).

where $E_{\mathcal{Z}}(\cdot)$ to the conditional expectation with respect to a sigma field \mathcal{Z} . The inequality (34) is a conditional version of Theorem 1 and equation (1.3) in Rio (1993), or Theorem 1.1 and Corollary 1.1. of Bosq (1996). The inequality (35) is a conditional version of the mixing inequality in Corollary A.2 of Hall and Heyde (1980, p. 278). Observe that both of the inequalities indicate $cov_{\mathcal{Z}}(x_t, x_{t+d}) \rightarrow 0$ as $d \rightarrow \infty$, so long as the sequence $\{x_t\}$ is conditionally α -mixing. We may obtain these inequalities by modifying the method used in Rio (1993) or Hall and Heyde (1980) with the conditional arguments. However, to do so requires $\alpha_{\mathcal{Z}}(d)$ to be measurable.¹³

It is difficult to derive the sufficient *and* necessary conditions that warrant \mathcal{Z} -measurability of the conditional mixing coefficient, $\alpha_{\mathcal{Z}}(d)$ or $\alpha_{\mathcal{Z}}(\mathcal{G}, \mathcal{H})$. Thus, we here only consider a sufficient condition. Stated formally:

Theorem 1 *Suppose that the sigma field \mathcal{Z} is generated by a countable partition $\Pi = \{\Pi_1, \dots, \Pi_i, \dots\}$ of Ω with $P(\Pi_i) > 0$ for all i . Then, $\alpha_{\mathcal{Z}}(\mathcal{G}, \mathcal{H})$ in (31) is measurable with respect to the sigma field \mathcal{Z} .*

When \mathcal{Z} is the sigma field generated by a time-invariant regressor z , the restriction on \mathcal{Z} imposed by Theorem 1 is satisfied if z is a discrete random variable, *i.e.*, the supports of z are countable. This condition would not be too restrictive in practice. In many empirical studies, time invariant regressors generally consist of dummy variables (such as gender, race, or region), or discrete variables (such as years of schooling). Such variables easily satisfy the requirement of Theorem 1.

4 Main Results

This section derives for the general model (1) the asymptotic distributions of the within, between, GLS estimators and the Hausman statistic. In Section 2, we have considered independently several simple models in which regressors are of particular characteristics. The general model we consider in this section contains all of the different types of regressors analyzed in Section 2. More detailed assumptions are introduced below.

From now on, the following notation is repeatedly used. The expression “ \rightarrow_p ” means “converges in probability,” while “ \Rightarrow ” means “converges in distribution” as in Section 2.2. For any matrix A , the norm $\|A\|$ signifies $\sqrt{\text{tr}(AA')}$. When B is a random matrix with $E\|B\|_p < \infty$, then $\|B\|_p$ denotes $(E\|B\|^p)^{1/p}$. We use $E_{\mathcal{F}}(\cdot)$ to denote the conditional expectation operator with respect to a sigma field \mathcal{F} . We also define $\|B\|_{\mathcal{F},p} = (E_{\mathcal{F}}\|B\|^p)^{1/p}$. The notation $x_N \sim a_N$ indicates that there exists n and finite constants d_1 and d_2 such that

¹³Admittedly, we here do not attempt to determine whether or not measurability of $\alpha_{\mathcal{Z}}(d)$ is a necessary condition for the conditional mixing inequalities (34) and (35). It might be possible to derive the inequalities with some alternative methods that do not require the measurability assumption. Thus, we would like to emphasize that measurability of $\alpha_{\mathcal{Z}}(d)$ is a sufficient, but not necessarily a necessary condition.

$\inf_{N \geq n} \frac{x_N}{a_N} \geq d_1$ and $\sup_{N \geq n} \frac{x_N}{a_N} \leq d_2$. We also use the following notation for relevant sigma-fields: $\mathcal{F}_{x_i} = \sigma(x_{i1}, \dots, x_{iT})$; $\mathcal{F}_{z_i} = \sigma(z_i)$; $\mathcal{F}_z = \sigma(\mathcal{F}_{z_1}, \dots, \mathcal{F}_{z_N})$; $\mathcal{F}_{w_i} = \sigma(\mathcal{F}_{x_i}, \mathcal{F}_{z_i})$; and $\mathcal{F}_w = \sigma(\mathcal{F}_{w_1}, \dots, \mathcal{F}_{w_N})$. The x_{it} and z_i are now $k \times 1$ and $g \times 1$ vectors, respectively.

As in Section 2, we assume that the regressors $(x'_{i1}, \dots, x'_{iT}, z'_i)'$ are independently distributed across different i . In addition, we make the following the assumption about the composite error terms u_i and v_{it} :

Assumption 1 (about u_i and v_{it}): For some $q > 1$,

- (i) the u_i are independent over different i with $\sup_i E |u_i|^{4q} < \infty$.
- (ii) The v_{it} are i.i.d. with mean zero and variance σ_v^2 across different i and t , and are independent of x_{is} , z_i and u_i , for all i , t , and s . Also, $\|v_{it}\|_{4q} \equiv \kappa_v$ is finite.

Assumption 1(i) is a standard regularity condition for error-components models. Assumption 1(ii) indicates that all of the regressors and individual effect are strictly exogenous with respect to the error terms v_{it} .¹⁴

We now make the assumptions about regressors. In Section 2, we have considered three different cases: CASEs 1, 2, and 3. Consistently with these cases, we partition the $k \times 1$ vector x_{it} into three subvectors, $x_{1,it}$, $x_{2,it}$, and $x_{3,it}$, which are $k_1 \times 1$, $k_2 \times 1$, and $k_3 \times 1$, respectively. The vector $x_{1,it}$ consists of the regressors with deterministic trends. We may think of three different types of trends: (i) cross-sectionally heterogeneous nonstochastic trends in mean (but not in variance or covariances); (ii) cross-sectionally homogeneous nonstochastic trends; and (iii) stochastic trends (trends in variance) such as unit-root time series. In Section 2, we have considered the first two cases as CASEs 1.1 and 1.2, respectively. The latter case is materially similar to CASE 2.2, except that the convergence rates of estimators and test statistics are different under these two cases. Thus, we here only consider the case (i). We do not cover the cases of stochastic trends (iii), leaving the analysis of such cases to future study.

The two subvectors $x_{2,it}$ and $x_{3,it}$ are random regressors with no trend in mean. The partition of $x_{2,it}$ and $x_{3,it}$ is made based on their correlatedness with z_i . Specifically, we assume that the $x_{2,it}$ are not correlated with z_i , while the $x_{3,it}$ are. In addition, in order to accommodate CASEs 2.1 and 2.2, we also partition the subvector $x_{2,it}$ into $x_{21,it}$ and $x_{22,it}$, which are $k_{21} \times 1$ and $k_{22} \times 1$, respectively. Similarly to CASE 2.1, the regressor vector $x_{21,it}$ is heterogeneous over different i , as well as different t , with different means $\Theta_{21,it}$. In contrast, $x_{21,it}$ is homogeneous cross-sectionally with means $\Theta_{22,t}$ for all i for given t . We also incorporate CASEs 3.1, 3.2 and 3.3 into the model by partitioning $x_{3,it}$ into $x_{31,it}$, $x_{32,it}$, and $x_{33,it}$, which are $k_{31} \times 1$, $k_{32} \times 1$, and $k_{33} \times 1$, respectively, depending on how fast their correlations with z_i decay over time. The more detailed assumptions on the regressors x_{it} and z_i follow:

¹⁴As discussed in Section 2.1, this assumption rules out weakly exogenous or predetermined regressors.

Assumption 2 (about $x_{1,it}$):

- (i) For some $q > 1$, $\kappa_{x_1} \equiv \sup_{i,t} \|x_{1,it} - Ex_{1,it}\|_{4q} < \infty$.
- (ii) Let $x_{h,1,it}$ be the h^{th} element of $x_{1,it}$. Then, $Ex_{h,1,it} \sim t^{m_{h,1}}$ for all i and $h = 1, \dots, k_1$, where $m_{h,1} > 0$.

Assumption 3 (about $x_{2,it}$): For some $q > 1$,

- (i) $\{x_{2,it} - Ex_{2,it}\}_t$ is an α -mixing process for all i , and is independent of \mathcal{F}_{z_i} for all i . Let α_i be the mixing coefficient of $x_{2,it}$. Then, $\sup_i \alpha_i(d)$ is of size $-3\frac{q}{q-1}$, i.e., $\sup_i \alpha_i(d) = O\left(d^{-p-3\frac{q}{q-1}}\right)$, for some $p > 0$.
- (ii) $E(x_{21,it}) = \Theta_{21,it}$ and $E(x_{22,it}) = \Theta_{22,t}$, where $\sup_{i,t} \|\Theta_{21,it}\|, \sup_t \|\Theta_{22,t}\| < \infty$, and $\Theta_{21,it} \neq \Theta_{21,jt}$ if $i \neq j$.
- (iii) $\kappa_{x_2} \equiv \sup_{i,t} \|x_{2,it} - Ex_{2,it}\|_{4q} < \infty$.

Assumption 4 (about $x_{3,it}$): For some $q > 1$,

- (i) $\{x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}\}_t$ is conditional α -mixing for all i . Let $\alpha_{\mathcal{F}_{z_i}}$ be the conditional α -mixing coefficient of $x_{3,it}$ on \mathcal{F}_{z_i} . Then, $\sup_i \alpha_{z_i}(d)$ is of size $-3\frac{q}{q-1}$ a.s., i.e., $\sup_i \alpha_{z_i}(d) = O\left(d^{-p-3\frac{q}{q-1}}\right)$ a.s., for some $p > 0$. Also, $E\left(\sum_{d=1}^{\infty} d^2 \sup_i \left(\alpha_{\mathcal{F}_{z_i}}(d)^{\frac{q-1}{q}}\right)^2\right) < \infty$.
- (ii) $E(x_{3,it}) = \Theta_{3,it}$, where $\sup_{i,t} \|\Theta_{3,it}\| < \infty$.
- (iii) $E\left(\sup_{i,t} \|x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}\|_{\mathcal{F}_{z_i}, 4q}^{8q}\right) < \infty$.
- (iv) Let $x_{h,3k,it}$ be the h^{th} element of $x_{3k,it}$, where $k = 1, 2, 3$. Then, conditional on z_i ,
 - (iv.1) $(E_{\mathcal{F}_{z_i}} x_{h,31,it} - Ex_{h,31,it}) \sim t^{-m_{h,31}}$ a.s., where $\frac{1}{2} < m_{h,31} \leq \infty$ for $h = 1, \dots, k_{31}$ (here, $m_{h,31} = \infty$ implies that $E_{\mathcal{F}_{z_i}} x_{h,31,it} - Ex_{h,31,it} = 0$ a.s.);
 - (iv.2) $(E_{\mathcal{F}_{z_i}} x_{h,32,it} - Ex_{h,32,it}) \sim t^{-\frac{1}{2}}$ a.s. for $h = 1, \dots, k_{32}$;
 - (iv.3) $(E_{\mathcal{F}_{z_i}} x_{h,33,it} - Ex_{h,33,it}) \sim t^{-m_{h,33}}$ a.s., where $0 \leq m_{h,33} < \frac{1}{2}$ for $h = 1, \dots, k_{33}$.

Assumption 5 (about z_i):

- (i) $\{z_i\}_i$ is i.i.d. over i with $E(z_i) = \Theta_z$, and $\|z_i\|_{4q} < \infty$ for some $q > 1$.
- (ii) The support of the density of z_i is countable for all i .

Two points may be worth mentioning here. First, Assumption 4(i) dictates that the regressors in $x_{3,it}$ are mixing conditional on \mathcal{F}_{z_i} . Alternately, we may assume that $x_{3,it}$ and z_i have some common factors, say f_i ($i = 1, \dots, N$), on which, conditionally, the variables in $x_{3,it}$ are mixing. This alternative assumption does not generate any materially different asymptotic result. Second, as discussed in Section 3, Assumption 5(ii) warrants that the conditional α -mixing coefficient $\alpha_{z_i}(d)$ is measurable.

Panel data estimators of individual coefficients have different convergence rates depending on the types of the corresponding regressors. To address these differences, we define:

$$\begin{aligned} D_{x,T} &= \text{diag}(D_{1T}, D_{2T}, D_{3T}); \\ D_T &= \text{diag}(D_{x,T}, I_g), \end{aligned}$$

where

$$\begin{aligned} D_{1T} &= \text{diag}(T^{-m_1}, \dots, T^{-m_{k_1}}); \\ D_{2T} &= \text{diag}(D_{21T}, D_{22T}) = \text{diag}(I_{k_{21}}, \sqrt{T}I_{k_{22}}); \\ D_{3T} &= \text{diag}(D_{31T}, D_{32T}, D_{33T}) \\ &= \text{diag}(\sqrt{T}I_{k_{31}}, \sqrt{T}I_{k_{32}}, T^{m_{1,33}}, \dots, T^{m_{k_{33},33}}). \end{aligned}$$

Observe that D_{1T} , D_{2T} , and D_{3T} are conformable to regressor vectors $x_{1,it}$, $x_{2,it}$, and $x_{3,it}$, respectively, while D_T and I_g are to x_{it} and z_i , respectively. The diagonal matrix D_T is chosen so that $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{w}_i' D_T$ is well defined and finite. For future use, we also define

$$\begin{aligned} G_{x,T} &= \text{diag}(D_{1T}, I_{k_{21}}, I_{k_{22}}, I_{k_3}); \\ J_{x,T} &= \text{diag}(I_{k_1}, I_{k_{21}}, D_{22T}, D_{3T}), \end{aligned}$$

so that

$$D_{x,T} = G_{x,T} J_{x,T}.$$

Using this notation, we make the following regularity assumptions on the unconditional and conditional means of regressors:

Assumption 6 (convergence as $T \rightarrow \infty$): Defining $t = [Tr]$, we assume that the following restrictions hold as $T \rightarrow \infty$.

- (i) Let $\tau_1(r) = \text{diag}(r^{m_{1,1}}, \dots, r^{m_{k_1,1}})$, where $m_{h,1}$ is defined in Assumption 2. Then,

$$D_{1T} E(x_{1,it}) \rightarrow \tau_1(r) \Theta_{1,i}$$

uniformly in i and $r \in [0, 1]$, for some $\Theta_{1,i} = (\Theta_{1,1,i}, \dots, \Theta_{k_1,1,i})'$ with $\sup_i \|\Theta_{1,i}\| < \infty$.

- (ii) $\Theta_{21,it} \rightarrow \Theta_{21,i}$ and $\Theta_{3,it} \rightarrow \Theta_{3,i}$ uniformly in i with $\sup_i \|\Theta_{21,i}\| < \infty$ and $\sup_i \|\Theta_{3,i}\| < \infty$.

(iii) Uniformly in i and $r \in [0, 1]$,

$$D_{31T} (E_{\mathcal{F}_{z_i}} x_{31,it} - Ex_{31,it}) \rightarrow 0_{k_{31} \times 1} \text{ a.s.};$$

$$D_{32T} (E_{\mathcal{F}_{z_i}} x_{32,it} - Ex_{32,it}) \rightarrow \frac{1}{\sqrt{r}} I_{k_{32}} g_{32,i}(z_i) \text{ a.s.};$$

$$D_{33T} (E_{\mathcal{F}_{z_i}} x_{33,it} - Ex_{33,it}) \rightarrow \tau_{33}(r) g_{33,i}(z_i) \text{ a.s.},$$

where

$$g_{32,i} = (g_{1,32,i}, \dots, g_{k_{32},32,i})'; \quad g_{33,i} = (g_{1,33,i}, \dots, g_{k_{33},33,i})',$$

and $g_{32,i}(z_i)$ and $g_{33,i}(z_i)$ are zero-mean functions of z_i with

$$0 < E \sup_i \|g_{3k,i}(z_i)\|^{4q} < \infty, \text{ for some } q > 1,$$

and $g_{3k,i} \neq g_{3k,j}$ for $i \neq j$, and $\tau_{33}(r) = \text{diag}(r^{-m_{1,33}}, \dots, r^{-m_{k_{33},33}})$.

(iv) There exist $\tilde{\tau}(r)$ and $\tilde{G}_i(z_i)$ such that

$$\|D_{3T} (E_{\mathcal{F}_{z_i}} x_{3,it} - Ex_{3,it})\| \leq \tilde{\tau}(r) \tilde{G}_i(z_i),$$

where $\int \tilde{\tau}(r)^{4q} dr < \infty$ and $E \sup_i \tilde{G}_i(z_i)^{4q} < \infty$ for some $q > 1$.

(v) Uniformly in (i, j) and $r \in [0, 1]$;

$$D_{31T} (Ex_{31,it} - Ex_{31,jt}) \rightarrow 0_{k_{31} \times 1},$$

$$D_{32T} (Ex_{32,it} - Ex_{32,jt}) \rightarrow \frac{1}{\sqrt{r}} I_{k_{32}} (\mu_{g_{32i}} - \mu_{g_{32j}}),$$

$$D_{33T} (Ex_{33,it} - Ex_{33,jt}) \rightarrow \tau_{33}(r) (\mu_{g_{33i}} - \mu_{g_{33j}}),$$

with $\sup_i \|\mu_{g_{32i}}\|, \sup_i \|\mu_{g_{33i}}\| < \infty$.

Some remarks would be useful to understand Assumption 6. First, to have an intuition about what the assumption implies, we consider, as an illustrative example, the simple model in CASE 3 in Section 2.2, in which $x_{3,it} = \Pi_i z_i / t^m + e_{it}$, where e_{it} is independent of z_i and *i.i.d.* across i . For this case,

$$D_{3T} (E_{\mathcal{F}_{z_i}} x_{3,it} - Ex_{3,it}) = D_{3T} \Pi_i (z_i - Ez_i) / t^m;$$

$$D_{3T} (Ex_{3,it} - Ex_{3,jt}) = D_{3T} (\Pi_i Ez_i - \Pi_j Ez_j) / t^m.$$

Thus,

$$g_{3k,i}(z_i) = \Pi_i (z_i - Ez_i);$$

$$\mu_{g_{3k,i}} = \Pi_i Ez_i.$$

Second, Assumption 6(iii) makes the restriction that $E \sup_i \|g_{3k,i}(z_i)\|^{4q}$ is strictly positive, for $k = 2, 3$. This restriction is made to warrant that $g_{3k,i}(z_i) \neq 0$ a.s. If $g_{3k,i}(z_i) = 0$ a.s.,¹⁵ then

$$D_{3kT} E_{\mathcal{F}_{z_i}} (x_{3k,it} - Ex_{3k,it}) \sim \tau_{3k}(r) g_{32,i}(z_i) = 0 \text{ a.s.},$$

¹⁵An example is the case in which $x_{3,it} = e_{it} \Pi_i z_i / t^m$, where e_{it} is independent of z_i with mean zero.

and the correlations between $x_{3,it}$ and z_i no longer play any important role in asymptotics. Assumption 6(iii) rules out such cases.

Assumption 6 is about the asymptotic properties of means of regressors as $T \rightarrow \infty$. We also need additional regularity assumptions on the means of regressors that apply as $N \rightarrow \infty$. Define

$$H_1 = \int_0^1 \tau_1(r) dr; \quad H_{32} = \left(\int_0^1 \frac{1}{\sqrt{r}} dr \right) I_{k_{32}}; \quad H_{33} = \int_0^1 \tau_{33}(r) dr;$$

and

$$E \left[\begin{pmatrix} H_{32}g_{32,i}(z_i) \\ H_{33}g_{33,i}(z_i) \\ z_i - Ez_i \end{pmatrix} \begin{pmatrix} H_{32}g_{32,i}(z_i) \\ H_{33}g_{33,i}(z_i) \\ z_i - Ez_i \end{pmatrix}' \right] = \begin{pmatrix} \Gamma_{g_{32},g_{32},i} & \Gamma_{g_{32},g_{33},i} & \Gamma_{g_{32},z,i} \\ \Gamma'_{g_{32},g_{33},i} & \Gamma_{g_{33},g_{33},i} & \Gamma_{g_{33},z,i} \\ \Gamma'_{g_{32},z,i} & \Gamma'_{g_{33},z,i} & \Gamma_{zz,i} \end{pmatrix}.$$

With this notation, we assume the followings:

Assumption 7 (convergence as $N \rightarrow \infty$): Define $\tilde{\Theta}_{1,i} = \Theta_{1,i} - \frac{1}{N} \sum_i \Theta_{1,i}$; $\tilde{\Theta}_{21,i} = \Theta_{21,i} - \frac{1}{N} \sum_i \Theta_{21,i}$; $\tilde{\mu}_{g_{32},i} = \mu_{g_{32},i} - \frac{1}{N} \sum_i \mu_{g_{32},i}$; and $\tilde{\mu}_{g_{33},i} = \mu_{g_{33},i} - \frac{1}{N} \sum_i \mu_{g_{33},i}$. As $N \rightarrow \infty$,

$$\begin{aligned} (i) \quad & \frac{1}{N} \sum_i \begin{pmatrix} H_1 \tilde{\Theta}_{1,i} \\ \tilde{\Theta}_{21,i} \\ H_{32} \tilde{\mu}_{g_{32},i} \\ H_{33} \tilde{\mu}_{g_{33},i} \end{pmatrix} \begin{pmatrix} H_1 \tilde{\Theta}_{1,i} \\ \tilde{\Theta}_{21,i} \\ H_{32} \tilde{\mu}_{g_{32},i} \\ H_{33} \tilde{\mu}_{g_{33},i} \end{pmatrix}' \rightarrow \begin{pmatrix} \Gamma_{\Theta_1, \Theta_1} & \Gamma_{\Theta_1, \Theta_{21}} & \Gamma_{\Theta_1, \mu_{32}} & \Gamma_{\Theta_1, \mu_{33}} \\ \Gamma'_{\Theta_1, \Theta_{21}} & \Gamma_{\Theta_{21}, \Theta_{21}} & \Gamma_{\Theta_{21}, \mu_{32}} & \Gamma_{\Theta_{21}, \mu_{33}} \\ \Gamma'_{\Theta_1, \mu_{32}} & \Gamma'_{\Theta_{21}, \mu_{32}} & \Gamma_{\mu_{32}, \mu_{32}} & \Gamma_{\mu_{32}, \mu_{33}} \\ \Gamma'_{\Theta_1, \mu_{33}} & \Gamma'_{\Theta_{21}, \mu_{33}} & \Gamma'_{\mu_{32}, \mu_{33}} & \Gamma_{\mu_{33}, \mu_{33}} \end{pmatrix}. \\ (ii) \quad & \frac{1}{N} \sum_i \begin{pmatrix} \Gamma_{g_{32},g_{32},i} & \Gamma_{g_{32},g_{33},i} & \Gamma_{g_{32},z,i} \\ \Gamma'_{g_{32},g_{33},i} & \Gamma_{g_{33},g_{33},i} & \Gamma_{g_{33},z,i} \\ \Gamma'_{g_{32},z,i} & \Gamma'_{g_{33},z,i} & \Gamma_{zz,i} \end{pmatrix} \rightarrow \begin{pmatrix} \Gamma_{g_{32},g_{32}} & \Gamma_{g_{32},g_{33}} & \Gamma_{g_{32},z} \\ \Gamma'_{g_{32},g_{33}} & \Gamma_{g_{33},g_{33}} & \Gamma_{g_{33},z} \\ \Gamma'_{g_{32},z} & \Gamma'_{g_{33},z} & \Gamma_{zz} \end{pmatrix}. \\ (iii) \quad & \text{The limit of } \frac{1}{N} \sum_i \Theta_{1,i} \Theta_{1,i}' \text{ exists.} \end{aligned}$$

Apparently, by Assumptions 6 and 7, we assume the sequential convergence of the means of regressors as $T \rightarrow \infty$ followed by $N \rightarrow \infty$. However, this by no means implies that our asymptotic analysis is a sequential one. Instead, the uniformity conditions in Assumption 6 allow us to obtain our asymptotic results using the joint limit approach that applies as $(N, T \rightarrow \infty)$ simultaneously.¹⁶ Joint limit results can be obtained under an alternative set of conditions that assume uniform limits of the means of regressors sequentially as $N \rightarrow \infty$ followed by $T \rightarrow \infty$. Nonetheless, we adopt Assumptions 6 and 7, because they are much more convenient to handle the trends in regressors $x_{1,it}$ and $x_{3,it}$ for asymptotics.

The following notation is for conditional or unconditional covariances among time-varying regressors. Define

$$\Gamma_i(t, s) = [\Gamma_{jl,i}(t, s)]_{jl},$$

¹⁶For the details on the relationship between the sequential and joint approaches, see Apostol (1974, Theorems 8.39 and 9.16) for the cases of double indexed real number sequences, and Phillips and Moon (1999) for the cases of random sequences.

where $\Gamma_{jl,i}(t, s) = E(x_{j,it} - E_{\mathcal{F}_{z_i}} x_{j,it})(x_{l,is} - E_{\mathcal{F}_{z_i}} x_{l,is})$, for $j, l = 2, 3$. Essentially, the Γ_i is the unconditional mean of the conditional variance-covariance matrix of $(x'_{2,it}, x'_{3,it})'$. We also define the unconditional variance-covariance matrix of $(x'_{1,it}, x'_{2,it}, x'_{3,it})'$ by

$$\tilde{\Gamma}_i(t, s) = [\tilde{\Gamma}_{jl,i}(t, s)]_{jl},$$

where $\tilde{\Gamma}_{jl,i}(t, s) = E(x_{j,it} - Ex_{j,it})(x_{l,is} - Ex_{l,is})$, for $j, l = 1, 2, 3$. Observe that $\Gamma_{22,i}(t, s) = \tilde{\Gamma}_{22,i}(t, s)$, since $x_{2,it}$ and z_i are independent. With this notation, we make the following assumption on the convergence of variances and covariances:

Assumption 8 (*convergence of covariances*): As $(N, T \rightarrow \infty)$,

$$(i) \frac{1}{N} \sum_i \frac{1}{T} \sum_t \sum_s \begin{pmatrix} \Gamma_{22,i}(t, s) & \Gamma_{23,i}(t, s) \\ \Gamma'_{23,i}(t, s) & \Gamma_{33,i}(t, s) \end{pmatrix} \rightarrow \begin{pmatrix} \Gamma_{22} & \Gamma_{23} \\ \Gamma'_{23} & \Gamma_{33} \end{pmatrix}.$$

$$(ii) \frac{1}{N} \sum_i \frac{1}{T} \sum_t \tilde{\Gamma}_i(t, t) \rightarrow \Phi.$$

Note that the variance matrix $[\Gamma_{jl}]_{j,l=2,3}$ is the cross section average of the long-run variance-covariance matrix of $(x'_{2,it}, x'_{3,it})'$. For future use, we partition the two limits in the assumption conformably to $(x'_{21,it}, x'_{22,it}, x'_{31,it}, x'_{32,it}, x'_{33,it})'$ as follows:

$$\begin{pmatrix} \Gamma_{22} & \Gamma_{23} \\ \Gamma'_{23} & \Gamma_{33} \end{pmatrix} = \begin{pmatrix} \Gamma_{21,21} & \Gamma_{21,22} & \Gamma_{21,31} & \Gamma_{21,32} & \Gamma_{21,33} \\ \Gamma'_{21,22} & \Gamma_{22,22} & \Gamma_{22,31} & \Gamma_{22,32} & \Gamma_{22,33} \\ \Gamma'_{21,31} & \Gamma'_{22,31} & \Gamma_{31,31} & \Gamma_{31,32} & \Gamma_{31,33} \\ \Gamma'_{21,32} & \Gamma'_{22,32} & \Gamma'_{31,32} & \Gamma_{32,32} & \Gamma_{32,33} \\ \Gamma'_{21,33} & \Gamma'_{22,33} & \Gamma'_{31,33} & \Gamma'_{32,33} & \Gamma_{33,33} \end{pmatrix};$$

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi'_{12} & \Phi_{22} & \Phi_{23} \\ \Phi'_{13} & \Phi'_{23} & \Phi_{33} \end{pmatrix}.$$

Finally, we make a formal definition of the random effects assumption, which is a more rigorous version of (3).

Assumption 9 (*random effects*): Conditional on \mathcal{F}_w , $\{u_i\}_{i=1,\dots,N}$ is i.i.d. with mean zero, variance σ_u^2 and finite $\kappa_u \equiv \|u_i\|_{\mathcal{F}_w, 4}$.

To investigate the power property of the Hausman test, we also need to define an alternative hypothesis which states a particular direction of model misspecification. Among many alternatives, we here consider a simpler one. Specifically, we consider an alternative hypothesis under which the conditional mean of u_i is a linear function of $D_T \tilde{w}_i$. Abusing the conventional definition of fixed effects (that indicates nonzero-correlations between $w_i = (x'_{it}, z'_i)'$ and u_i), we refer to this alternative as the fixed effects assumption:

Assumption 10 (*fixed effects*): Conditional on \mathcal{F}_w , the $\{u_i\}_{i=1,\dots,N}$ is i.i.d. with mean $\tilde{w}'_i D_T \lambda$ and variance σ_w^2 , where λ is a $(k+g) \times 1$ nonrandom nonzero vector.

Here, $D_T \tilde{w}_i = [(D_{x,T} \tilde{x}_i)', \tilde{z}_i]$ can be viewed as a vector of detrended regressors. Thus, Assumption 10 indicates non-zero correlations between the effect u_i and detrended regressors. The term $\tilde{w}_i' D_T \lambda$ can be replaced by $\lambda_o + \bar{w}_i' D_T \lambda$, where λ_o is any constant scalar. We use the term $\tilde{w}_i' D_T \lambda$ instead of $\lambda_o + \bar{w}_i' D_T \lambda$ simply for convenience.

A sequence of local versions of the fixed effects hypothesis is given:

Assumption 11 (*local alternatives to random effects*): Conditional on \mathcal{F}_w , the sequence $\{u_i\}_{i=1, \dots, N}$ is i.i.d. with mean $\tilde{w}_i' D_T \lambda / \sqrt{N}$, variance σ_u^2 , and $\kappa_u^4 = E_{\mathcal{F}_w} (u_i - E_{\mathcal{F}_w} u_i)^4 < \infty$, where $\lambda \neq 0_{(k+g) \times 1}$ is a nonrandom vector in \mathbb{R}^{k+g} .

Under this Assumption, $E(D_T \tilde{w}_i u_i) = \frac{1}{\sqrt{N}} E(D_T \tilde{w}_i \tilde{w}_i' D_T) \lambda \rightarrow 0_{(k+g) \times 1}$, as $(N, T \rightarrow \infty)$.

Although Assumptions 10 and 11 are convenient to analyze the power properties of the Hausman test, they are somewhat restrictive. Specifically, under these alternative hypotheses, the Hausman test lacks power to detect correlations between the effect u_i and the time invariant regressors z_i . To see this, suppose we partition λ into $(\lambda'_x, \lambda'_z)'$ corresponding to $\tilde{w}_i = (\tilde{x}'_i, \tilde{z}'_i)'$. Assume that $\lambda_x = 0_{k \times 1}$; that is, the x_{it} ($t = 1, \dots, T$) are not correlated with u_i , conditional on z_i . For simplicity, assume that T is fixed. For this case, under the fixed effects assumption, the between estimators of β and γ , $\hat{\beta}_b$ and $\hat{\gamma}_b$, are equivalent to least squares on the model

$$\tilde{y}_i = \beta' \tilde{x}_i + (\gamma + \lambda_z)' \tilde{z}_i + (u_i^* + \tilde{v}_i),$$

where $u_i^* = u_i - \tilde{w}_i' D_T \lambda = u_i - \lambda'_z \tilde{z}_i$. From this, we can easily see that $\hat{\beta}_b$ and $\hat{\gamma}_b$ are asymptotically unbiased estimators of β and $(\gamma + \lambda_z)$, respectively. That is, $\hat{\gamma}_b$ is not an asymptotically unbiased estimator of γ . As we have discussed in Section 2.2, the asymptotic distribution of the Hausman statistic depends on that of $\hat{\beta}_b$, not of $\hat{\gamma}_b$. Thus, the Hausman test does not have power to detect the violations of the random effects assumption that do not bias $\hat{\beta}_b$ (regardless of the size of T). Accordingly, under our fixed effects and the local alternative assumptions (Assumptions 10 and 11), the Hausman test possesses no power to detect nonzero correlations between z_i and u_i . This problem arises of course because we assume that the conditional mean of the effect \tilde{u}_i is a linear function of \tilde{w}_i . When the conditional mean of the effect is a nonlinear function of \tilde{w}_i , the Hausman test can possess power to detect nonzero correlations between u_i and z_i .¹⁷

The following lemmas provide some results that are useful to derive the asymptotic distributions of the within, between, and GLS estimators of β and γ .

¹⁷Even if the conditional mean of u_i is linear in \tilde{w}_i , the Hausman test may have power to detect non-zero λ_z , if λ_x and λ_z are not functionally independent. For example, consider a model with scalar x_{it} and z_i . Suppose that x_{it} and z_i have a common factor f_i ; that is, $x_{it} = f_i + e_{it}$ and $z_i = f_i + \eta_i$. (This is the case discussed below Assumption 5.) Assume $E(u_i | f_i, \eta_i, \bar{e}_i) = c\eta_i$. Assume that f_i, η_i and e_{it} are normal, mutually independent, and i.i.d. over different i and t with zero means, and variances $\sigma_f^2, \sigma_\eta^2$, and σ_e^2 , respectively. Note that under given assumptions, x_{it} is not correlated with u_i , while z_i is. For this case, however,

Lemma 2 Under Assumptions 1-8, we obtain the following results as $(N, T \rightarrow \infty)$. For some positive semidefinite matrices Ψ_x and Ξ (defined in the Appendix),

- (a) $\frac{1}{N} \sum_i \frac{1}{T} \sum_t G_{x,T} \tilde{x}_{it} \tilde{x}'_{it} G_{x,T} \rightarrow_p \Psi_x$;
- (b) $\frac{1}{\sqrt{N}} \sum_i \frac{1}{\sqrt{T}} \sum_t G_{x,T} \tilde{x}_{it} \tilde{v}_{it} \Rightarrow N(0, \sigma_v^2 \Psi_x)$;
- (c) $\frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{w}'_i D_T \rightarrow_p \Xi$;
- (d) $\frac{1}{\sqrt{N}} \sum_i D_T \tilde{w}_i \tilde{v}_i \rightarrow_p 0_{(k+g) \times 1}$.

Lemma 3 Under Assumptions 1-8 and Assumption 11 (local alternatives to random effects), as $(N, T \rightarrow \infty)$,

$$\frac{1}{\sqrt{N}} \sum_i D_T \tilde{w}_i \tilde{u}_i \Rightarrow N(\Xi \lambda, \sigma_u^2 \Xi).$$

Lemma 4 Under Assumptions 1-8 and Assumption 10 (fixed effects),

$$\frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{u}_i \rightarrow_p \Xi \lambda,$$

as $(N, T \rightarrow \infty)$.

The following assumption is required for identification of the within and between estimators of β and γ .

Assumption 12 The matrices Ψ_x and Ξ are positive definite.

Two remarks on this assumption follow. First, this assumption is also sufficient for identification of the GLS estimation. Second, while the positive definiteness of the matrix Ξ is required for identification of the between estimators, it is not a necessary condition for the asymptotic distribution of the Hausman statistic obtained below. We can obtain the same asymptotic results for the Hausman test even if we alternatively assume that within estimation can identify β (positive definite Ψ_x) and between estimation can identify γ given β (the part of Ξ corresponding to \tilde{z}_i is positive definite).¹⁸ Nonetheless, we assume that Ξ is invertible for convenience.

We now consider the asymptotic distributions of the within, between and GLS estimators of β and γ :

we can show that

$$E(u_i | \bar{x}_i, z_i) = \bar{x}_i \lambda_x + z_i \lambda_z,$$

where $d = (\sigma_f^2 + \sigma_e^2/T)(\sigma_f^2 + \sigma_\eta^2) - \sigma_f^4$, $\lambda_x = -c\sigma_f^2\sigma_\eta^2/d$, $\lambda_z = c(\sigma_f^2 + \sigma_e^2/T)\sigma_\eta^2/d$. Observe that $\lambda_x \neq 0$, if $\lambda_z \neq 0$ ($c \neq 0$). Thus, λ_x is functionally related to λ_z . In addition, it is easy to show that $\text{plim}_{N \rightarrow \infty} \hat{\beta}_b = \beta + \lambda_x \neq \beta$, if $\lambda_z \neq 0$ ($c \neq 0$). Thus, nonzero λ_z biases $\hat{\beta}_b$.

¹⁸This claim can be checked with the following simple example. Consider a simple model with one time-varying regressor x_{it} and one time invariant regressor z_i . Assume that $x_{it} = az_i + e_{it}$, where the e_{it} are i.i.d. over different i and t . For this model, it is straightforward to show that the matrix Ξ fails to be invertible. Nonetheless, under the random effects assumption, the Hausman statistic can be shown to follow a χ^2 distribution with the degree of freedom equal to one.

Theorem 5 (asymptotic distribution of the within estimator): Under Assumptions 1-8 and Assumption 12, as $(N, T \rightarrow \infty)$,

$$\sqrt{NT}G_{x,T}^{-1}(\hat{\beta}_w - \beta) \Rightarrow N(0, \sigma_v^2 \Psi_x^{-1}).$$

Theorem 6 (asymptotic distribution of the between estimator): Suppose that Assumption 1-8 and 12 hold. As $(N, T \rightarrow \infty)$,

(a) under Assumption 9 (random effects),

$$D_T^{-1} \sqrt{N} \begin{pmatrix} \hat{\beta}_b - \beta \\ \hat{\gamma}_b - \gamma \end{pmatrix} = \begin{pmatrix} D_{x,T}^{-1} \sqrt{N} (\hat{\beta}_b - \beta) \\ \sqrt{N} (\hat{\gamma}_b - \gamma) \end{pmatrix} \Rightarrow N(0, \sigma_u^2 \Xi^{-1});$$

(b) under Assumption 11 (local alternatives to random effects),

$$D_T^{-1} \sqrt{N} \begin{pmatrix} \hat{\beta}_b - \beta \\ \hat{\gamma}_b - \gamma \end{pmatrix} = \begin{pmatrix} D_{x,T}^{-1} \sqrt{N} (\hat{\beta}_b - \beta) \\ \sqrt{N} (\hat{\gamma}_b - \gamma) \end{pmatrix} \Rightarrow N(\Xi \lambda, \sigma_u^2 \Xi^{-1}).$$

Theorem 7 (asymptotic distribution of the GLS estimator of β): Suppose that Assumptions 1-8 and 12 hold.

(a) Under Assumption 11 (local alternatives to random effects),

$$\sqrt{NT}G_{x,T}^{-1}(\hat{\beta}_g - \beta) = \sqrt{NT}G_{x,T}^{-1}(\hat{\beta}_w - \beta) + o_p(1),$$

as $(N, T \rightarrow \infty)$.

(b) Suppose that Assumption 10 (fixed effects) holds. Partition $\lambda = (\lambda'_x, \lambda'_z)'$ conformably to the sizes of x_{it} and z_i . Assume that $\lambda_x \neq 0_{k \times 1}$. If $N/T \rightarrow c < \infty$ and the included regressors are only of the $x_{22,it}$ - and $x_{3,it}$ -types (no trends and no cross-sectional heteroskedasticity in x_{it}), then

$$\sqrt{NT}G_{x,T}^{-1}(\hat{\beta}_g - \beta) = \sqrt{NT}G_{x,T}^{-1}(\hat{\beta}_w - \beta) + o_p(1).$$

Theorem 8 (asymptotic distribution of the GLS estimator of γ): Suppose that Assumptions 1-8 and 12 hold. Define $l'_z = (0_{g \times k}; I_g)$. Then, the following statements hold as $(N, T \rightarrow \infty)$.

(a) Under Assumption 11 (local alternatives to random effects),

$$\begin{aligned} \sqrt{N}(\hat{\gamma}_g - \gamma) &= \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_i \tilde{z}_i \tilde{u}_i \right) + o_p(1) \\ &\Rightarrow N \left((l'_z \Xi l_z)^{-1} l'_z \Xi \lambda, \sigma_u^2 (l'_z \Xi l_z)^{-1} \right). \end{aligned}$$

(b) Under Assumption 10 (fixed effects),

$$(\hat{\gamma}_g - \gamma) \rightarrow_p (l'_z \Xi l_z)^{-1} l'_z \Xi \lambda.$$

Several remarks follow. First, all of the asymptotic results given in Theorems 5-8 except for Theorem 7(b) hold as $(N, T \rightarrow \infty)$, without any particular restriction on the convergence rates of N and T . The relative size of N and T does not matter for the results, so long as both N and T are large. Second, one can easily check that the convergence rates of the panel data estimates of individual β coefficients reported in Theorems 5-8 are consistent with those from Section 2.2. Third, Theorem 6 shows that under Assumption 9 (random effects), the between estimator of γ , $\hat{\gamma}_b$, is \sqrt{N} -consistent regardless of the characteristics of time-varying regressors. Fourth, both the between estimators of β and γ are asymptotically biased under the sequence of local alternatives (Assumption 11). Fifth, as Theorem 7(a) indicates, the within and GLS estimators of β are asymptotically equivalent not only under the random effects assumption, but also under the local alternatives. Furthermore, the GLS estimator of β is asymptotically unbiased under the local alternatives, while the between estimator of β is not. The asymptotic equivalence between the within and GLS estimation under the random effects assumption is nothing new. Previous studies have shown this equivalence based on a naive sequential limit method ($T \rightarrow \infty$ followed by $N \rightarrow \infty$) and some strong assumptions such as fixed regressors. Theorem 7(a) and (b) confirm the same equivalence result, but with more rigorous joint limit approach as $(N, T \rightarrow \infty)$ simultaneously. It is also intriguing to see that the GLS and within estimators are equivalent even under the local alternative hypotheses.

Sixth, somewhat surprisingly, as Theorem 7(b) indicates, even under the fixed effects assumption (Assumption 10), the GLS estimator of β could be asymptotically unbiased (and consistent) and equivalent to the within counterpart, (i) if the size (N) of the cross section units does not dominate excessively the size (T) of time series in the limit ($N/T \rightarrow c < \infty$), and (ii) if the model does not contain trended or cross-sectionally heterogeneous time-varying regressors. This result indicates that when the two conditions are satisfied, the biases in GLS caused by fixed effects are generally much smaller than those in between. If at least one of these two conditions is violated, that is, if $N/T \rightarrow \infty$, or if the other types of regressors are included, the limit of $(\hat{\beta}_g - \hat{\beta}_w)$ is determined by how fast $N/T \rightarrow \infty$ and how fast the trends in the regressors increase or decrease.¹⁹

Finally, Theorem 8(a) indicates that under the local alternative hypotheses, the GLS estimator $\hat{\gamma}_g$ is \sqrt{N} -consistent and asymptotically normal, but asymptotically biased. The limiting distribution of $\hat{\gamma}_g$, in this case, is equivalent to the limiting distribution of the OLS estimator of γ in the panel model with the known coefficients of the time-varying regressors x_{it} (OLS on $\tilde{y}_{it} - \beta' \tilde{x}_{it} = \gamma' \tilde{z}_i + (u_i + \tilde{v}_{it})$). Clearly, the GLS estimator $\hat{\gamma}_g$ is asymptotically more efficient than the between estimator $\hat{\gamma}_b$. On the other hand, under the fixed effect assumption, unlike the GLS estimator of β , $\hat{\beta}_g$, the GLS estimator $\hat{\gamma}_g$ is not consistent as $(N, T \rightarrow \infty)$. The asymptotic bias of $\hat{\gamma}_g$ is given in Theorem 8(b).

¹⁹In this case, without specific assumptions on the convergence rates of N/T and the trends, it is hard to generalize the limits of the difference of the within and the GLS estimators.

Lastly, the following theorem finds the asymptotic distribution of the Hausman test statistic under the random effect assumption and the local alternatives:

Theorem 9 *Suppose that Assumptions 1-8 and 12 hold. Corresponding to the size of $(\bar{x}'_i, z'_i)'$, partition Ξ and λ , respectively, as follows:*

$$\Xi = \begin{pmatrix} \Xi_{xx} & \Xi_{xz} \\ \Xi'_{xz} & \Xi_{zz} \end{pmatrix}; \lambda = \begin{pmatrix} \lambda_x \\ \lambda_z \end{pmatrix}.$$

Then, as $(N, T \rightarrow \infty)$,

(a) under Assumption 9 (random effects),

$$\mathcal{HM}_{NT} \Rightarrow \chi_k^2;$$

(b) under Assumption 11 (local alternatives to random effects),

$$\mathcal{HM}_{NT} \Rightarrow \chi_k^2(\eta),$$

where $\eta = \lambda'_x (\Xi_{xx} - \Xi_{xz} \Xi_{zz}^{-1} \Xi'_{xz}) \lambda_x / \sigma_u^2$ is the noncentral parameter.

Theorem 9 shows that under the random effects assumption, the Hausman statistic is asymptotically χ^2 -distributed with degrees of freedom equal to k (the number of the time-varying regressors). Furthermore, Theorem 9 (ii) shows that the Hausman statistic has significant local power to detect any correlation between the time-varying regressors x_{it} and the effect u_i . This is so, because the noncentral parameter η equals zero if, and only if, $\lambda_x = 0_{k \times 1}$. In contrast, the noncentrality parameter η does not depend on λ_z , indicating that the Hausman test has no power to detect nonzero correlations between time invariant regressors z_i and the individual effect u_i in the direction of our local alternative hypotheses (Assumption 11). This result holds even if T is finite and fixed. As discussed earlier, this is due to the fact that the conditional mean of the effect u_i is a linear function of regressors under our local alternative hypotheses. When the conditional mean is not linear, the Hausman test could have a power to detect nonzero correlations of the effect u_i with the time invariant regressors. However, the power of the Hausman test to such correlations is generally limited. This is so because the Hausman test can detect such correlations only if they can cause a large bias in the between estimator of β , the coefficient vector on time-varying regressors (see Ahn and Low, 1996).

5 Conclusion

This paper has considered the asymptotic properties of the popular panel data estimators and the Hausman test. We find that the convergence rates of the estimators and the test statistic are sensitive to data generating process. In particular, the convergence rates of the between estimator crucially depend on whether the data are cross-sectionally heteroskedastic or homoskedastic. Despite the different convergence rates, however, the estimators are consistent and

asymptotically normal under the random effects assumption. The conventional Hausman test is also well defined. The Hausman test, which is based on the difference between the GLS and within estimators, has significant local power to detect violations of the random effects assumption (in particular, non-zero correlations between the time-varying regressors and unobservable individual effects), despite the fact that the two estimators are asymptotically identical under a sequence of local alternative hypotheses.

In this paper, we have restricted our attention to the asymptotic properties of the existing estimators and tests when panel data contain both large numbers of cross section and time series observations. Apparently, thus, this paper does not provide any new estimator or test. However, this paper makes several contributions to the literature. First, our findings have pedagogical values for future studies. For example, we find that asymptotics as $(N, T \rightarrow \infty)$ are much more sensitive to data generating processes than asymptotics as either $N \rightarrow \infty$ or $T \rightarrow \infty$ are. However, previous studies have often assumed that data are cross-sectionally i.i.d.. Our findings suggest that future studies should pay more attention to cross-sectional heterogeneity. Second, we consider the cases in which the time series of time-varying regressors are not ergodic due to their correlations with time invariant regressors. For such cases, we have shown that the limits of averages of panel data can be derived under the assumption of conditional α -mixing. It would also be interesting to see how this conditional α -mixing concept can be refined and generalized to other more sophisticated panel data models. Finally, differently from many other previous studies, we avoid making any particular restriction on the relative sizes of N and T . We do so using a more rigorous joint limit instead of other simple sequential limit methods. Thus we are confident that our theoretical results apply to a broader range of panel data.

An obvious extension of our paper is the instrumental variables estimation of Hausman and Taylor (1981), Amemiya and MaCurdy (1986), and Breusch, Mizon and Schmidt (1989). For an intermediate model between fixed effects and random effects, these studies propose several instrumental variables estimators by which both the coefficients on time-varying and time invariant regressors can be consistently estimated. It would be interesting to investigate the large N and large T properties of these instrumental variables estimators and the Hausman tests based on these estimators.

6 Appendix A: Preliminary Results

We here provide some preliminary lemmas that are useful to prove the main results in Section 4. From now on, we use the notation M to denote a generic constant, if no explanation follows.

Lemma 10 *Let $f_{i,T}$, f_i , and g_i are integrable functions in probability space (Ω, \mathcal{F}, P) . If $f_{i,T} \rightarrow f_i$ a.s. uniformly in i as $T \rightarrow \infty$, and there exists g_i such that $|f_{i,T}| \leq g_i$ for all i and T with $E \sup_i g_i < \infty$, then $E f_{i,T} \rightarrow E f_i$ uniformly in i as $T \rightarrow \infty$.*

Proof

Let $h_{i,T} = |f_{i,T} - f_i|$. Under given assumptions, $0 \leq \sup_i h_{i,T} \leq 2 \sup_i g_i$ and $h_{i,T} \rightarrow 0$ uniformly in i as $T \rightarrow \infty$. Then, by Fatou's Lemma

$$\begin{aligned} 2E \sup_i g_i &= E \left(\liminf_{T \rightarrow \infty} \left(2 \sup_i g_i - \sup_i h_{i,T} \right) \right) \\ &\leq \liminf_{T \rightarrow \infty} E \left(2 \sup_i g_i - \sup_i h_{i,T} \right) \\ &= 2E \sup_i g_i - \limsup_{T \rightarrow \infty} E \sup_i h_{i,T}, \end{aligned}$$

from which we can deduce

$$\limsup_{T \rightarrow \infty} E \sup_i h_{i,T} \leq 0.$$

Then, since

$$0 \leq \limsup_{T \rightarrow \infty} \sup_i |E(f_{i,T} - f_i)| \leq \limsup_{T \rightarrow \infty} \sup_i E h_{i,T} \leq \limsup_{T \rightarrow \infty} E \sup_i h_{i,T} \leq 0,$$

we have the required result: $E f_{i,T} \rightarrow E f_i$ uniformly in i as $T \rightarrow \infty$. ■

The following lemma is a uniform version of the Toeplitz lemma.

Lemma 11 *Let a_{it} be a sequence of real numbers such that $a_{it} \rightarrow a_i$ uniformly in i as $t \rightarrow \infty$ with $\sup_i |a_i| < M$. Then, (a) $\frac{1}{T} \sum_t a_{it} \rightarrow a_i$ uniformly in i , and (b) $\frac{1}{T} \sum_t a_{it}^2 \rightarrow a_i^2$ uniformly in i .*

Proof

From the uniform convergence of a_{it} , for a given $\varepsilon > 0$, we can choose t_0 such that $t \geq t_0$ implies that

$$\sup_i |a_{it} - a_i| < \varepsilon.$$

Then, Part (a) follows because $t \geq t_0$ implies

$$\sup_i \left| \frac{1}{T} \sum_t (a_{it} - a_i) \right| \leq \frac{1}{T} \sum_t \sup_i |a_{it} - a_i| < \varepsilon.$$

For Part (b), notice that

$$\begin{aligned} \sup_i |a_{it}^2 - a_i^2| &\leq \sup_i |a_{it} + a_i| \sup |a_{it} - a_i| \\ &\leq M \sup |a_{it} - a_i| \rightarrow 0 \end{aligned}$$

as $t \rightarrow \infty$, where the last inequality holds because $\sup_i |a_i| < \infty$ and $\sup |a_{it} - a_i| \rightarrow 0$. Then, by Part (a), we can obtain the desired result. ■

Lemma 12 *Suppose that $X_{i,T}$ and X_i are sequences of random vectors. Suppose that $X_{i,T} \rightarrow X_i$ in probability (or almost surely) uniformly in i as $T \rightarrow \infty$, and $\frac{1}{N} \sum_i X_i \rightarrow X$ in probability (or almost surely) as $N \rightarrow \infty$. Then, as $(N, T \rightarrow \infty)$,*

$$\frac{1}{N} \sum_i X_{i,T} \rightarrow X$$

in probability (or almost surely).

Proof

We only prove the lemma for the case of convergence in probability, because the almost sure convergence case can be proven by the similar fashion. Since $\frac{1}{N} \sum_i X_i \rightarrow_p X$ as $N \rightarrow \infty$ and $X \rightarrow_p X_i$ uniformly in i as $T \rightarrow \infty$, for given $\varepsilon, \delta > 0$, we can choose N_0 and T_0 such that

$$\begin{aligned} P \left\{ \left\| \frac{1}{N} \sum_i X_i - X \right\| > \frac{\varepsilon}{2} \right\} &\leq \frac{\delta}{2}; \\ P \left\{ \sup_i \|X_{i,T} - X_i\| > \frac{\varepsilon}{2} \right\} &\leq \frac{\delta}{2}, \end{aligned}$$

whenever $N \geq N_0$ and $T \geq T_0$. Now, suppose that $N \geq N_0$ and $T \geq T_0$. Then,

$$\begin{aligned} &P \left\{ \left\| \frac{1}{N} \sum_i X_{i,T} - X \right\| > \varepsilon \right\} \\ &\leq P \left\{ \left\| \frac{1}{N} \sum_i (X_{i,T} - X_i) \right\| > \frac{\varepsilon}{2} \right\} + P \left\{ \left\| \frac{1}{N} \sum_i X_i - X \right\| > \frac{\varepsilon}{2} \right\} \\ &\leq P \left\{ \sup_i \|(X_{i,T} - X_i)\| > \frac{\varepsilon}{2} \right\} + \frac{\delta}{2} = \delta. \quad \blacksquare \end{aligned}$$

Lemma 13 *Suppose that a sequence of random vectors Z_i is independently distributed across i . Let $\mathcal{F}_{z_i} = \sigma(Z_i)$. Assume that Q_{iT} ($k \times k$) is a sequence of independent random matrices across i satisfying*

$$\sup_{i,T} E_{\mathcal{F}_{z_i}} \|Q_{iT}\| \mathbf{1} \{ \|Q_{iT}\| > M \} \rightarrow 0 \text{ a.s.}, \quad (36)$$

as $M \rightarrow \infty$, where $1(\cdot)$ is the indicator function that equals one if the argument in the parenthesis is correct, and otherwise equals zero. Then, as $(N, T \rightarrow \infty)$,

$$\frac{1}{N} \sum_i (Q_{iT} - E_{\mathcal{F}_{z_i}} Q_{iT}) \rightarrow_p 0.$$

In fact, Lemma 13 still holds even if we replace the conditional mean operator $E_{\mathcal{F}_{z_i}}(\cdot)$ by the unconditional operator $E(\cdot)$. Thus, we have the following corollary.

Corollary 14 *Suppose that Q_{iT} ($k \times k$) is a sequence of independent random matrices across i satisfying*

$$\sup_{i,T} E \|Q_{iT}\| 1\{\|Q_{iT}\| > M\} \rightarrow 0 \text{ a.s.}, \quad (37)$$

as $M \rightarrow \infty$. Then, as $(N, T \rightarrow \infty)$,

$$\frac{1}{N} \sum_i (Q_{iT} - EQ_{iT}) \rightarrow_p 0.$$

Proof of Lemma 13

Let $\mathcal{F}_z = \sigma(\mathcal{F}_{z_1}, \dots, \mathcal{F}_{z_N})$ and $E_{\mathcal{F}_z}$ denote the conditional expectation on \mathcal{F}_z . For any $\varepsilon > 0$, we need to show that

$$P \left\{ \left\| \frac{1}{N} \sum_i (Q_{iT} - E_{\mathcal{F}_{z_i}} Q_{iT}) \right\| > \varepsilon \right\} \rightarrow 0.$$

This follows from the dominated convergence theorem if we can show that

$$P_{\mathcal{F}_z} \left\{ \left\| \frac{1}{N} \sum_i (Q_{iT} - E_{\mathcal{F}_{z_i}} Q_{iT}) \right\| > \varepsilon \right\} \rightarrow 0 \text{ a.s..}$$

But, this in turn follows from the conditional Markov inequality if we can show that

$$E_{\mathcal{F}_z} \left\| \frac{1}{N} \sum_i (Q_{iT} - E_{\mathcal{F}_{z_i}} Q_{iT}) \right\| \rightarrow 0 \text{ a.s..} \quad (38)$$

To show (38), define:

$$P_{iT} = Q_{iT} 1\{\|Q_{iT}\| \leq M\}; R_{iT} = Q_{iT} 1\{\|Q_{iT}\| > M\}.$$

Then, almost surely,

$$\begin{aligned} & E_{\mathcal{F}_z} \left\| \frac{1}{N} \sum_i (Q_{iT} - E_{\mathcal{F}_{z_i}} Q_{iT}) \right\| \\ & \leq E_{\mathcal{F}_z} \left\| \frac{1}{N} \sum_i (P_{iT} - E_{\mathcal{F}_{z_i}} P_{iT}) \right\| + E_{\mathcal{F}_z} \left\| \frac{1}{N} \sum_i (R_{iT} - E_{\mathcal{F}_{z_i}} R_{iT}) \right\| \\ & \leq \left(E_{\mathcal{F}_z} \left\| \frac{1}{N} \sum_i (P_{iT} - E_{\mathcal{F}_{z_i}} P_{iT}) \right\|^2 \right)^{\frac{1}{2}} + 2 \frac{1}{N} \sum_i E_{\mathcal{F}_{z_i}} \|R_{iT}\|. \end{aligned}$$

By definition,

$$\begin{aligned} & \frac{1}{N} \sum_i E_{\mathcal{F}_{z_i}} \|R_{iT}\| \\ & \leq \sup_{i,T} E_{\mathcal{F}_{z_i}} \|R_{iT}\| \leq \sup_{i,T} E_{\mathcal{F}_{z_i}} \|Q_{iT}\| \mathbb{1}\{\|Q_{iT}\| > M\}. \end{aligned} \quad (39)$$

Also,

$$\begin{aligned} & \left(E_{\mathcal{F}_z} \left\| \frac{1}{N} \sum_i (P_{iT} - E_{\mathcal{F}_{z_i}} P_{iT}) \right\| \right)^{\frac{1}{2}} \\ & = \left(\frac{1}{N^2} \sum_i E_{\mathcal{F}_{z_i}} \|P_{iT} - E_{\mathcal{F}_{z_i}} P_{iT}\|^2 \right)^{\frac{1}{2}} \leq \frac{1}{\sqrt{N}} \left(\sup_{i,T} E_{\mathcal{F}_{z_i}} \|P_{iT}\|^2 \right)^{\frac{1}{2}} \\ & \leq \frac{1}{\sqrt{N}} M. \end{aligned} \quad (40)$$

Choose $M = N^a$, where $0 < a < \frac{1}{2}$. Then, (39), (40) $\rightarrow 0$ a.s.. Consequently, we have (38). ■

Lemma 15 *Suppose that Assumptions 1-8 hold. Let $\mathcal{F}_z^\infty = \sigma(z_1, \dots, z_N, \dots)$. For a generic constant M that is independent of N and T , and for some \mathcal{F}_z^∞ -measurable function M_z , the followings hold:*

- (a) $\sup_{i,T} \left\| \frac{1}{T} \sum_t (x_{1,it} - E x_{1,it}) \right\|_4 < M$;
- (b) $\sup_{i,T} \left\| \frac{1}{\sqrt{T}} \sum_t (x_{2,it} - E x_{2,it}) \right\|_4 < M$;
- (c) $\sup_{i,T} \left\| \frac{1}{\sqrt{T}} \sum_t (x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}) \right\|_{\mathcal{F}_{z_i}, 4} < M_z$, a.s.;
- (d) $\sup_{i,T} \left\| \frac{1}{\sqrt{T}} \sum_t (x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}) \right\|_4 < M$.

Proof

We here use q to denote the real number used in Assumptions 1-6, which is strictly greater than 1.

Part (a)

Note that

$$\begin{aligned} \sup_{i,T} \left\| \frac{1}{T} \sum_t (x_{1,it} - E x_{1,it}) \right\|_4 & \leq \sup_{i,T} \|x_{1,it} - E x_{1,it}\|_4 \\ & \leq \sup_{i,T} \|x_{1,it} - E x_{1,it}\|_{4q} = \kappa_{x_1} < \infty, \end{aligned}$$

where the first inequality holds by Minkowski's inequality, the second inequality holds by Liapunov's inequality, and the last inequality holds by Assumption 2. Choose $M = \kappa_{x_1}$. ■

Part (b)

Let $x_{h,2,it}$ be the h^{th} element of $x_{h,2,it}$. The required result follows if we can show that

$$\sup_{i,T} \left\| \frac{1}{\sqrt{T}} \sum_t (x_{h,2,it} - E x_{h,2,it}) \right\|_4 < M \text{ for all } h. \quad (41)$$

The proof of (41) is similar to the proof of Lemma 1 of Andrews (1991). This proof relies on the following α -mixing inequality presented in Hall and Heyde (1980, p. 278). Suppose that Y and W are random variables that are \mathcal{G} -measurable and \mathcal{H} -measurable, respectively, with $E|Y|^p < \infty$ and $E|W|^q < \infty$, where $p, q > 1$ with $1/p + 1/q < 1$. Then,

$$|E(Y - EY)(W - EW)| \leq 8 \|Y\|_p \|W\|_q [\alpha(\mathcal{G}, \mathcal{H})]^{1-1/p-1/q}, \quad (42)$$

where $\alpha(\mathcal{G}, \mathcal{H})$ is the α -mixing coefficient between the sigma fields \mathcal{G} and \mathcal{H} .

Now, let $X_{it} = x_{h,2,it} - E x_{h,2,it}$. Notice that

$$\begin{aligned} & \sup_{i,T} E \left(\frac{1}{\sqrt{T}} \sum_t X_{it} \right)^4 \\ & \leq \sup_{i,T} \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{p=1}^T \sum_{k=1}^T |E(X_{it} X_{is} X_{ip} X_{ik})| \\ & \leq 4! \sup_{i,T} \frac{1}{T^2} \sum_t \sum_{s=0}^{T-t} \sum_{p=0}^{T-s} \sum_{k=0}^{T-s-p} |E(X_{it} X_{i,t+s} X_{i,t+s+p} X_{i,t+s+p+k})| \\ & \leq 4! \sup_{i,T} \frac{1}{T^2} \sum_t \sum_{\substack{0 \leq p, k \leq s \\ 0 \leq p+k+s \leq T-t}} |E(X_{it} (X_{i,t+s} X_{i,t+s+p} X_{i,t+s+p+k}))| \\ & \quad + 4! \sup_{i,T} \frac{1}{T^2} \sum_t \sum_{\substack{0 \leq s, k \leq p \\ 0 \leq p+k+s \leq T-t}} \left| \begin{array}{l} E[(X_{it} X_{i,t+s})(X_{i,t+s+p} X_{i,t+s+p+k})] \\ - E(X_{it} X_{i,t+s}) E(X_{i,t+s+p} X_{i,t+s+p+k}) \end{array} \right| \\ & \quad + 4! \sup_{i,T} \frac{1}{T^2} \sum_t \sum_{\substack{0 \leq s, k \leq p \\ 0 \leq p+k+s \leq T-t}} |E(X_{it} X_{i,t+s}) E(X_{i,t+s+p} X_{i,t+s+p+k})| \\ & \quad + 4! \sup_{i,T} \frac{1}{T^2} \sum_t \sum_{\substack{0 \leq s, p \leq k \\ 0 \leq p+k+s \leq T-t}} |E((X_{it} X_{i,t+s} X_{i,t+s+p}) X_{i,t+s+p+k})| \\ & = I + II + III + IV, \text{ say.} \end{aligned}$$

By applying the inequality of (42) to $X_{i,t+s} X_{i,t+s+p} X_{i,t+s+p+k}$ and X_{it} and then by the Hölder inequality, we have

$$\begin{aligned} I & \leq 4! 8 \sup_{i,T} \frac{1}{T^2} \sum_{t=1}^T \sum_{0 \leq p, k \leq s \leq T-t} \|X_{it}\|_{4q} \|X_{i,t+s}\|_{4q} \|X_{i,t+s+p}\|_{4q} \\ & \quad \times \|X_{i,t+s+p+k}\|_{4q} \alpha_i(s)^{\frac{q-1}{q}} \end{aligned}$$

$$\begin{aligned}
&\leq 4!8 \left(\sup_{i,t} \|X_{it}\|_{4q} \right)^4 \sup_{i,T} \frac{1}{T} \sum_{s=1}^{T-1} \sum_{p=0}^s \sum_{k=0}^s \alpha_i(s)^{\frac{q-1}{q}} \\
&\leq 4!8 \left(\sup_{i,t} \|X_{it}\|_{4q} \right)^4 \sum_{s=1}^{\infty} (s+1)^2 \sup_i \alpha_i(s)^{\frac{q-1}{q}} < M, \tag{43}
\end{aligned}$$

where the last bound holds because $\sup_i \alpha_i(s)$ is of size $-3\frac{q}{q-1}$ (see Assumption 3(i)). By the similar fashion, we can also show that

$$II, III, IV \leq 4!8 \left(\sup_{i,t} \|X_{it}\|_{4q} \right)^4 \sum_{s=1}^{\infty} (s+1)^2 \sup_i \alpha_i(s)^{\frac{q-1}{q}} < M,$$

and we have all the required result. ■

Part (c)

Let $Y_{it} = x_{h,3,it} - E_{\mathcal{F}_{z_i}} x_{h,3,it}$. Using the arguments similar to those used in the proof of Part (b), we can show that under Assumption 4,

$$\begin{aligned}
&\sup_{i,T} \left(E_{\mathcal{F}_{z_i}} \left(\frac{1}{\sqrt{T}} \sum_t Y_{it} \right)^4 \right) \\
&\leq 2M \left(\sup_{i,T} \|Y_{it}\|_{\mathcal{F}_{z_i},4q}^4 \right) \sum_{s=1}^{\infty} s^2 \left(\sup_i \alpha_{\mathcal{F}_{z_i}}(s)^{\frac{q-1}{q}} \right) \text{ a.s.}, \tag{44}
\end{aligned}$$

for some constant M . By Assumption 4(i), $\sum_{s=1}^{\infty} s^2 \sup_i \alpha_{\mathcal{F}_{z_i}}(s)^{\frac{q-1}{q}} < \infty$ a.s.. Finally, since the terms $\sup_{i,t} \|Y_{it}\|_{\mathcal{F}_{z_i},4q}^4$ and $\sum_{s=1}^{\infty} s^2 \sup_i \alpha_{\mathcal{F}_{z_i}}(s)^{\frac{q-1}{q}}$ are \mathcal{F}_z^∞ -measurable, we have the desired result by choosing

$$M_z = 2M \left(\sup_{i,T} \|Y_{it}\|_{\mathcal{F}_{z_i},4q}^4 \right) \sum_{s=1}^{\infty} s^2 \left(\sup_i \alpha_{\mathcal{F}_{z_i}}(s)^{\frac{q-1}{q}} \right). \blacksquare$$

Part (d)

Again, let $Y_{it} = x_{h,3,it} - E_{\mathcal{F}_{z_i}} x_{h,3,it}$. From (44), we have

$$\begin{aligned}
&\sup_{i,T} \left(E \left(\frac{1}{\sqrt{T}} \sum_t Y_{it} \right)^4 \right) \\
&= \sup_{i,T} E \left(E_{\mathcal{F}_{z_i}} \left(\frac{1}{\sqrt{T}} \sum_t Y_{it} \right)^4 \right) \\
&\leq 2ME \left[\left(\sup_{i,t} \|Y_{it}\|_{\mathcal{F}_{z_i},4q}^4 \right) \sum_{s=1}^{\infty} s^2 \left(\sup_i \alpha_{\mathcal{F}_{z_i}}(s)^{\frac{q-1}{q}} \right) \right] \\
&\leq 2M \left[E \left(\sup_{i,t} \|Y_{it}\|_{\mathcal{F}_{z_i},4q}^8 \right) \right]^{\frac{1}{2}} \left[E \left\{ \sum_{s=1}^{\infty} s^2 \left(\sup_i \alpha_{\mathcal{F}_{z_i}}(s)^{\frac{q-1}{q}} \right) \right\}^2 \right]^{\frac{1}{2}} \tag{45}
\end{aligned}$$

for some finite constant M . But Assumption 4(i) and (iii) imply that the right-hand side of the last inequality in (45) is finite. This completes the proof. ■

7 Appendix B: Proofs of Main Results

Proof of Theorem 1

For the desired result, it is enough to prove that

$$\{\omega \in \Omega | \alpha_{\mathcal{Z}}(\mathcal{G}, \mathcal{H})(\omega) \leq x\} \in \mathcal{Z}, \quad (46)$$

for all $x \in \mathbb{R}$. Since the partition $\Pi = \{\Pi_1, \dots, \Pi_i, \dots\}$ of Ω generates the sigma field \mathcal{Z} , we have

$$\begin{aligned} & \sup_{G \in \mathcal{G}, H \in \mathcal{H}} |(P_{\mathcal{Z}}(G \cap H))(\omega) - (P_{\mathcal{Z}}G)(\omega)(P_{\mathcal{Z}}H)(\omega)| \\ &= \sum_{i \in \mathbb{I}} \sup_{G \in \mathcal{G}, H \in \mathcal{H}} \left| \frac{P(G \cap H \cap \Pi_i)}{P\Pi_i} - \frac{P(G \cap \Pi_i)}{P\Pi_i} \frac{P(H \cap \Pi_i)}{P\Pi_i} \right| 1_{\Pi_i}(\omega), \end{aligned}$$

where $1_{\Pi_i}(\omega)$ denotes the indicator function that equals one if $\omega \in \Pi_i$, and otherwise equals zero. Let $\mathbb{I} = \{1, 2, \dots, i, \dots\}$ be the set of positive integers, and let

$$\mathbb{I}_x = \left\{ i \in \mathbb{I} \mid \sup_{G \in \mathcal{G}, H \in \mathcal{H}} \left| \frac{P(G \cap H \cap \Pi_i)}{P\Pi_i} - \frac{P(G \cap \Pi_i)}{P\Pi_i} \frac{P(H \cap \Pi_i)}{P\Pi_i} \right| \leq x \right\}.$$

Then, we have (46) because

$$\{\omega \in \Omega | \alpha_{\mathcal{Z}}(\mathcal{G}, \mathcal{H})(\omega) \leq x\} = \cup_{i \in \mathbb{I}_x} \Pi_i \in \mathcal{Z}. \quad \blacksquare$$

Before we start proving the lemmas and theorems in Section 4, we introduce some additional lemmas that are used repeatedly below. Recall that $w_{it} = (x'_{1,it}, x'_{2,it}, x'_{3,it}, z'_i)'$. We also repeatedly use the diagonal matrix D_T defined in Section 4.

Lemma 16 *Suppose that Assumptions 1-8 hold. Define $\Xi = \Xi_1 + \Xi_2$, where*

$$\begin{aligned} \Xi_1 &= \text{diag} \left(0_{k_1}, 0_{k_{21}}, \left(\begin{array}{ccc} \Gamma_{22,22} & \Gamma_{22,31} & \Gamma_{22,32} \\ \Gamma'_{22,31} & \Gamma_{31,31} & \Gamma_{31,32} \\ \Gamma'_{22,32} & \Gamma'_{31,32} & \Gamma_{32,32} \end{array} \right), 0_{k_{33}}, 0_{k_z} \right); \\ \Xi_2 &= \left(\begin{array}{ccccccc} \Gamma_{\Theta_1, \Theta_1} & \Gamma_{\Theta_1, \Theta_{21}} & 0 & 0 & \Gamma_{\Theta_1, \mu_{32}} & \Gamma_{\Theta_1, \mu_{33}} & 0 \\ \Gamma'_{\Theta_1, \Theta_{21}} & \Gamma_{\Theta_{21}, \Theta_{21}} & 0 & 0 & \Gamma_{\Theta_{21}, \mu_{32}} & \Gamma_{\Theta_{21}, \mu_{33}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \Gamma'_{\Theta_1, \mu_{32}} & \Gamma'_{\Theta_{21}, \mu_{32}} & 0 & 0 & \Gamma_{g_{32}, g_{32}} & \Gamma_{g_{32}, g_{33}} & \Gamma_{g_{32}, z} \\ & & & & +\Gamma_{\mu_{32}, \mu_{32}} & +\Gamma_{\mu_{32}, \mu_{33}} & \\ \Gamma'_{\Theta_1, \mu_{33}} & \Gamma'_{\Theta_{21}, \mu_{33}} & 0 & 0 & \Gamma'_{g_{32}, g_{33}} & \Gamma_{g_{33}, g_{33}} & \Gamma_{g_{33}, z} \\ & & & & +\Gamma'_{\mu_{32}, \mu_{32}} & +\Gamma_{\mu_{33}, \mu_{33}} & \\ 0 & 0 & 0 & 0 & \Gamma'_{g_{32}, z} & \Gamma'_{g_{33}, z} & \Gamma_{z, z} \end{array} \right). \end{aligned}$$

Then, under Assumption 11, as $(N, T \rightarrow \infty)$, the followings hold.

- (a) $\frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{w}'_i D_T \rightarrow_p \Xi$.
- (b) $\sup_{N, T} \sup_{1 \leq i \leq N} E \|D_T \tilde{w}_i\|^4 < M$, for some constant $M < \infty$.
- (c) $\frac{1}{\sqrt{N}} \sum_i D_T \tilde{w}_i \tilde{u}_i \Rightarrow N (\Xi \lambda, \sigma_u^2 \Xi)$.
- (d) $\frac{1}{\sqrt{N}} \sum_i D_T \tilde{w}_i \tilde{v}_i \rightarrow_p 0$.

Proof

Part (a)

To find the joint limit of

$$\frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{w}'_i D_T = \frac{1}{N} \sum_i D_T (\bar{w}_i - \bar{w}) (\bar{w}_i - \bar{w})' D_T,$$

we define

$$\begin{aligned} E_i^* \bar{w}_i &= (E \bar{x}'_{1,i}, E \bar{x}'_{2,i}, E_{\mathcal{F}_{z_i}} \bar{x}'_{3,i}, z'_i)'; \\ E^* \bar{w} &= (E \bar{x}'_1, E \bar{x}'_2, E_{\mathcal{F}_z} \bar{x}'_3, \bar{z}'). \end{aligned}$$

With this notation, we write

$$\begin{aligned} & \frac{1}{N} \sum_i D_T (\bar{w}_i - \bar{w}) (\bar{w}_i - \bar{w})' D_T \\ &= \frac{1}{N} \sum_i D_T [(\bar{w}_i - E_i^* \bar{w}_i) + (E_i^* \bar{w}_i - E^* \bar{w}) + (E^* \bar{w} - \bar{w})] \\ & \quad \times [(\bar{w}_i - E_i^* \bar{w}_i) + (E_i^* \bar{w}_i - E^* \bar{w}) + (E^* \bar{w} - \bar{w})]' D_T \\ &= \frac{1}{N} \sum_i (I_{1,i,NT} + I_{2,i,NT} + I_{3,NT}) (I_{1,i,NT} + I_{2,i,NT} + I_{3,NT})', \text{ say.} \quad (47) \end{aligned}$$

We complete the proof by showing the following:

$$\frac{1}{N} \sum_i I_{1,i,NT} I'_{1,i,NT} \rightarrow_p \Xi_1; \quad (48)$$

$$\frac{1}{N} \sum_i I_{2,i,NT} I'_{2,i,NT} \rightarrow_p \Xi_2; \quad (49)$$

$$I_{3,NT} \rightarrow_p 0; \quad (50)$$

and

$$\frac{1}{N} \sum_i I_{1,i,NT} I'_{2,i,NT}, \frac{1}{N} \sum_i I_{1,i,NT} I'_{3,NT}, \frac{1}{N} \sum_i I_{2,i,NT} I'_{3,NT} \rightarrow_p 0. \quad (51)$$

Proof of (48): Write

$$\frac{1}{N} \sum_i I_{1,i,NT} I'_{1,i,NT}$$

$$= \begin{pmatrix} (I_{1,1}) & (I_{1,21} & I_{1,22} & I_{1,31} & I_{1,32} & I_{1,33}) & 0 \\ \left(\begin{array}{c} I'_{1,21} \\ I'_{1,22} \\ I'_{1,31} \\ I'_{1,32} \\ I'_{1,33} \\ 0 \end{array} \right) & \left(\begin{array}{ccccc} I'_{21,21} & I'_{21,22} & I'_{21,31} & I'_{21,32} & I'_{21,33} \\ I'_{22,21} & I'_{22,22} & I'_{22,31} & I'_{22,32} & I'_{22,33} \\ I'_{21,31} & I'_{22,31} & I'_{31,31} & I'_{31,32} & I'_{31,32} \\ I'_{21,32} & I'_{22,32} & I'_{31,32} & I'_{32,32} & I'_{32,33} \\ I'_{21,33} & I'_{22,33} & I'_{31,33} & I'_{32,33} & I'_{33,33} \end{array} \right) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where the partition is made conformable to the size of

$$(x'_{1,it}, x'_{21,it}, x'_{22,it}, x'_{31,it}, x'_{32,it}, x'_{33,it}, z'_i)'$$

We now consider each element in $\frac{1}{N} \sum_i I_{1,i,NT} I'_{1,i,NT}$. For $I_{1,1}$, notice that by Lemma 15(a),

$$\begin{aligned} & \left\| \frac{1}{N} \sum_i (\bar{x}_{1,i} - E\bar{x}_{1,i}) (\bar{x}_{1,i} - E\bar{x}_{1,i})' \right\| \\ & \leq \frac{1}{N} \sum_i \|\bar{x}_{1,i} - E\bar{x}_{1,i}\|^2 \leq \frac{1}{N} \sum_i \frac{1}{T} \sum_t \|x_{1,it} - Ex_{1,it}\|^2 = O_p(1). \end{aligned}$$

Since each element in the diagonal matrix D_{1T} tends to zero, we have

$$\begin{aligned} I_{1,1} &= \frac{1}{N} \sum_i D_{1T} (\bar{x}_{1,i} - E\bar{x}_{1,i}) (\bar{x}_{1,i} - E\bar{x}_{1,i})' D_{1T} \\ &= D_{1T} O_p(1) D_{1T} = o_p(1). \end{aligned}$$

Next we consider the second diagonal block of $\frac{1}{N} \sum_i I_{1,i,NT} I'_{1,i,NT}$. Define

$$q_{iT} = \sqrt{T} \begin{pmatrix} \bar{x}_{2,i} - E\bar{x}_{2,i} \\ \bar{x}_{3,i} - E_{\mathcal{F}_{z_i}} \bar{x}_{3,i} \end{pmatrix};$$

$$Q_{iT} = q_{iT} q'_{iT}.$$

Then, by Lemma 15 (b) and (d), for some constant $q > 1$,

$$\sup_{i,T} E \|Q_{iT}\|^{2q} = \sup_{i,T} E \|q_{iT}\|^{4q} < \infty,$$

which verifies the condition (37) of Corollary 14. In consequence, from Corollary 14 and Assumption 8(i), we have

$$\begin{aligned} & \frac{T}{N} \sum_i \begin{pmatrix} (\bar{x}_{2,i} - E\bar{x}_{2,i})(\bar{x}_{2,i} - E\bar{x}_{2,i})' & (\bar{x}_{2,i} - E\bar{x}_{2,i})(\bar{x}_{3,i} - E_{\mathcal{F}_{z_i}} \bar{x}'_{3,i})' \\ (\bar{x}_{3,i} - E_{\mathcal{F}_{z_i}} \bar{x}_{3,i})(\bar{x}_{2,i} - E\bar{x}_{2,i})' & (\bar{x}_{3,i} - E_{\mathcal{F}_{z_i}} \bar{x}_{3,i})(\bar{x}_{3,i} - E_{\mathcal{F}_{z_i}} \bar{x}'_{3,i})' \end{pmatrix} \\ &= \frac{1}{N} \sum_i \frac{1}{T} \sum_t \sum_s \begin{pmatrix} (x_{2,it} - Ex_{2,it}) & (x_{2,it} - Ex_{2,it}) \\ \times (x_{2,is} - Ex_{2,is})' & \times (x_{3,is} - E_{\mathcal{F}_{z_i}} x_{3,is})' \\ (x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}) & (x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}) \\ \times (x_{2,is} - Ex_{2,is})' & \times (x_{3,is} - E_{\mathcal{F}_{z_i}} x_{3,is})' \end{pmatrix} \\ &= \frac{1}{N} \sum_i Q_{iT} \rightarrow_p \lim_N \frac{1}{N} \sum_i EQ_{iT} = \begin{pmatrix} \Gamma_{22} & \Gamma_{23} \\ \Gamma'_{23} & \Gamma_{33} \end{pmatrix}, \end{aligned}$$

as $(N, T \rightarrow \infty)$. Now, recall that

$$\begin{aligned} D_{2T} &= \text{diag}(D_{21T}, D_{22T}) = \text{diag}\left(I_{k_{21}}, \sqrt{T}I_{K_{22}}\right); \\ D_{3T} &= \text{diag}(D_{31T}, D_{32T}, D_{33T}) = \text{diag}\left(\sqrt{T}I_{k_{31}}, \sqrt{T}I_{k_{32}}, (T^{m_{h,33}})_h\right), \end{aligned}$$

where $0 \leq m_{h,33} < \frac{1}{2}$ for all $h = 1, \dots, k_{33}$. Therefore, as $(N, T \rightarrow \infty)$,

$$\begin{pmatrix} I'_{21,21} & I'_{21,22} & I'_{21,31} & I'_{21,32} & I'_{21,33} \\ I'_{21,22} & I'_{22,22} & I'_{22,31} & I'_{22,32} & I'_{22,33} \\ I'_{21,31} & I'_{22,31} & I'_{31,31} & I'_{31,32} & I'_{31,32} \\ I'_{21,32} & I'_{22,32} & I'_{31,32} & I'_{32,32} & I'_{32,33} \\ I'_{21,33} & I'_{22,33} & I'_{31,33} & I'_{32,33} & I'_{33,33} \end{pmatrix} \rightarrow_p \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \Gamma_{22,22} & \Gamma_{22,31} & \Gamma_{22,32} & 0 \\ 0 & \Gamma'_{22,31} & \Gamma_{31,31} & \Gamma_{31,32} & 0 \\ 0 & \Gamma'_{22,32} & \Gamma'_{31,32} & \Gamma_{32,32} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Finally, by the Cauchy-Schwarz inequality, the off-diagonal block component

$$\begin{pmatrix} I_{1,21} & I_{1,22} & I_{1,31} & I_{1,32} & I_{1,33} \end{pmatrix} \rightarrow_p 0,$$

as $(N, T \rightarrow \infty)$.

Proof of (49): Recall that

$$\begin{aligned} E_i^* \bar{w}_i &= (E\bar{x}'_{1,i}, E\bar{x}'_{2,i}, E_{\mathcal{F}_{z_i}} \bar{x}'_{3,i}, z'_i)' ; \\ E^* \bar{w} &= (E\bar{x}'_1, E\bar{x}'_2, E_{\mathcal{F}_z} \bar{x}'_3, \bar{z}')'. \end{aligned}$$

Write

$$\begin{aligned} & I_{2,i,NT} \\ &= \begin{pmatrix} D_{1T} (E\bar{x}_{1,i} - E\bar{x}_1) \\ E\bar{x}_{21,i} - E\bar{x}_{21} \\ 0 \\ D_{31T} \left((E_{\mathcal{F}_{z_i}} \bar{x}_{31,i} - E\bar{x}_{31,i}) - (E_{\mathcal{F}_z} \bar{x}_{31} - E\bar{x}_{31}) + (E\bar{x}_{31,i} - E\bar{x}_{31}) \right) \\ D_{32T} \left((E_{\mathcal{F}_{z_i}} \bar{x}_{32,i} - E\bar{x}_{32,i}) - (E_{\mathcal{F}_z} \bar{x}_{32} - E\bar{x}_{32}) + (E\bar{x}_{32,i} - E\bar{x}_{32}) \right) \\ D_{33T} \left((E_{\mathcal{F}_{z_i}} \bar{x}_{33,i} - E\bar{x}_{33,i}) - (E_{\mathcal{F}_z} \bar{x}_{33} - E\bar{x}_{33}) + (E\bar{x}_{33,i} - E\bar{x}_{33}) \right) \\ z_i - z \end{pmatrix}. \end{aligned}$$

To obtain the required result, we use Lemma 12. By Assumption 6, as $T \rightarrow \infty$, we have

$$I_{2,i,NT} \rightarrow I_{2,i,N} \text{ a.s. uniformly in } i, \quad (52)$$

where

$$I_{21,i,N} = \begin{pmatrix} H_1 \tilde{\Theta}_{1,i} \\ \tilde{\Theta}_{21,i} \\ 0 \\ 0 \\ H_{32} \left[g_{32,i}(z_i) - \frac{1}{N} \sum_i g_{32,i}(z_i) \right] + H_{32} \left[\mu_{g_{32,i}} - \frac{1}{N} \sum_i \mu_{g_{32,i}} \right] \\ H_{33} \left[g_{33,i}(z_i) - \frac{1}{N} \sum_i g_{33,i}(z_i) \right] + H_{33} \left[\mu_{g_{33,i}} - \frac{1}{N} \sum_i \mu_{g_{33,i}} \right] \\ z_i - \frac{1}{N} \sum_i z_i \end{pmatrix}.$$

Write

$$\begin{aligned}
& \frac{1}{N} \sum_i I_{2,i,N} I'_{2,i,N} \\
&= \begin{pmatrix} II_{1,1} & II_{1,21} & 0 & 0 & II_{1,32} & II_{1,33} & II_{1,z} \\ II'_{1,21} & II_{21,21} & 0 & 0 & II_{21,32} & II_{21,33} & II_{21,z} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ II'_{1,32} & II'_{21,32} & 0 & 0 & II_{32,32} & II_{32,33} & II_{32,z} \\ II'_{1,33} & II'_{21,33} & 0 & 0 & II'_{32,33} & II_{33,33} & II_{33,z} \\ II'_{1,z} & II'_{21,z} & 0 & 0 & II'_{32,z} & II'_{33,z} & II_{z,z} \end{pmatrix}. \quad (53)
\end{aligned}$$

By Assumption 7(i), as $N \rightarrow \infty$,

$$\begin{pmatrix} II_{1,1} & II_{1,21} \\ II'_{1,21} & II_{21,21} \end{pmatrix} \rightarrow \begin{pmatrix} \Gamma_{\Theta_1, \Theta_1} & \Gamma_{\Theta_1, \Theta_{21}} \\ \Gamma'_{\Theta_1, \Theta_{21}} & \Gamma_{\Theta_{21}, \Theta_{21}} \end{pmatrix}. \quad (54)$$

By the weak law of large numbers (WLLN) and Assumption 7, as $N \rightarrow \infty$, we have

$$\begin{aligned}
& \begin{pmatrix} II_{32,32} & II_{32,33} & II_{32,z} \\ II'_{32,33} & II_{33,33} & II_{33,z} \\ II'_{32,z} & II'_{33,z} & II_{z,z} \end{pmatrix} \\
& \rightarrow_p \begin{pmatrix} \Gamma_{g_{32}, g_{32}} + \Gamma_{\mu_{32}, \mu_{32}} & \Gamma_{g_{32}, g_{33}} + \Gamma_{\mu_{32}, \mu_{32}} & \Gamma_{g_{32}, z} \\ \Gamma'_{g_{32}, g_{33}} + \Gamma'_{\mu_{32}, \mu_{32}} & \Gamma_{g_{33}, g_{33}} + \Gamma_{\mu_{33}, \mu_{33}} & \Gamma_{g_{33}, z} \\ \Gamma'_{g_{32}, z} & \Gamma'_{g_{33}, z} & \Gamma_{z, z} \end{pmatrix}. \quad (55)
\end{aligned}$$

From Assumption 7 and WLLN with the assumption that $Eg_{32,i}(z_i) = Eg_{33,i}(z_i) = 0$, it follows that

$$\begin{pmatrix} II_{1,32} & II_{1,33} \\ II_{21,32} & II_{21,33} \end{pmatrix} \rightarrow_p \begin{pmatrix} \Gamma_{\Theta_1, \mu_{32}} & \Gamma_{\Theta_1, \mu_{33}} \\ \Gamma_{\Theta_{21}, \mu_{32}} & \Gamma_{\Theta_{21}, \mu_{33}} \end{pmatrix}, \quad (56)$$

as $N \rightarrow \infty$. In addition, by WLLN with Assumption 5,

$$\begin{pmatrix} II_{1,z} \\ II_{21,z} \end{pmatrix} \rightarrow_p \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (57)$$

as $N \rightarrow \infty$. The results (53)-(57) indicate that

$$\frac{1}{N} \sum_i I_{2,i,N} I'_{2,i,N} \rightarrow_p \Xi_2, \quad (58)$$

as $N \rightarrow \infty$. Finally, by Lemma 12, (52) and (58) imply (49).

Proof of (50): Notice that

$$E \left\| \frac{1}{N} \sum_i D_{1T} \frac{1}{T} \sum_t (x_{1,it} - Ex_{1,it}) \right\|^2$$

$$\begin{aligned}
&= \frac{1}{N^2} \sum_i E \left\| D_{1T} \frac{1}{T} \sum_t (x_{1,it} - E x_{1,it}) \right\|^2 \\
&\leq \frac{1}{N} \|D_{1T}\|^2 \sup_{i,t} E \left\| \frac{1}{T} \sum_t (x_{1,it} - E x_{1,it}) \right\|^2 \rightarrow 0, \tag{59}
\end{aligned}$$

where the last convergence result holds by Lemma 15(a). Similarly, by Lemma 15(b),

$$\begin{aligned}
&E \left\| \frac{1}{N} \sum_i D_{2kT} \frac{1}{T} \sum_t (x_{2k,it} - E x_{2k,it}) \right\|^2 \\
&\leq \frac{1}{N} \sup_{i,t} E \left\| \frac{1}{\sqrt{T}} \sum_t (x_{2k,it} - E x_{2k,it}) \right\|^2 \rightarrow 0, \tag{60}
\end{aligned}$$

for $k = 1, 2$. Finally, by Lemma 15(d), we have

$$\begin{aligned}
&E \left\| \frac{1}{N} \sum_i D_{3kT} \frac{1}{T} \sum_t (x_{3k,it} - E_{\mathcal{F}_{z_i}} x_{3k,it}) \right\|^2 \\
&\leq \frac{1}{N} \sup_{i,T} \left\| \frac{1}{\sqrt{T}} D_{3kT} \right\|^2 E \left\| \frac{1}{\sqrt{T}} \sum_t (x_{3k,it} - E_{\mathcal{F}_{z_i}} x_{3k,it}) \right\|^2 \\
&\rightarrow 0, \tag{61}
\end{aligned}$$

for $k = 1, 2, 3$. The results (59), (60) and (61) imply that $I_{3,NT} \rightarrow_p 0$.

Proof of (51): Notice that by (48) and (50),

$$\left\| \left(\frac{1}{N} \sum_i I_{1,i,NT} \right) I'_{3,NT} \right\|^2 = \left\| \frac{1}{N} \sum_i I_{1,i,NT} \right\|^2 \|I_{3,NT}\|^2 = O_p(1) o_p(1);$$

$$\left\| \left(\frac{1}{N} \sum_i I_{2,i,NT} \right) I'_{3,NT} \right\|^2 = \left\| \frac{1}{N} \sum_i I_{2,i,NT} \right\|^2 \|I_{3,NT}\|^2 = O_p(1) o_p(1).$$

Thus, as $(N, T \rightarrow \infty)$,

$$\left(\frac{1}{N} \sum_i I_{1,i,NT} \right) I'_{3,NT}, \quad \left(\frac{1}{N} \sum_i I_{2,i,NT} \right) I'_{3,NT} \rightarrow_p 0.$$

We now consider the $(k, l)^{th}$ term of $\frac{1}{N} \sum_i I_{1,i,NT} I'_{2,i,NT}$. By the Cauchy-Schwarz inequality,

$$\left[\left(\frac{1}{N} \sum_i I_{1,i,NT} I'_{2,i,NT} \right)_{k,l} \right]^2 \leq \left(\frac{1}{N} \sum_i [(I_{1,i,NT})_k]^2 \right) \left(\frac{1}{N} \sum_i [(I_{2,i,NT})_l]^2 \right),$$

where $A_{k,l}$ and a_k denote the $(k, l)^{th}$ and the k^{th} elements of matrix A and vector a , respectively. In view of (48) and (49) (the limits of $\frac{1}{N} \sum_i I_{1,i,NT} I'_{1,i,NT}$ and $\frac{1}{N} \sum_i I_{2,i,NT} I'_{2,i,NT}$), we can see that all of the elements in $\frac{1}{N} \sum_i I_{1,i,NT} I'_{2,i,NT}$ converge to zero, except for

$$\frac{1}{N} \sum_i D_{32T} (\bar{x}_{32,i} - E_{\mathcal{F}_{z_i}} \bar{x}_{32,i}) (E_{\mathcal{F}_{z_i}} \bar{x}_{32,i} - E_{\mathcal{F}_z} \bar{x}_{32})' D_{32T}.$$

Thus, we can complete the proof by showing that this term converges in probability to a conformable zero matrix.

Let

$$\begin{aligned} Q_{1,iT} &= D_{32T} (\bar{x}_{32,i} - E_{\mathcal{F}_{z_i}} \bar{x}_{32,i}); \\ Q_{2,iT} &= D_{32T} (E_{\mathcal{F}_{z_i}} \bar{x}_{32,i} - E_{\mathcal{F}_z} \bar{x}_{32}); \\ Q_{2,i} &= H_{32} \left(g_{32,i}(z_i) - \frac{1}{N} \sum_i g_{32,i}(z_i) + \mu_{g_{32,i}} - \frac{1}{N} \sum_i \mu_{g_{32,i}} \right). \end{aligned}$$

By the Cauchy-Schwarz inequality and Lemma 12,

$$\begin{aligned} & \left\| \frac{1}{N} \sum_i Q_{1,iT} (Q_{2,iT} - Q_{2,i})' \right\|^2 \\ & \leq \left(\frac{1}{N} \sum_i \|Q_{1,iT}\|^2 \right) \left(\frac{1}{N} \sum_i \|Q_{2,iT} - Q_{2,i}\|^2 \right) \\ & \leq \left(\frac{1}{N} \sum_i \|Q_{1,iT}\|^2 \right) \sup_i \|Q_{2,iT} - Q_{2,i}\|^2 \\ & = O_p(1) o(1) = o_p(1), \end{aligned}$$

where the last line holds since

$$E \left(\frac{1}{N} \sum_i \|Q_{1,iT}\|^2 \right) < M$$

by Lemma 15(d) and by Assumption 6. Thus,

$$\frac{1}{N} \sum_i Q_{1,iT} Q'_{2,iT} = \frac{1}{N} \sum_i Q_{1,iT} Q'_{2,i} + o_p(1). \quad (62)$$

Next, notice that $E Q_{1,iT} Q'_{2,i} = 0$. Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned} & E \left\| \left(\frac{1}{N} \sum_i \text{vec}(Q_{1,iT} Q'_{2,i}) \right) \right\|^2 \\ & = E \left\| \left(\frac{1}{N} \sum_i (Q_{2,i} \otimes Q_{1,iT}) \right) \right\|^2 = \frac{1}{N^2} \sum_i E (Q'_{2,i} Q_{2,i} Q'_{1,iT} Q_{1,iT}) \\ & \leq \frac{1}{N^2} \sum_i [E \|Q_{2,i}\|^4 E \|Q_{1,iT}\|^4]^{1/2}. \end{aligned}$$

By Lemma 15(d),

$$\sup_{i,T} E \|Q_{1,iT}\|^4 < M,$$

and by Assumption 6(iii) and (v),

$$E \|Q_{2,i}\|^4 < M.$$

Thus,

$$\frac{1}{N^2} \sum_i \left[E \|Q_{2,i}\|^4 E \|Q_{1,iT}\|^4 \right]^{1/2} \rightarrow 0,$$

which implies

$$E \left\| \left(\frac{1}{N} \sum_i \text{vec} (Q_{1,iT} Q'_{2,i}) \right) \right\|^2 \rightarrow 0.$$

By the Chebychev's inequality, then, we have

$$\frac{1}{N} \sum_i Q_{1,iT} Q'_{2,i} \rightarrow_p 0,$$

as $(N, T \rightarrow \infty)$. Finally, in view of (62) we have the desired result that as $(N, T \rightarrow \infty)$,

$$\frac{1}{N} \sum_i Q_{1,iT} Q'_{2,iT} \rightarrow_p 0. \blacksquare$$

Part (b)

From (47), we write

$$\begin{aligned} & \sup_{N,T} \sup_{1 \leq i \leq N} E \|D_T (\bar{w}_i - \bar{w})\|^4 \\ &= \sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{1,i,NT} + I_{2,i,NT} + I_{3,i,NT}\|^4 \\ &\leq M_1 \left(\begin{aligned} & \sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{1,i,NT}\|^4 \\ & + \sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{2,i,NT}\|^4 \\ & + \sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{3,i,NT}\|^4 \end{aligned} \right), \end{aligned} \quad (63)$$

for some constant M_1 . Thus, we can complete the proof by showing that each of the three terms in the right-hand side of the inequality (63) is bounded.

For some constant M_2 ,

$$\begin{aligned} & \sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{1,i,NT}\|^4 \\ &\leq M_2 \left(\begin{aligned} & \sup_{i,T} E \|D_{1T} (\bar{x}_{1,i} - E\bar{x}_{1,i})\|^4 \\ & + \sup_{i,T} E \left\| \frac{D_{2T}}{\sqrt{T}} \sqrt{T} (\bar{x}_{2,i} - E\bar{x}_{2,i}) \right\|^4 \\ & + \sup_{i,T} E \left\| \frac{D_{3T}}{\sqrt{T}} \sqrt{T} (\bar{x}_{3,i} - E_{\mathcal{F}_{z_i}} \bar{x}_{3,i}) \right\|^4 \end{aligned} \right). \end{aligned} \quad (64)$$

By Lemma 15 and the definitions of D_{1T} , D_{2T} , and D_{3T} , each term in (64) is finite. Thus,

$$\sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{1,i,NT}\|^4 < \infty.$$

Next, we consider the second term in the right-hand side of the inequality (63). For some constant M_3 ,

$$\begin{aligned} & \sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{2,i,NT}\|^4 \\ \leq & M_3 \left(\begin{aligned} & \sup_{N,T} \sup_{1 \leq i \leq N} \|D_{1T}(E\bar{x}_{1,i} - E\bar{x}_1)\|^4 \\ & + \sup_{N,T} \sup_{1 \leq i \leq N} \|E\bar{x}_{21,i} - E\bar{x}_{21}\|^4 \\ & + \sup_{i,T} E \|D_{3T}(E_{\mathcal{F}_{z_i}}\bar{x}_{3,i} - E\bar{x}_{3,i})\|^4 \\ & + \sup_{N,T} \sup_{1 \leq i \leq N} E \|D_{3T}(E_{\mathcal{F}_z}\bar{x}_3 - E\bar{x}_3)\|^4 \\ & + \sup_{N,T} \sup_{1 \leq i \leq N} \|D_{3T}(E\bar{x}_{3,i} - E\bar{x}_3)\|^4 \\ & + \sup_N \sup_{1 \leq i \leq N} E \|z_i - \bar{z}\|^4 \end{aligned} \right). \quad (65) \end{aligned}$$

For the required result, we need to show that all of the terms in the right-hand side of the inequality (65) are bounded. Notice that for some finite constant M_4 ,

$$\begin{aligned} & \|D_{1T}(E\bar{x}_{1,i} - E\bar{x}_1)\|^4 \\ \rightarrow & \left\| H_1 \left(\Theta_{1,i} - \frac{1}{N} \sum_i \Theta_{1,i} \right) \right\|^4 \quad (\text{uniformly in } i \text{ as } T \rightarrow \infty) \\ \leq & 2 \|H_1\| \sup_i \|\Theta_{1,i}\| < M_4, \end{aligned}$$

and

$$\begin{aligned} & \|E\bar{x}_{21,i} - E\bar{x}_{21}\|^4 \\ \rightarrow & \left\| \Theta_{21,i} - \frac{1}{N} \sum_i \Theta_{21,i} \right\|^4 \quad (\text{uniformly in } i \text{ as } T \rightarrow \infty) \\ \leq & 2 \sup_i \|\Theta_{21,i}\|^4 < M_4. \end{aligned}$$

So,

$$\sup_{N,T} \sup_{1 \leq i \leq N} \|D_{1T}(E\bar{x}_{1,i} - E\bar{x}_1)\|^4, \sup_{N,T} \sup_{1 \leq i \leq N} \|E\bar{x}_{21,i} - E\bar{x}_{21}\|^4 < \infty. \quad (66)$$

By Assumption 6(iii), the followings hold uniformly in i almost surely as $T \rightarrow \infty$:

$$\begin{aligned} & \|D_{31T}(E_{\mathcal{F}_{z_i}}\bar{x}_{31,i} - E\bar{x}_{31,i})\|^4 \rightarrow 0, \\ & \|D_{32T}(E_{\mathcal{F}_{z_i}}\bar{x}_{31,i} - E\bar{x}_{31,i})\|^4 \rightarrow \|H_{32g_{32,i}}(z_i)\|^4, \\ & \|D_{33T}(E_{\mathcal{F}_{z_i}}\bar{x}_{33,i} - E\bar{x}_{33,i})\|^4 \rightarrow \|H_{33g_{33,i}}(z_i)\|^4. \end{aligned}$$

Notice that since $\|D_{3T}(E_{\mathcal{F}_{z_i}} x_{3,it} - E x_{3,it})\|^4 \leq \tilde{\tau}(r)^4 \tilde{G}(z_i)^4$ by Assumption 6(iv), we have $\sup_i \|D_{3kT}(E_{\mathcal{F}_{z_i}} \bar{x}_{3k,i} - E \bar{x}_{3k,i})\|^4 \leq \left(\int_0^1 \tilde{\tau}(r)^4 dr\right) \sup_i \tilde{G}(z_i)^4$ for $k = 1, 2$, and 3. Therefore, by Lemma 10, we have

$$\begin{aligned} & E \|D_{3kT}(E_{\mathcal{F}_{z_i}} \bar{x}_{3k,i} - E \bar{x}_{3k,i})\|^4 \\ \rightarrow & E \left\| g_{3k,i}(z_i) \int_0^1 \tau_{3k}(r) dr \right\|^4 \quad (\text{uniformly in } i) \\ \leq & \left(E \sup_i \|g_{3k,i}(z_i)\|^4 \right) \left(\left\| \int_0^1 \tau_{3k}(r) dr \right\|^4 \right) < \infty, \text{ for } k = 1, 2, 3, \end{aligned}$$

where $\tau_{31}(r) = 0$. So,

$$\sup_{i,T} E \|D_{3T}(E_{\mathcal{F}_{z_i}} \bar{x}_{3,i} - E \bar{x}_{3,i})\|^4 < \infty. \quad (67)$$

Similarly, it follows that

$$\sup_{N,T} \sup_{1 \leq i \leq N} E \|D_{3T}(E_{\mathcal{F}_z} \bar{x}_3 - E \bar{x}_{3,i})\|^4 < \infty. \quad (68)$$

In addition, notice that

$$\begin{aligned} & \sup_{N,T} \sup_{1 \leq i \leq N} \|D_{3T}(E \bar{x}_{3,i} - E \bar{x}_3)\| \\ \leq & \sup_{N,T} \sup_{1 \leq i \leq N} \left\| \begin{pmatrix} D_{31T}(E \bar{x}_{31,i} - E \bar{x}_{31}) \\ D_{32T}(E \bar{x}_{32,i} - E \bar{x}_{32}) - H_{32} \tilde{\mu}_{g_{32,i}} \\ D_{33T}(E \bar{x}_{33,i} - E \bar{x}_{33}) - H_{33} \tilde{\mu}_{g_{33,i}} \end{pmatrix} \right\| \\ & + \sup_N \sup_{1 \leq i \leq N} \left\| \begin{pmatrix} 0 \\ H_{32} \tilde{\mu}_{g_{32,i}} \\ H_{33} \tilde{\mu}_{g_{33,i}} \end{pmatrix} \right\|. \end{aligned} \quad (69)$$

By Assumption 6(v), as $T \rightarrow \infty$,

$$\begin{aligned} & \sup_N \sup_{1 \leq i \leq N} \left\| \begin{pmatrix} D_{31T}(E \bar{x}_{31,i} - E \bar{x}_{31}) \\ D_{32T}(E \bar{x}_{32,i} - E \bar{x}_{32}) - H_{32} \tilde{\mu}_{g_{32,i}} \\ D_{33T}(E \bar{x}_{33,i} - E \bar{x}_{33}) - H_{33} \tilde{\mu}_{g_{33,i}} \end{pmatrix} \right\| \\ = & \sup_N \sup_{1 \leq i \leq N} \frac{1}{N} \sum_{j=1}^N \left\| \begin{pmatrix} D_{31T}(E \bar{x}_{31,i} - E \bar{x}_{31,j}) \\ D_{32T}(E \bar{x}_{32,i} - E \bar{x}_{32,j}) - H_{32} \left(\mu_{g_{32,i}} - \mu_{g_{32,j}} \right) \\ D_{33T}(E \bar{x}_{33,i} - E \bar{x}_{33,j}) - H_{33} \left(\mu_{g_{33,i}} - \mu_{g_{33,j}} \right) \end{pmatrix} \right\| \\ \leq & \sup_{i,j} \left\| \begin{pmatrix} D_{31T}(E \bar{x}_{31,i} - E \bar{x}_{31,j}) \\ D_{32T}(E \bar{x}_{32,i} - E \bar{x}_{32,j}) - H_{32} \left(\mu_{g_{32,i}} - \mu_{g_{32,j}} \right) \\ D_{33T}(E \bar{x}_{33,i} - E \bar{x}_{33,j}) - H_{33} \left(\mu_{g_{33,i}} - \mu_{g_{33,j}} \right) \end{pmatrix} \right\| \rightarrow 0. \end{aligned}$$

Therefore, the first term (69) is finite. The second term in (69) is also finite since $\sup_i \|\mu_{g32,i}\|, \sup_i \|\mu_{g33,i}\| < \infty$ as assumed in Assumption 6(v). Therefore,

$$\sup_{N,T} \sup_{1 \leq i \leq N} \|D_{3T} (E\bar{x}_{3,i} - E\bar{x}_3)\| < \infty.$$

In addition, by Assumption 5,

$$\sup_N \sup_{1 \leq i \leq N} E \|z_i - \bar{z}\|^4 < \infty. \quad (70)$$

Therefore, from (66) – (70) we have

$$\sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{2,i,NT}\|^4 < \infty.$$

Finally, since $I_{3,NT} = \frac{1}{N} \sum_i I_{1,i,NT}$,

$$\begin{aligned} & \sup_{N,T} E \|I_{3,NT}\|^4 \\ & \leq \sup_{N,T} E \left(\frac{1}{N} \sum_i \|I_{1,i,NT}\| \right)^4 \quad (\text{by triangle inequality}) \\ & \leq \sup_{N,T} \sup_{1 \leq i \leq N} E \|I_{1,i,NT}\|^4 \quad (\text{by Holder's inequality}) \\ & < \infty. \quad \blacksquare \end{aligned}$$

Part (c)

Recall that under Assumption 11, conditional on \mathcal{F}_w , u_i is independently distributed with mean $\frac{\tilde{w}'_i D_T}{\sqrt{N}} \lambda$, variance σ_u^2 , and $\kappa_u^4 = E_{\mathcal{F}_w} (u_i - E_{\mathcal{F}_w} u_i)^4 < \infty$, where λ is a nonrandom vector in \mathbb{R}^{k+g} . So, $E(D_T \tilde{w}_i u_i) = \frac{1}{\sqrt{N}} E(D_T \tilde{w}_i \tilde{w}'_i D_T) \lambda$. Define $Q_{i,T} = D_T \tilde{w}_i (u_i - E_{\mathcal{F}_w} u_i)$; and let $\iota \in \mathbb{R}^{k+g}$ with $\|\iota\| = 1$. Then, we can complete the proof by showing that as $(N, T \rightarrow \infty)$,

$$\frac{1}{\sqrt{N}} \sum_i \iota' Q_{i,T} \Rightarrow N(0, \sigma_u^2 \iota' \Xi \iota). \quad (71)$$

This is so because this condition, together with the Cramer-Wold device, Assumption 11 and Part (a), implies that

$$\begin{aligned} & \frac{1}{\sqrt{N}} \sum_i D_T \tilde{w}_i u_i = \frac{1}{\sqrt{N}} \sum_i D_T \tilde{w}_i E_{\mathcal{F}_w} u_i + \frac{1}{\sqrt{N}} \sum_i Q_{i,T} \\ & = \frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{w}'_i D_T \lambda + \frac{1}{\sqrt{N}} \sum_i Q_{i,T} \\ & \Rightarrow N(\Xi \lambda, \sigma_u^2 \Xi). \end{aligned}$$

Now we start the proof of (71). Let $s_{i,T}^2 = E (\iota' Q_{i,T})^2$ and $S_{N,T}^2 = \sum_i s_{i,T}^2$. Under Assumption 11, we have

$$\begin{aligned} s_{i,T}^2 &= E (\iota' Q_{i,T})^2 \\ &= \iota' E \left[\left(E_{\mathcal{F}_w} (u_i - E_{\mathcal{F}_w} u_i)^2 \right) D_T \tilde{w}_i \tilde{w}_i' D_T \right] \iota \\ &= \sigma_u^2 E [\iota' D_T \tilde{w}_i \tilde{w}_i' D_T \iota]. \end{aligned}$$

By Part (a),

$$\frac{1}{N} \sum_i \iota' D_T \tilde{w}_i \tilde{w}_i' D_T \iota \rightarrow_p \iota' \Xi \iota > 0,$$

as $(N, T \rightarrow \infty)$. By Part (b),

$$\sup_{N,T} \sup_{1 \leq i \leq N} E \|\iota' D_T \tilde{w}_i\|^2 1_{\{\|\iota' D_T \tilde{w}_i\| > M\}} \rightarrow 0,$$

as $M \rightarrow \infty$, and so $\|\iota' D_T \tilde{w}_i\|^2$ is uniformly integrable in N, T . Then, by Vitali's lemma, it follows that

$$\frac{1}{N} S_{N,T}^2 \rightarrow \sigma_u^2 \iota' \Xi \iota > 0,$$

as $(N, T \rightarrow \infty)$. Thus, for our required result of (71), it is sufficient to show

$$\sum_i \frac{\iota' Q_{i,T}}{S_{N,T}} \Rightarrow N(0, 1), \quad (72)$$

as $(N, T \rightarrow \infty)$. Let $P_{i,NT} = \frac{\iota' Q_{i,T}}{S_{N,T}}$. Note that, under Assumption 9, $E(P_{i,NT}) = 0$ and $\sum_i E P_{i,NT}^2 = 1$. According to Theorem 2 of Phillips and Moon (1999), the weak convergence in (72) follows if we can show that

$$\sum_i E P_{i,NT}^2 1_{\{|P_{i,NT}^2| > \varepsilon\}} \rightarrow 0 \text{ for all } \varepsilon > 0, \quad (73)$$

as $(N, T \rightarrow \infty)$. Since

$$\begin{aligned} &\sup_{N,T} \sup_{1 \leq i \leq N} E \|Q_{i,T}\|^4 \\ &\leq \|\iota\|^8 \sup_{N,T} \sup_{1 \leq i \leq N} E \left(\|D_T \tilde{w}_i\|^4 \left(E_{\mathcal{F}_w} (u_i - E_{\mathcal{F}_w} u_i)^4 \right) \right) \\ &\leq \kappa_u^4 \sup_{N,T} \sup_{1 \leq i \leq N} E \|D_T \tilde{w}_i\|^4 < \infty, \end{aligned}$$

the Lindeberg-Feller condition (73) follows, and we have all the desired results. \blacksquare

Part (d)

Since \tilde{v}_i is independent of \tilde{w}_i , we have

$$\begin{aligned}
& E \left\| \frac{1}{\sqrt{N}} \sum_i D_T \tilde{w}_i \tilde{v}_i \right\|^2 = E \left\| \frac{1}{\sqrt{N}} \sum_i D_T \tilde{w}_i \bar{v}_i \right\|^2 \\
&= \frac{\sigma_v^2}{T} \text{tr} \left(\frac{1}{N} \sum_i E D_T \tilde{w}_i \tilde{w}_i' D_T \right) \\
&\leq \frac{\sigma_v^2}{T} \sup_{N,T} \sup_{1 \leq i \leq N} E \|D_T \tilde{w}_i\|^2 \rightarrow 0,
\end{aligned}$$

as $(N, T \rightarrow \infty)$, where the last convergence holds because Part (b) warrants

$$\sup_{N,T} \sup_{1 \leq i \leq N} E \|D_T \tilde{w}_i\|^2 < \infty. \blacksquare$$

Proof of Lemma 2

Part (a)

By definition, we have

$$\begin{aligned}
& \frac{1}{N} \sum_i \frac{1}{T} \sum_t G_{xT} (x_{it} - \bar{x}_i) (x_{it} - \bar{x}_i)' G_{xT} \\
&= \frac{1}{N} \sum_i \frac{1}{T} \sum_t G_{xT} (x_{it} - E_i x_{it} + E_i x_{it} - E_i \bar{x}_i + E_i \bar{x}_i - \bar{x}_i) \\
&\quad \times (x_{it} - E_i x_{it} + E_i x_{it} - E_i \bar{x}_i + E_i \bar{x}_i - \bar{x}_i)' G_{xT}, \\
&= \frac{1}{N} \sum_i \frac{1}{T} \sum_t (III_{1,iT} + III_{2,iT} + III_{3,iT}) (III_{1,iT} + III_{2,iT} + III_{3,iT})', \text{ say.}
\end{aligned}$$

First, we show that

$$\begin{aligned}
& \frac{1}{N} \sum_i \frac{1}{T} \sum_t III_{1,iT} III_{1,iT}' \\
&= \frac{1}{N} \sum_i \frac{1}{T} \sum_t \begin{pmatrix} D_{1T} (x_{1,it} - E x_{1,it}) \\ x_{2,it} - E x_{2,it} \\ x_{3,it} - E x_{3,it} \end{pmatrix} \begin{pmatrix} D_{1T} (x_{1,it} - E x_{1,it}) \\ x_{2,it} - E x_{2,it} \\ x_{3,it} - E x_{3,it} \end{pmatrix}' \\
&\rightarrow_p \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Phi_{22} & \Phi_{23} \\ 0 & \Phi_{23}' & \Phi_{33} \end{pmatrix}. \tag{74}
\end{aligned}$$

For this, set

$$Q_{iT} = \frac{1}{T} \sum_t \begin{pmatrix} D_{1T} (x_{1,it} - E x_{1,it}) \\ x_{2,it} - E x_{2,it} \\ x_{3,it} - E x_{3,it} \end{pmatrix} \begin{pmatrix} D_{1T} (x_{1,it} - E x_{1,it}) \\ x_{2,it} - E x_{2,it} \\ x_{3,it} - E x_{3,it} \end{pmatrix}'.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} \|Q_{iT}\|^2 &\leq \left[\frac{1}{T} \sum_t \left\| \begin{pmatrix} D_{1T}(x_{1,it} - Ex_{1,it}) \\ x_{2,it} - Ex_{2,it} \\ x_{3,it} - Ex_{3,it} \end{pmatrix} \right\|^2 \right]^2 \\ &\leq M_1 \left(\frac{1}{T} \sum_t \left\{ \|D_{1T}(x_{1,it} - Ex_{1,it})\|^4 + \|x_{2,it} - Ex_{2,it}\|^4 \right. \right. \\ &\quad \left. \left. + \|x_{3,it} - Ex_{3,it}\|^4 \right\} \right). \end{aligned}$$

Notice that by Assumptions 2(i), 3(iii),

$$\sup_{i,t} E \|x_{1,it} - Ex_{1,it}\|^4, \quad \sup_{i,t} E \|x_{2,it} - Ex_{2,it}\|^4 < \infty, \quad (75)$$

and by Assumption 4(iii) and Assumption 6(iii),

$$\begin{aligned} &\sup_{i,t} E \|x_{3,it} - Ex_{3,it}\|^4 \\ &\leq M_1 \left(\sup_{i,t} E_{\mathcal{F}_{z_i}} \|x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}\|^4 \right) + M_1 E \|E_{\mathcal{F}_{z_i}} x_{3,it} - Ex_{3,it}\|^4 \\ &< \infty. \end{aligned} \quad (76)$$

Thus, we have

$$\sup_{i,T} E \|Q_{iT}\|^2 < \infty.$$

From this,

$$\sup_{i,T} E \|Q_{iT}\| \mathbb{1} \{ \|Q_{iT}\| > M \} \leq \frac{\sup_{i,T} E \|Q_{iT}\|^2}{M} \rightarrow 0,$$

as $M \rightarrow \infty$. By Corollary 14 and Assumption 8(ii), then,

$$\begin{aligned} &\frac{1}{N} \sum_i \frac{1}{T} \sum_t III_{1,iT} III'_{1,iT} \\ &= \frac{1}{N} \sum_i Q_{iT} \rightarrow_p \lim_{N,T} \frac{1}{N} \sum_i EQ_{iT} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Phi_{22} & \Phi_{23} \\ 0 & \Phi'_{23} & \Phi_{33} \end{pmatrix}. \end{aligned}$$

Next, by Assumption 6(i) and (ii), Lemma 11, and Lemma 12, we have

$$\begin{aligned} &\frac{1}{N} \sum_i \frac{1}{T} \sum_t III_{2,iT} III'_{2,iT} \\ &= \frac{1}{N} \sum_i \frac{1}{T} \sum_t \begin{pmatrix} D_{1T}(Ex_{1,it} - E\bar{x}_{1,i}) \\ Ex_{2,it} - E\bar{x}_{2,i} \\ Ex_{3,it} - E\bar{x}_{3,i} \end{pmatrix} \begin{pmatrix} D_{1T}(Ex_{1,it} - E\bar{x}_{1,i}) \\ Ex_{2,it} - E\bar{x}_{2,i} \\ Ex_{3,it} - E\bar{x}_{3,i} \end{pmatrix}' \\ &\rightarrow \begin{pmatrix} \int_0^1 (\tau_1 - \int \tau_1) (\lim_N \frac{1}{N} \sum_i \Theta_{1,i} \Theta'_{1,i}) (\tau_1 - \int \tau_1)' dr & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (77)$$

In addition, we have

$$\begin{aligned}
& E \left\| \frac{1}{N} \sum_i \frac{1}{T} \sum_t III_{3,iT} III'_{3,iT} \right\| \\
&= E \left\| \frac{1}{N} \sum_i \frac{1}{T} \sum_t \begin{pmatrix} D_{1T}(x_{1,it} - Ex_{1,it}) \\ (x_{2,it} - Ex_{2,it}) \\ (x_{3,it} - Ex_{3,it}) \end{pmatrix} \right\|^2 \\
&\leq \frac{1}{N^2} \sum_i \frac{1}{T} \sum_t E \left\| \begin{pmatrix} D_{1T}(x_{1,it} - Ex_{1,it}) \\ (x_{2,it} - Ex_{2,it}) \\ (x_{3,it} - Ex_{3,it}) \end{pmatrix} \right\|^4 \\
&\rightarrow_p 0, \tag{78}
\end{aligned}$$

as $(N, T \rightarrow \infty)$, as shown in (75) and (76). Thus,

$$\frac{1}{N} \sum_i \frac{1}{T} \sum_t III_{3,iT} III'_{3,iT} \rightarrow_p 0.$$

From the Cauchy-Schwarz inequality, (74), (77) and (78), we have

$$\begin{aligned}
\frac{1}{N} \sum_i \frac{1}{T} \sum_t III_{1,iT} III'_{2,iT} &\rightarrow_p 0; \quad \frac{1}{N} \sum_i \frac{1}{T} \sum_t III_{1,iT} III'_{3,iT} \rightarrow_p 0; \\
\frac{1}{N} \sum_i \frac{1}{T} \sum_t III_{2,iT} III'_{3,iT} &\rightarrow_p 0,
\end{aligned}$$

as $(N, T \rightarrow \infty)$, Combining all of these, we have

$$\begin{aligned}
& \frac{1}{N} \sum_i \frac{1}{T} \sum_t G_{xT}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' G_{xT} \\
&\rightarrow_p \begin{pmatrix} \int_0^1 (\tau_1 - \int \tau_1) (\lim_N \frac{1}{N} \sum_i \Theta_{1,i} \Theta'_{1,i}) (\tau_1 - \int \tau_1)' dr & 0 & 0 \\ 0 & \Phi_{22} & \Phi_{23} \\ 0 & \Phi'_{23} & \Phi_{33} \end{pmatrix} \\
&\equiv \Psi_x.
\end{aligned}$$

as $(N, T \rightarrow \infty)$.

Part (b)

First, let $Q_{i,T} = \frac{1}{\sqrt{T}} \sum_t G_{xT}(x_{it} - \bar{x}_i) v_{it}$; and let $\iota \in \mathbb{R}^k$ with $\|\iota\| = 1$. If we can show that as $(N, T \rightarrow \infty)$,

$$\frac{1}{\sqrt{N}} \sum_i \iota' Q_{i,T} \Rightarrow N(0, \sigma_v^2 \iota' \Psi_x \iota), \tag{79}$$

then, the Cramer-Wold device implies our desired result. Now let $s_{i,T}^2 = E(\iota' Q_{i,T})^2$ and $S_{NT}^2 = \sum_i s_{i,T}^2$. Using similar arguments for (74) – (78), it

is possible to show that

$$\begin{aligned} \frac{1}{N} S_{NT}^2 &= \sigma_v^2 l' E \left(\frac{1}{N} \sum_i \frac{1}{T} \sum_t G_{xT} (x_{it} - \bar{x}_i) (x_{it} - \bar{x}_i)' G_{xT} \right) l_k \\ &\rightarrow \sigma_v^2 l' \Psi_x l > 0, \end{aligned} \quad (80)$$

as $(N, T \rightarrow \infty)$. So, the asymptotic normality in (79) holds if

$$\sum_i \frac{l' Q_{i,T}}{S_{NT}} \Rightarrow N(0, 1), \quad (81)$$

as $(N, T \rightarrow \infty)$. Let $P_{i,NT} = \frac{l' Q_{i,T}}{S_{NT}}$. Then, $E(P_{i,NT}) = 0$ and $\sum_i E P_{i,NT}^2 = 1$. Thus, by the central limit theorem of the double indexed process (*e.g.*, see Theorem 2 in Phillips and Moon, 1999), we can claim that (81) holds, if we can show that

$$\sum_i E P_{i,NT}^2 1\{|P_{i,NT}^2| > \varepsilon\} \rightarrow 0 \text{ for all } \varepsilon > 0, \quad (82)$$

as $(N, T \rightarrow \infty)$.

Now, in view of (80), condition (82) holds if

$$\sup_{i,T} E \|l' Q_{i,T}\|^4 \leq \sup_{i,T} E \|Q_{i,T}\|^4 < \infty. \quad (83)$$

Note for some constant M_1 that

$$\begin{aligned} &\sup_{i,T} E \|Q_{i,T}\|^4 \\ &= \sup_{i,T} E \left\| \frac{1}{\sqrt{T}} \sum_t G_{x,T} (x_{it} - \bar{x}_i) v_{it} \right\|^4 \\ &= \sup_{i,T} E \left[\text{tr} \left(\frac{1}{T^2} \sum_{t,s,p,q} \left(G_{x,T} (x_{it} - \bar{x}_i) (x_{is} - \bar{x}_i)' G_{x,T} \right) \right. \right. \\ &\quad \left. \left. \otimes G_{x,T} (x_{ip} - \bar{x}_i) (x_{iq} - \bar{x}_i)' G_{x,T} \right) \right. \\ &\quad \left. \times E(v_{it} v_{is} v_{ip} v_{iq}) \right] \\ &\leq M_1 \sup_{i,T} \left| E \left(\frac{1}{T^2} \sum_{t,s} \text{tr} \left(G_{x,T} (x_{it} - \bar{x}_i) (x_{is} - \bar{x}_i)' G_{x,T} \right) \right) \right. \\ &\quad \left. \times E(v_{it}^2 v_{is}^2) \right| \\ &\quad + M_1 \sup_{i,T} \left| E \left(\frac{1}{T^2} \sum_{t,s} \text{tr} \left(G_{x,T} (x_{it} - \bar{x}_i) (x_{is} - \bar{x}_i)' G_{x,T} \right) \right) \right. \\ &\quad \left. \otimes G_{x,T} (x_{is} - \bar{x}_i) (x_{it} - \bar{x}_i)' G_{x,T} \right) \right. \\ &\quad \left. \times E(v_{it}^2 v_{is}^2) \right| \\ &\quad + M_1 \sup_{i,T} \left| E \left(\frac{1}{T^2} \sum_{t,s} \text{tr} \left(G_{x,T} (x_{it} - \bar{x}_i) (x_{it} - \bar{x}_i)' G_{x,T} \right) \right) \right. \\ &\quad \left. \otimes G_{x,T} (x_{is} - \bar{x}_i) (x_{is} - \bar{x}_i)' G_{x,T} \right) \right. \\ &\quad \left. \times E(v_{it}^2 v_{is}^2) \right|. \end{aligned} \quad (84)$$

Using the fact that $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$ and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} & \frac{1}{T^2} \sum_t \sum_s \text{tr} \begin{pmatrix} G_{x,T}(x_{it} - \bar{x}_i)(x_{is} - \bar{x}_i)' G_{x,T} \\ \otimes G_{x,T}(x_{it} - \bar{x}_i)(x_{is} - \bar{x}_i)' G_{x,T} \end{pmatrix} \\ &= \frac{1}{T^2} \sum_t \sum_s \left\{ \text{tr} [G_{x,T}(x_{it} - \bar{x}_i)(x_{is} - \bar{x}_i)' G_{x,T}] \right\}^2 \\ &\leq \left[\frac{1}{T} \sum_t \|G_{x,T}(x_{it} - \bar{x}_i)\|^2 \right]^2. \end{aligned}$$

Similarly,

$$\begin{aligned} & \frac{1}{T^2} \sum_t \sum_s \text{tr} \begin{pmatrix} G_{x,T}(x_{it} - \bar{x}_i)(x_{is} - \bar{x}_i)' G_{x,T} \\ \otimes G_{x,T}(x_{is} - \bar{x}_i)(x_{it} - \bar{x}_i)' G_{x,T} \end{pmatrix} \\ &\leq \left[\frac{1}{T} \sum_t \|G_{x,T}(x_{it} - \bar{x}_i)\|^2 \right]^2, \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{T^2} \sum_t \sum_s \text{tr} \begin{pmatrix} G_{x,T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' G_{x,T} \\ \otimes G_{x,T}(x_{is} - \bar{x}_i)(x_{is} - \bar{x}_i)' G_{x,T} \end{pmatrix} \\ &= \left(\frac{1}{T} \sum_t \text{tr} (G_{x,T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' G_{x,T}) \right)^2 \\ &= \left[\frac{1}{T} \sum_t \|G_{x,T}(x_{it} - \bar{x}_i)\|^2 \right]^2. \end{aligned}$$

Thus, the right hand side of (84) is less than or equal to

$$3M_1\kappa_v^4 \sup_{i,T} \sup_{1 \leq t \leq T} E \left[\frac{1}{T} \sum_t \|G_{x,T}(x_{it} - \bar{x}_i)\|^2 \right]^2.$$

Note that

$$\begin{aligned} & \sup_{i,T} \sup_{1 \leq t \leq T} E \left[\frac{1}{T} \sum_t \|G_{x,T}(x_{it} - \bar{x}_i)\|^2 \right]^2 \\ &\leq \sup_{i,T} \sup_{1 \leq t \leq T} E \left[\frac{1}{T} \sum_t \|G_{x,T}x_{it}\|^2 \right]^2 \\ &\leq M_2 \left(\sup_{i,T} \sup_{1 \leq t \leq T} E \|D_{1T}x_{1,it}\|^4 + \sup_{i,T} \sup_{1 \leq t \leq T} E \|x_{2,it}\|^4 \right. \\ &\quad \left. + \sup_{i,T} \sup_{1 \leq t \leq T} E \|x_{3,it}\|^4 \right), \end{aligned}$$

for some constant $M_2 < \infty$. By Assumptions 2 and 6(i), for $t = [Tr]$ and finite constant M_2 , we have

$$\begin{aligned} & \sup_{i,T} \sup_{1 \leq t \leq T} E \|D_{1T} x_{1,it}\|^4 \\ & \leq \sup_{i,T} \sup_{1 \leq t \leq T} M_3 \left(\|D_{1T}\|^4 E \|(x_{1,it} - Ex_{1,it})\|^4 + \|D_{1T} Ex_{1,it}\|^4 \right) \\ & \rightarrow \sup_i \sup_{r \in [0,1]} \|\tau_1(r) \Theta_{1,i}\|^4 < \infty. \end{aligned}$$

Next, similarly, by Assumptions 3(ii) and (iii),

$$\begin{aligned} \sup_{i,T} \sup_{1 \leq t \leq T} E \|x_{2,it}\|^4 & \leq \sup_{i,T} \sup_{1 \leq t \leq T} M_3 \left[E \|x_{2,it} - Ex_{2,it}\|^4 + \|Ex_{2,it}\|^4 \right] \\ & < \infty, \end{aligned}$$

and by Assumptions 4(ii) and (iii) and 6(iii) and (iv),

$$\begin{aligned} \sup_{i,T} \sup_{1 \leq t \leq T} E \|x_{3,it}\|^4 & \leq \sup_{i,T} \sup_{1 \leq t \leq T} M_3 \left(E \|x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}\|^4 + \|Ex_{3,it}\|^4 \right) \\ & \leq M_2 \left[\begin{aligned} & E \left(\sup_{i,t} \|x_{3,it} - E_{\mathcal{F}_{z_i}} x_{3,it}\|^4 \right) \\ & + \sup_{i,t} E \|E_{\mathcal{F}_{z_i}} x_{3,it} - Ex_{3,it}\|^4 \\ & + \sup_{i,t} \|Ex_{3,it}\|^4 \end{aligned} \right] \\ & < \infty. \end{aligned}$$

Therefore,

$$\sup_{i,T} \sup_{1 \leq t \leq T} E \left[\frac{1}{T} \sum_t \|G_{x,T}(x_{it} - \bar{x}_i)\|^2 \right]^2 < M,$$

which yields (83). ■

Part (c)

By Lemma 16(a). ■

Part (d)

By Lemma 16(d). ■

Proof of Lemma 3

By Lemma 16(c). ■

Proof of Lemma 4

Write

$$\frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{u}_i = \frac{1}{N} \sum_i D_T \tilde{w}_i (u_i - E_{\mathcal{F}_w} u_i) + \left(\frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{w}_i' D_T \right) \lambda. \quad (85)$$

Notice that by Assumption 10,

$$E \left(E_{\mathcal{F}_w} \left\| \frac{1}{N} \sum_i D_T \tilde{w}_i (u_i - E_{\mathcal{F}_w} u_i) \right\|^2 \right) = \frac{\sigma_u^2}{N} \text{tr} E \left(\frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{w}_i' D_T \right).$$

But, by Lemma 16(b),

$$\text{tr}E \left(\frac{1}{N} \sum_i D_T \tilde{w}_i \tilde{w}'_i D_T \right) < M,$$

for some constant M . This implies that

$$E \left\| \frac{1}{N} \sum_i D_T \tilde{w}_i (u_i - E_{\mathcal{F}_w} u_i) \right\|^2 \rightarrow 0,$$

as $(N, T \rightarrow \infty)$. Thus, by Chebyshev's inequality,

$$\frac{1}{N} \sum_i D_T \tilde{w}_i (u_i - E_{\mathcal{F}_w} u_i) \rightarrow_p 0, \quad (86)$$

as $(N, T \rightarrow \infty)$. Then, Lemma 16(a), (85) and (86) imply our desired result. ■

Proof of Theorem 5

By Lemma 2(a) and (b). ■

Proof of Theorem 6

By Lemma 2(c), (d), and Lemma 3. ■

Before we prove the rest of the theorems given in Section 4, we introduce the following notation:

$$\begin{aligned} A_1 &= \frac{1}{N} \sum_i \frac{1}{T} \sum_t (x_{it} - \bar{x}_i) (x_{it} - \bar{x}_i)'; \\ A_2 &= \frac{1}{N} \sum_i \frac{1}{T} \sum_t (x_{it} - \bar{x}_i) (v_{it} - \bar{v}_i); \\ A_3 &= \frac{1}{N} \sum_i \tilde{x}_i \tilde{x}'_i; \quad A_4 = \frac{1}{N} \sum_i \tilde{x}_i \tilde{u}_i; \quad A_5 = \frac{1}{N} \sum_i \tilde{x}_i \tilde{v}_i; \\ B_3 &= \frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i; \quad B_4 = \frac{1}{N} \sum_i \tilde{z}_i \tilde{u}_i; \quad B_5 = \frac{1}{N} \sum_i \tilde{z}_i \tilde{v}_i; \\ C &= \frac{1}{N} \sum_i \tilde{x}_i \tilde{z}'_i; \\ F_1 &= A_3 - CB_3^{-1}C'; \quad F_2 = A_4 + A_5 - CB_3^{-1}(B_4 + B_5). \end{aligned} \quad (87)$$

Proof of Theorem 7

Using the notation given in (87), we can express the GLS estimator $\hat{\beta}_g$ by

$$\begin{aligned} & \sqrt{NT} G_{x,T}^{-1} (\hat{\beta}_g - \beta) \\ &= [G_{x,T} A_1 G_{x,T} + \theta_T^2 G_{x,T} \{A_3 - CB_3^{-1}C'\} G_{x,T}]^{-1} \\ & \quad \times \sqrt{NT} G_{x,T} \{A_2 + \theta_T^2 [(A_4 + A_5) - CB_3^{-1}(B_4 + B_5)]\}, \end{aligned} \quad (88)$$

where $\theta_T = \sqrt{\sigma_v^2 / (T\sigma_u^2 + \sigma_v^2)}$.

Part (a)

First, consider

$$\begin{aligned} & \theta_T^2 G_{x,T} \{A_3 - CB_3^{-1}C'\} G_{x,T} \\ &= \theta_T^2 G_{x,T} D_{x,T}^{-1} \left\{ - \left(\frac{1}{N} \sum_i D_{x,T} \tilde{x}_i \tilde{z}'_i \right) \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{x}'_i D_{x,T} \right) \right\} \\ & \quad \times D_{x,T}^{-1} G_{x,T}. \end{aligned}$$

By Lemma 2,

$$\left\{ - \left(\frac{1}{N} \sum_i D_{x,T} \tilde{x}_i \tilde{z}'_i \right) \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{x}'_i D_{x,T} \right) \right\} = O_p(1),$$

and by definition,

$$G_{x,T} D_{x,T}^{-1} = J_{x,T}^{-1} = O(1),$$

as $(N, T \rightarrow \infty)$. Thus,

$$\theta_T^2 G_{x,T} \{A_3 - CB_3^{-1} C'\} G_{x,T} = O_p(\theta_T^2) = o_p(1). \quad (89)$$

Next, consider

$$\begin{aligned} & \sqrt{NT} G_{x,T} \{ \theta_T^2 [(A_4 + A_5) - CB_3^{-1} (B_4 + B_5)] \} \\ = & \theta_T^2 \sqrt{T} G_{x,T} D_{x,T}^{-1} \left\{ \begin{aligned} & \frac{1}{\sqrt{N}} \sum_i D_{x,T} \tilde{x}_i (\tilde{u}_i + \tilde{v}_i) \\ & - \left(\frac{1}{N} \sum_i D_{x,T} \tilde{x}_i \tilde{z}'_i \right) \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_i \tilde{z}_i (\tilde{u}_i + \tilde{v}_i) \right) \end{aligned} \right\}. \end{aligned}$$

By Lemma 2, under the local alternatives to random effects (Assumption 11),

$$\left\{ \begin{aligned} & \frac{1}{\sqrt{N}} \sum_i D_{x,T} \tilde{x}_i (\tilde{u}_i + \tilde{v}_i) \\ & - \left(\frac{1}{N} \sum_i D_{x,T} \tilde{x}_i \tilde{z}'_i \right) \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_i \tilde{z}_i (\tilde{u}_i + \tilde{v}_i) \right) \end{aligned} \right\} = O_p(1).$$

By definition,

$$\theta_T^2 \sqrt{T} G_{x,T} D_{x,T}^{-1} = O\left(\frac{1}{\sqrt{T}}\right).$$

Thus,

$$\sqrt{NT} G_{x,T} \{ \theta_T^2 [(A_4 + A_5) - CB_3^{-1} (B_4 + B_5)] \} = O_p\left(\frac{1}{\sqrt{T}}\right) = o_p(1). \quad (90)$$

Substituting (89) and (90) into (88), we have

$$\begin{aligned} \sqrt{NT}(\hat{\beta}_g - \beta) &= [G_{x,T} A_1 G_{x,T} + o_p(1)]^{-1} [\sqrt{NT} G_{x,T} A_2 + o_p(1)] \\ &= \sqrt{NT}(\hat{\beta}_w - \beta) + o_p(1). \end{aligned}$$

The last equality results from Lemma 2(a), (b) and Theorem 5. ■

Part (b)

Similarly to Part (a), we can easily show that under the assumptions given in Part (b), the denominator in (88) is

$$\frac{1}{NT} \sum_i \sum_t G_{x,T} (x_{it} - \bar{x}_i) (x_{it} - \bar{x}_i)' G_{x,T} + o_p(1). \quad (91)$$

Consider the second term of the numerator of (88):

$$\theta^2 \sqrt{T} G_{x,T} \left\{ \begin{aligned} & \frac{1}{\sqrt{N}} \sum_i \tilde{x}_i (\tilde{u}_i + \tilde{v}_i) \\ & - \left(\frac{1}{N} \sum_i \tilde{x}_i \tilde{z}'_i \right) \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_i \tilde{z}_i (\tilde{u}_i + \tilde{v}_i) \right) \end{aligned} \right\}. \quad (92)$$

Notice that by Lemmas 2 and 4, under the fixed effect assumption (Assumption 10), the first term of (92) is

$$\begin{aligned}
& \theta^2 \sqrt{NT} G_{x,T} \frac{1}{N} \sum_i \tilde{x}_i (\tilde{u}_i + \tilde{v}_i) \\
&= \theta^2 \sqrt{NT} G_{x,T} D_{x,T}^{-1} \frac{1}{N} \sum_i D_{x,T} \tilde{x}_i (\tilde{u}_i + \tilde{v}_i) \\
&= \theta^2 \sqrt{NT} G_{x,T} D_{x,T}^{-1} \{ \Xi_x \lambda + o_p(1) \},
\end{aligned}$$

where we partition

$$\Xi = \begin{pmatrix} \Xi_{xx} & \Xi_{xz} \\ \Xi_{zx} & \Xi_{zz} \end{pmatrix}$$

conformably to the sizes of x_{it} and z_i , and set $\Xi_x = (\Xi_{xx}, \Xi_{xz})$. Similarly, by Lemmas 2 and 4, under the fixed effect assumption, the second term of (92) is

$$\begin{aligned}
& \theta_T^2 \sqrt{NT} G_{x,T} D_{x,T}^{-1} \left\{ \left(\frac{1}{N} \sum_i D_{x,T} \tilde{x}_i \tilde{z}'_i \right) \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{N} \sum_i \tilde{z}_i (\tilde{u}_i + \tilde{v}_i) \right) \right\} \\
&= \theta_T^2 \sqrt{NT} G_{x,T} D_{x,T}^{-1} \{ \Xi_{xz} \Xi_{zz}^{-1} \Xi_z \lambda + o_p(1) \}.
\end{aligned}$$

Therefore, the limit of (92) is

$$\left(\theta_T^2 \sqrt{NT} G_{x,T} D_{x,T}^{-1} \right) \left[\begin{pmatrix} \Xi_{xx} - \Xi_{xz} \Xi_{zz}^{-1} \Xi_{zx} & \vdots & 0 \end{pmatrix} \lambda + o_p(1) \right].$$

Recall that it is assumed that $\frac{N}{T} \rightarrow c < \infty$. Also, recall that under the restrictions given in the theorem, $G_{x,T} = \text{diag}(I_{k_{22}}, I_{k_3})$ and $D_{x,T} = \text{diag}(\sqrt{T} I_{k_{22}}, D_{3T})$. Then, letting $\lambda_{\max}(A)$ denote the maximum eigenvalue of matrix A , we can have

$$\begin{aligned}
& \lambda_{\max} \left(\theta_T^2 \sqrt{NT} G_{x,T} D_{x,T}^{-1} \right) \\
&= O(1) \sqrt{\frac{N}{T}} \lambda_{\max} \left(\frac{1}{\sqrt{T}} I_{k_{22}}, D_{3T}^{-1} \right) \rightarrow 0.
\end{aligned}$$

So, under the assumptions of Part (b), the probability limit of the numerator of (88) is

$$\frac{1}{\sqrt{NT}} \sum_i \sum_t G_{x,T} \tilde{x}_{it} \tilde{v}_{it} + o_p(1). \quad (93)$$

Combining (91) and (93), we can obtain Part (b). ■

Proof of Theorem 8

Using the notation in (87), we can express the GLS estimator $\hat{\gamma}_g$ by

$$\hat{\gamma}_g - \gamma$$

$$\begin{aligned}
&= \left[B_3 - C' \left(\frac{1}{\theta_T^2} A_1 + A_3 \right)^{-1} C \right]^{-1} \\
&\quad \times \left[(B_4 + B_5) - C' \left(\frac{1}{\theta_T^2} A_1 + A_3 \right)^{-1} \left(\frac{1}{\theta_T^2} A_2 + (A_4 + A_5) \right) \right]. \quad (94)
\end{aligned}$$

Part (a)

Notice that

$$\begin{aligned}
&C' \left(\frac{1}{\theta_T^2} A_1 + A_3 \right)^{-1} C \\
&= C' D_{x,T} \left(\frac{1}{T \theta_T^2} \sqrt{T} J_{x,T} G_{x,T} A_1 G_{x,T} J_{x,T} \sqrt{T} + D_{x,T} A_3 D_{x,T} \right)^{-1} D_{x,T} C \\
&= T \theta_T^2 (C' D_{x,T}) \frac{J_{x,T}^{-1}}{\sqrt{T}} \left(\begin{array}{c} G_{x,T} A_1 G_{x,T} \\ + \frac{J_{x,T}^{-1}}{\sqrt{T}} D_{x,T} A_3 D_{x,T} \frac{J_{x,T}^{-1}}{\sqrt{T}} \end{array} \right)^{-1} \frac{J_{x,T}^{-1}}{\sqrt{T}} (D_{x,T} C) \\
&= O_p \left(\frac{J_{x,T}^{-2}}{T} \right) = o_p(1).
\end{aligned}$$

The third equality holds because the limit of $G_{x,T} A_1 G_{x,T}$ is positive definite (by Lemma 2(a) and Assumption 12), $T \theta_T^2 = O(1)$, and

$$D_{x,T} C, G_{x,T} A_1 G_{x,T}, D_{x,T} A_3 D_{x,T} = O_p(1)$$

(by Lemma 2(a), (c)). The last equality results from the fact that

$$\frac{J_{x,T}^{-1}}{\sqrt{T}} = o(1).$$

Thus, as $(N, T \rightarrow \infty)$, the denominator of (94) is

$$B_3 - C' \left(\frac{1}{\theta_T^2} A_1 + A_3 \right)^{-1} C = B_3 + o_p(1). \quad (95)$$

Next, under both the random effects assumption (Assumption 9) and the local alternatives (Assumption 11), it follows from Lemmas 2 and 3 that the second term in the numerator of (94) is

$$\begin{aligned}
&C' \left(\frac{1}{\theta^2} A_1 + A_3 \right)^{-1} \left(\frac{1}{\theta^2} \sqrt{N} A_2 + \sqrt{N} (A_4 + A_5) \right) \\
&= C' D_{x,T} \left(\frac{1}{\theta_T^2} \sqrt{T} J_{x,T} G_{x,T} A_1 G_{x,T} J_{x,T} \sqrt{T} + D_{x,T} A_3 D_{x,T} \right)^{-1} \\
&\quad \times \left(\frac{\sqrt{T}}{T \theta^2} J_{x,T} \sqrt{N T} G_{x,T} A_2 + \sqrt{N} D_{x,T} (A_4 + A_5) \right)
\end{aligned}$$

$$\begin{aligned}
&= T\theta^2 (C'D_{x,T}) \frac{J_{x,T}^{-1}}{\sqrt{T}} \left(G_{x,T}A_1G_{x,T} + \frac{J_{x,T}^{-1}}{\sqrt{T}} D_{x,T}A_3D_{x,T} \frac{J_{x,T}^{-1}}{\sqrt{T}} \right)^{-1} \\
&\quad \times \left(\frac{1}{T\theta^2} \sqrt{NT} G_{x,T}A_2 + \frac{J_{x,T}^{-1}}{\sqrt{T}} \sqrt{N} D_{x,T} (A_4 + A_5) \right) \\
&= O_p \left(\frac{J_{x,T}^{-1}}{\sqrt{T}} \right) = o_p(1),
\end{aligned}$$

as $(N, T \rightarrow \infty)$. Also, by Lemma 2(d), as $(N, T \rightarrow \infty)$,

$$\sqrt{N}B_5 = \frac{1}{\sqrt{N}} \sum_i \tilde{z}_i \tilde{v}_i = o_p(1).$$

Therefore, the numerator of (94) is

$$\begin{aligned}
&\sqrt{N} \left((B_4 + B_5) - C' \left(\frac{1}{\theta_T^2} A_1 + A_3 \right)^{-1} \left(\frac{1}{\theta_T^2} A_2 + (A_4 + A_5) \right) \right) \\
&= \sqrt{N}B_4 + o_p(1),
\end{aligned} \tag{96}$$

as $(N, T \rightarrow \infty)$. In view of (94), (95) and (96), we have

$$\sqrt{N}(\hat{\gamma}_g - \gamma) = \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_i \tilde{z}_i \tilde{u}_i \right) + o_p(1),$$

as $(N, T \rightarrow \infty)$.

Finally, by Lemma 2(c) and Lemma 3, as $(N, T \rightarrow \infty)$,

$$\begin{aligned}
\sqrt{N}(\hat{\gamma}_g - \gamma) &= \left(\frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_i \tilde{z}_i \tilde{u}_i \right) \\
&\Rightarrow N \left((l'_z \Xi l_z)^{-1} (l'_z \Xi \lambda), (l'_z \Xi l_z)^{-1} \right),
\end{aligned}$$

as required. ■

Part (b)

Under the assumptions in Part (b), as shown for the denominator of (94),

$$B_3 - C' \left(\frac{1}{\theta_T^2} A_1 + A_3 \right)^{-1} C = \frac{1}{N} \sum_i \tilde{z}_i \tilde{z}'_i + o_p(1) \rightarrow_p l'_z \Xi l_z, \tag{97}$$

as $(N, T \rightarrow \infty)$. Next, consider the numerator of (94),

$$\begin{aligned}
&\left[(B_4 + B_5) - C' \left(\frac{1}{\theta^2} A_1 + A_3 \right)^{-1} \left(\frac{1}{\theta^2} A_2 + (A_4 + A_5) \right) \right] \\
&= (B_4 + B_5)
\end{aligned}$$

$$\begin{aligned}
& -T\theta^2 (C' D_{x,T}) \frac{J_{x,T}^{-1}}{\sqrt{T}} \left(G_{x,T} A_1 G_{x,T} + \frac{J_{x,T}^{-1}}{\sqrt{T}} D_{x,T} A_3 D_{x,T} \frac{J_{x,T}^{-1}}{\sqrt{T}} \right)^{-1} \\
& \times \left(\frac{1}{\sqrt{NT}\theta^2} \sqrt{NT} G_{x,T} A_2 + \frac{J_{x,T}^{-1}}{\sqrt{T}} D_{x,T} (A_4 + A_5) \right).
\end{aligned}$$

By Lemma 2(d), (b), and (d),

$$\begin{aligned}
B_5 &= o_p(1), \\
\sqrt{NT} G_{x,T} A_2 &= O_p(1), \\
D_{x,T} A_5 &= O_p(1),
\end{aligned}$$

respectively. Under the fixed effect assumption (Assumption 10), Lemma 4 implies that

$$D_{x,T} A_4 = O_p(1),$$

as $(N, T \rightarrow \infty)$. Since

$$\frac{1}{\sqrt{NT}\theta_T^2}, \frac{J_{x,T}^{-1}}{\sqrt{T}} = o(1),$$

and

$$T\theta^2 (C' D_{x,T}) \frac{J_{x,T}^{-1}}{\sqrt{T}} \left(G_{x,T} A_1 G_{x,T} + \frac{J_{x,T}^{-1}}{\sqrt{T}} D_{x,T} A_3 D_{x,T} \frac{J_{x,T}^{-1}}{\sqrt{T}} \right)^{-1} = o_p(1)$$

(as shown in Part (a)), we have

$$(B_4 + B_5) - C' \left(\frac{1}{\theta^2} A_1 + A_3 \right)^{-1} \left(\frac{1}{\theta^2} A_2 + (A_4 + A_5) \right) = B_4 + o_p(1). \quad (98)$$

But, according to Lemma 4,

$$B_4 = \frac{1}{N} \sum_i \tilde{z}_i \tilde{u}_i \rightarrow_p l'_z \Xi \lambda. \quad (99)$$

Therefore, (94), (97), (98) and (99) imply

$$\hat{\gamma}_g \rightarrow_p \gamma + (l'_z \Xi l_z)^{-1} l'_z \Xi \lambda,$$

as $(N, T \rightarrow \infty)$. ■

Proof of Theorem 9

Using the notation in (87), we can express the Hausman test statistic by

$$\begin{aligned}
\mathcal{HM}_{NT} &= \left[(A_1 + \theta_T^2 F_1)^{-1} \sqrt{NT} (A_2 + \theta_T^2 F_2) - A_1^{-1} \sqrt{NT} A_2 \right]' \\
&\times \left[\sigma_v^2 A_1^{-1} - \sigma_v^2 (A_1 + \theta_T^2 F_1)^{-1} \right]^{-1} \\
&\times \left[(A_1 + \theta_T^2 F_1)^{-1} \sqrt{NT} (A_2 + \theta_T^2 F_2) - A_1^{-1} \sqrt{NT} A_2 \right].
\end{aligned}$$

Observe that for any conformable matrices P and Q , we have

$$(P + Q)^{-1} - P^{-1} = -P^{-1}QP^{-1} + (P + Q)^{-1}QP^{-1}QP^{-1}.$$

Using this fact, we write

$$(A_1 + \theta_T^2 F_1)^{-1} - A_1^{-1} = -\theta_T^2 A_1^{-1} F_1 A_1^{-1} + \theta_T^4 R_1, \quad (100)$$

where $R_1 = (A_1 + \theta_T^2 F_1)^{-1} F_1 A_1^{-1} F_1 A_1^{-1}$. Define

$$\mathcal{Q} = (A_1 + \theta_T^2 F_1)^{-1} \sqrt{NT} (A_2 + \theta_T^2 F_2) - A_1^{-1} \sqrt{NT} A_2.$$

Then,

$$\begin{aligned} & \mathcal{Q} \\ &= (A_1 + \theta_T^2 F_1)^{-1} \sqrt{NT} (A_2 + \theta_T^2 F_2) - A_1^{-1} \sqrt{NT} (A_2 + \theta_T^2 F_2) \\ & \quad + A_1^{-1} \sqrt{NT} \theta_T^2 F_2 \\ &= \left\{ (A_1 + \theta_T^2 F_1)^{-1} - A_1^{-1} \right\} \sqrt{NT} \{A_2 + \theta_T^2 F_2\} + A_1^{-1} \sqrt{NT} \theta_T^2 F_2 \\ &= -A_1^{-1} (\theta_T^2 F_1) A_1^{-1} \sqrt{NT} \{A_2 + \theta_T^2 F_2\} + A_1^{-1} \sqrt{NT} \theta_T^2 F_2 \\ & \quad + \theta_T^4 R_1 \sqrt{NT} \{A_2 + \theta_T^2 F_2\} \\ &= -\theta_T^2 \sqrt{NT} [A_1^{-1} F_1 A_1^{-1} A_2 - A_1^{-1} F_2] - \theta_T^2 \sqrt{NT} \left[\begin{array}{c} \theta_T^2 A_1^{-1} F_1 A_1^{-1} F_2 \\ -\theta_T^2 R_1 \{A_2 + \theta_T^2 F_2\} \end{array} \right] \\ &= -\theta_T^2 \sqrt{NT} [A_1^{-1} F_1 A_1^{-1} A_2 - A_1^{-1} F_2] - \theta_T^4 \sqrt{NT} R_2, \end{aligned} \quad (101)$$

where $R_2 = A_1^{-1} F_1 A_1^{-1} F_2 - R_1 \{A_2 + \theta_T^2 F_2\}$.

In view of (100) and (101), we now can rewrite the Hausman statistic as

$$\begin{aligned} \mathcal{H}\mathcal{M}_{NT} &= \mathcal{Q}' \left[\sigma_v^2 A_1^{-1} - \sigma_v^2 (A_1 + \theta_T^2 F_1)^{-1} \right]^{-1} \mathcal{Q} \\ &= \theta_T \sqrt{NT} [A_1^{-1} F_1 A_1^{-1} A_2 - A_1^{-1} F_2 + \theta_T^2 R_2]' \\ & \quad \times [\sigma_v^2 A_1^{-1} F_1 A_1^{-1} - \sigma_v^2 \theta_T^2 R_1]^{-1} \\ & \quad \times \theta_T \sqrt{NT} [A_1^{-1} F_1 A_1^{-1} A_2 - A_1^{-1} F_2 + \theta_T^2 R_2]; \end{aligned}$$

or equivalently,

$$\begin{aligned} & \mathcal{H}\mathcal{M}_{NT} \\ &= \theta_T \sqrt{NT} \left[\begin{array}{c} G_1 \left(J_{x,T}^{-1} D_{x,T} F_1 D_{x,T} J_{x,T}^{-1} \right) G_1 G_{x,T} A_2 \\ -G_1 J_{x,T}^{-1} D_{x,T} F_2 + \theta_T^2 G_{x,T}^{-1} R_2 \end{array} \right]' \\ & \quad \times \left[\sigma_v^2 G_1 \left(J_{x,T}^{-1} D_{x,T} F_1 D_{x,T} J_{x,T}^{-1} \right) G_1 + \sigma_v^2 \theta_T^2 \left(G_{x,T}^{-1} R_1 G_{x,T}^{-1} \right) \right]^{-1} \\ & \quad \times \theta_T \sqrt{NT} \left[\begin{array}{c} G_1 \left(J_{x,T}^{-1} D_{x,T} F_1 D_{x,T} J_{x,T}^{-1} \right) G_1 G_{x,T} A_2 \\ -G_1 J_{x,T}^{-1} D_{x,T} F_2 + \theta_T^2 G_{x,T}^{-1} R_2 \end{array} \right], \end{aligned}$$

where $G_1 = G_{x,T}^{-1} A_1^{-1} G_{x,T}^{-1}$.

Notice that

$$\begin{aligned} \theta_T^2 \left(G_{x,T}^{-1} R_1 G_{x,T}^{-1} \right) &= \theta_T^2 G_{x,T}^{-1} (A_1 + \theta^2 F_1)^{-1} G_{x,T}^{-1} \\ &\quad \times G_{x,T} F_1 G_{x,T} G_1 G_{x,T} F_1 G_{x,T} G_1. \end{aligned}$$

Lemma 2(a) and Assumption 12 imply

$$G_1 = O_p(1).$$

In addition, by Lemma 2(c)

$$G_{x,T} F_1 G_{x,T} = J_{x,T}^{-1} D_{x,T} F_1 D_{x,T} J_{x,T}^{-1} = O_p(1), \quad (102)$$

since $J_{x,T}^{-1} = O(1)$. Thus,

$$\sigma_v^2 \theta_T^2 \left(G_{x,T}^{-1} R_1 G_{x,T}^{-1} \right) = O_p(\theta_T^2) = o_p(1). \quad (103)$$

Now, consider

$$\begin{aligned} &\theta_T \sqrt{NT} \theta_T^2 G_{x,T}^{-1} R_2 \\ &= \theta_T \sqrt{T} \left[\sqrt{N} \theta_T^2 G_{x,T}^{-1} (A_1^{-1} F_1 A_1^{-1} F_2) + \sqrt{N} \theta_T^2 G_{x,T}^{-1} R_1 (A_2 + \theta_T^2 F_2) \right] \\ &= \theta_T \sqrt{T} \left[\begin{aligned} &\theta_T^2 \left(G_{x,T}^{-1} A_1^{-1} G_{x,T}^{-1} \right) (G_{x,T} F_1 G_{x,T}) \left(G_{x,T}^{-1} A_1^{-1} G_{x,T}^{-1} \right) G_{x,T} \sqrt{N} F_2 \\ &+ \theta_T^2 \left(G_{x,T}^{-1} R_1 G_{x,T}^{-1} \right) \left(\sqrt{N} G_{x,T} A_2 + \theta_T^2 \sqrt{N} G_{x,T} F_2 \right) \end{aligned} \right]. \end{aligned}$$

Under the local alternatives (Assumption 11), we may deduce from Lemmas 2 and 3 that

$$\begin{aligned} G_1, G_{x,T} F_1 G_{x,T}, G_{x,T} \sqrt{N} F_2 &= O_p(1); \\ \theta_T^2 \left(G_{x,T}^{-1} R_1 G_{x,T}^{-1} \right), \sqrt{N} G_{x,T} A_2 &= o_p(1). \end{aligned}$$

Since $\theta_T \sqrt{T} = O(1)$, we have

$$\theta_T \sqrt{NT} \theta_T^2 G_{x,T}^{-1} R_2 = O(1) [\theta_T^2 O_p(1) + o_p(1)] = o_p(1). \quad (104)$$

Using the results (103) and (104), we now can approximate the Hausman statistic as follows:

$$\begin{aligned} &\mathcal{HM}_{NT} \\ &= \frac{\theta_T}{\sigma_v} \sqrt{NT} \left[\left(J_{x,T}^{-1} D_{x,T} F_1 D_{x,T} J_{x,T}^{-1} \right) (G_1 G_{x,T} A_2) - J_{x,T}^{-1} D_{x,T} F_2 + o_p(1) \right]' \\ &\quad \times \left(J_{x,T}^{-1} D_{x,T} F_1 D_{x,T} J_{x,T}^{-1} + o_p(1) \right)^{-1} \\ &\quad \times \frac{\theta_T}{\sigma_v} \sqrt{NT} \left[\left(J_{x,T}^{-1} D_{x,T} F_1 D_{x,T} J_{x,T}^{-1} \right) (G_1 G_{x,T} A_2) - J_{x,T}^{-1} D_{x,T} F_2 + o_p(1) \right] \\ &= \frac{\theta_T}{\sigma_v} \sqrt{NT} (D_{x,T} F_2)' (D_{x,T} F_1 D_{x,T})^{-1} \frac{\theta_T}{\sigma_v} \sqrt{NT} (D_{x,T} F_2) \\ &\quad + o_p(1). \end{aligned}$$

Here, the last line holds because under the local alternative hypotheses,

$$J_{x,T}^{-1}D_{x,T}F_1D_{x,T}J_{x,T}^{-1} = O_p(1),$$

by (102), and

$$\frac{\theta_T}{\sigma_v}G_1\left(\sqrt{NT}G_{x,T}A_2\right) = O_p(\theta_T) = o_p(1),$$

by Lemma 2(a), (b) and Assumption 12.

Finally, by Lemma 2(c), as $(N, T \rightarrow \infty)$,

$$\begin{aligned} & D_{x,T}F_1D_{x,T} \\ &= D_{x,T}A_3D_{x,T} - D_{x,T}CB_3^{-1}C'D_{x,T} \\ &= \frac{1}{N}\sum_i D_{x,T}\tilde{x}_i\tilde{x}'_iD_{x,T} \\ &\quad - \left(\frac{1}{N}\sum_i D_{x,T}\tilde{x}_i\tilde{z}'_i\right)\left(\frac{1}{N}\sum_i \tilde{z}_i\tilde{z}'_i\right)^{-1}\left(\frac{1}{N}\sum_i D_{x,T}\tilde{x}_i\tilde{z}'_i\right)' \\ &\rightarrow {}_p\Xi_{xx} - \Xi_{xz}\Xi_{zz}^{-1}\Xi_{zx} > 0. \end{aligned}$$

Also, Lemmas 2(d) and 3 imply that under the local alternative hypotheses, as $(N, T \rightarrow \infty)$,

$$\begin{aligned} & \sqrt{N}D_{x,T}F_2 \\ &= \sqrt{N}D_{x,T}A_4 + \sqrt{N}D_{x,T}A_5 - D_{x,T}CB_3^{-1}\sqrt{N}(B_4 + B_5) \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^N D_{x,T}\tilde{x}_i\tilde{u}_i \\ &\quad - \left(\frac{1}{N}\sum_{i=1}^N D_{x,T}\tilde{x}_i\tilde{z}'_i\right)\left(\frac{1}{N}\sum_{i=1}^N \tilde{z}_i\tilde{z}'_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{z}_i\tilde{u}_i\right) + o_p(1) \\ &\Rightarrow \left(I_k \quad -\Xi_{xz}\Xi_{zz}^{-1} \right) N(\Xi\lambda, \sigma_u^2\Xi). \end{aligned}$$

As $(N, T \rightarrow \infty)$, $\frac{\theta_T}{\sigma_v}\sqrt{T} \rightarrow \frac{1}{\sigma_u}$. Therefore, under the hypothesis of random effects,

$$\mathcal{HM}_{NT} \Rightarrow \chi_k^2,$$

a χ^2 distribution with the degrees of freedom equal to k . In contrast, under the local alternative hypotheses,

$$\mathcal{HM}_{NT} \Rightarrow \chi_k^2(\eta),$$

where η is the noncentral parameter.

References

- [1] Ahn, S.C., and S. Low, 1996, A reformulation of the Hausman test for regression models with cross-section-time-series data, *Journal of Econometrics*, 71, 291-307.
- [2] Alvarez, J., and M. Arellano, 1998, The time series and cross-section asymptotics of dynamic panel data estimators, *mimeo*, CEMFI, Spain.
- [3] Amemiya T., and T.E. MaCurdy, 1986, Instrumental-variables estimation of an error components model, *Econometrica*, 54, 869-880.
- [4] Andrews, D., 1991, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59, 817-858
- [5] Apostol, T., 1974, *Mathematical Analysis*, 2nd Ed., Addison-Welsey, Reading, MA.
- [6] Balestra, P. and M. Nerlove, 1966, Pooling cross-section and time-series data in the estimation of a dynamic model: The demand for natural gas, *Econometrica*, 34, 585- 612.
- [7] Baltagi, B. and S. Khanti-Akom, 1990, On efficient estimation with panel data: An empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics*, 5, 401-406.
- [8] Baltagi, B., 1995, *Econometric Analysis of Panel Data*, John Wiley & Sons Ltd., New York, New York.
- [9] Bosq, 1996, *Nonparametric Statistics for Stochastic processes*, Springer-Verlag, New York.
- [10] Breusch, T.S., G.E. Mizon, and P. Schmidt, 1989, Efficient estimation using panel data, *Econometrica*, 57, 695-700.
- [11] Choi, I., 1998, Asymptotic analysis of a nonstationary error component model, *mimeo*, Kookmin University, Korea.
- [12] Cornwell, C. and P. Rupert, 1988, Efficient estimation with panel data: An empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics*, 3, 149-155.
- [13] Hahn, J. and G. Kuersteiner, 2000, Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, *mimeo*.
- [14] Hausman, J.A., 1978, Specification tests in econometrics, *Econometrica*, 46, 1251-1272.
- [15] Hausman, J.A. and W.E. Taylor, 1981, Panel data and unobservable individual effects, *Econometrica*, 49, 1377-1398.

- [16] Hall and Heyde, 1980, *Martingale Limit Theory and Its Application*, Academic Press, San Diego.
- [17] Higgins, M. and E. Zakrajšek, 1999, Purchasing power parity: Three stakes through the heart of the unit root null, NBER working paper.
- [18] Hsiao, C., 1986, *Analysis of Panel Data*, New York: Cambridge University Press.
- [19] Im, K.S., M.H. Pesaran, and Y. Shin, 1997, Testing for unit roots in Heterogeneous panels, *mimeo*.
- [20] Kao, C., 1999, Spurious regression and residual-based tests for cointegration in panel data, *Journal of Econometrics*, 90, 1-44.
- [21] Levin, A., C-F Lin, 1992, Unit root tests in panel data: Asymptotic and finite-sample properties, *mimeo*, University of California at San Diego.
- [22] Levin, A., C-F Lin, 1993, Unit root tests in panel data: New results, *mimeo*, University of California at San Diego.
- [23] Maddala, G.S., 1971, The use of variance components models in pooling cross section and time series data, *Econometrica*, 39, 341-358.
- [24] Mátyás, L. and P. Sevestre, 1992, *The econometrics of Panel Data*, Dordrecht: Kluwer Academic Publishers.
- [25] Phillips, P.C.B., and H. Moon, 1999, Linear regression limit theory for nonstationary panel data, *Econometrica*, 67, 1057-1111.
- [26] Quah, D., 1994, Exploiting Cross-section variations for unit root inference in dynamic data, *Economics Letters*, 44, 9-19.
- [27] van der Vaart and Wellner, 1996, *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.