Econometric Theory, **29**, 2013, 1079–1135.
doi:10.1017/S0266466613000017

TESTING HOMOGENEITY IN PANEL DATA MODELS WITH INTERACTIVE FIXED EFFECTS

LIANGJUN SU
Singapore Management University

QIHUI CHEN
University of California, San Diego

This paper proposes a residual-based Lagrange Multiplier (LM) test for slope homogeneity in large-dimensional panel data models with interactive fixed effects. We first run the panel regression under the null to obtain the restricted residuals and then use them to construct our LM test statistic. We show that after being appropriately centered and scaled, our test statistic is asymptotically normally distributed under the null and a sequence of Pitman local alternatives. The asymptotic distributional theories are established under fairly general conditions that allow for both lagged dependent variables and conditional heteroskedasticity of unknown form by relying on the concept of conditional strong mixing. To improve the finite-sample performance of the test, we also propose a bootstrap procedure to obtain the bootstrap p-values and justify its validity. Monte Carlo simulations suggest that the test has correct size and satisfactory power. We apply our test to study the Organization for Economic Cooperation and Development economic growth model.

1. INTRODUCTION

Recently large-dimensional panel data models with interactive fixed effects have attracted a great deal of attention in econometrics. Pesaran (2006) proposes the common correlated effects (CCE) estimators for heterogeneous panels and derives their asymptotic normal distributions under fairly general conditions. Bai (2009a) studies the asymptotic properties of principal component analysis (PCA) estimators and demonstrates that they are $\sqrt{NT}$ consistent, where N and T refer to the individual and time series dimensions, respectively. Kapetanios and Pesaran (2007) propose a factor-augmented estimator by augmenting a linear panel data model with estimated common factors to account for cross-sectional dependence

The authors gratefully thank the co-editor, Guido Kuersteiner, and three anonymous referees for their many constructive comments on the previous version of the paper. They also express their sincere appreciation to Hashem Pesaran for discussions on the subject matter. They sincerely thank the participants at the 2011 Tsinghua International Conference in Econometrics and the seminar at the University of Adelaide who provided valuable suggestions and discussion. The first author gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. correspondence to Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; e-mail: ljsu@smu.edu.sg.

© Cambridge University Press 2013

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

and study its finite-sample properties via Monte Carlo simulations. Greenaway-McGrevy, Han, and Sul (2012) formally establish the asymptotic distribution of this estimator and provide specific conditions under which the estimated factors can be used in place of the latent factors in the regression. Moon and Weidner (2010b) (MW hereafter) reinvestigate the PCA estimation of Bai (2009) in the framework of quasi-maximum likelihood estimation (QMLE) of dynamic linear panel data models with interactive fixed effects and find that there are two sources of asymptotic bias: one is due to the presence of serial correlation or heteroskedasticity of the idiosyncratic error term, and the other is due to the presence of predetermined regressors. In addition, Moon and Weidner (2010a) discuss the validity of PCA estimation for panel data models when the number of factors as interactive fixed effects is unknown and has to be chosen according to certain information criteria. Pesaran and Tosetti (2011) consider estimation of panel data models with a multifactor error structure and spatial error correlations and find that Pesaran's CCE procedure continues to yield consistent and asymptotically normal estimates of the slope coefficients.

Panel data models with interactive fixed effects are useful modeling paradigms. In macroeconomics, incorporating interactive effects can account for the heterogeneous impact of unobservable common shocks, and the regressors can be such inputs as labor and capital. In finance, combination of unobserved factors and observed covariates can explain the excess returns of assets. In microeconomics, panel data models with interactive fixed effects can incorporate unmeasured skills or unobservable characteristics to study the individual wage rate. Nevertheless, in most empirical studies it is commonly assumed that the coefficients of the observed regressors are homogeneous. In fact, most of the literature reviewed previously is developed for homogeneous panel data models with interactive fixed effects. The only exceptions are Pesaran (2006), Kapetanios and Pesaran (2007), and Pesaran and Tosetti (2011), which are applicable to heterogeneous panels but typically require certain rank conditions to estimate individual slopes.[1] Su and Jin (2012) extend Pesaran (2006) to nonparametric regression with a multifactor error structure.

Slope homogeneity assumption greatly simplifies the estimation and inference process, and the proposed estimator can be efficient if there is no heterogeneity in the individual slopes. Nevertheless, if the slope homogeneity assumption is not true, estimates based on panel data models with homogeneous slopes can be inconsistent and lead to misleading statistical inference; see, e.g., Hsiao (2003, Ch. 6) and Baltagi, Bresson, and Pirotte (2008). So it is necessary and prudent to test for slope homogeneity before imposing it.

There are many studies on testing for slope homogeneity and poolability in the panel data literature; see Pesaran, Smith, and Im (1996), Phillips and Sul (2003), Pesaran and Yamagata (2008) (PY hereafter), Blomquist (2010), Lin (2010), and Jin and Su (2013), among others. Pesaran et al. (1996) propose a Hausman-type test by comparing the standard fixed effects estimator with the mean group estimator. Phillips and Sul (2003) also propose a Hausman-type test for slope

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

homogeneity for AR(1) panel data models in the presence of cross-sectional dependence. Recently, PY have developed a standardized version of Swamy's test for slope homogeneity in large panel data models with fixed effects and unconditional heteroskedasticity, and Blomquist (2010) proposes a bootstrap version of PY's Swamy test that is claimed to be robust to general forms of cross-sectional dependence and serial correlation. Lin (2010) proposes a test for slope homogeneity in linear panel data models with fixed effects and conditional heteroskedasticity. Jin and Su (2013) propose a nonparametric test for poolability in nonparametric regression models with a multifactor error structure. Nevertheless, to the best of our knowledge, there is no available test of slope homogeneity for large-dimensional panel data models with interactive fixed effects.

In this paper we consider a residual-based Lagrange Multiplier (LM) test for slope homogeneity in large-dimensional panel data models with interactive fixed effects where both lagged dependent variables and conditional heteroskedasticity of unknown form may be present. Under the null hypothesis of homogeneous slopes, the observable regressors should not contain any useful information about the residuals from the Bai (2009a) PCA estimation. This motivates us to construct a residual-based test. We first estimate a restricted model by imposing slope homogeneity. Then we consider heterogeneous regression of the restricted residuals on the observable regressors and test whether the slope coefficients in this regression are identically zero based on the LM principle. We study the asymptotic distribution of the LM test statistic under a set of fairly general conditions that allow for both dynamics and conditional heteroskedasticity of unknown form. We show that after being appropriately standardized, the LM test statistic is asymptotically normally distributed under the null hypothesis and a sequence of Pitman local alternatives. We also propose a bootstrap method to obtain the bootstrap p-values to improve the finite-sample performance of our test and justify its asymptotic validity. In the Monte Carlo experiments, we show that the test has correct size and satisfactory power. We apply our test to the Organization for Economic Cooperation and Development (OECD) economic growth data and reject the null of homogeneous slopes.

To sum up, our residual-based LM test has several advantages. First, the intuition as detailed previously is clear. It is consistent and has power in detecting local alternatives converging to the null at the usual $N^{-1/4}T^{-1/2}$ rate, which is also obtained by PY. Second, unlike PY's test, which requires estimation under both the null and alternative, we only require estimation of the panel data models under the null hypothesis. This is extremely important because the Bai (2009a) PCA estimation (or equivalently MW's QMLE) is only applicable to homogeneous large-dimensional panels with interactive fixed effects. The Pesaran (2006) CCE procedure can be used to estimate the models under both the null and alternative, but it would require certain rank conditions that are not needed here. Third, it is feasible to study the local asymptotic behavior of our test statistic. To analyze the asymptotic local power property of our test, we need to extend MW's asymptotic distribution theory from the case of homogeneous slopes to the case where local

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

deviations from the null are allowed (see eqn. (3.5) in section 3.3). As demonstrated in the supplementary appendix, this extension is nontrivial. The local deviations affect the asymptotic behavior of the estimator of the dominant component, i.e., $\boldsymbol{\beta}$ in eqn. (3.5), in the heterogeneous slope parameters and the asymptotic mean of our test statistic in a fairly complicated but tractable manner.

The remainder of the paper is organized as follows. In Section 2, we introduce the hypotheses and the test statistic. In Section 3 we derive the asymptotic distributions of our test statistic under both the null and a sequence of local Pitman alternatives and propose a bootstrap procedure to obtain the p-values for our test. We also remark on the other potential applications and extensions of our test. In Section 4, we conduct Monte Carlo experiments to evaluate the finite-sample performance of our test and apply it to the OECD economic growth data. Section 5 concludes. All proofs are relegated to Appendix B (online supplementary material).

To proceed, we adopt the following notation. For an $m \times n$ real matrix A, we denote its transpose as A', its Frobenius norm as $\|A\|_F \left(\equiv \left[\mathrm{tr}\left(AA'\right)\right]^{1/2}\right)$, its spectral norm as $\|A\| (\equiv \sqrt{\mu_1 (A'A)})$, where $\equiv$ means "is defined as" and $\mu_1(\cdot)$ denotes the largest eigenvalue of a real symmetric matrix. Note that the two norms are equal when A is a vector, and they can be used interchangeably. More generally, we use $\mu_s(\cdot)$ to denote the sth largest eigenvalue of a real symmetric matrix by counting multiple eigenvalues multiple times. Here $P_A \equiv A\left(A'A\right)^{-1}A'$ and $M_A \equiv I_m - P_A$, where I_m denotes an $m \times m$ identity matrix. When A is symmetric, we use $\mu_{\min}(A)$ to denote its minimum eigenvalue and $A > 0$ to denote that A is positive definite. Let $\mathbf{i}_T$ denote a $T \times 1$ vector of ones. Moreover, the operator $\xrightarrow{P}$ denotes convergence in probability, and $\xrightarrow{D}$ denotes convergence in distribution. We use $(N, T) \to \infty$ to denote the joint convergence of N and T when N and T pass to infinity simultaneously.

2. BASIC FRAMEWORK

In this section, we first specify the null and alternative hypotheses and then introduce the estimation of the restricted model under the null. Finally, we propose a residual-based LM test statistic.

2.1. The Model and Hypotheses

Consider the heterogeneous panel data model with interactive fixed effects

$$Y_{it} = \boldsymbol{\beta}_i^{0\prime} X_{it} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{2.1}$$

where X_{it} is a $K \times 1$ vector of strictly exogenous regressors, $\boldsymbol{\beta}_i^0$ is an $K \times 1$ vector of unknown slope coefficients, λ_i^0 is an $r \times 1$ vector of factor loadings, F_t^0 is an $r \times 1$ vector of common factors, ε_{it} is an idiosyncratic error term, and $\boldsymbol{\beta}_i^0$, λ_i^0, F_t^0,

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

and ε_{it} are unobserved. Here $\{\lambda_i^0\}$ and $\{F_t^0\}$ may be potentially correlated with $\{X_{it}\}$.

The null hypothesis of interest is

$$\mathbb{H}_0: \boldsymbol{\beta}_i^0 = \boldsymbol{\beta}^0 \quad \text{for some } \boldsymbol{\beta}^0 \in \mathbb{R}^K \quad \text{for all } i = 1, \ldots, N. \tag{2.2}$$

The alternative hypothesis is

$$\mathbb{H}_1: \boldsymbol{\beta}_i^0 \neq \boldsymbol{\beta}_j^0 \quad \text{for some } i \neq j. \tag{2.3}$$

To construct a residual-based test for the preceding null hypothesis, we propose to estimate the model under the null hypothesis and obtain the residuals from the regression. The vector X_{it} should not contain any useful information on such residuals under the null and should contain some under the alternative.

2.2. Estimation of the Restricted Model

To proceed, let $X_{it,k}$ denote the kth element of X_{it} for $k = 1, \ldots, K$. Define

$$Y_i \equiv (Y_{i1}, \ldots, Y_{iT})', \qquad X_i \equiv (X_{i1}, \ldots, X_{iT})', \qquad \varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{iT})',$$
$$F^0 \equiv \left(F_1^0, \ldots, F_T^0\right)', \qquad \lambda^0 \equiv \left(\lambda_1^0, \ldots, \lambda_N^0\right)', \qquad X_{i,\cdot k} \equiv (X_{i1,k}, \ldots, X_{iT,k})',$$
$$\mathbf{Y} \equiv (Y_1, \ldots, Y_N)', \qquad \mathbf{X}_k \equiv \left(X_{1,\cdot k}, \ldots, X_{N,\cdot k}\right)', \quad \text{and} \quad \varepsilon \equiv (\varepsilon_1, \ldots, \varepsilon_N)'.$$

Apparently $\mathbf{Y}$, $\mathbf{X}_k$, and ε all denote $N \times T$ matrices. Then under $\mathbb{H}_0$, $Y_i = X_i\boldsymbol{\beta}^0 + F^0\lambda_i^0 + \varepsilon_i$, and we can write the model (2.1) in matrix form:

$$\mathbf{Y} = \sum_{k=1}^{K} \beta_k^0 \mathbf{X}_k + \lambda^0 F^{0\prime} + \varepsilon, \tag{2.4}$$

where $\boldsymbol{\beta}^0 = \left(\beta_1^0, \ldots, \beta_K^0\right)'$.

For the restricted model in (2.4), under the identification restrictions that $F'F/T = I_r$ and $\lambda'\lambda =$ diagonal matrix, Bai (2009a) studies the PCA estimates of $\boldsymbol{\beta}^0$, λ^0, and F^0, which are given by the solutions to the following set of nonlinear equations:

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} X_i' M_{\tilde{F}} X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_{\tilde{F}} Y_i, \tag{2.5}$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\tilde{\boldsymbol{\beta}}\right)\left(Y_i - X_i\tilde{\boldsymbol{\beta}}\right)'\right]\tilde{F} = \tilde{F}V_{NT}, \tag{2.6}$$

and

$$\tilde{\lambda}' = \frac{1}{T}\left[\tilde{F}'\left(Y_1 - X_1\tilde{\boldsymbol{\beta}}\right), \ldots, \tilde{F}'\left(Y_N - X_N\tilde{\boldsymbol{\beta}}\right)\right], \tag{2.7}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

where V_{NT} is a diagonal matrix that consists of the r largest eigenvalues of the bracketed matrix in (2.6), arranged in decreasing order. See Bai (2009a) for a robust iteration scheme to obtain the estimate $(\tilde{\boldsymbol{\beta}}, \tilde{F}, \tilde{\lambda})$.

Moon and Weidner (2010a, 2010b) reinvestigate the Bai (2009a) PCA estimation and put it in the framework of the Gaussian QMLE. Let

$$\left(\hat{\boldsymbol{\beta}}, \hat{\lambda}, \hat{F}\right) = \underset{(\boldsymbol{\beta}, \lambda, F)}{\arg\min}\, \mathcal{L}_{NT}(\boldsymbol{\beta}, \lambda, F), \tag{2.8}$$

where

$$\mathcal{L}_{NT}(\boldsymbol{\beta}, \lambda, F) \equiv \frac{1}{NT}\mathrm{tr}\left[\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k \mathbf{X}_k - \lambda F'\right)'\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k \mathbf{X}_k - \lambda F'\right)\right], \tag{2.9}$$

$\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_K)'$, $F \equiv (F_1, \ldots, F_T)'$, and $\lambda \equiv (\lambda_1, \ldots, \lambda_N)'$. In particular, $\boldsymbol{\beta}^0$ can be estimated by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min} L_{NT}(\boldsymbol{\beta}), \tag{2.10}$$

where the negative profile quasi log-likelihood function $L_{NT}(\boldsymbol{\beta})$ is given by

$$\begin{aligned} L_{NT}(\boldsymbol{\beta}) &= \min_{\lambda, F} \mathcal{L}_{NT}(\boldsymbol{\beta}, \lambda, F) \\ &= \min_{F} \frac{1}{NT}\mathrm{tr}\left[\left(\boldsymbol{Y} - \sum_{k=1}^{K}\beta_k \mathbf{X}_k\right) M_F \left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k \mathbf{X}_k\right)'\right] \\ &= \frac{1}{NT}\sum_{t=r+1}^{T}\mu_t\left[\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k \mathbf{X}_k\right)'\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k \mathbf{X}_k\right)\right]. \end{aligned} \tag{2.11}$$

See MW for the demonstration of the equivalence of the last three expressions.

Clearly, nothing ensures that $L_{NT}(\boldsymbol{\beta})$ is a convex function, and there is no closed form solution to minimizing it. One has to adopt numerical optimization to obtain the QMLE of $\boldsymbol{\beta}^0$. If both $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are global solutions, then they should be identical because the objective functions considered by Bai (2009a) and Moon and Weidner (2010a, 2010b) are identical. Through simulations we find that it is desirable to use the Bai (2009a) estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}^0$ as an initial estimator in the numerical optimization procedure. After one obtains the QMLE $\hat{\boldsymbol{\beta}}$, one could secure the QMLE $(\hat{\lambda}, \hat{F})$ of $\left(\lambda^0, F^0\right)$ according to (2.6) and (2.7). Then we can estimate ε_i by $\hat{\varepsilon}_i = Y_i - X_i\hat{\boldsymbol{\beta}} - \hat{F}\hat{\lambda}_i$ under the null, where $\hat{F} = (\hat{F}_1, \hat{F}_2, \ldots \hat{F}_T)'$ and $\hat{\varepsilon}_i = \left(\hat{\varepsilon}_{i1}, \hat{\varepsilon}_{i2}, \ldots, \hat{\varepsilon}_{iT}\right)'$. It is easy to verify that

$$\hat{\varepsilon}_i = M_{\hat{F}}\varepsilon_i + M_{\hat{F}}X_i\left(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}\right) + M_{\hat{F}}F^0\lambda_i^0 + M_{\hat{F}}X_i\left(\boldsymbol{\beta}_i^0 - \boldsymbol{\beta}^0\right). \tag{2.12}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

2.3. An LM Test for Slope Homogeneity

To motivate our LM test for slope homogeneity, we consider a working auxiliary regression model for $\hat{\varepsilon}_{it}$:

$$\hat{\varepsilon}_{it} = \phi_i' X_{it} + \eta_{it}, \tag{2.13}$$

where for each i, ϕ_i can be regarded as the slope parameter in the time series regression of $\hat{\varepsilon}_{it}$ on X_{it} without an intercept term[2] and η_{it} is an error term that has zero mean. Under $\mathbb{H}_0$, we expect $\phi_i = 0$ in (2.13) for each i because

$$\hat{\varepsilon}_{it} = \left(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}\right)' X_{it} + \lambda_i^{0\prime} F_t^0 - \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it},$$

where $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 \xrightarrow{P} 0$ and $\lambda_i^{0\prime} F_t^0 - \hat{\lambda}_i' \hat{F}_t \xrightarrow{P} 0$ under $\mathbb{H}_0$. Under $\mathbb{H}_1$, $\boldsymbol{\beta}_i^0 - \hat{\boldsymbol{\beta}}$ does not converge to 0 in probability for some i, implying that

$$\hat{\varepsilon}_{it} = \left(\boldsymbol{\beta}_i^0 - \hat{\boldsymbol{\beta}}\right)' X_{it} + \lambda_i^{0\prime} F_t^0 - \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it}$$

will contain some useful information about X_{it} so that we expect $\phi_i \neq 0$ in (2.13) for some i. Therefore we can test $\mathbb{H}_0$ by testing whether

$$\mathbb{H}_0^* : \phi_i = 0 \quad \text{for all } i = 1, \ldots, N \tag{2.14}$$

holds for the auxiliary regression model (2.13).

Pretending η_{it} are independent and identically distributed (i.i.d.) (denoted IID in the mathematical expression in this paper) according to $N\left(0, \sigma^2\right)$ across i and t in (2.13), maximizing the Gaussian quasi log-likelihood of $\hat{\varepsilon}_{it}$ is equivalent to minimizing the following criterion function:

$$\ell(\phi) = \sum_{i=1}^{N} \left(\hat{\varepsilon}_i - X_i \phi_i\right)' \left(\hat{\varepsilon}_i - X_i \phi_i\right),$$

where $\phi = (\phi_1', \ldots, \phi_N')'$. The test of $\mathbb{H}_0^*$ can be based on the LM statistic defined by[3]

$$LM_{NT} = \left(T^{-1/2} \frac{\partial \ell(0)}{\partial \phi}\right)' \left(-T^{-1} \frac{\partial^2 \ell(0)}{\partial \phi \partial \phi'}\right)^{-1} \left(T^{-1/2} \frac{\partial \ell(0)}{\partial \phi}\right), \tag{2.15}$$

where $-T^{-1}(\partial^2 \ell(0))/(\partial \phi \partial \phi')$ serves as an estimate of the information matrix under $\mathbb{H}_0^*$. Noting that $\partial \ell(0)/\partial \phi_i = X_i' \hat{\varepsilon}_i$, and $\partial^2 \ell(0)/\partial \phi_i \partial \phi_j' = -X_i' X_j 1\{i = j\}$ where $1\{\cdot\}$ is the usual indicator function, we have

$$LM_{NT} = \sum_{i=1}^{N} \hat{\varepsilon}_i' X_i \left(X_i' X_i\right)^{-1} X_i' \hat{\varepsilon}_i. \tag{2.16}$$

We will show that after being appropriately scaled and centered, LM_{NT} is asymptotically normally distributed under the null and a sequence of Pitman local alternatives.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Remark 1. In an early version of the paper, we followed the lead of Su and Ullah (2013) and motivated our test statistic through the average of goodness-of-fit statistics (R^2). To this goal, we considered the time series linear regression model

$$\hat{\varepsilon}_{it} = \alpha_i + \phi_i' X_{it} + \eta_{it}, \qquad t = 1, \ldots, T, \tag{2.17}$$

for each cross-sectional unit $i = 1, \ldots, N$, where η_{it} is the error term. As before, under $\mathbb{H}_0$ X_{it} cannot explain the total variation in $\hat{\varepsilon}_{it}$ so that the goodness-of-fit measure R_i^2 for the preceding time series regression should be close to 0, and it deviates from 0 for some cross-sectional units under $\mathbb{H}_1$. So one could base a test on the average of these cross-sectional R_i^2's:

$$\bar{R}_{NT}^2 = \frac{1}{N} \sum_{i=1}^{N} R_i^2. \tag{2.18}$$

A close examination of the asymptotic analysis there suggests that one could consider the time series regression in (2.17) without the intercept term α_i. In this case, $\bar{R}_{NT}^2$ can be interpreted as another LM statistic that takes into account unconditional *cross-sectional* heteroskedasticity in (2.13) explicitly. Let $\sigma_i^2 = \mathbb{E}\left(\varepsilon_{it}^2\right)$ for $i = 1, \ldots, N$. Then one minimizes the following criterion function:

$$\ell_1(\phi, \sigma_1^2, \ldots, \sigma_N^2) = \sum_{i=1}^{N} \left(\hat{\varepsilon}_i - X_i \phi_i\right)' \left(\hat{\varepsilon}_i - X_i \phi_i\right) / \sigma_i^2,$$

and the corresponding LM statistic for testing (2.14) is given by

$$LM_{1NT} \equiv \sum_{i=1}^{N} \hat{\varepsilon}_i' X_i \left(X_i' X_i\right)^{-1} X_i' \hat{\varepsilon}_i / \hat{\sigma}_i^2, \tag{2.19}$$

where $\hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2$. The term LM_{1NT} can be written as $N \bar{R}_{NT}^2$ if one obtains (uncentered) R_i^2 without the intercept term in (2.17). If one allows the intercept term in (2.17), it is easy to verify that the corresponding LM statistic becomes

$$LM_{2NT} \equiv \sum_{i=1}^{N} \hat{\varepsilon}_i' M_0 X_i \left(X_i' M_0 X_i\right)^{-1} X_i' M_0 \hat{\varepsilon}_i / \breve{\sigma}_i^2, \tag{2.20}$$

where $M_0 \equiv I_T - T^{-1} \mathbf{i}_T \mathbf{i}_T'$ and $\breve{\sigma}_i^2 = T^{-1} \hat{\varepsilon}_i' M_0 \hat{\varepsilon}_i$.

As kindly pointed out by an anonymous referee, conditional heteroskedasticity of unknown form is an important concern in empirical applications. So we focus on the analysis of the LM test statistic LM_{NT} in (2.16) by allowing conditional heteroskedasticity of unknown form rather than unconditional cross-sectional heteroskedasticity.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

2.4. Alternative Approaches

Alternatively, we can consider estimating the model (2.1) under the null and alternative hypotheses, respectively, and comparing the restricted and unrestricted estimators of $\boldsymbol{\beta}_i$ in the spirit of the Hausman test. Nevertheless, the Bai (2009a) iterative PCA method is not applicable to heterogeneous panel data models, and we have to resort to the Pesaran (2006) CCE method to obtain the unrestricted estimators of $\boldsymbol{\beta}_i$, $i = 1, \ldots, N$.

PY propose a test of slope homogeneity for large panel data models with fixed effects. Specifically, they consider testing the null that $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i in the following conventional fixed effects panel data model:

$$Y_{it} = \alpha_i + \boldsymbol{\beta}_i' X_{it} + \varepsilon_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T. \tag{2.21}$$

To construct their test statistic, one needs to run both restricted and unrestricted regressions. Let $\hat{\boldsymbol{\beta}}_i \equiv \left(X_i' M_0 X_i\right)^{-1} X_i' M_0 Y_i$, $\hat{\boldsymbol{\beta}}_{FE} \equiv \left(\sum_{i=1}^N X_i' M_0 X_i\right)^{-1} \sum_{i=1}^N X_i' M_0 Y_i$, and $\tilde{\boldsymbol{\beta}}_{WFE} \equiv \left(\sum_{i=1}^N \tilde{\sigma}_{i,\mathbf{PY}}^{-2} X_i' M_0 X_i\right)^{-1} \sum_{i=1}^N \tilde{\sigma}_{i,\mathbf{PY}}^{-2} X_i' M_0 Y_i$, where $\tilde{\sigma}_{i,\mathbf{PY}}^2 = (T-1)^{-1} \left(Y_i - X_i' \hat{\boldsymbol{\beta}}_{FE}\right)' M_0 \left(Y_i - X_i' \hat{\boldsymbol{\beta}}_{FE}\right)$. PY's standardized Swamy test statistic is

$$\tilde{\Delta}_{adj}^{\mathbf{PY}} \equiv \sqrt{\frac{N(T+1)}{T-K-1}} \left(\frac{N^{-1} \tilde{S}^{\mathbf{PY}} - K}{\sqrt{2K}} \right), \tag{2.22}$$

where $\tilde{S}^{\mathbf{PY}} \equiv \sum_{i=1}^N \left(\hat{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\beta}}_{WFE}\right)' X_i' M_0 X_i \left(\hat{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\beta}}_{WFE}\right) / \tilde{\sigma}_{i,\mathbf{PY}}^2$. PY prove that $\tilde{\Delta}_{adj}^{\mathbf{PY}} \xrightarrow{d} N(0,1)$ under certain regularity conditions.

Here, we can also apply PY's method to test $\phi_i = \phi = 0$ for all i in (2.17). In this case, we only need to obtain the unrestricted estimate of ϕ_i by $\hat{\phi}_i = \left(X_i' M_0 X_i\right)^{-1} X_i' M_0 \hat{\varepsilon}_i$ because the analogue of either $\hat{\boldsymbol{\beta}}_{FE}$ or $\tilde{\boldsymbol{\beta}}_{WFE}$ is given by 0. Let $\tilde{\sigma}_i^2 \equiv (T-1)^{-1} \hat{\varepsilon}_i' M_0 \hat{\varepsilon}_i$. Then we can consider the following analogue of $\tilde{S}^{\mathbf{PY}}$:

$$\tilde{S} = \sum_{i=1}^N \left(\hat{\phi}_i - 0\right)' \frac{X_i' M_0 X_i}{\tilde{\sigma}_i^2} \left(\hat{\phi}_i - 0\right) = \sum_{i=1}^N \hat{\varepsilon}_i' M_0 X_i \left(X_i' M_0 X_i\right)^{-1} X_i' M_0 \hat{\varepsilon}_i / \tilde{\sigma}_i^2,$$

which differs from the LM statistic in (2.20) only in the estimation of σ_i^2.

3. ASYMPTOTIC DISTRIBUTIONS

In this section we first present a set of assumptions that are necessary for asymptotic analyses and then study the asymptotic distributions of LM_{NT} under the null hypothesis and a sequence of Pitman local alternatives. We also propose a bootstrap procedure to obtain the bootstrap p-values for our test.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

3.1. Assumptions

Let $\mathcal{D} \equiv \left(F^0, \lambda^0\right)$, $\mathbb{E}_{\mathcal{D}}(A) \equiv \mathbb{E}(A|\mathcal{D})$, and $\|A\|_{q,\mathcal{D}} \equiv [\mathbb{E}_{\mathcal{D}}(\|A\|_F^q)]^{1/q}$. Define $\mathcal{F}_{NT,t} \equiv \sigma\left(\mathcal{D}, \{X_{i,t+1}, X_{it}, \varepsilon_{it}, X_{i,t-1}, \varepsilon_{i,t-1}, \ldots\}_{i=1}^N\right)$. We make the following assumptions.

Assumption A.1.

(a) $T^{-1}\sum_{t=1}^T \mathbb{E}||F_t^0||^{8+4\sigma} = O(1)$ for some $\sigma > 0$ and $T^{-1}F^{0\prime}F^0 \xrightarrow{P} \Sigma_F > 0$ for some $r \times r$ matrix Σ_F as $T \to \infty$.

(b) $N^{-1}\sum_{i=1}^N \mathbb{E}||\lambda_i^0||^{8+4\sigma} = O(1)$ and $N^{-1}\lambda^{0\prime}\lambda^0 \xrightarrow{P} \Sigma_\lambda > 0$ for some $r \times r$ matrix Σ_λ as $N \to \infty$.

(c) $(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\|\zeta_{it}\|^{8+4\sigma} = O(1)$ for $\zeta_{it} = X_{it}$, ε_{it}, and $X_{it}\varepsilon_{it}$.

(d) $\min_{1\le i\le N} \mu_{\min}\left(\mathbf{\Omega}_i\right) \ge \underline{c}_X$ a.s. for some $\underline{c}_X > 0$ where $\mathbf{\Omega}_i \equiv T^{-1}\mathbb{E}_{\mathcal{D}}(X_i'X_i)$.

(e) Let $\mathbf{X}_{(\alpha)} = \sum_{k=1}^K \alpha_k \mathbf{X}_k$ such that $\|\alpha\| = 1$ where $\alpha = (\alpha_1, \ldots, \alpha_K)'$. There exists a finite constant $C_r > 0$ such that $\min_{\{\alpha\in\mathbb{R}^K:\|\alpha\|=1\}} \sum_{t=2r+1}^T \mu_t\left(\mathbf{X}_{(\alpha)}'\mathbf{X}_{(\alpha)}\right) \ge C_r$ with probability approaching 1.

(f) $\|\boldsymbol{\varepsilon}\| = O_P(\max(\sqrt{N}, \sqrt{T}))$.

Assumption A.2.

(a) For each $i = 1, \ldots, N$, $\{(X_{it}, \varepsilon_{it}) : t = 1, 2, \ldots\}$ is conditionally strong mixing given $\mathcal{D}$ with mixing coefficients $\left\{\alpha_{NT,i}^{\mathcal{D}}(\cdot)\right\}$. Here $\alpha_{\mathcal{D}}(\cdot) \equiv \alpha_{NT}^{\mathcal{D}}(\cdot) \equiv \max_{1\le i\le N} \alpha_{NT,i}^{\mathcal{D}}(\cdot)$ satisfies $\alpha_{\mathcal{D}}(s) = O_{a.s.}\left(s^{-\rho}\right)$ where $\rho = 3(2+\sigma)/\sigma + \epsilon$ for some arbitrarily small $\epsilon > 0$ and σ is as defined in Assumption A.1(a). In addition, there exist integers τ_0, $\tau_* \in (1, T)$ such that $NT\alpha_{\mathcal{D}}(\tau_0) = o_{a.s.}(1)$, $T(T + N^{1/2})\alpha_{\mathcal{D}}(\tau_*)^{(1+\sigma)/(2+\sigma)} = o_{a.s.}(1)$, and $N^{1/2}T^{-1}\tau_*^2 = o(1)$.

(b) (ε_i, X_i), $i = 1, \ldots, N$, are mutually independent of each other conditional on $\mathcal{D}$.

(c) For each $i = 1, \ldots, N$, $\mathbb{E}\left(\varepsilon_{it}|\mathcal{F}_{NT,t-1}\right) = 0$ a.s.

Assumption A.3.

(a) As $(N, T) \to \infty$, $N^{3/4}/T \to 0$ and $T^{2/3}/N \to 0$.

(b) As $(N, T) \to \infty$, $(NT)^{1/(8+4\sigma)}T^{1/2}N^{-1} \to 0$ and $N^{1/8}(NT)^{3/(8+4\sigma)}\log(NT)/T \to 0$.

A.1(a)–(c) mainly impose moment conditions on F_t^0, λ_i^0, X_{it}, and ε_{it}. Note that we require finite-eighth plus moments for F_t^0, λ_i^0, X_{it}, ε_{it}, and $X_{it}\varepsilon_{it}$ to derive the asymptotic distribution of our feasible test statistic in what follows. Some of the moment conditions can be weakened for the proof of Theorem 3.1. Admittedly, our moment conditions are generally stronger than those assumed

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

in the literature for the purpose of estimation (e.g., Bai, 2009a) or for testing slope homogeneity in conventional panel data models with additive fixed effects (e.g., PY). For example, Bai (2009a) only requires finite fourth moments for F_t^0, λ_i^0, and X_{it} and finite eighth moments for ε_{it}; he assumes independence between ε_{it} and $(X_{js}, F_s^0, \lambda_j^0)$ for all i, j, t, s and thus does not need conditions on the cross product $X_{it}\varepsilon_{it}$. PY assume finite second and ninth moments for X_{it} and ε_{it}, respectively. A.1(d) requires that $\mathbb{E}_{\mathcal{D}}(X_i'X_i)/T$ be positive definite a.s. uniformly in i. A.1(e) and (f) are identical to Assumption $2(ii)$ and Assumption $1(ii)$ in Moon and Weidner (2010a), respectively. As remarked by the latter authors, A.1(e) imposes the usual noncollinearity condition on $\mathbf{X}_k$ and A.1(f) can be satisfied for various error processes. With more complicated analysis, it is possible to relax either assumption.

A.2(a) requires that each individual time series $\{(X_{it}, \varepsilon_{it}) : t = 1, 2, \ldots\}$ be strong mixing conditional on $\mathcal{D}$ (or $\mathcal{D}$-strong-mixing). See Appendix A (online supplementary material) for the definition of conditional strong mixing. To acknowledge the fact that the conditioning set $\mathcal{D}$ depends on the sample sizes N and T, we use $\alpha_{NT,i}^{\mathcal{D}}(\cdot)$ to denote the $\mathcal{D}$-strong-mixing coefficient for the ith individual time series. Prakasa Rao (2009) extends the concept of (unconditional) strong mixing to conditional strong mixing for a sequence of random variables. In analogy with the relationship between independence and strong mixing (asymptotic independence), conditional strong mixing generalizes the concept of conditional independence and requires variables that lie far apart in time to be approximately independent given the conditional information. It is well known that neither conditional independence nor independence implies the other. Similarly, conditional strong mixing does not imply strong mixing for a sequence of random variables or vice versa. To appreciate the importance of conditioning, consider the simple AR(1) panel data model with interactive fixed effects

$$Y_{it} = \rho_0 Y_{i,t-1} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T. \tag{3.1}$$

Even if $\left\{\left(\varepsilon_{it}, F_t^0\right), t \geq 1\right\}$ is a strong mixing process, $\{Y_{it}, t \geq 1\}$ is generally not unless λ_i^0 is nonstochastic. For this reason, Hahn and Kuersteiner (2011) assume that the individual fixed effects are nonrandom and uniformly bounded in their study of nonlinear dynamic panel data models. In the case of random fixed effects, they suggest adopting the concept of conditional strong mixing where the mixing coefficient is defined by conditioning on the fixed effects. Here we define the conditional strong mixing processes by conditioning on $\mathcal{D} = \left(F^0, \lambda^0\right)$, which, in conjunction with A.2(b), will greatly simplify the proofs of some technical lemmas in Appendix A (online supplementary material) and Proposition B.1 in various places. Note that we only require that the mixing coefficients decay at an algebraic rate, which is weaker than the geometric decay rate imposed by Hahn and Kuersteiner (2011). The dependence of the mixing rate on σ in A.2(a) and A.1 reflects the trade-off between the degree of dependence and the moment bounds of the process $\{(X_{it}, \varepsilon_{it}), t \geq 1\}$. The last set of conditions in A.2(a) can easily be met. For

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

clarity, assume for the moment that $N/T \to \kappa \in (0,\infty)$ as $(N,T) \to \infty$. These conditions will be satisfied by taking $\tau_0 = T^{\nu_0}$ and $\tau_* = T^{\nu_*}$ for some $\nu_0 \in (2/\rho, 1)$ and $\nu_* \in (2(2+\sigma)/[\rho(1+\sigma)], \frac{1}{4})$ provided $2(2+\sigma)/[\rho(1+\sigma)] < \frac{1}{4}$, which is satisfied if $\sigma \leq 3/5$ or ϵ is not too small in A.2(a). If the process is strong mixing with a geometric mixing rate, the conditions on $\alpha_{\mathcal{D}}(\cdot)$ can easily be met by specifying $\tau_0 = \tau_* = \lfloor C_\tau \log T \rfloor$ for some sufficiently large C_τ, where $\lfloor a \rfloor$ denotes the integer part of a.

It is worth mentioning that Assumption A.2(b) does not rule out cross-sectional dependence among $(X_{it}, \varepsilon_{it})$. When $X_{it} = Y_{i,t-1}$ and ε_{it} exhibits conditional heteroskedasticity (e.g., $\varepsilon_{it} = \sigma_0(Y_{i,t-1})\epsilon_{it}$ where $\epsilon_{it} \backsim \text{IID}(0,1)$ and $\sigma_0(\cdot)$ is an unknown smooth function) as in (3.1), $(X_{it}, \varepsilon_{it})$ are not independent across i because of the presence of common factors irrespective of whether one allows λ_i^0 to be independent across i or not. Nevertheless, conditional on $\mathcal{D}$, it is possible that $(X_{it}, \varepsilon_{it})$ is independent across i such that A.2(b) is still satisfied. Here the cross-sectional dependence is similar to the type of cross-sectional dependence generated by common shocks studied by Andrews (2005). The difference is that Andrews (2005) assumes i.i.d. observations conditional on the σ-field generated by the common shocks in a cross-section framework, whereas we have *conditionally independent but nonidentically distributed* observations across the individual dimension in a panel framework.[4]

A.2(c) requires that the error term ε_{it} be a martingale difference sequence (m.d.s.) with respect to the filter $\mathcal{F}_{NT,t}$ which allows for lagged dependent variables in X_{it} and conditional heteroskedasticity, skewness, or kurtosis of unknown form in ε_{it}. In sharp contrast, both Bai (2009a) and Pesaran (2006) assume that ε_{it} is independent of X_{js}, λ_j, and F_s for all i, t, j, and s. MW allow dynamics but assume that ε_{it}'s are independent across both i and t. As a referee kindly points out, allowing lagged dependent variables broadens the potential applicability of our test. It can be used to potentially ameliorate problems caused by serial dependence in the error term. To see this, consider a static model of the form

$$Y_{it} = \boldsymbol{\beta}^{0\prime} X_{it} + \lambda_i^{0\prime} F_t^0 + e_{it}, \qquad i = 1,\ldots,N, \qquad t = 1,\ldots,T, \tag{3.2}$$

where $e_{it} = \rho^0 e_{i,t-1} + \varepsilon_{it}$ and ε_{it} is an m.d.s. Noting that

$$Y_{it} = \boldsymbol{\beta}^{0\prime} X_{it} + \rho^0 Y_{i,t-1} - \rho^0 \boldsymbol{\beta}^{0\prime} X_{i,t-1} + \lambda_i^{0\prime} F_t^0 - \rho^0 \lambda_i^{0\prime} F_{t-1}^0 + \varepsilon_{it}, \tag{3.3}$$

we can obtain a consistent estimate of $\boldsymbol{\beta}^0$ by considering the regression of Y_{it} on X_{it}, $Y_{i,t-1}$, and $X_{i,t-1}$ with interactive fixed effects characterized by $2r$ unobservable factors. Even though such an approach does not impose the restrictions on the parameters in (3.3) and may result in some efficiency loss, it provides a straightforward solution to the problem of first-order serial correlation in the error process. The extension to the case of general AR(p) error process is also feasible. See Greenaway-McGrevy et al. (2012) for a similar approach in the literature on interactive fixed effects.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

A.3(a) and (b) impose conditions on the rates at which N and T pass to infinity and on the interaction between (N,T) and σ. It is worth mentioning that MW only consider the distributional theory under the assumption that N and T pass to infinity at the same rate, whereas Bai (2009a) also considers the case where $T/N \to 0$ or $N/T \to 0$ in the absence of serial correlation and heteroskedasticity (see Bai, 2009a, Thm. 2). Here we allow N and T to pass to infinity at either identical or suitably restricted different rates. If the conditional mixing process $\{(X_{it}, \varepsilon_{it}), t \geq 1\}$ has geometric decay rate, one can take σ in A.1 arbitrarily small. In this case A.3(b) puts the following most stringent restrictions on (N,T) by passing $\sigma \to 0 : N^{4/5}/T \to 0$ and $T^{5/7}/N \to 0$ as $(N,T) \to \infty$, ignoring the logarithm term. On the other hand, if $\sigma \geq 0.5$ in A.1, then A.3(b) becomes redundant under A.3(a), which specifies the minimum requirement on (N,T). Note that A.3(a) is stronger than the minimum requirement ($T/N^2 \to 0$ and $N/T^2 \to 0$) in Bai (2003) for $\sqrt{N}$- and $\sqrt{T}$-consistent estimation of factors and factor loadings, respectively. It reflects the asymmetric roles played by N and T in the construction of our test statistic. In the case of conventional panel data models with strictly exogenous regressors only, PY require that either $\sqrt{N}/T \to 0$ or $\sqrt{N}/T^2 \to 0$ for two of their tests, but for stationary dynamic panel data models, they prove the asymptotic validity of their test only under the condition that $N/T \to \kappa \in [0, \infty)$.

3.2. Asymptotic Null Distribution

Let $h_{i,ts}$ denote the (t,s)th element of $H_i \equiv M_{F^0} P_{X_i} M_{F^0}$. Let $b_{it} \equiv X_{it} - T^{-1} \sum_{s=1}^{T} F_t^{0\prime} \left(F^{0\prime} F^0 / T\right)^{-1} F_s^0 \, \mathbb{E}_{\mathcal{D}}(X_{is})$ and $\bar{b}_{it} \equiv \Omega_i^{-1/2} b_{it}$. Define

$$B_{NT} \equiv N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2 h_{i,tt} \quad \text{and}$$

$$V_{NT} \equiv 4T^{-2} N^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} \mathbb{E}_{\mathcal{D}} \left[\varepsilon_{it} \bar{b}_{it}^{\prime} \sum_{s=1}^{t-1} \bar{b}_{is} \varepsilon_{is} \right]^2. \tag{3.4}$$

The following theorem states the asymptotic null distribution of the infeasible statistic LM_{NT}.

THEOREM 3.1. *Suppose Assumptions A.1–A.3 hold. Then under* $\mathbb{H}_0$,

$$J_{NT} \equiv \left(N^{-1/2} LM_{NT} - B_{NT} \right) / \sqrt{V_{NT}} \xrightarrow{D} N(0,1).$$

Remark 2. The proof of the preceding theorem is tedious and is relegated to Appendix B (online supplementary material). The key step in the proof is to show that under $\mathbb{H}_0$, $\sqrt{V_{NT}} J_{NT} = A_{NT} + o_P(1)$, where $A_{NT} \equiv \sum_{t=2}^{T} Z_{NT,t}$ and $Z_{NT,t} \equiv 2T^{-1} N^{-1/2} \sum_{i=1}^{N} \sum_{s=1}^{t-1} \varepsilon_{it} \varepsilon_{is} \bar{b}_{it}^{\prime} \bar{b}_{is}$. By construction, $\{Z_{NT,t}, \mathcal{F}_{NT,t}\}$ is an m.d.s. so that we can apply the martingale central limit theorem (CLT) to show that $A_{NT}/\sqrt{V_{NT}} \xrightarrow{D} N(0,1)$ under Assumptions A.1–A.3.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

To implement the test, we need consistent estimates of both B_{NT} and V_{NT}. We propose to estimate them, respectively, by

$$\hat{B}_{NT} = N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 \hat{h}_{i,tt} \quad \text{and} \quad \hat{V}_{NT} = 4T^{-2}N^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left[\hat{\varepsilon}_{it} \hat{b}_{it}' \sum_{s=1}^{t-1} \hat{b}_{is} \hat{\varepsilon}_{is} \right]^2,$$

where $\hat{h}_{i,ts}$ denotes the (t,s)th element of $\hat{H}_i \equiv M_{\hat{F}} P_{X_i} M_{\hat{F}}$, $\hat{b}_{it} = \hat{\Omega}_i^{-1/2}(X_{it} - T^{-1}\sum_{s=1}^{T} \hat{F}_t' \hat{F}_s X_{is})$, and $\hat{\Omega}_i = T^{-1} X_i' X_i$.[5] Then we can define a feasible test statistic:

$$\hat{J}_{NT} \equiv \left(N^{-1/2} L M_{NT} - \hat{B}_{NT} \right) / \sqrt{\hat{V}_{NT}}.$$

The following theorem establishes the consistency of $\hat{B}_{NT}$ and $\hat{V}_{NT}$ and the asymptotic distribution of $\hat{J}_{NT}$ under $\mathbb{H}_0$.

THEOREM 3.2. *Suppose Assumptions A.1–A.3 hold. Then under* $\mathbb{H}_0$, $\hat{B}_{NT} = B_{NT} + o_P(1)$, $\hat{V}_{NT} = V_{NT} + o_P(1)$, *and* $\hat{J}_{NT} \xrightarrow{d} N(0,1)$.

Remark 3. Theorem 3.2 implies that the test statistic $\hat{J}_{NT}$ is asymptotically pivotal. We can compare $\hat{J}_{NT}$ with the one-sided critical value z_α, i.e., the upper αth percentile from the standard normal distribution, and reject the null when $\hat{J}_{NT} > z_\alpha$ at the asymptotic α significance level.

Remark 4. We obtain the preceding distributional results despite the fact that the unobserved factors and factor loadings can only be estimated at slower rates (uniformly $N^{-1/2}$ for the former and uniformly $T^{-1/2}$ for the latter) than that at which the homogeneous slope parameter $\boldsymbol{\beta}$ can be estimated under the null under the conditions that $N/T^2 \to 0$ and $T^2/N \to 0$ (see Bai, 2003). The slow convergence rates of these factor and factor loadings estimates do not have adverse asymptotic effects on the estimation of the bias term B_{NT}, the variance term V_{NT}, and the asymptotic distribution of $\hat{J}_{NT}$. Nevertheless, they can play an important role in finite samples. For this reason, we will also propose a residual-based bootstrap procedure to obtain the bootstrap p-values for the $\hat{J}_{NT}$ test.

3.3. Asymptotic Local Power Property

To examine the asymptotic local power property of our test, we consider the following sequence of Pitman local alternatives:

$$\mathbb{H}_1(\gamma_{NT}) : \boldsymbol{\beta}_i^0 = \boldsymbol{\beta}^0 + \gamma_{NT}\delta_i \quad \text{for } i = 1, 2, \ldots, N, \tag{3.5}$$

where the δ_i's are $K \times 1$ vectors of fixed constants such that $\|\delta_i\| < C_\delta$ for all i and $\delta_i \neq \delta_j$ for some pair $i \neq j$.

Let $\alpha_{ik} \equiv \lambda_i^{0\prime} \left(\lambda^{0\prime} \lambda^0 / N \right)^{-1} \lambda_k^0$ and $\tilde{X}_i \equiv M_{F^0} X_i - N^{-1} \sum_{j=1}^{N} \alpha_{ij} M_{F^0} X_j$. Let D_{NT} denote a $K \times K$ matrix whose (k_1, k_2)th element is given by[6]

$$D_{NT,k_1k_2} = (NT)^{-1} \operatorname{tr}\left(M_{\lambda^0} \mathbf{X}_{k_1} M_{F^0} \mathbf{X}_{k_2}' \right). \tag{3.6}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Let Π_{NT} be a $K \times 1$ vector whose kth element is given by

$$\Pi_{NT,k} = (NT)^{-1} \operatorname{tr}\left(M_{\lambda^0} \mathbf{X}_k M_{F^0} \boldsymbol{\Delta}'\right), \tag{3.7}$$

where $\boldsymbol{\Delta}$ is an $N \times T$ matrix whose (i,t)th element is given by $X_{it}' \delta_i$. Following the remark after Lemma A.2 in Appendix A (online supplementary material) we have that under $\mathbb{H}_1(\gamma_{NT})$ with $\gamma_{NT} = N^{-1/4} T^{-1/2}$,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \gamma_{NT} D_{NT}^{-1} \Pi_{NT} + o_P(\gamma_{NT}) = O_P(\gamma_{NT}). \tag{3.8}$$

Define

$$\Theta_{NT} = \frac{1}{NT} \sum_{l=1}^{N} \left(M_{F^0} X_i \delta_i - \tilde{X}_i D_{NT}^{-1} \Pi_{NT}\right)' P_{X_i} \left(M_{F^0} X_i \delta_i - \tilde{X}_i D_{NT}^{-1} \Pi_{NT}\right). \tag{3.9}$$

The following theorem establishes the asymptotic distribution of $\hat{J}_{NT}$ under $\mathbb{H}_1\left(N^{-1/4} T^{-1/2}\right)$.

THEOREM 3.3. *Suppose Assumptions A.1–A.3 hold. Suppose that* $\Theta_0 \equiv \operatorname{plim}_{(N,T)\to\infty} \Theta_{NT}$ *and* $V_0 \equiv \operatorname{plim}_{(N,T)\to\infty} V_{NT} > 0$ *exist. Then under* $\mathbb{H}_1\left(N^{-1/4} T^{-1/2}\right)$ *we have* $\hat{J}_{NT} \xrightarrow{D} N(\Theta_0/\sqrt{V_0}, 1)$.

Remark 5. Theorem 3.3 implies that our test has nontrivial asymptotic power against the sequence of local alternatives that deviate from the null at the rate $N^{-1/4} T^{-1/2}$ provided $\Theta_0 > 0$, and the asymptotic local power function is given by $P\left(\hat{J}_{NT} > z | \mathbb{H}_1\left(N^{-1/4} T^{-1/2}\right)\right) \to 1 - \Phi\left(z - \Theta_0/\sqrt{V_0}\right)$ where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. As either N or T increases, the power of our test will increase, but it is expected to increase faster as $T \to \infty$ than as $N \to \infty$. The rate $N^{-1/4} T^{-1/2}$ is the same as that obtained by PY, indicating that the estimation of factors and factor loadings does not affect the rate at which our test can detect the local alternatives.

Remark 6. The requirement $\Theta_0 > 0$ imposes some restrictions on the degree of slope heterogeneity under the local alternatives and on the interactions between the heterogeneity parameters δ_i, the observed regressors X_i, and the unobserved factors F_t^0. In terms of the degree of slope heterogeneity, it requires that $\boldsymbol{\beta}_i^0$ and $\boldsymbol{\beta}_j^0$ differ from each other for a "large" number of pairs (i, j) with $i \neq j$. In particular, it rules out the case where only a fixed number of slope parameters are distinct from a finite number of others (e.g., only $\boldsymbol{\beta}_1^0$ is different from a finite number of other slope coefficients) and the case where the distinct number of elements in $\left\{\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \ldots, \boldsymbol{\beta}_N^0\right\}$ is diverging to infinity as $N \to \infty$ but at a rate slower than N. It is worth mentioning that our test has power in the case where individual slopes can be classified into a finite number of groups, e.g.,

$$\boldsymbol{\beta}_i^0 = \begin{cases} \boldsymbol{\beta}_{(1)}^0 \text{ if } i \in G_1 \\ \boldsymbol{\beta}_{(2)}^0 \text{ if } i \in G_2, \end{cases}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

where G_1 and G_2 form a partition for $\{1,2,\ldots,N\}$. In terms of interactions between δ_i, X_i, and F_t^0, the expression of Θ_{NT} in (3.9) is too complicated to analyze. Using the expressions for $\tilde{X}_i$ and Π_{NT}, we can rewrite Θ_{NT} as $\Theta_{NT} = (NT)^{-1}\sum_{i=1}^N G_i' H_i G_i$ where

$$G_i \equiv X_i\delta_i - X_i D_{NT}^{-1}\frac{1}{NT}\sum_{k=1}^N \tilde{X}_k' X_k\delta_k - \frac{1}{N}\sum_{k=1}^N a_{ik}X_k\delta_k,$$

$D_{NT}^{-1}\frac{1}{NT}\sum_{k=1}^N \tilde{X}_k' X_k\delta_k$ can be viewed as a weighted average of δ_k's, and $\frac{1}{N}\sum_{k=1}^N a_{ik}X_k\delta_k$ is a weighted average of $X_k\delta_k$. Apparently, Θ_{NT} is a quadratic function of $(\delta_1,\ldots,\delta_N)$, and it is 0 under $\mathbb{H}_0$ and no less than 0 otherwise. To simplify the expression for Θ_{NT}, we hypothesize that F_t^0 is either observable or absent from the model. If F_t^0 were observable, then following Bai (2009a) $D_{NT} = D_{F^0} \equiv (NT)^{-1}\sum_{i=1}^N X_i' M_{F^0} X_i$ and $\tilde{X}_k = M_{F^0}X_k$ so that Θ_0 would reduce to the probability limit of

$$\frac{1}{NT}\sum_{i=1}^N \left\{X_i\delta_i - X_i D_{F^0}^{-1}\frac{1}{NT}\sum_{k=1}^N X_k' M_{F^0}X_k\delta_k\right\}' \times H_i \left\{X_i\delta_i - X_i D_{F^0}^{-1}\frac{1}{NT}\sum_{k=1}^N X_k' M_{F^0}X_k\delta_k\right\},$$

where $M_{F^0}X_i\delta_i - M_{F^0}X_i D_{F^0}^{-1}\frac{1}{NT}\sum_{k=1}^N X_k' M_{F^0}X_k\delta_k$ denotes the residual from the $\mathcal{L}_2$ projection of $M_{F^0}X_i\delta_i$ on the space spanned by the columns of $M_{F^0}X_i$. If F_t^0 were absent in the model, then Θ_0 further reduces to the probability limit of $\frac{1}{NT}\sum_{i=1}^N \delta_i' X_i' X_i\delta_i$, and apparently the requirement that Θ_0 be strictly positive does not seem stringent at all.

Remark 7. We motivate our LM test statistics by considering the regression model in (2.13), which does not contain an intercept term. Alternatively, as a referee suggests, we could include an intercept term in the preceding regression. In this case, the LM statistic becomes

$$\widetilde{LM}_{NT} = \sum_{i=1}^N \hat{\varepsilon}_i' M_0 X_i \left(X_i' M_0 X_i\right)^{-1} X_i' M_0\hat{\varepsilon}_i = \sum_{i=1}^N \hat{\varepsilon}_i' P_{M_0X_i}\hat{\varepsilon}_i, \tag{3.10}$$

where $P_{M_0X_i} = M_0X_i\left(X_i'M_0X_i\right)^{-1}X_i'M_0$. The presence of the demeaned operator M_0 along the time dimension would complicate the asymptotic analysis a great deal because it would introduce another layer of summation whenever it appears. Let $\tilde{h}_{i,ts}$ denote the (t,s)th element of $\tilde{H}_i \equiv M_{F^0}P_{M_0X_i}M_{F^0}$. Let $\overrightarrow{b}_{it} = \left[X_{it} - \mathbb{E}_{\mathcal{D}}(\bar{X}_{i\cdot})\right] - T^{-1}\sum_{s=1}^T F_t^{0\prime}\left(F^{0\prime}F^0/T\right)^{-1}F_s^0\left[\mathbb{E}_{\mathcal{D}}(X_{is}) - \mathbb{E}_{\mathcal{D}}(\bar{X}_{i\cdot})\right]$ and $\tilde{\Omega}_i = T^{-1}\sum_{t=1}^T \mathrm{Var}_{\mathcal{D}}\left(X_{it} - \bar{X}_{i\cdot}\right)$ where $\bar{X}_{i\cdot} = T^{-1}\sum_{s=1}^T X_{is}$. Let $\tilde{b}_{it} =$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$\tilde{\Omega}_i^{-1/2}\overrightarrow{b}_{it}$. Define the asymptotic bias and variance terms, respectively, as

$$\tilde{B}_{NT} = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2\tilde{h}_{i,tt} \quad \text{and}$$

$$\tilde{V}_{NT} = 4T^{-2}N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbb{E}_D\left[\varepsilon_{it}\tilde{b}_{it}'\sum_{s=1}^{t-1}\tilde{b}_{is}\varepsilon_{is}\right]^2. \tag{3.11}$$

They can be estimated, respectively, by

$$\breve{B}_{NT} = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\breve{h}_{i,tt} \quad \text{and} \quad \breve{V}_{NT} = 4T^{-2}N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\hat{\varepsilon}_{it}\breve{b}_{it}'\sum_{s=1}^{t-1}\breve{b}_{is}\hat{\varepsilon}_{is}\right]^2, \tag{3.12}$$

where $\breve{h}_{i,ts}$ denote the (t,s)th element of $\breve{H}_i \equiv M_{\hat{F}}P_{M_0X_i}M_{\hat{F}}$, $\breve{b}_{it} = \breve{\Omega}_i^{-1/2}\left[(X_{it} - \bar{X}_{i\cdot}) - T^{-1}\sum_{s=1}^{T}\hat{F}_t'\hat{F}_s\,(X_{is} - \bar{X}_{i\cdot})\right]$, and $\breve{\Omega}_i = T^{-1}X_i'M_0X_i$. Define

$$\breve{J}_{NT} \equiv \left(N^{-1/2}\widetilde{LM}_{NT} - \breve{B}_{NT}\right)/\sqrt{\breve{V}_{NT}}. \tag{3.13}$$

Following the asymptotic analyses in the Appendix B (online supplementary material) we can show that

$$\breve{J}_{NT} \xrightarrow{d} N\left(\sqrt{\Theta_\infty/V_\infty}, 1\right) \quad \text{under } \mathbb{H}_1\left(N^{-1/4}T^{-1/2}\right), \tag{3.14}$$

where $\tilde{\Theta}_{NT} = (NT)^{-1}\sum_{l=1}^{N}(M_{F^0}X_i\delta_i - \tilde{X}_iD_{NT}^{-1}\Pi_{NT})'P_{M_0X_i}(M_{F^0}X_i\delta_i - \tilde{X}_iD_{NT}^{-1}\Pi_{NT})$, and we assume that both $\operatorname{plim}_{(N,T)\to\infty}\tilde{\Theta}_{NT} = \Theta_\infty$ and $\operatorname{plim}_{(N,T)\to\infty}\breve{V}_{NT} = V_\infty$ exist. If $\bar{X}_{i\cdot} = 0$, i.e., $\{X_{it}, t = 1,\ldots,T\}$ is a demeaned process for each i, then we can demonstrate that the asymptotic local power function for $\breve{J}_{NT}$ is the same as that for $\hat{J}_{NT}$. In fact, if we demean X_{it} along the time dimension for each i before calculating $\breve{J}_{NT}$ and $\hat{J}_{NT}$, the two test statistics are identical and thus have the same asymptotic properties. In the general case, it is hard to compare the two tests in terms of asymptotic local power. Our limited simulation results suggest that in general $\breve{J}_{NT}$ is less powerful than $\hat{J}_{NT}$, and thus we only focus on the study of $\hat{J}_{NT}$ in this paper.

Remark 8. Under the global alternative $\mathbb{H}_1$, we can define the pseudo-true parameter $\boldsymbol{\beta}^*$ as the probability limit of $\hat{\boldsymbol{\beta}}$. Let $\bar{\boldsymbol{\Delta}}$ denote an $N \times T$ matrix whose (i,t)th element is given by $\bar{\Delta}_{it} \equiv X_{it}'(\boldsymbol{\beta}_i^0 - \boldsymbol{\beta}^*)$. Unless $\left\|\bar{\boldsymbol{\Delta}}\right\| = o_P(\sqrt{NT})$ the proof of Lemma A.2 breaks down so that a rigorous treatment of the asymptotic behavior of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$ seems impossible under general global alternative. Let $\bar{\Delta}_i \equiv (\bar{\Delta}_{i1},\ldots,\bar{\Delta}_{iT})'$. Heuristically, one expects that $\hat{\varepsilon}_{it} = \varepsilon_{it} + \bar{\Delta}_{it} + o_P(1)$ and

$$\begin{aligned} N^{-1}T^{-1}LM_{NT} &= N^{-1}T^{-1}\sum_{i=1}^{N}\left(\varepsilon_i + \bar{\Delta}_i\right)'P_{X_i}\left(\varepsilon_i + \bar{\Delta}_i\right) + o_P(1) \\ &= N^{-1}T^{-1}\sum_{i=1}^{N}\bar{\Delta}_i'P_{X_i}\bar{\Delta}_i + o_P(1), \end{aligned}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

which has a positive probability limit under some suitable conditions. This, together with the fact that $\hat{B}_{NT} = O_P(N^{1/2})$ and $\hat{V}_{NT} = O_P(1)$ under $\mathbb{H}_1$, implies that $\hat{J}_{NT} = \left(N^{-1/2}LM_{NT} - \hat{B}_{NT}\right)/\sqrt{\hat{V}_{NT}}$ would diverge to infinity for fixed alternatives at rate $N^{1/2}T$ as $(N,T) \to \infty$ provided $\text{plim}_{(N,T)\to\infty} N^{-1}T^{-1}$ $\sum_{i=1}^{N} \bar{\Delta}_i' P_{X_i} \bar{\Delta}_i > 0$. This suggests that $\hat{J}_{NT}$ is consistent and is expected to diverge to infinity at rate $N^{1/2}T$ for general global alternatives.

3.4. A Bootstrap Version of the Test

As mentioned previously, because of the slow convergence rates of the factors and factor loadings estimates, the asymptotic normal null distribution of our test statistic may not approximate its finite-sample distribution well in practice. Therefore it is worthwhile to propose a bootstrap procedure to improve the finite-sample performance of our test. Subsequently we propose a fixed-design wild bootstrap (WB) method to obtain the bootstrap p-values for out test. The procedure goes as follows:

1. Estimate the restricted model in (2.4) and obtain the residuals $\hat{\varepsilon}_{it} = Y_{it} - \hat{\boldsymbol{\beta}}' X_{it} - \hat{\lambda}_i' \hat{F}_t$, where $\hat{\boldsymbol{\beta}}$, $\hat{\lambda}_i$, and $\hat{F}_t$ are estimates under $\mathbb{H}_0$. Calculate the test statistic $\hat{J}_{NT}$ based on $\{\hat{\varepsilon}_{it}, X_{it}, \hat{F}_t\}$.
2. For $i = 1, \ldots, N$ and $t = 1, 2, \ldots, T$, obtain the bootstrap error $\varepsilon_{it}^* = \hat{\varepsilon}_{it}\varsigma_{it}$ where ς_{it} are i.i.d. $N(0,1)$ across i and t. Generate the bootstrap analogue Y_{it}^* of Y_{it} by holding $(X_{it}, \hat{F}_t, \hat{\lambda}_i)$ as fixed:[7] $Y_{it}^* = \hat{\boldsymbol{\beta}}' X_{it} + \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it}^*$ for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$.
3. Given the bootstrap resample $\{Y_{it}^*, X_{it}\}$, run the restricted model estimation and obtain the bootstrap residuals $\hat{\varepsilon}_{it}^* = Y_{it}^* - \hat{\boldsymbol{\beta}}^{*\prime} X_{it} - \hat{\lambda}_i^{*\prime} \hat{F}_t^*$, where $\hat{\boldsymbol{\beta}}^*$, $\hat{\lambda}_i^*$, and $\hat{F}_t^*$ are the Gaussian QMLEs of $\hat{\boldsymbol{\beta}}$, $\hat{\lambda}_i$, and $\hat{F}_t$, respectively. Calculate the test statistic $\hat{J}_{NT}^*$ based on $\{\hat{\varepsilon}_{it}^*, X_{it}, \hat{F}_t^*\}$.
4. Repeat steps 2 and 3 for B times and index the bootstrap test statistics as $\{\hat{J}_{NT,l}^*\}_{l=1}^B$. The bootstrap p-value is calculated by $p^* \equiv B^{-1} \sum_{l=1}^B 1\{\hat{J}_{NT,l}^* > \hat{J}_{NT}\}$, where $1\{\cdot\}$ is the usual indicator function.

Remark 9. It is straightforward to implement the preceding bootstrap procedure. The idea of fixed-design WB is not new; see, e.g., Hansen (2000) and Gonçalves and Kilian (2004). The latter authors consider both fixed- and recursive-design WB for autoregressions with conditional heteroskedasticity of unknown form, but their simulations suggest that neither WB method dominates the other. Because the theoretical justification for the asymptotic validity of fixed-design WB is much easier than that of the recursive-design WB, we adopt the fixed-design WB here. Note that in the bootstrap world, $(X_{it}, \hat{\lambda}_i, \hat{F}_t)$ is nonrandom and thus independent of ε_{js}^* for all i, t, j, s given the data so that the asymptotic variance formula can be simplified in this case. Even so, we continue to use the formula defined in Section 3.3.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

The following theorem states the main result of Section 3.4.

THEOREM 3.4. *Suppose that Assumptions A.1–A.3 hold. Then* $\hat{J}_{NT}^{*} \xrightarrow{D^*} N(0,1)$ *in probability, where* $\xrightarrow{D^*}$ *denotes weak convergence under the bootstrap probability measure conditional on the observed sample* $\mathcal{W}_{NT} \equiv \{(X_1, Y_1), \ldots, (X_N, Y_N)\}$:

The preceding theorem shows that the bootstrap provides an asymptotic valid approximation to the limit null distribution of $\hat{J}_{NT}$. This holds as long as we generate the bootstrap data by imposing the null hypothesis. If the null hypothesis does not hold in the observed sample, then we expect $\hat{J}_{NT}$ to explode at the rate $N^{1/4}T^{1/2}$, which delivers the consistency of the bootstrap-based test $\hat{J}_{NT}^{*}$.

3.5. Discussions and Extensions

The focus of this paper is to design a test for slope homogeneity in large-dimensional panel data models with interactive fixed effects. It turns out that our test statistic LM_{NT} or $\hat{J}_{NT}$ can be used for other testing purposes after suitable modifications.

3.5.1. Test of Model (2.1) against a Pure Factor Model. First, we can test the specification of the model (2.1) against a pure factor model. Specifically, we can test the null hypothesis $\mathbb{H}_{00} : \boldsymbol{\beta}_i^0 = \mathbf{0}_{K\times 1}$ for all $i = 1, \ldots, N$ against the alternative hypothesis $\mathbb{H}_{10} : \boldsymbol{\beta}_i^0 \neq \mathbf{0}_{K\times 1}$ for some $i = 1, \ldots, N$, where $\mathbf{0}_{K\times 1}$ is a $K \times 1$ vector of zeros. Under $\mathbb{H}_{00}$, $\boldsymbol{\beta}_i$ is a constant that does not vary across i, and it is identically equal to 0, implying that the regressor X_{it} has no explanatory power for Y_{it}. Under $\mathbb{H}_{10}$, we may have either heterogeneous slopes or homogeneous nonzero slopes.

There are various areas where such a test is applicable. Here we focus on a potential application to the asset returns in finance. With the advance of the capital asset pricing model and the arbitrage pricing theory, factor models have become one of the most important tools in modern finance. The traditional factor model specifies the excess returns of asset i at time t as

$$R_{it} = \lambda_i^{0\prime} F_t^0 + \eta_{it}, \tag{3.15}$$

where λ_i^0 is an $r \times 1$ vector of factor loadings, F_t^0 is an $r \times 1$ vector of latent factors, and η_{it} is the usual idiosyncratic error term. Even though the development of the asset pricing theory can proceed without a complete specification of how many and what factors are required, empirical testing does not have this luxury. For this reason, some authors (e.g., Lehmann and Modest, 1988; Connor and Korajzcyk, 1998) use estimated factors to test the asset pricing theory despite the drawback that the statistically estimated factors do not have immediate economic interpretation. A more popular approach is to rely on economic intuition

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

and theory as a guideline to come up with a list of observed variables/factors G_t to serve as proxies of the unobservable factors F_t^0. The most eminent example is the three observable risk factors discussed in Fama and French (1993) the market excess return, small minus big factor, and high minus low factor. Then an appealing question is whether these observable factors are, in fact, the underlying latent factors. Bai and Ng (2006) consider statistics to determine if the observed and latent factors are exactly the same and apply their tests to assess how well the Fama and French factors and several business cycle indicators can approximate the latent factors in portfolio and stock returns.

Here we offer an alternative approach by considering the following model:

$$R_{it} = \boldsymbol{\beta}_i^{0\prime} G_t + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \tag{3.16}$$

where G_t denotes a $K \times 1$ vector of observable factors and plays the role of X_{it} in (2.1). Clearly, we cannot estimate the preceding model by using any existing method. Nevertheless, as Bai (2009b) demonstrates, the model is identified under the null

$$\mathbb{H}_{01} : \boldsymbol{\beta}_i^0 = \boldsymbol{\beta}^0 \quad \text{for all } i = 1, \ldots, N \tag{3.17}$$

provided $T^{-1} G' M_{F^0} G > 0$ where $G \equiv (G_1, G_2, \ldots, G_T)'$, i.e., there is no multicollinearity between G and $F^0 \equiv (F_1^0, F_2^0, \ldots, F_T^0)'$. Let $G_{t,k}$ denote the kth element of G_t, $k = 1, \ldots, K$. If there exists an $r \times 1$ vector α_k such that $G_{t,k} = \alpha_k' F_t^0$ for all t, we can say that $G_{t,k}$ is an exact factor. If the kth column of G lies in the space spanned by the column vectors of F^0, which is the case when $G_{t,k}$ is an exact factor, then we cannot estimate the restricted model under $\mathbb{H}_{01}$. This motivates us to consider the following null instead:

$$\mathbb{H}_{02} : \boldsymbol{\beta}_i^0 = \mathbf{0}_{K\times 1} \quad \text{for all } i = 1, \ldots, N. \tag{3.18}$$

Intuitively speaking, $\mathbb{H}_{02}$ indicates that given the r latent factors in F_t^0, the K observable risk factors in G_t are redundant in explaining the asset returns in (3.16). In the case when we reject $\mathbb{H}_{02}$, it means that the r latent factors in F_t^0 cannot span the space of the K observable factors. Various reasons can cause the latter to occur. One reason is that the K observable factors are all relevant but $r < K$. If this is the case, we should observe the change from rejecting $\mathbb{H}_{02}$ to failing to reject $\mathbb{H}_{02}$ as we increase r. Another reason is that the observable factors in G_t are bad proxies for the latent factors. This suggests the importance of testing $\mathbb{H}_{02}$ against its alternative $\mathbb{H}_{12} : \boldsymbol{\beta}_i^0 \neq \mathbf{0}_{K\times 1}$ for some $i = 1, \ldots, N$. Note that we allow heterogeneous factor loadings for the observable factors under $\mathbb{H}_{12}$.

As a referee kindly points out, the LM principle can be applied to a situation like this. Our LM_{NT} or $\hat{J}_{NT}$ statistic can be used to test $\mathbb{H}_{02}$ against $\mathbb{H}_{12}$ with minor modifications. Under $\mathbb{H}_{02}$, we have a pure factor model so that both the latent factors F_t^0 and the factor loadings λ_i^0 can be estimated, say, by $\hat{F}_t$ and $\hat{\lambda}_i$, respectively. Let $\hat{\varepsilon}_{it} = R_{it} - \hat{\lambda}_i' \hat{F}_t$. Then we can construct the statistic as before. It is easy to see that the asymptotic distribution theory in the preceding analysis continues to hold in this case.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

3.5.2. Test of the Linear Functional Form in (2.1). We can also test the correct specification of the linear functional form in (2.1) by considering a nonparametric heterogeneous panel data model with interactive fixed effects:

$$Y_{it} = m_i(X_{it}) + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{3.19}$$

where $m_i(\cdot)$, $i = 1, \ldots, N$, are unknown but smooth functions. The null hypothesis is

$$\mathbb{H}_0^{(1)} : m_i(x) = \boldsymbol{\beta}_i^{0\prime} x \quad \text{for some } \boldsymbol{\beta}_i^0 \in \mathbb{R}^K \quad \text{and all } i = 1, \ldots, N.$$

Under $\mathbb{H}_0^{(1)}$ and certain rank conditions, we can estimate the heterogeneous linear panel in (2.1) by the Pesaran (2006) CCE method, obtain the residuals, and run the time series regression of these residuals on X_{it} nonparametrically to construct a test statistic similar to our LM statistic. In the case of rejection, one can follow Su and Jin (2012) to consider nonparametric estimation of $m_i(\cdot)$.

Alternatively, we can consider the Bai (2009a) canonical model

$$Y_{it} = \boldsymbol{\beta}^{0\prime} X_{it} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{3.20}$$

and test whether the preceding linear model is correctly specified. The model under the alternative is obtained by replacing $\boldsymbol{\beta}^{0\prime} X_{it}$ in the preceding model by $m(X_{it})$, where $m(\cdot)$ is an unknown but smooth function. In this case, we can obtain the residuals $\hat{\varepsilon}_{it}$ from the model (3.20) and obtain a nonparametric analogue of the LM test statistic studied previously. We leave the details for the future research.

4. MONTE CARLO SIMULATION AND APPLICATION

In this section, we first conduct a small set of Monte Carlo simulations to evaluate the finite-sample performance of our test and then apply it to the OECD real gross domestic product (GDP) growth data.

4.1. Simulation

4.1.1. Data Generating Processes. We consider the following eight data generating processes (DGPs):

DGP 1: $Y_{it} = \rho^0 Y_{i,t-1} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 2: $Y_{it} = \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 3: $Y_{it} = \rho_1^0 Y_{i,t-1} + \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 4: $Y_{it} = \rho_1^0 Y_{i,t-1} + \rho_2^0 Y_{i,t-2} + \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 5: $Y_{it} = \rho_i^0 Y_{i,t-1} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 6: $Y_{it} = \beta_{i1}^0 X_{it,1} + \beta_{i2}^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

DGP 7: $Y_{it} = \rho_{i1}^0 Y_{i,t-1} + \beta_{i1}^0 X_{it,1} + \beta_{i2}^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 8: $Y_{it} = \rho_{i1}^0 Y_{i,t-1} + \rho_{i2}^0 Y_{i,t-2} + \beta_{i1}^0 X_{it,1} + \beta_{i2}^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,

where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, $\left(\rho^0, \rho_1^0, \rho_2^0, \beta_1^0, \beta_2^0\right) = (0.6,\ 0.5,\ 0.25,\ 1, 3)$, $\rho_i^0 \backsim \text{IID } U(0.45, 0.75)$, $\rho_{i1}^0 \backsim \text{IID } U(0.45, 0.55)$, $\rho_{i2}^0 \backsim \text{IID } U(0.2, 0.3)$, $\beta_{i1}^0 \backsim \text{IID } U(0.9, 1.1)$, and $\beta_{i2}^0 \backsim \text{IID } U(2.7, 3.3)$. Here $\lambda_i^0 = \left(\lambda_{i1}^0, \lambda_{i2}^0\right)'$, $F_t^0 = \left(F_{t1}^0, F_{t2}^0\right)'$, and the regressors are generated according to

$$X_{it,1} = \mu_1 + c_1 \lambda_i^{0\prime} F_t^0 + \eta_{it,1},$$
$$X_{it,2} = \mu_2 + c_2 \lambda_i^{0\prime} F_t^0 + \eta_{it,2},$$

where the variables λ_{ij}^0, F_{tj}^0, and $\eta_{it,j}$, $j = 1, 2$, are all i.i.d. $N(0, 1)$ and mutually independent of each other. Clearly, the regressors $X_{it,1}$ and $X_{it,2}$ are correlated with λ_i^0 and F_t^0. We set $\mu_1 = c_1 = 0.25$ and $\mu_2 = c_2 = 0.5$. Note that DGPs 1–4 are used for the level study and DGPs 5–8 for the power study. For the dynamic models (DGPs 1, 3, 4, 5, 7, and 8), we discard the first 100 observations along the time dimension when generating the data. For the heterogeneous slope parameters in DGPs 5–8, they are generated once and then fixed across replications.

For the idiosyncratic error term ε_{it}, we consider both the cases of conditional homoskedasticity and heteroskedasticity. In the former case, three standardized distributions are used to draw ε_{it} (independently from λ_{ij}^0, F_{tj}^0, and $\eta_{it,j}$, $j = 1, 2$) to ensure it has mean 0 and variance 1:

$$(i)\ \ \varepsilon_{it} \backsim \text{IID } N(0, 1), \qquad (ii)\ \varepsilon_{it} \backsim \text{IID student } t_9/\sqrt{9/7},$$
$$(iii)\ \ \varepsilon_{it} \backsim \text{IID } \left(\chi_4^2 - 4\right)/\sqrt{8}. \tag{4.1}$$

The choice of the latter two distributions satisfies the moment conditions on ε_{it} and serves to provide evidence on the effects of fat tailedness and skewness on our test. In the latter case, the error terms ε_{it}'s are generated from the process

$$\varepsilon_{it} = \sigma_{it}\epsilon_{it}, \qquad \sigma_{it} = \left(0.25 + 0.1 \sum_{k=1}^{K} X_{it,k}^2/K\right)^{1/2}, \tag{4.2}$$

where $X_{it,k}$ denotes the kth element of X_{it}, X_{it} signifies the $K \times 1$ vector of regressors in the corresponding DGPs, and ϵ_{it}'s are drawn from the same three standardized distributions used before:

$$\text{(i)}\ \ \epsilon_{it} \backsim \text{IID } N(0, 1), \qquad \text{(ii)}\ \epsilon_{it} \backsim \text{IID student } t_9/\sqrt{9/7},$$
$$\text{(iii)}\ \ \epsilon_{it} \backsim \text{IID } \left(\chi_4^2 - 4\right)/\sqrt{8}. \tag{4.3}$$

4.1.2. Test Results. We consider our $\hat{J}_{NT}$ test based on both asymptotic normal critical values and the bootstrap p-values. We consider N, $T = 20$, 40, 60.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

For each combination of N, T, and error distributions in (4.1) or (4.3), we consider 2,000 simulations for the nonbootstrap version of the test. For the bootstrap version of the test, we use 500 replications for each scenario and $B = 250$ bootstrap resamples for each replication.

We first consider the nonbootstrap version of our test. Tables 1 and 2 report the finite-sample properties of our test under the null in the case of conditional homoskedasticity and heteroskedasticity, respectively. We focus on the sample mean (mean), standard error (s.e.), and rejection frequency (rej. freq.) at 0.05 nominal level for our test across 2,000 simulations. Note that asymptotic theory suggests that $\hat{J}_{NT}$ has asymptotic mean 0 and standard error 1, respectively, when the null hypothesis of slope homogeneity is satisfied. Table 1 indicates that the sample mean of $\hat{J}_{NT}$ tends to be positive in finite samples, and it can be as large as 1.23 for certain DGPs; see, e.g., the case $(N, T) = (60, 20)$ in DGP 4. The is true irrespective of the distributions of the error terms. Similarly, the sample s.e. of $\hat{J}_{NT}$ is generally larger than the theoretical value 1 in all DGPs for all error distributions under investigation. In some case, the sample s.e. can be as large as 3.14 (see the case $(N, T) = (60, 60)$ in DGP 2) despite the fact that admittedly the larger sample s.e.'s tend to be driven by several outliers in the simulations and if one eliminates these outliers among the 2,000 replications, then s.e.'s would be significantly reduced. In terms of rejection frequency at 0.05 nominal level, we find that (1) the $\hat{J}_{NT}$ test tends to be oversized, (2) the size distortion tends to increase as N increases for fixed T and it becomes most severe in the case when N/T is largest, (3) the size distortion is only mild when $T/N \geq 1$, and (4) the fat-tailedness or skewness of the error terms does not play an important role. As a referee points out, the size distortion in the case when N is large relative to T is closely related to the well-known incidental parameter problem in panel data models. The results in Table 2 for the case of conditional heteroskedasticity are largely similar to those in Table 1. This is due to the fact that the asymptotic bias and variance formulas of our test automatically take into account the potential presence of conditional heteroskedasticity of unknown form.

Tables 3 and 4 report the finite-sample power rejection frequency of our test based on the bootstrap p-values under the null and alternative, respectively. We summarize some important findings from these tables. First, Table 3 suggests that the level of the bootstrap version of our test tends to be well-behaved across all DGPs under investigation. This is true regardless of the presence of conditional heteroskedasticity or not and whether the error terms exhibit fat-tailedness or skewness or not. Second, Table 4 suggests that the finite-sample power behavior of the bootstrap version of our test is quite satisfactory for DGPs 5–8. As either N or T increases, the power of our test increases, and as asymptotic theory predicts, it increases faster as T increases for fixed N than as N increases for fixed T. In addition, we find that the presence of conditional heteroskedasticity generally makes it harder to detect the deviations from slope homogeneity when the signal-noise ratio is controlled.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

TABLE 1. Finite-sample properties of the LM test under the null (conditional homoskedasticity case, nominal level 0.05)

			$\varepsilon_{it} \backsim N(0,1)$			$\varepsilon_{it} \backsim t_9/\sqrt{\frac{9}{7}}$			$\varepsilon_{it} \backsim (\chi^2_4 - 4)/\sqrt{8}$		
DGP	N	T	mean	s.e.	rej. freq.	mean	s.e.	rej. freq.	mean	s.e.	rej. freq.
1	20	20	0.166	1.306	0.073	0.158	1.202	0.072	0.209	1.614	0.085
		40	−0.023	1.081	0.060	−0.031	1.092	0.061	−0.021	1.044	0.055
		60	−0.075	1.096	0.044	−0.077	1.013	0.041	−0.084	1.070	0.045
	40	20	0.318	1.200	0.095	0.368	1.624	0.104	0.302	1.144	0.100
		40	0.112	1.024	0.064	0.124	1.103	0.065	0.068	1.077	0.063
		60	−0.015	1.053	0.054	−0.028	1.033	0.052	0.028	1.049	0.056
	60	20	0.423	1.352	0.113	0.421	1.315	0.111	0.475	1.667	0.109
		40	0.143	1.073	0.071	0.160	1.043	0.068	0.120	1.089	0.064
		60	0.077	1.048	0.060	0.076	1.044	0.063	0.118	0.985	0.053
2	20	20	0.265	1.519	0.084	0.270	1.515	0.083	0.249	1.235	0.090
		40	−0.035	1.242	0.055	0.006	2.024	0.052	−0.072	1.043	0.055
		60	−0.111	1.331	0.046	−0.077	1.734	0.048	−0.148	1.065	0.041
	40	20	0.526	1.775	0.131	0.534	1.876	0.135	0.521	1.624	0.135
		40	0.065	1.025	0.062	0.082	1.098	0.066	0.128	1.782	0.067
		60	0.051	1.798	0.060	0.072	2.417	0.059	−0.002	1.005	0.047
	60	20	0.758	2.396	0.169	0.707	1.969	0.170	0.762	2.321	0.164
		40	0.264	1.990	0.077	0.217	1.030	0.072	0.219	1.003	0.071
		60	0.107	1.011	0.062	0.113	1.008	0.059	0.119	3.143	0.055
3	20	20	0.389	1.438	0.107	0.407	1.431	0.112	0.409	1.470	0.115
		40	0.026	1.228	0.049	0.040	1.264	0.056	0.035	1.196	0.053
		60	−0.050	2.006	0.045	−0.107	1.019	0.046	−0.072	1.606	0.044
	40	20	0.718	1.735	0.166	0.800	2.148	0.167	0.770	2.091	0.168
		40	0.201	1.298	0.085	0.226	1.782	0.083	0.205	1.735	0.075
		60	0.052	1.094	0.054	0.043	1.001	0.059	0.133	2.477	0.065
	60	20	1.021	2.134	0.246	0.995	1.981	0.235	1.049	2.704	0.225
		40	0.361	1.123	0.103	0.460	2.702	0.104	0.379	2.614	0.090
		60	0.304	3.263	0.070	0.186	1.533	0.064	0.265	2.699	0.071
4	20	20	0.531	1.526	0.131	0.544	1.569	0.132	0.527	1.392	0.144
		40	0.091	1.688	0.066	0.081	1.503	0.066	0.125	1.706	0.065
		60	−0.069	1.396	0.049	−0.036	1.793	0.044	0.032	2.380	0.053
	40	20	0.909	1.787	0.215	0.922	1.809	0.217	0.891	1.608	0.210
		40	0.362	2.044	0.096	0.408	2.324	0.104	0.344	2.376	0.090
		60	0.114	1.560	0.064	0.202	2.554	0.070	0.216	2.935	0.074
	60	20	1.226	2.019	0.294	1.231	2.064	0.296	1.237	2.394	0.285
		40	0.514	2.147	0.119	0.467	1.554	0.121	0.450	2.283	0.100
		60	0.324	2.923	0.079	0.251	1.957	0.076	0.238	1.311	0.080

Note: For each error distribution, mean, s.e., and rej. freq. refer to the sample mean, standard error, and 0.05 nominal level rejection frequency of the LM test based on 2,000 replications, respectively.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

TABLE 2. Finite-sample properties of the LM test under the null (conditional heteroskedasticity case, nominal level 0.05)

			$\varepsilon_{it} \backsim N(0,1)$			$\varepsilon_{it} \backsim t_9/\sqrt{\frac{9}{7}}$			$\varepsilon_{it} \backsim (\chi^2_4 - 4)/\sqrt{8}$		
DGP	N	T	mean	s.e.	rej. freq.	mean	s.e.	rej. freq.	mean	s.e.	rej. freq.
1	20	20	0.140	1.310	0.071	0.247	1.849	0.076	0.155	1.346	0.075
		40	−0.032	1.410	0.053	−0.027	1.049	0.047	−0.076	0.998	0.050
		60	−0.099	1.016	0.044	−0.082	1.009	0.042	−0.083	1.025	0.044
	40	20	0.292	1.242	0.090	0.319	1.384	0.081	0.357	1.297	0.083
		40	0.011	1.024	0.047	−0.010	0.995	0.046	−0.006	1.122	0.047
		60	−0.074	1.192	0.033	−0.080	0.950	0.040	−0.124	0.944	0.030
	60	20	0.358	1.815	0.087	0.408	1.724	0.091	0.333	1.265	0.089
		40	0.035	1.007	0.047	0.007	1.044	0.043	0.012	0.936	0.043
		60	−0.081	0.970	0.035	−0.062	0.980	0.042	−0.060	1.054	0.038
2	20	20	0.329	1.276	0.091	0.349	1.530	0.095	0.275	1.072	0.098
		40	0.033	1.297	0.055	−0.036	1.046	0.052	−0.037	1.361	0.053
		60	−0.098	1.688	0.042	−0.083	2.074	0.042	−0.028	2.746	0.047
	40	20	0.639	1.282	0.162	0.649	1.543	0.158	0.630	1.660	0.149
		40	0.258	2.750	0.083	0.175	1.044	0.076	0.225	1.035	0.083
		60	0.033	1.032	0.061	0.122	2.462	0.059	0.038	1.037	0.060
	60	20	0.946	2.728	0.213	0.872	1.914	0.210	0.912	2.416	0.192
		40	0.320	1.025	0.099	0.320	0.998	0.087	0.330	0.998	0.094
		60	0.134	1.013	0.070	0.156	1.035	0.076	0.138	0.982	0.070
3	20	20	0.493	1.209	0.140	0.588	1.416	0.155	0.610	1.570	0.145
		40	0.187	1.423	0.090	0.150	1.356	0.077	0.201	1.409	0.088
		60	0.116	2.384	0.059	0.110	1.481	0.079	0.109	1.740	0.067
	40	20	0.965	2.140	0.214	0.975	2.029	0.201	0.931	1.946	0.201
		40	0.365	1.951	0.096	0.335	1.664	0.094	0.347	2.039	0.089
		60	0.156	1.976	0.070	0.235	2.713	0.070	0.146	1.960	0.067
	60	20	1.110	2.125	0.248	1.251	2.703	0.252	1.198	2.343	0.244
		40	0.409	1.672	0.101	0.478	2.273	0.103	0.468	1.724	0.123
		60	0.119	1.075	0.070	0.438	3.596	0.081	0.427	3.722	0.086
4	20	20	0.672	1.415	0.173	0.642	1.289	0.175	0.636	1.320	0.171
		40	0.319	1.308	0.111	0.285	1.229	0.099	0.325	1.413	0.109
		60	0.162	1.250	0.078	0.237	1.273	0.099	0.240	1.586	0.094
	40	20	0.942	1.449	0.218	0.996	1.471	0.241	1.044	1.699	0.256
		40	0.530	1.734	0.124	0.404	1.496	0.110	0.426	1.484	0.101
		60	0.236	1.441	0.078	0.248	1.791	0.082	0.245	1.813	0.081
	60	20	1.259	1.872	0.292	1.262	1.907	0.286	1.240	1.777	0.279
		40	0.448	1.670	0.101	0.397	1.555	0.100	0.517	1.834	0.111
		60	0.098	1.327	0.057	0.193	1.524	0.065	0.256	1.805	0.078

Note: For each error distribution, mean, s.e., and rej. freq. refer to the sample mean, standard error, and 0.05 nominal level rejection frequency of the LM test based on 2,000 replications, respectively.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

TABLE 3. Finite-sample rejection frequency for the bootstrap version of our test under the null (nominal level 0.05)

			Conditional homoskedasticity			Conditional heteroskedasticity		
DGP	N	T	$N(0,1)$	$\varepsilon_{it} \sim t_9/\sqrt{\frac{9}{7}}$	$(\chi^2_4-4)/\sqrt{8}$	$N(0,1)$	$\epsilon_{it} \sim t_9/\sqrt{\frac{9}{7}}$	$(\chi^2_4-4)/\sqrt{8}$
1	20	20	0.056	0.054	0.076	0.030	0.050	0.036
		40	0.066	0.072	0.052	0.048	0.034	0.040
		60	0.058	0.044	0.046	0.042	0.020	0.036
	40	20	0.052	0.050	0.054	0.040	0.030	0.034
		40	0.054	0.052	0.060	0.020	0.042	0.026
		60	0.050	0.044	0.042	0.028	0.046	0.026
	60	20	0.060	0.044	0.056	0.020	0.026	0.036
		40	0.046	0.040	0.048	0.028	0.028	0.022
		60	0.042	0.040	0.074	0.028	0.030	0.042
2	20	20	0.064	0.058	0.040	0.040	0.052	0.084
		40	0.060	0.054	0.048	0.066	0.056	0.054
		60	0.064	0.060	0.064	0.044	0.056	0.070
	40	20	0.050	0.064	0.058	0.080	0.070	0.066
		40	0.040	0.058	0.058	0.056	0.060	0.066
		60	0.064	0.066	0.054	0.044	0.064	0.060
	60	20	0.062	0.066	0.058	0.068	0.086	0.070
		40	0.054	0.042	0.054	0.054	0.056	0.054
		60	0.036	0.044	0.058	0.038	0.068	0.056
3	20	20	0.068	0.070	0.052	0.050	0.072	0.050
		40	0.048	0.048	0.040	0.058	0.056	0.068
		60	0.066	0.058	0.040	0.048	0.060	0.042
	40	20	0.050	0.052	0.058	0.046	0.050	0.054
		40	0.062	0.060	0.048	0.044	0.036	0.062
		60	0.054	0.042	0.076	0.046	0.048	0.036
	60	20	0.066	0.058	0.076	0.054	0.054	0.050
		40	0.040	0.038	0.042	0.036	0.036	0.038
		60	0.054	0.048	0.058	0.044	0.030	0.052
4	20	20	0.054	0.056	0.058	0.082	0.084	0.034
		40	0.074	0.070	0.064	0.082	0.078	0.074
		60	0.060	0.052	0.038	0.056	0.078	0.074
	40	20	0.066	0.050	0.048	0.032	0.058	0.078
		40	0.060	0.066	0.042	0.070	0.046	0.054
		60	0.060	0.048	0.052	0.044	0.036	0.046
	60	20	0.070	0.076	0.072	0.046	0.040	0.050
		40	0.032	0.050	0.056	0.034	0.034	0.036
		60	0.064	0.064	0.060	0.032	0.024	0.022

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

TABLE 4. Finite-sample rejection frequency for the bootstrap version of our test under the alternative (nominal level 0.05)

			Conditional homoskedasticity			Conditional heteroskedasticity		
DGP	N	T	$N(0,1)$	$\varepsilon_{it} \sim t_9/\sqrt{\frac{9}{7}}$	$(\chi^2_4-4)/\sqrt{8}$	$N(0,1)$	$\epsilon_{it} \sim t_9/\sqrt{\frac{9}{7}}$	$(\chi^2_4-4)/\sqrt{8}$
5	20	20	0.248	0.366	0.350	0.090	0.124	0.134
		40	0.846	0.752	0.766	0.390	0.358	0.394
		60	0.788	0.944	0.932	0.412	0.552	0.528
	40	20	0.528	0.580	0.456	0.188	0.204	0.162
		40	0.996	0.966	0.908	0.714	0.550	0.420
		60	0.998	0.996	0.992	0.786	0.752	0.596
	60	20	0.752	0.716	0.766	0.238	0.222	0.234
		40	0.994	0.994	0.994	0.644	0.630	0.732
		60	1.000	1.000	0.992	0.776	0.862	0.886
6	20	20	0.256	0.276	0.234	0.630	0.652	0.624
		40	0.662	0.672	0.694	0.980	0.984	0.986
		60	0.890	0.884	0.938	1.000	0.998	1.000
	40	20	0.438	0.424	0.384	0.936	0.876	0.806
		40	0.806	0.882	0.918	0.998	1.000	1.000
		60	0.994	0.994	0.996	1.000	1.000	1.000
	60	20	0.710	0.578	0.632	0.986	0.970	0.994
		40	0.984	0.968	0.990	1.000	1.000	1.000
		60	1.000	1.000	1.000	1.000	1.000	1.000
7	20	20	0.356	0.456	0.564	0.130	0.206	0.198
		40	0.980	0.948	0.734	0.418	0.372	0.250
		60	0.994	0.996	0.984	0.494	0.612	0.494
	40	20	0.826	0.672	0.846	0.208	0.222	0.262
		40	1.000	0.988	1.000	0.632	0.456	0.634
		60	1.000	1.000	1.000	0.810	0.754	0.810
	60	20	0.920	0.830	0.872	0.270	0.270	0.238
		40	1.000	0.998	1.000	0.812	0.580	0.702
		60	1.000	1.000	1.000	0.870	0.896	0.918
8	20	20	0.888	0.878	0.882	0.142	0.156	0.114
		40	0.998	0.996	0.998	0.360	0.294	0.264
		60	1.000	1.000	1.000	0.516	0.492	0.542
	40	20	0.958	0.958	0.958	0.158	0.162	0.158
		40	1.000	0.998	1.000	0.392	0.308	0.466
		60	1.000	1.000	1.000	0.552	0.514	0.754
	60	20	0.974	0.972	0.962	0.172	0.180	0.184
		40	1.000	1.000	1.000	0.518	0.414	0.486
		60	1.000	1.000	1.000	0.722	0.776	0.806

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

4.2. An Application to the OECD Economic Growth Data

The economic growth model has been a key issue over many decades in macroeconomics. It is interesting to incorporate interactive fixed effects in panel model study, which can account for heterogeneous impact of unobservable common shocks. However, the slope homogeneity assumption of Bai (2009a) can be restrictive in empirical work. For classical panel data models there have been a number of studies suggesting that the slope homogeneity assumption may be too restrictive in studying economic growth; see Basssanini and Scarpetta (2002), Bond, Leblebicioglu, and Schiantarelli (2010), and Eberhardt and Teal (2011), among others. Bassanini and Scarpetta (2002) estimate a standard growth equation using the annual data for 21 OECD countries from 1971 to 1998 and conduct the Hausman test for the long-run slope homogeneity hypothesis. They find that the homogeneity restriction can be rejected at the 5% level when some time dummies are added to the model. Bond et al. (2010) present evidence of a positive relationship between investment as a share of GDP and the long-run growth rate of DGP per worker and find that allowing for heterogeneity across countries in model parameters suggests that growth rates are typically less persistent than suggested by pooled instrumental variables estimates. Eberhardt and Teal (2011) develop a central argument that cross-country heterogeneity in the impacts of observables and unobservables on output and growth rates is important for reliable empirical analysis.

If the slope homogeneity assumption is not true, estimates based on it can be inconsistent, and its associated inference can be misleading. Therefore it is prudent to test whether the impacts of labor and capital on economic growth are homogeneous across countries after modeling the heterogeneous impacts of unobservable common shocks such as technological shocks and financial crises. Here we apply our test to the OECD economic growth data that are analyzed in Zhang, Su, and Phillips (2012) for a different modeling strategy. The data set consists of four economic variables for $N = 16$ OECD countries, whereas are GDP, Capital stock (K), Labor input (L), and Human capital (H). The first three are seasonally adjusted quarterly data from 1975Q4 to 2010Q3 ($T = 140$), whereas we use linear interpolation to obtain the quarterly observations for Human capital because there are only 5-year census data available.

We consider the following economic growth model:

$$\Delta \ln GDP_{it} = \sum_{l=1}^{L} \rho_{i,l} \Delta \ln GDP_{i,t-l} + \beta_{i,1} \Delta \ln K_{it} + \beta_{i,2} \Delta \ln L_{it} + \beta_{i,3} \Delta \ln H_{it} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it},$$

where F_t^0 is an $r \times 1$ vector that represents common shocks such as technological shocks and financial crises, λ_i^0 represents the heterogeneous impacts of common shocks on country i, and $\Delta \ln Z_{it} = \ln Z_{it} - \ln Z_{it-1}$ for $Z = GDP$, K, L, and H. The terms $\beta_{i,1}, \beta_{i,2}$, and $\beta_{i,3}$ are coefficients of growth rates of K, L, and H, respectively. We consider five values for the number L of lagged dependent

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

TABLE 5. Test statistics and bootstrap p-values for the application to the OECD GDP growth data

Model	1	2	3	4	5	6	7	8
0	5.40	3.13	1.94	2.62	3.81	2.43	3.13	3.78
	(0.000)	(0.021)	(0.043)	(0.003)	(0.006)	(0.023)	(0.038)	(0.051)
1	4.44	3.37	3.96	6.50	4.39	4.29	4.67	4.78
	(0.000)	(0.048)	(0.045)	(0.013)	(0.025)	(0.027)	(0.054)	(0.066)
2	3.22	4.16	3.68	6.86	4.34	4.17	5.22	5.29
	(0.000)	(0.001)	(0.076)	(0.017)	(0.008)	(0.014)	(0.014)	(0.023)
3	2.90	3.96	3.76	4.35	4.78	4.45	5.75	5.65
	(0.000)	(0.003)	(0.016)	(0.000)	(0.000)	(0.007)	(0.014)	(0.042)
4	2.58	4.87	3.40	4.41	4.28	4.78	5.60	9.31
	(0.002)	(0.000)	(0.038)	(0.005)	(0.004)	(0.009)	(0.023)	(0.007)

Note: The numbers in parentheses are bootstrap p-values where the bootstrap number B is 1,000.

variables, namely, $L = 0, 1, \ldots, 4$, and name the corresponding model as model L. Model 0 ($L = 0$) is a static panel, data model with interactive fixed effects. In models 1–4, $\rho_{i,l}$ represents the impact of previous quarters' GDP growth rate on the current one in country i. We are interested in testing for homogeneous coefficients for the 16 OECD countries.

We consider $r = 1, 2, \ldots, 8$ to capture the interactive fixed effects in the growth model.[8] Table 5 reports the test statistics and the bootstrap p-values for our test of slope homogeneity. From the table, we see that the bootstrap p-values for all numbers of factors under investigation are uniformly much smaller than 0.10 in all cases and smaller than 0.05 in most cases. So we can reject the null hypothesis of homogeneous slopes at the 5% level for all models for a majority of values of r. The results imply that the slope homogeneity assumption may not be plausible at all despite the fact it is commonly assumed in the literature (cf. Eberhardt and Teal, 2011, p. 109). This implies that we have to resort to the Pesaran (2006) CCE method to obtain the heterogeneous impacts of labor and capital on economic growth across OECD countries.

5. CONCLUSIONS

In this paper we propose an LM test for slope homogeneity in large-dimensional dynamic panel data models with interactive fixed effects and conditional heteroskedasticity of unknown form. We first estimate the model under the null to obtain the restricted residuals, which are then used to construct the test statistic. We demonstrate that after being appropriately normalized, it is asymptotically normally distributed under the null hypothesis of homogeneous slopes and it has

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

power to detect Pitman local alternatives at the rate of $T^{-1/2}N^{-1/4}$. We also propose a wild bootstrap procedure to obtain the bootstrap p-values. Simulations demonstrate that the bootstrap version of our test behaves reasonably well in finite samples. The application to the OECD economic growth data indicates that the commonly imposed slope homogeneity assumption is rather fragile.

When the null hypothesis of homogeneous slopes is rejected, we may consider applying the Pesaran (2006) CCE method to obtain consistent estimates of both individual slopes and their cross-sectional average under certain rank conditions. If some prior information is available, one can divide the cross-sectional units into several groups, test the slope homogeneity within each group, and estimate the homogeneous slopes within each individual group in the case of failure of rejection. Alternatively, a panel structure model in the spirit of Sun (2005) may be considered.

Note to readers. Some of the mathematical derivations require rather long and tedious calculations. In the Appendixes that follow, we shall therefore omit several details that can be found in the Supplementary Material at Cambridge Journals Online (journals.cambridge.org/ect).

NOTES

1. The rank condition must also be satisfied when estimating the homogeneous model.
2. Under the standard assumption that $\mathbb{E}(\varepsilon_{it}) = 0$ for each i, $\hat{\varepsilon}_{it}$ also centers around 0 for each i under the null in the sense $\hat{\varepsilon}_{it} = \varepsilon_{it} + o_P(1)$ so that an intercept term in the regression (2.13) is unnecessary.
3. For an excellent survey on LM-principle-based misspecification tests, see Godfrey (1988).
4. Alternatively, cross-sectional dependence can be generated via the specification of a spatial weight matrix, which is regularly used in the spatial econometrics literature; see, e.g., Anselin (1988). But this type of cross-sectional dependence is local in nature.
5. Let $\hat{b}_i \equiv (\hat{b}_{i1}, \ldots, \hat{b}_{iT})'$. Noting that $\hat{F}'\hat{F}/T = I_R$, we have $\hat{b}_i = M_{\hat{F}} X_i \hat{\Omega}_i^{-1/2}$.
6. An alternative expression for D_{NT} is given by $D_{NT} \equiv D_{NT}(F^0) = 1/NT \sum_{i=1}^N X_i' M_{F^0} X_i - 1/T \left(1/N^2 \sum_{i=1}^N \sum_{k=1}^N X_i' M_{F^0} X_k a_{ik}\right) = 1/NT \sum_{i=1}^N \tilde{X}_i' \tilde{X}_i$, which is used by Bai (2009a).
7. This is the case even if X_{it} contains lagged dependent variables, say, $Y_{i,t-1}$ and $Y_{i,t-2}$.
8. Alternatively, one can use the information criteria proposed by Bai and Ng (2002) to determine the number r of factors. But it is well known that their criteria tend to fail when the cross-sectional unit N is small, which is the case here.

REFERENCES

Andrews, D.W.K. (2005) Cross-section regression with common shocks. *Econometrica* 73, 1551–1585.

Anselin, L. (1988) *Spatial Econometric Methods and Models*. Kluwer.

Bai, J. (2009a) Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.

Bai, J. (2009b) Supplement to "Panel data models with interactive fixed effects": Technical details and proofs. *Econometrica Supplementary Material*.

Bai, J. & S. Ng (2002) Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Bai, J. (2003) Inferential theory for factors models of large dimensions. *Econometrica* 71, 135-173.
Bai, J. & S. Ng (2006) Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* 131, 507–537.
Baltagi, B.H., G. Bresson, & A. Pirotte (2008) To pool or not to pool? In L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data*, pp. 517–546, Springer-Verlag.
Bassanini, A. & S. Scarpetta (2002) Does human capital matter for growth in OECD countries? A pooled mean-group approach. *Economics Letter* 74, 399–405.
Bernstein, D.S. (2005) *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press.
Blomquist, J. (2010) A Panel Bootstrap Test for Slope Homogeneity. Working paper, Lund University.
Bond, S., A. Leblebicioglu, & F. Schiantarelli (2010) Capital accumulation and growth: A new look at the empirical evidence. *Journal of Applied Econometrics* 25, 1073–1099.
Connor, G. & R. Korajzcyk (1998) Risk and return in an equilibrium APT application of a new test methodology. *Journal of Financial Economics* 21, 225–289.
Eberhardt, M. & F. Teal (2011) Econometrics for grumblers: A new look at the literature on cross-country growth empirics. *Journal of Economic Surveys* 25, 109–155.
Fama, E.F. & K. French (1993) Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 18, 61–90.
Godfrey, L.G. (1988) *Misspecification Tests in Econometrics: The Lagrangian Multiplier Principle and Other Approaches*. Cambridge University Press.
Gonçalves, S. & L. Kilian (2004) Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89–120.
Greenaway-McGrevy, R., C. Han, & D. Sul (2012) Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics* 169, 48–53.
Hahn, J. & G. Kuersteiner (2011) Reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.
Hansen, B.E. (2000) Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.
Hsiao, C. (2003) *Analysis of Panel Data*. Cambridge University Press.
Jin, S. & L. Su (2013) A nonparametric poolability test for panel data models with cross section dependence. *Econometric Reviews* 32, 469–512.
Kapetanios, G. & M.H. Pesaran (2007) Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset return. In G. Phillips and E. Tzavalis (eds.), *The Refinement of Econometric Estimation and Testing Procedures: Finite Sample and Asymptotic Analysis*. Cambridge University Press, pp. 239–281.
Lee, A.J. (1990) *U-statistics: Theory and Practice*. Marcel Dekker.
Lehmann, B. & D. Modest (1988) The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics* 21, 213–254.
Lin, C.-C. (2010) Testing for Slope Homogeneity in a Linear Panel Model with Fixed Effects and Conditional Heteroskedasticity. Working paper, Institute of Economics, Academia Sinica.
Moon, H. & M. Weidner (2010a) Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects. Working paper, University of Southern California.
Moon, H. & M. Weidner (2010b) Dynamic Linear Panel Regression Models with Interactive Fixed Effects. Working paper, University of Southern California.
Pesaran, H., R. Smith, & K.S. Im (1996) Dynamic linear models for heterogeneous panels. In L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, 2nd rev. ed., pp. 145–195. Kluwer Academic Publishers.
Pesaran, M.H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74, 967–1012.
Pesaran, M.H. & E. Tosetti (2011) Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182–202.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Pesaran, M.H. & T. Yamagata (2008) Testing slope homogeneity in large panels. *Journal of Econometrics* 142, 50–93.

Phillips, P.C.B. & D. Sul (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6, 217–259.

Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer-Verlag.

Prakasa Rao, B.L.S. (2009) Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics* 61, 441–460.

Roussas, G.G. (2008) On conditional independence, mixing and association. *Stochastic Analysis and Applications* 26, 1274–1309.

Su, L. & S. Jin (2012) Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169, 34–47.

Su, L. & A. Ullah (2013) A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. *Econometric Theory*, 29, 187–212.

Sun, Y. (2005) Estimation and Inference in Panel Structure Models. Working paper, Department of Economics, University of California, San Diego.

Zhang, Y., L. Su, & P.C.B. Phillips (2012) Testing for common trends in semiparametric panel data models with fixed effects. *Econometrics Journal* 15, 56–100.

APPENDIX A: Some Technical Lemmas

In this Appendix we provide some technical lemmas. In Appendix B we use them to prove the main results in the paper. The proofs of these lemmas and of Theorem 3.2 are provided online at Cambridge Journals Online in supplementary material to this paper. Readers may refer to the supplementary material associated with this paper, available at Cambridge Journals Online (journals.cambridge.org/ect).

To proceed, we first provide the definition for conditional strong mixing processes and then to state some technical lemmas that are used in the proofs of the main results in the paper.

DEFINITION A.1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let $P_{\mathcal{B}}(\cdot) \equiv P(\cdot|\mathcal{B})$. Let $\{\xi_t, t \geq 1\}$ be a sequence of random variables defined on $(\Omega, \mathcal{A}, P)$. The sequence $\{\xi_t, t \geq 1\}$ is said to be conditionally strong mixing given $\mathcal{B}$ (or $\mathcal{B}$-strong-mixing) if there exists a nonnegative $\mathcal{B}$-measurable random variable $\alpha^{\mathcal{B}}(t)$ converging to 0 a.s. as $t \to \infty$ such that*

$$\left|P_{\mathcal{B}}(A \cap B) - P_{\mathcal{B}}(A) P_{\mathcal{B}}(B)\right| \leq \alpha^{\mathcal{B}}(t) \quad a.s. \tag{A.1}$$

for all $A \in \sigma(\xi_1, \ldots, \xi_k)$, $B \in \sigma\left(\xi_{k+t}, \xi_{k+t+1}, \ldots\right)$, and $k \geq 1$, $t \geq 1$.

The preceding definition is due to Prakasa Rao (2009); see also Roussas (2008). When one takes $\alpha^{\mathcal{B}}(t)$ as the supremum of the left-hand-side object in (A.1) over the set $\{A \in \sigma(\xi_1, \ldots, \xi_k),\ B \in \sigma\left(\xi_{k+t}, \xi_{k+t+1}, \ldots\right),\ k \geq 1\}$, we refer to it as the $\mathcal{B}$-strong-mixing coefficient.

Let C signify a generic constant whose exact value may vary from case to case. Let $C_{\mathcal{D}}$ denote an $O_P(1)$ object that depends on $\mathcal{D} \equiv \left\{F^0, \lambda^0\right\}$. Let $\delta_{NT} \equiv \min(\sqrt{N}, \sqrt{T})$ and $\gamma_{NT} \equiv N^{-1/4}T^{-1/2}$. Let $P_{\mathcal{D}}(\cdot) \equiv P(\cdot|\mathcal{D})$. Let $\mathbb{E}_{\mathcal{D}}(\cdot)$ and $\mathrm{Var}_{\mathcal{D}}(\cdot)$ denote the conditional expectation and variance given $\mathcal{D}$, respectively. Let $\|A\|_{q,\mathcal{D}} \equiv [\mathbb{E}_{\mathcal{D}}(\|A\|^q)]^{1/q}$. Let $\alpha_{ik} \equiv \lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\lambda_k^0$ and $F_{ts} \equiv F_t^{0\prime}\left(F^{0\prime}F^0/T\right)^{-1}F_s^0$. Let $\Phi_1 \equiv \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}$, $\Phi_2 \equiv F^0\left(F^{0\prime}F^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}$, and $\Phi_3 \equiv \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}$.

Let $e_{it} \equiv Y_{it} - X_{it}'\boldsymbol{\beta}^0 - F_t^{0\prime}\lambda_i^0$, $e_i \equiv (e_{i1},\ldots,e_{iT})'$, and $\mathbf{e} \equiv (e_1,\ldots,e_N)'$. Note that $e_{it} = \varepsilon_{it} + \gamma_{NT}X_{it}'\delta_i$ under $\mathbb{H}_1(\gamma_{NT})$. Let $C_{NT}^{(1)}$ and $C_{NT}^{(2)}$ denote $K\times 1$ vectors whose kth elements are given, respectively, by

$$C_{NT,k}^{(1)} = \frac{1}{NT\gamma_{NT}}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{e}'\right) \quad \text{and} \tag{A.2}$$

$$C_{NT,k}^{(2)} = -\frac{1}{NT\gamma_{NT}}\mathrm{tr}\left(\mathbf{e}M_{F^0}\mathbf{e}'M_{\lambda^0}\mathbf{X}_k\Phi_1' + \mathbf{e}'M_{\lambda^0}\mathbf{e}'M_{F^0}\mathbf{X}_k\Phi_1 + \mathbf{e}'M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{e}'\Phi_1\right). \tag{A.3}$$

The following lemma studies the asymptotic property of $\hat{\boldsymbol{\beta}}$ under $\mathbb{H}_1(\gamma_{NT})$.

LEMMA A.2. *Suppose Assumptions A.1–A.3 hold. Then under* $\mathbb{H}_1(\gamma_{NT})$

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \gamma_{NT}D_{NT}^{-1}\left(C_{NT}^{(1)} + C_{NT}^{(2)}\right) + O_P\left\{\left[\gamma_{NT}^2\left(\delta_{NT}^{-1} + \gamma_{NT}\right) + \gamma_{NT}\delta_{NT}^{-3}\right]^{1/2}\right\}.$$

Remark. Noting that under $\mathbb{H}_1(\gamma_{NT})$ with $\gamma_{NT} = N^{-1/4}T^{-1/2}$, $C_{NT,k}^{(1)} = \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\boldsymbol{\Delta}'\right) + \frac{1}{NT\gamma_{NT}}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}'\right) = (NT)^{-1}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\boldsymbol{\Delta}'\right) + O_P\left(\left[(NT)^{-1/2} + T^{-1}\right]\gamma_{NT}^{-1}\right) = O_P(1)$ and similarly $C_{NT,k}^{(2)} = O_P\left(T^{-1}\gamma_{NT}^{-1} + \gamma_{NT}\right) = O_P(1)$, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \gamma_{NT}D_{NT}^{-1}\Pi_{NT} + o_P(\gamma_{NT}) \quad \text{under Assumption A.3,} \tag{A.4}$$

where Π_{NT} is defined in (3.7). This means that $C_{NT}^{(2)}$ and the second term in $C_{NT}^{(1)}$ are asymptotically smaller than the first term in $C_{NT}^{(1)}$, and so the convergence rate of $\hat{\boldsymbol{\beta}}$ mainly hinges on the rate of local alternatives that converge to the null.

Let $\{\xi_t, t\geq 1\}$ be an l-dimensional conditional strong mixing process with mixing coefficient $\alpha^{\mathcal{D}}(\cdot)$ and distribution function $F_t(\cdot|\mathcal{D})$ given $\mathcal{D}$. The following lemma extends Davydov's inequality from the unconditional version to a conditional version.

LEMMA A.3. *Suppose that* A_1 *and* A_2 *are random variables that are measurable with respect to* $\sigma(\xi_1,\ldots,\xi_s)$ *and* $\sigma(\xi_{s+\tau},\ldots,\xi_T)$, *respectively, and that* $\|A_1\|_{p,\mathcal{D}}$ *and* $\|A_2\|_{q,\mathcal{D}}$ *are bounded in probability, where* p, $q > 1$ *and* $p^{-1}+q^{-1} < 1$. *Then* $\left|\mathbb{E}_{\mathcal{D}}(A_1A_2) - \mathbb{E}_{\mathcal{D}}(A_1)\mathbb{E}_{\mathcal{D}}(A_2)\right| \leq 8\|A_1\|_{p,\mathcal{D}}\|A_2\|_{q,\mathcal{D}}\alpha^{\mathcal{D}}(\tau)^{1-p^{-1}-q^{-1}}$.

The following lemma extends the Bernstein-type inequality for unconditional strong mixing processes to that for conditional strong mixing processes.

LEMMA A.4. *Suppose that the conditional strong mixing process* $\{\xi_t, t\geq 1\}$ *has zero mean given* $\mathcal{D}$, $\sup_{t\geq 1}|\xi_t| \leq M_0$, *and* $\sup_{t\geq 1}\left|\mathrm{Var}_{\mathcal{D}}(\xi_t)\right| \leq M_{\mathcal{D}}$. *Then for any* $\epsilon > 0$ *and* $\tau \leq T$,

$$P_{\mathcal{D}}\left(\left|T^{-1}\sum_{t=1}^{T}\xi_t\right| \geq \epsilon\right) \leq 2\tau\exp\left(-\frac{T\epsilon^2}{4\tau M_{\mathcal{D}} + 2M_0\epsilon\tau/3}\right) + 2T\alpha^{\mathcal{D}}(\tau).$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Define the mth-order U-statistic $\mathcal{U}_T = \binom{T}{m}^{-1} \sum_{1 \le t_1 < \ldots < t_m \le T} \vartheta\left(\xi_{t_1}, \ldots, \xi_{t_m}\right)$ where ϑ is symmetric in its arguments. Let $\vartheta_{(0)} = \int \ldots \int \vartheta\left(v_{t_1}, \ldots, v_{t_m}\right) \Pi_{s=1}^{m} dF_{t_s}\left(v_{t_s}|\mathcal{D}\right)$ and $\vartheta_{(c)}(v_1, \ldots, v_c) = \int \ldots \int \vartheta(v_1, \ldots, v_c, \ v_{t_{c+1}}, \ldots, v_{t_m}) \ \Pi_{s=c+1}^{m} dF_{t_s}\left(v_{t_s}|\mathcal{D}\right)$ for $c = 1, \ldots, m$. Let $h^{(1)}(v) = \vartheta_{(1)}(v) - \vartheta_{(0)}$ and $h^{(c)}(v_1, \ldots, v_c) = \vartheta_{(c)}(v_1, \ldots, v_c) - \sum_{j=1}^{c-1} \sum_{(c,j)} h^{(j)}\left(v_{t_1}, \ldots, v_{t_j}\right) - \vartheta_{(0)}$ for $c = 2, \ldots, m$, where the sum $\sum_{(c,j)}$ is taken over all subsets $1 \le t_1 < t_2 < \cdots < t_j \le c$ of $\{1, 2, \ldots, c\}$. Let $\mathcal{H}_T^{(c)} = \binom{T}{c}^{-1} \sum_{1 \le t_1 < \ldots < t_c \le T} h^{(c)}\left(\xi_{t_1}, \ldots, \xi_{t_c}\right)$. Then by Theorem 1 in Lee (1990, p. 26), we have the following Hoeffding decomposition:

$$\mathcal{U}_T = \vartheta_{(0)} + \sum_{c=1}^{m} \binom{m}{c} \mathcal{H}_T^{(c)}. \tag{A.5}$$

To study the second moment of $\mathcal{H}_T^{(c)}$ for $3 \le c \le m$, we need the following lemma.

LEMMA A.5. *Let $\{\xi_t, t \ge 1\}$ be an l-dimensional strong mixing process conditional on $\mathcal{D}$ with mixing coefficient $\alpha^{\mathcal{D}}(\cdot)$ and distribution function $F_t(\cdot|\mathcal{D})$. Let the integers $(t_1, \ldots, t_m)$ be such that $1 \le t_1 < t_2 < \cdots < t_m \le T$. Suppose that $\max\{\int |\vartheta(v_1, \ldots, v_m)|^{1+\tilde{\sigma}} \, dF_{t_1,\ldots,t_m}(v_1, \ldots, v_m|\mathcal{D}), \int |\vartheta(v_1, \ldots, v_m)|^{1+\tilde{\sigma}} \, dF_{t_1,\ldots,t_j}(v_1, \ldots, v_j|\mathcal{D}) \, dF_{t_{j+1},\ldots,t_m}\left(v_{j+1}, \ldots, v_m|\mathcal{D}\right)\} \le C_{\mathcal{D}}(t_1, \ldots, t_m)$ for some $\tilde{\sigma} > 0$, where, e.g., $F_{t_1,\ldots,t_m}(v_1, \ldots, v_m|\mathcal{D})$ denotes the distribution function of $\left(\xi_{t_1}, \ldots, \xi_{t_m}\right)$ given $\mathcal{D}$. Then*

$$\left| \int \vartheta(v_1, \ldots, v_m) \, dF_{t_1,\ldots,t_m}(v_1, \ldots, v_m|\mathcal{D}) - \int \vartheta(v_1, \ldots, v_m) \, dF^{(1)}_{t_1,\ldots,t_j}\left(v_1, \ldots, v_j|\mathcal{D}\right) \right.$$
$$\left. \times \, dF_{t_{j+1},\ldots,t_m}\left(v_{j+1}, \ldots, v_m\right) \right| \le 4 C_{\mathcal{D}}(t_1, \ldots, t_m)^{1/(1+\tilde{\sigma})} \alpha^{\mathcal{D}}\left(t_{j+1} - t_j\right)^{\tilde{\sigma}/(1+\tilde{\sigma})}.$$

LEMMA A.6. *Let $\{\xi_t, t \ge 1\}$ be an l-dimensional strong mixing process conditional on $\mathcal{D}$ with mixing coefficient $\alpha^{\mathcal{D}}(\cdot)$ and distribution function $F_t(\cdot|\mathcal{D})$. Suppose that $\alpha^{\mathcal{D}}(s) = O_{a.s.}\left(s^{-3(2+\sigma)/\sigma - \epsilon}\right)$. If there exists $\sigma > 0$ such that*

$$L_T \equiv \max\left\{ \int \left|\vartheta\left(v_{t_1}, \ldots, v_{t_m}\right)\right|^{2+\sigma} \ \Pi_{s=1}^{m} dF_{t_s}\left(v_{t_s}|\mathcal{D}\right), \ \mathbb{E}_{\mathcal{D}} \left|\vartheta\left(\xi_{t_1}, \ldots, \xi_{t_m}\right)\right|^{2+\sigma} \right\}$$
$$\le \sum_{q=1}^{m} C_{q\mathcal{D}}\left(t_q\right)$$

and $T^{-1} \sum_{q=1}^{m} \sum_{t_q=1}^{T} C_{q\mathcal{D}}\left(t_q\right) = O_P(1)$, then $\mathbb{E}_{\mathcal{D}}[\mathcal{H}_T^{(c)}]^2 = O_P\left(T^{-3}\right)$ for $3 \le c \le m$.

LEMMA A.7. *Recall from Assumption A.1 that $\Omega_i \equiv \mathbb{E}_{\mathcal{D}}\left(X_i' X_i\right)/T$ and $\hat{\Omega}_i \equiv X_i' X_i / T$. Suppose Assumptions A.1–A.3 hold. Then*

(i) $\mu_1(\hat{\Omega}_i) \le \mu_1(\Omega_i) + O_P\left(T^{-1/2}\right)$,

(ii) $\mu_{\min}(\hat{\Omega}_i) \ge \mu_{\min}(\Omega_i) - O_P\left(T^{-1/2}\right)$,

(iii) $\max_{1 \le i \le N} ||\hat{\Omega}_i - \Omega_i||_F = O_P(a_{NT})$, *and*

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

(iv) $\max_{1\leq i\leq N}||\hat{\Omega}_i^{-1}-\Omega_i^{-1}||_F = O_P(a_{NT})$, *where* $a_{NT}\equiv \max\{(NT)^{1/(4+2\sigma)}\log(NT)/T, (\log(NT)/T)^{1/2}\}$.

LEMMA A.8. *Recall from Section 3.2 that* $H_i \equiv M_{F^0}P_{X_i}M_{F^0}$ *and* $h_{i,ts}$ *denotes the* (t,s)*th element of* H_i : $h_{i,ts}=\sum_{r=1}^T\sum_{q=1}^T \eta_{tr}X_{ir}'\left(X_i'X_i\right)^{-1}X_{iq}\eta_{qs}$ *where* η_{tr} *denotes the* (t,r)*th element of* M_{F^0}. *Let* $\bar{h}_{i,ts}\equiv T^{-1}\sum_{r=1}^T\sum_{q=1}^T\eta_{tr}X_{ir}'\,\Omega_i^{-1}X_{iq}\eta_{qs}$. *Suppose Assumptions A.1–A.3 hold. Then* $D_{1NT}\equiv N^{-1/2}\sum_{i=1}^N\sum_{1\leq s\neq t\leq T}\varepsilon_{it}\varepsilon_{is}\left(h_{i,ts}-\bar{h}_{i,ts}\right)=o_P(1)$.

LEMMA A.9. *Suppose Assumptions A.1–A.3 hold. Then*

(i) $D_{2NT,1}\equiv T^{-2}N^{-1/2}\sum_{i=1}^N\sum_{1\leq s<t\leq T}\sum_{r=1}^T\varepsilon_{it}\varepsilon_{is}F_{tr}\left[X_{ir}-\mathbb{E}_{\mathcal{D}}(X_{ir})\right]'\Omega_i^{-1}X_{is}=o_P(1)$,

(ii) $D_{2NT,2}\equiv T^{-3}N^{-1/2}\sum_{i=1}^N\sum_{1\leq s<t\leq T}\sum_{r=1}^T\sum_{q=1}^T\varepsilon_{it}\varepsilon_{is}F_{tr}\left[X_{ir}-\mathbb{E}_{\mathcal{D}}(X_{ir})\right]'\Omega_i^{-1}\left[X_{iq}-\mathbb{E}_{\mathcal{D}}\left(X_{iq}\right)\right]\times F_{qs}=o_P(1)$,

(iii) $D_{2NT,3}\equiv T^{-3}N^{-1/2}\sum_{i=1}^N\sum_{1\leq s<t\leq T}\sum_{r=1}^T\sum_{q=1}^T\varepsilon_{it}\varepsilon_{is}F_{tr}\left[X_{ir}-\mathbb{E}_{\mathcal{D}}(X_{ir})\right]'\Omega_i^{-1}\mathbb{E}_{\mathcal{D}}\left(X_{iq}\right)F_{qs}=o_P(1)$.

LEMMA A.10. *Suppose Assumptions A.1–A.3 hold. Then* $D_{3NT}\equiv\gamma_{NT}N^{-1/2}\sum_{i=1}^N\varepsilon_i'M_{F^0}P_{X_i}M_{F^0}X_i\delta_i=o_P(1)$.

APPENDIX B: Proofs of the Results in Section 3

Proof of Theorem 3.1. The proof is a special case of that of Theorem 3.3 and thus is omitted. ■

Proof of Theorem 3.2. The theorem can be proved under $\mathbb{H}_1(\gamma_{NT})$. The proof is quite involved and is given in the supplementary Appendix. ■

Proof of Theorem 3.3. Following Moon and Weidner (2010a), we can readily show that

$$M_{\hat{F}}=M_{F^0}+\sum_{k=1}^K\left(\beta_k^0-\hat{\beta}_k\right)M_k^{(0)}+M^{(1)}+M^{(2)}+M^{(rem)}, \tag{B.1}$$

where

$$M_k^{(0)}=-M_{F^0}\mathbf{X}_k'\Phi_1-\Phi_1'\mathbf{X}_kM_{F^0}\quad\text{for } k=1,\ldots,K,$$

$$M^{(1)}=-M_{F^0}\boldsymbol{\varepsilon}'\Phi_1-\Phi_1'\boldsymbol{\varepsilon}M_{F^0},$$

$$M^{(2)}=M_{F^0}\boldsymbol{\varepsilon}'\Phi_1\boldsymbol{\varepsilon}'\Phi_1+\Phi_1'\boldsymbol{\varepsilon}\Phi_1'\boldsymbol{\varepsilon}M_{F^0}-M_{F^0}\boldsymbol{\varepsilon}'M_{\lambda^0}\boldsymbol{\varepsilon}\Phi_2-\Phi_2\boldsymbol{\varepsilon}'M_{\lambda^0}\boldsymbol{\varepsilon}M_{F^0}$$
$$-M_{F^0}\boldsymbol{\varepsilon}'\Phi_3\boldsymbol{\varepsilon}M_{F^0}+\Phi_1'\boldsymbol{\varepsilon}M_{F^0}\boldsymbol{\varepsilon}'\Phi_1,$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

and the remainder $M^{(rem)}$ satisfies

$$\begin{aligned}\left\|M^{(rem)}\right\|_F &= O_P\left(\left(\delta_{NT}^{-1}+\gamma_{NT}+\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right\|\right)\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right\|\right.\\ &\qquad \left.+(NT)^{-3/2}\max\left(\sqrt{N},\sqrt{T}\right)^3+\gamma_{NT}^3\right)\\ &= O_P\left(\delta_{NT}^{-1}\gamma_{NT}+\delta_{NT}^{-3}\right)\\ &= O_P\left(\delta_{NT}^{-1}\gamma_{NT}\right) \text{ under Assumption A.3(a).} \qquad \textbf{(B.2)}\end{aligned}$$

It is straightforward to show that

$$\left\|M_k^{(0)}\right\|_F = O_P(1) \quad \text{for } k=1,\ldots,K, \qquad \left\|M^{(1)}\right\|_F = O_P(N^{-1/2}), \quad \text{and}$$
$$\left\|M^{(2)}\right\|_F = O_P\left(\delta_{NT}^{-2}\right). \qquad \textbf{(B.3)}$$

Combining (B.1) with (2.12) yields

$$\begin{aligned}\hat{\varepsilon}_i &= \left[M_{F^0}(\varepsilon_i+c_i)+\sum_{k=1}^K\left(\beta_k^0-\hat{\beta}_k\right)M_k^{(0)}\left(\varepsilon_i+F^0\lambda_i^0+c_i\right)\right]+M^{(1)}\left(\varepsilon_i+F^0\lambda_i^0+c_i\right)\\ &\quad+\left(M^{(2)}+M^{(rem)}\right)\left(\varepsilon_i+F^0\lambda_i^0+c_i\right)\\ &\equiv d_{1i}+d_{2i}+d_{3i}, \quad \text{say,} \qquad \textbf{(B.4)}\end{aligned}$$

where $c_i \equiv X_i(\boldsymbol{\beta}^0-\hat{\boldsymbol{\beta}})+X_i\left(\boldsymbol{\beta}_i^0-\boldsymbol{\beta}^0\right)=X_i(\boldsymbol{\beta}^0-\hat{\boldsymbol{\beta}})+\gamma_{NT}X_i\delta_i$ satisfies

$$\|c_i\|_F = O_P(\gamma_{NT})\|X_i\|_F. \qquad \textbf{(B.5)}$$

It follows that

$$\begin{aligned}\sqrt{V_{NT}}J_{NT} &= N^{-1/2}\sum_{i=1}^N(d_{1i}+d_{2i}+d_{3i})'P_{X_i}(d_{1i}+d_{2i}+d_{3i})-B_{NT}\\ &= \left\{N^{-1/2}\sum_{i=1}^N d_{1i}'P_{X_i}d_{1i}-B_{NT}\right\}+N^{-1/2}\sum_{i=1}^N d_{2i}'P_{X_i}d_{2i}\\ &\quad+N^{-1/2}\sum_{i=1}^N d_{3i}'P_{X_i}d_{3i}+2N^{-1/2}\sum_{i=1}^N d_{1i}'P_{X_i}d_{2i}+2N^{-1/2}\sum_{i=1}^N d_{1i}'P_{X_i}d_{3i}\\ &\quad+2N^{-1/2}\sum_{i=1}^N d_{2i}'P_{X_i}d_{3i}\\ &\equiv (A_{1NT}-B_{NT})+A_{2NT}+A_{3NT}+2A_{4NT}+2A_{5NT}+2A_{6NT}, \quad \text{say.}\end{aligned}$$

We complete the proof by showing that under $\mathbb{H}_1(\gamma_{NT})$, (a) $A_{1NT}-B_{NT}-\Theta_{NT}\xrightarrow{D}N(0,V_0)$, (b) $A_{sNT}=o_P(1)$ for $s=2,\ldots,6$, where Θ_{NT} is defined in (3.9). We prove (a) in Proposition B.1 and (b) in Propositions B.2–B.6, which follow.

PROPOSITION B.1. *$A_{1NT}-B_{NT}-\Theta_{NT}\xrightarrow{D}N(0,V_0)$ under $\mathbb{H}_1(\gamma_{NT})$.*

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Proof. Observe that $A_{1NT} = N^{-1/2}\sum_{i=1}^{N} d_{1i}' P_{X_i} d_{1i} = A_{1NT,1} + A_{1NT,2} + 2A_{1NT,3}$, where

$$A_{1NT,1} = N^{-1/2}\sum_{i=1}^{N}\left(\varepsilon_i' + c_i'\right) M_{F^0} P_{X_i} M_{F^0}\left(\varepsilon_i + c_i\right),$$

$$A_{1NT,2} = N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\beta_k^0 - \hat{\beta}_k\right)\left(\varepsilon_i' + \lambda_i^{0\prime} F^{0\prime} + c_i'\right) M_k^{(0)} P_{X_i} \sum_{l=1}^{K}\left(\beta_l^0 - \hat{\beta}_l\right) M_l^{(0)}$$
$$\times\left(\varepsilon_i + F^0\lambda_i^0 + c_i\right),$$

$$A_{1NT,3} = N^{-1/2}\sum_{i=1}^{N}\left(\varepsilon_i' + c_i'\right) M_{F^0} P_{X_i} \sum_{k=1}^{K}\left(\beta_k^0 - \hat{\beta}_k\right) M_k^{(0)}\left(\varepsilon_i + F^0\lambda_i^0 + c_i\right).$$

We prove the proposition by showing that (i) $A_{1NT,1} - B_{NT} - \Theta_{1NT} \xrightarrow{D} N(0, V_0)$, (ii) $A_{1NT,2} = \Theta_{2NT} + o_P(1)$, and (iii) $A_{1NT,3} = \Theta_{3NT} + o_P(1)$, where

$$\Theta_{1NT} \equiv (NT)^{-1}\sum_{i=1}^{N}\left(\delta_i - D_{NT}^{-1}\Pi_{NT}\right)' X_i' M_{F^0} P_{X_i} M_{F^0} X_i\left(\delta_i - D_{NT}^{-1}\Pi_{NT}\right),$$

$$\Theta_{2NT} \equiv (NT)^{-1}\sum_{i=1}^{N}\left(D_{NT}^{-1}\Pi_{NT}\right)' \bar{X}_i' P_{X_i}\bar{X}_i\left(D_{NT}^{-1}\Pi_{NT}\right),$$

$$\Theta_{3NT} \equiv (NT)^{-1}\sum_{i=1}^{N}\left(\delta_i - D_{NT}^{-1}\Pi_{NT}\right)' X_i' M_{F^0} P_{X_i}\bar{X}_i\left(D_{NT}^{-1}\Pi_{NT}\right),$$

and $\bar{X}_i \equiv N^{-1}\sum_{l=1}^{N}\alpha_{il} M_{F^0} X_l$. The result follows because in view of the fact that $M_{F^0}X_i\delta_i - M_{F^0}X_i D_{NT}^{-1} \times \Pi_{NT} + \bar{X}_i D_{NT}^{-1}\Pi_{NT} = M_{F^0}X_i\delta_i - (M_{F^0}X_i - N^{-1}\sum_{l=1}^{N}\alpha_{il}M_{F^0}X_l)D_{NT}^{-1}\Pi_{NT} = M_{F^0}X_i\delta_i - \tilde{X}_i D_{NT}^{-1}\Pi_{NT}$, we have $\Theta_{1NT} + \Theta_{2NT} + 2\Theta_{3NT} = (NT)^{-1}\sum_{l=1}^{N}(M_{F^0}X_i\delta_i - \tilde{X}_i D_{NT}^{-1}\Pi_{NT})' P_{X_i}(M_{F^0}X_i\delta_i - \tilde{X}_i D_{NT}^{-1}\Pi_{NT}) = \Theta_{NT}$.

Step 1. We prove (i) $A_{1NT,1} - B_{NT} - \Theta_{1NT} \xrightarrow{D} N(0, V_0)$ under $\mathbb{H}_1(\gamma_{NT})$. Observe that

$$\begin{aligned} A_{1NT,1} - B_{NT} - \Theta_{1NT} &= \left(N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' M_{F^0} P_{X_i} M_{F^0}\varepsilon_i - B_{NT}\right) \\ &\quad + \left(N^{-1/2}\sum_{i=1}^{N} c_i' M_{F^0} P_{X_i} M_{F^0} c_i - \Theta_{1NT}\right) \\ &\quad + 2N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' M_{F^0} P_{X_i} M_{F^0} c_i \\ &\equiv A_{1NT,11} + A_{1NT,12} + 2A_{1NT,13}, \quad \text{say.} \end{aligned}$$

It suffices to show that (i1) $A_{1NT,11} \xrightarrow{D} N(0, V_0)$, (i2) $A_{1NT,12} = o_P(1)$, and (i3) $A_{1NT,13} = o_P(1)$.

First, we show (i1). Recall from Lemma A.8 that $H_i = M_{F^0} P_{X_i} M_{F^0}$, $h_{i,ts}$ denotes the (t,s)th element of H_i : $h_{i,ts} = \sum_{r=1}^{T}\sum_{q=1}^{T}\eta_{tr}\ X_{ir}'\left(X_i'X_i\right)^{-1} X_{iq}\eta_{qs}$ and

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$\bar{h}_{i,ts} \equiv T^{-1}\sum_{r=1}^{T}\sum_{q=1}^{T}\eta_{tr}\, X_{ir}'\Omega_i^{-1}X_{iq}\eta_{qs}$. Then we have

$$A_{1NT,11} = \frac{2}{\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\varepsilon_{it}\varepsilon_{is}\bar{h}_{i,ts} + \frac{2}{\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\varepsilon_{it}\varepsilon_{is}\left(h_{i,ts}-\bar{h}_{i,ts}\right)$$

$$\equiv A_{1NT,111} + A_{1NT,112}, \quad \text{say.}$$

By Lemma A.8, $A_{1NT,112} = o_P(1)$. Using $\eta_{tr} = 1_{tr} - T^{-1}F_{tr}$, we have

$$\begin{aligned}
A_{1NT,111} &= \frac{2}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}\varepsilon_{it}\varepsilon_{is}\eta_{tr}X_{ir}'\Omega_i^{-1}X_{iq}\eta_{qs}\\
&= \frac{2}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\varepsilon_{it}\varepsilon_{is}X_{it}'\Omega_i^{-1}X_{is}\\
&\quad - \frac{4}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\varepsilon_{it}\varepsilon_{is}F_{tr}\mathbb{E}_{\mathcal{D}}\left(X_{ir}'\right)\Omega_i^{-1}X_{is}\\
&\quad + \frac{2}{T^3\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}\varepsilon_{it}\varepsilon_{is}F_{tr}\mathbb{E}_{\mathcal{D}}\left(X_{ir}'\right)\Omega_i^{-1}\mathbb{E}_{\mathcal{D}}\left(X_{iq}\right)F_{qs}\\
&\quad - \frac{4}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\varepsilon_{it}\varepsilon_{is}F_{tr}\left[X_{ir}-\mathbb{E}_{\mathcal{D}}\left(X_{ir}\right)\right]'\Omega_i^{-1}X_{is}\\
&\quad + \frac{2}{T^3\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}\varepsilon_{it}\varepsilon_{is}F_{tr}\left[X_{ir}-\mathbb{E}_{\mathcal{D}}\left(X_{ir}\right)\right]'\Omega_i^{-1}\\
&\quad \times\left[X_{iq}-\mathbb{E}_{\mathcal{D}}\left(X_{iq}\right)\right]F_{qs}\\
&\quad + \frac{4}{T^3\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}\varepsilon_{it}\varepsilon_{is}F_{tr}\left[X_{ir}-\mathbb{E}_{\mathcal{D}}\left(X_{ir}\right)\right]'\Omega_i^{-1}\\
&\quad \times\mathbb{E}_{\mathcal{D}}\left(X_{iq}\right)F_{qs}\\
&\equiv A_{1NT,111a} + A_{1NT,111b} + A_{1NT,111c} + A_{1NT,111d} + A_{1NT,111e}\\
&\quad + A_{1NT,111f}, \quad \text{say.}
\end{aligned}$$

By Lemma A.9, $A_{1NT,111d} + A_{1NT,111e} + A_{1NT,111f} = o_P(1)$. We are left to show that $Z_{NT} \equiv A_{1NT,111a} + A_{1NT,111b} + A_{1NT,111c} \xrightarrow{D} N(0, V_0)$. Observe that

$$\begin{aligned}
Z_{NT} &= \frac{2}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\varepsilon_{it}\varepsilon_{is}\left(X_{it}-T^{-1}\sum_{r=1}^{T}F_{tr}\mathbb{E}_{\mathcal{D}}\left(X_{ir}\right)\right)'\Omega_i^{-1}\\
&\quad\times\left(X_{is}-T^{-1}\sum_{q=1}^{T}\mathbb{E}_{\mathcal{D}}\left(X_{iq}\right)F_{qs}\right)\\
&= \sum_{t=2}^{T}Z_{NT,t},
\end{aligned}$$

where $Z_{NT,t} \equiv 2T^{-1}N^{-1/2}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\varepsilon_{it}\varepsilon_{is}\bar{b}_{it}'\bar{b}_{is}$, $\bar{b}_{it} \equiv \Omega_i^{-1/2}b_{it}$ and $b_{it} \equiv X_{it} - T^{-1}\sum_{r=1}^{T}F_{tr}\mathbb{E}_{\mathcal{D}}\left(X_{ir}\right)$. By Assumption A.2(c)

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$\mathbb{E}\left(Z_{NT,t}|\mathcal{F}_{NT,t-1}\right) \equiv 2T^{-1}N^{-1/2}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\varepsilon_{is}\bar{b}'_{it}\bar{b}_{is}\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{NT,t-1}) = 0.$$

That is, $\{Z_{NT,t},\ \mathcal{F}_{NT,t}\}$ is an m.d.s. By the martingale CLT (e.g., Pollard, 1984, p. 171), it suffices to show that

$$\mathcal{Z} \equiv \sum_{t=2}^{T}\mathbb{E}_{\mathcal{F}_{NT,t-1}}\left|Z_{NT,t}\right|^4 = o_P(1) \quad \text{and} \quad \sum_{t=2}^{T}Z_{NT,t}^2 - V_{NT} = o_P(1), \tag{B.6}$$

where $\mathbb{E}_{\mathcal{F}_{NT,t-1}}$ denotes expectation conditional on $\mathcal{F}_{NT,t-1}$. Observing that $\mathcal{Z} \geq 0$, it suffices to show $\mathcal{Z} = o_P(1)$ by showing that $\mathbb{E}_{\mathcal{D}}(\mathcal{Z}) = o_P(1)$ by Markov's inequality. Noting that $(\varepsilon_{it}, X_{it})$ are independent across i given $\mathcal{D}$ by Assumption A.2(b), and $\{\varepsilon_{it}, \mathcal{F}_{NT,t}\}$ is an m.d.s. by Assumption A.2(c), we have

$$\begin{aligned}\mathbb{E}_{\mathcal{D}}(\mathcal{Z}) &= \frac{16}{T^4N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{1\leq r,s,q,v\leq t-1}\\ &\quad\times\mathbb{E}_{\mathcal{D}}\left(\bar{b}'_{it}\bar{b}_{is}\bar{b}'_{jt}\bar{b}_{jr}\bar{b}'_{kt}\bar{b}_{kq}\bar{b}'_{lt}\bar{b}_{lv}\varepsilon_{is}\varepsilon_{jr}\varepsilon_{kq}\varepsilon_{lv}\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}\right)\\ &= 48\mathcal{Z}_1 + 16\mathcal{Z}_2,\end{aligned}$$

where

$$\begin{aligned}\mathcal{Z}_1 &\equiv \frac{1}{T^4N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{1\leq r,s,q,v\leq t-1}\mathbb{E}_{\mathcal{D}}\left(\bar{b}'_{it}\bar{b}_{is}\bar{b}'_{it}\bar{b}_{ir}\varepsilon_{is}\varepsilon_{ir}\varepsilon_{it}^2\right)\\ &\quad\times\mathbb{E}_{\mathcal{D}}\left(\bar{b}'_{jt}\bar{b}_{jq}\bar{b}'_{jt}\bar{b}_{jv}\varepsilon_{jq}\varepsilon_{jv}\varepsilon_{jt}^2\right),\end{aligned} \tag{B.7}$$

$$\mathcal{Z}_2 \equiv \frac{1}{T^4N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{1\leq r,s,q,v\leq t-1}\mathbb{E}_{\mathcal{D}}\left(\bar{b}'_{it}\bar{b}_{is}\bar{b}'_{it}\bar{b}_{ir}\bar{b}'_{it}\bar{b}_{iq}\bar{b}'_{it}\bar{b}_{iv}\varepsilon_{is}\varepsilon_{ir}\varepsilon_{iq}\varepsilon_{iv}\varepsilon_{it}^4\right). \tag{B.8}$$

For the moment we assume that $K = 1$ so that we can treat the $K \times 1$ vector $\bar{b}_{it}$ as a scalar. (The general case follows from the Slutsky lemma and the fact that $\bar{b}'_{it}\bar{b}_{is}\bar{b}'_{it}\bar{b}_{ir} = \sum_{k=1}^{K}\sum_{l=1}^{K}\bar{b}_{it,k}\bar{b}_{is,k}\bar{b}_{it,l}\bar{b}_{ir,l}$ where $\bar{b}_{it,k}$ denotes the kth element of $\bar{b}_{it}$.) To bound the summation in (B.7), we consider three cases for the time indexes in $S \equiv \{r,s,q,v,t-1\}$: (a) $\#S = 5$, (b) $\#S = 4$, and (c) $\#S \leq 3$. We use EZ_{1a}, EZ_{1b}, and EZ_{1c} to denote the corresponding summations when the time indices are restricted to be cases (a), (b) and (c), respectively. In case (a), using Davydov's inequality in Lemma A.3 yields

$$\left|\mathbb{E}_{\mathcal{D}}\left(\bar{b}_{it}\bar{b}_{is}\bar{b}_{it}\bar{b}_{ir}\varepsilon_{is}\varepsilon_{ir}\varepsilon_{it}^2\right)\right| \leq 8C_{9i\mathcal{D}}(t,s,r)\,\alpha_{\mathcal{D}}(t-1-(s\vee r))^{(1+\sigma)/(2+\sigma)}, \tag{B.9}$$

where $a \vee b \equiv \max(a,b)$ and $C_{9i\mathcal{D}}(t,s,r) \equiv \left\|\bar{b}_{is}\bar{b}_{ir}\varepsilon_{is}\varepsilon_{ir}\right\|_{4+2\sigma,\mathcal{D}}\left\|\bar{b}_{it}^2\varepsilon_{it}^2\right\|_{4+2\sigma,\mathcal{D}}$. A similar inequality holds for $\mathbb{E}_{\mathcal{D}}(\bar{b}_{jt}\bar{b}_{jq}\bar{b}_{jt}\bar{b}_{jv}\varepsilon_{jq}\varepsilon_{jv}\varepsilon_{jt}^2)$. By the repeated use of the Cauchy–Schwarz and Jensen inequalities,

$$\begin{aligned}\left|C_{9i\mathcal{D}}(t_1,t_2,t_3)\right| &\leq \frac{1}{2}\left[\left\|\bar{b}_{it_2}\bar{b}_{it_3}\varepsilon_{it_2}\varepsilon_{it_3}\right\|_{4+2\sigma,\mathcal{D}}^2 + \left\|\bar{b}_{it_1}^2\varepsilon_{it_1}^2\right\|_{4+2\sigma,\mathcal{D}}^2\right]\\ &\leq \frac{1}{4}\left\{\left\|\bar{b}_{is}\varepsilon_{is}\right\|_{8+4\sigma,\mathcal{D}}^2 + \left\|\bar{b}_{it_3}\varepsilon_{it_3}\right\|_{8+4\sigma,\mathcal{D}}^2 + 2\left\|\bar{b}_{it_1}^2\varepsilon_{it_1}^2\right\|_{4+2\sigma,\mathcal{D}}^2\right\}\\ &\leq \sum_{q=1}^{3}\bar{C}_{9i\mathcal{D}}(t_q),\end{aligned}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

where $\bar{C}_{9i\mathcal{D}}(t) = \frac{1}{2}\left\{\left\|\bar{b}_{it}\varepsilon_{it}\right\|^2_{8+4\sigma,\mathcal{D}} + \left\|\bar{b}^2_{it}\varepsilon^2_{it}\right\|^2_{4+2\sigma,\mathcal{D}}\right\}$. With this, we can readily show that

$$\begin{aligned}\mathbb{E}Z_{1a} &\le \frac{C}{T^4}\sum_{t_1=2}^{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\sum_{1\le t_2,t_3\le t_1-1}\left[\sum_{q=1}^{3}\bar{C}_{9i\mathcal{D}}(t_q)\right]\alpha_{\mathcal{D}}(t_1-1-(t_2\vee t_3))^{(1+\sigma)/(2+\sigma)}\right\}^2 \\ &= O_P\left(T^{-1}\right).\end{aligned}$$

In case (b), we consider two subcases: (b1) one and only one of r,s,q,v equals $t-1$; (b2) $\#\{r,s,q,v\}=3$. We use EZ_{1b1} and EZ_{1b2} to denote the corresponding summations when the individual indexes are restricted to subcases (b1) and (b2), respectively. In subcase (b1), without loss of generality we assume that $v=t-1$ and apply that

$$\left|\mathbb{E}_{\mathcal{D}}\left(\bar{b}_{jt}\bar{b}_{jq}\bar{b}_{jt}\bar{b}_{j,t-1}\varepsilon_{jq}\varepsilon_{j,t-1}\varepsilon^2_{jt}\right)\right| \le 8C_{10j\mathcal{D}}(t,q)\,\alpha_{\mathcal{D}}(t-1-q)^{(1+\sigma)/(2+\sigma)}$$

for $C_{10j\mathcal{D}}(t,q) \equiv \left\|\bar{b}_{jq}\varepsilon_{jq}\right\|_{8+4\sigma,\mathcal{D}}\left\|\bar{b}^2_{jt}\bar{b}_{j,t-1}\varepsilon_{j,t-1}\varepsilon^2_{jt}\right\|_{(8+4\sigma)/3,\mathcal{D}} \le \bar{C}_{10i\mathcal{D}}(t) + \bar{C}_{10i\mathcal{D}}(q)$ with $\bar{C}_{10i\mathcal{D}}(t) \equiv \left\|\bar{b}_{jt}\varepsilon_{jt}\right\|^2_{8+4\sigma,\mathcal{D}} + \left\|\bar{b}^2_{jt}\bar{b}_{j,t-1}\varepsilon_{j,t-1}\varepsilon^2_{jt}\right\|^2_{(8+4\sigma)/3,\mathcal{D}}$ and (B.9) to obtain

$$\begin{aligned}\mathbb{E}Z_{1b1} \le \frac{C}{T^4}\sum_{t_1=2}^{T}&\left\{\frac{1}{N}\sum_{i=1}^{N}\sum_{1\le t_2,t_3\le t_1-1}\left[\sum_{q=1}^{3}\bar{C}_{9i\mathcal{D}}(t_q)\right]\right. \\ &\left.\times\,\alpha_{\mathcal{D}}(t_1-1-(t_2\vee t_3))^{(1+\sigma)/(2+\sigma)}\right\} \\ &\times\left\{\frac{1}{N}\sum_{j=1}^{N}\sum_{1\le t_4\le t_1-1}\left[\bar{C}_{10i\mathcal{D}}(t_1)+\bar{C}_{10i\mathcal{D}}(t_4)\right]\alpha_{\mathcal{D}}(t_1-1-t_4)^{(1+\sigma)/(2+\sigma)}\right\} \\ &= O_P\left(T^{-1}\right).\end{aligned}$$

In subcase (b2), without loss of generality we assume that $q=v$ and $r<s<t-1$. We consider two subsubcases: (b21) either $t-1-s>\tau_*$ or $s-r>\tau_*$, (b22) $t-1-s\le\tau_*$ and $s-t\le\tau_*$. In the first case, we have

$$\left|\mathbb{E}_{\mathcal{D}}\left(\bar{b}_{it}\bar{b}_{is}\bar{b}_{it}\bar{b}_{ir}\varepsilon_{is}\varepsilon_{ir}\varepsilon^2_{it}\right)\right| \le \begin{cases} 8C_{11i\mathcal{D}}(t,s,r)\,\alpha_{\mathcal{D}}(\tau_*)^{(1+\sigma)/(2+\sigma)} & \text{if } t-1-s>\tau_* \\ 8C_{12i\mathcal{D}}(t,s,r)\,\alpha_{\mathcal{D}}(\tau_*)^{(1+\sigma)/(2+\sigma)} & \text{if } s-r>\tau_* \end{cases},$$

where $C_{11i\mathcal{D}}(t,s,r) \equiv \left\|\bar{b}_{it}\bar{b}_{it}\varepsilon^2_{it}\right\|_{4+2\sigma,\mathcal{D}}\left\|\bar{b}_{is}\bar{b}_{ir}\varepsilon_{is}\varepsilon_{ir}\right\|_{4+2\sigma,\mathcal{D}}$ and $C_{12i\mathcal{D}}(t,s,r) \equiv \left\|\bar{b}_{it}\bar{b}_{is}\bar{b}_{it}\varepsilon_{is}\varepsilon^2_{it}\right\|_{(8+4\sigma)/3,\mathcal{D}}\left\|\bar{b}_{ir}\varepsilon_{ir}\right\|_{8+4\sigma,\mathcal{D}}$. These results, in conjunction with the fact that the total number of terms in the summation in subcases (b22) is of order $O\left(N^2T^3\tau^2_*\right)$, imply that

$$\begin{aligned}\mathbb{E}Z_{1b2} &\le O_P\left(T^2\alpha_{\mathcal{D}}(\tau_*)^{(1+\sigma)/(2+\sigma)}\right) + CT^{-4}N^{-2}O_P\left(N^2T^3\tau^2_*\right) \\ &= O_P\left(T^2\alpha_{\mathcal{D}}(\tau_*)^{(1+\sigma)/(2+\sigma)} + T^{-1}\tau^2_*\right) = o_P(1) \quad \text{by Assumption A.2(a)}.\end{aligned}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Consequently, $\mathbb{E}Z_{1b} = o_P(1)$. In case (c), we have $\mathbb{E}Z_{1c} = O_P\left(T^{-1}\right)$ as the number of terms in the summation is $O\left(N^2T^3\right)$ and each term in absolute value has bounded expectation. It follows that $\mathcal{Z}_1 = o_P(1)$.

To bound $\mathcal{Z}_2$, we consider two cases for the set of indexes $S \equiv \{r,s,q,v,t-1\}$, (a) $\#S = 5$ and (b) all the other cases. We use $\mathbb{E}Z_{2a}$ and $\mathbb{E}Z_{2b}$ to denote the corresponding summations when the individual indexes are restricted to subcases (a) and (b), respectively. In the first case, letting $c = \max(s,r,q,v)$ we have

$$\left|\mathbb{E}_{\mathcal{D}}\left(\bar{b}_{it}^4\bar{b}_{is}\bar{b}_{ir}\bar{b}_{iq}\bar{b}_{iv}\varepsilon_{is}\varepsilon_{ir}\varepsilon_{iq}\varepsilon_{iv}\varepsilon_{it}^4\right)\right| \leq 8C_{13i\mathcal{D}}(t,s,r,q,v)\,\alpha_{\mathcal{D}}(t-1-c)^{\sigma/(2+\sigma)},$$

where $C_{13i\mathcal{D}}(t,s,r,q,v) \equiv \left\|\bar{b}_{is}\bar{b}_{ir}\bar{b}_{iq}\bar{b}_{iv}\varepsilon_{is}\varepsilon_{ir}\varepsilon_{iq}\varepsilon_{iv}\right\|_{2+\sigma,\mathcal{D}}\left\|\bar{b}_{it}^4\varepsilon_{it}^4\right\|_{2+\sigma,\mathcal{D}}$. It is easy to verify that $C_{13i\mathcal{D}}(t_1,t_2,\ t_3,t_4,t_5) \leq \sum_{q=1}^5 \bar{C}_{13i\mathcal{D}}\left(t_q\right)$ where $\bar{C}_{13i\mathcal{D}}(t) \equiv \left\|\bar{b}_{it}\varepsilon_{it}\right\|_{8+4\sigma,\mathcal{D}}^2$. Then $\mathbb{E}Z_{2a} \leq CN^{-2}T^{-1}\sum_{i=1}^N\sum_{t=1}^T \bar{C}_{13i\mathcal{D}}(t)\sum_{s=1}^T \alpha_{\mathcal{D}}(s)^{\sigma/(2+\sigma)} = O_P\left(N^{-1}\right)$. In case (b), we have $\mathbb{E}Z_{2b} = O_P\left(N^{-1}\right)$. It follows that $\mathcal{Z}_2 = O_P\left(N^{-1}\right)$, and thus $\mathcal{Z} = o_P(1)$. Consequently the first part of (B.6) follows.

For the second part of (B.6), noting that $(\varepsilon_{it}, X_{it})$ are independent across i given $\mathcal{D}$ by Assumption A.2(b), and $\left\{\varepsilon_{it}, \mathcal{F}_{NT,t}\right\}$ is an m.d.s. by Assumption A.2(c), we have by the law of iterated expectations that

$$\sum_{t=2}^T \mathbb{E}_{\mathcal{D}}\left(Z_{NT,t}^2\right) = 4T^{-2}N^{-1}\sum_{t=2}^T \mathbb{E}_{\mathcal{D}}\left[\sum_{i=1}^N\sum_{s=1}^{t-1}\varepsilon_{it}\varepsilon_{is}\bar{b}_{it}'\bar{b}_{is}\right]^2$$

$$= 4T^{-2}N^{-1}\sum_{t=2}^T\sum_{i=1}^N\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\mathbb{E}_{\mathcal{D}}\left(\varepsilon_{it}^2\varepsilon_{is}\varepsilon_{ir}\bar{b}_{it}'\bar{b}_{is}\bar{b}_{it}'\bar{b}_{ir}\right) = V_{NT}.$$

In addition, we can show by straightforward moment calculations that $\mathbb{E}_{\mathcal{D}}\left(\sum_{t=2}^T Z_{NT,t}^2\right)^2 = V_{NT}^2 + o_P(1)$. Thus $\mathrm{Var}_{\mathcal{D}}\left(\sum_{t=2}^T Z_{NT,t}^2\right) = o_P(1)$, and the second part of (B.6) follows.

Next we show (i2). Let $\tilde{c}_i \equiv \gamma_{NT}X_i\left(\delta_i - D_{NT}^{-1}\Pi_{NT}\right)$. Then by (A.4)

$$c_i = \gamma_{NT}X_i\left(\delta_i - D_{NT}^{-1}\Pi_{NT}\right) + o_P(\gamma_{NT})X_i = \tilde{c}_i + o_P(\gamma_{NT})X_i. \tag{B.10}$$

Noting that $N^{-1/2}\sum_{i=1}^N \tilde{c}_i'M_{F^0}P_{X_i}M_{F^0}\tilde{c}_i = (NT)^{-1}\sum_{i=1}^N\left(\delta_i - D_{NT}^{-1}\Pi_{NT}\right)X_i'M_{F^0}P_{X_i}M_{F^0}X_i(\delta_i - X_iD_{NT}^{-1}\Pi_{NT}) = \Theta_{1NT}$, we have

$$A_{1NT,12} = N^{-1/2}\sum_{i=1}^N (c_i - \tilde{c}_i)'M_{F^0}P_{X_i}M_{F^0}(c_i - \tilde{c}_i)$$

$$+2N^{-1/2}\sum_{i=1}^N \tilde{c}_i'M_{F^0}P_{X_i}M_{F^0}(c_i - \tilde{c}_i)$$

$$\equiv A_{1NT,121} + 2A_{1NT,122}, \quad \text{say.}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

By (B.10), the fact that $\sum_{i=1}^{N}\|X_i\|_F^2 = O_P(NT)$, $\left\|M_{F^0}\right\| = 1$, and $\left\|P_{X_i}\right\| = 1$,

$$\begin{aligned}\left|A_{1NT,121}\right| &\le N^{-1/2}\sum_{i=1}^{N}\|c_i-\tilde{c}_i\|^2 = o_P\left(\gamma_{NT}^2\right)N^{-1/2}\sum_{i=1}^{N}\|X_i\|_F^2\\ &= o_P\left(\gamma_{NT}^2\right)O_P\left(N^{1/2}T\right) = o_P(1).\end{aligned}$$

Similarly, we can show that $A_{1NT,122} = o_P(1)$. This completes the proof of (i2).

Now we show (i3). We decompose $A_{1NT,13}$ as follows:

$$\begin{aligned}A_{1NT,13} &= \gamma_{NT}N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' M_{F^0}P_{X_i}M_{F^0}X_i\delta_i + N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' M_{F^0}P_{X_i}M_{F^0}X_i\left(\boldsymbol{\beta}^0-\hat{\boldsymbol{\beta}}\right)\\ &\equiv \gamma_{NT}A_{1NT,131} + A_{1NT,132}\left(\boldsymbol{\beta}^0-\hat{\boldsymbol{\beta}}\right), \quad \text{say.}\end{aligned}$$

In view of the fact that $||\boldsymbol{\beta}^0-\hat{\boldsymbol{\beta}}|| = O_P(\gamma_{NT})$, we can prove $A_{1NT,13} = o_P(1)$ by showing that (i3a) $\gamma_{NT}A_{1NT,131} = o_P(1)$ and (i3b) $\gamma_{NT}A_{1NT,132} = o_P(1)$. (i3a) is proved in Lemma A.10, and (i3b) can be proved analogously, say, by taking δ_i as a $K\times 1$ vector of ones. This completes the proof of (i3).

Step 2. **We prove** (ii) $A_{1NT,2} = \Theta_{2NT} + o_P(1)$ under $\mathbb{H}_1(\gamma_{NT})$. First, we decompose $A_{1NT,2}$ as follows:

$$\begin{aligned}A_{1NT,2} &= \sum_{k=1}^{K}\left(\beta_k^0-\hat{\beta}_k\right)\sum_{l=1}^{K}\left(\beta_l^0-\hat{\beta}_l\right)N^{-1/2}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_k^{(0)}P_{X_i}M_l^{(0)}F^0\lambda_i^0\\ &\quad+\sum_{k=1}^{K}\left(\beta_k^0-\hat{\beta}_k\right)\sum_{l=1}^{K}\left(\beta_l^0-\hat{\beta}_l\right)N^{-1/2}\\ &\quad\times\sum_{i=1}^{N}\left\{\varepsilon_i' M_k^{(0)}P_{X_i}M_l^{(0)}\varepsilon_i + c_i' M_k^{(0)}P_{X_i}M_l^{(0)}c_i + 2\varepsilon_i' M_k^{(0)}P_{X_i}M_l^{(0)}F^0\lambda_i^0\right.\\ &\qquad\left.+2\varepsilon_i' M_k^{(0)}P_{X_i}M_l^{(0)}c_i + 2\lambda_i' F^{0\prime}M_k^{(0)}P_{X_i}M_l^{(0)}c_i\right\}\\ &\equiv A_{1NT,21} + A_{1NT,22}, \quad \text{say.}\end{aligned}$$

We want to show that (ii1) $A_{1NT,21} = \Theta_{2NT} + o_P(1)$ and (ii2) $A_{1NT,2} = o_P(1)$. (ii1) follows because

$$\begin{aligned}A_{1NT,21} &= N^{-1/2}\sum_{k=1}^{K}\left(\beta_k^0-\hat{\beta}_k\right)\sum_{l=1}^{K}\left(\beta_l^0-\hat{\beta}_l\right)\\ &\quad\times\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}\Phi_1'\mathbf{X}_k M_{F^0}P_{X_i}M_{F^0}\mathbf{X}_l'\Phi_1F^0\lambda_i^0\\ &= N^{-1/2}\sum_{k=1}^{K}\left(\beta_k^0-\hat{\beta}_k\right)\sum_{l=1}^{K}\left(\beta_l^0-\hat{\beta}_l\right)\\ &\quad\times\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\mathbf{X}_k M_{F^0}P_{X_i}M_{F^0}\mathbf{X}_l'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_i^0\end{aligned}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$= \frac{1}{NT}\sum_{k=1}^{K} \iota'_{Kk} D_{NT}^{-1}\Pi_{NT} \sum_{l=1}^{K} \iota'_{Kl} D_{NT}^{-1}\Pi_{NT}$$
$$\times \sum_{i=1}^{N} \lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\mathbf{X}_k M_{F^0} P_{X_i} M_{F^0}\mathbf{X}'_l\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_i^0 + o_P(1)$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{k=1}^{K} \iota'_{Kk} D_{NT}^{-1}\Pi_{NT}\bar{X}'_{k,\cdot i} P_{X_i} \sum_{l=1}^{K} \iota'_{Kl} D_{NT}^{-1}\Pi_{NT}\bar{X}_{l,\cdot i} + o_P(1)$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\left(D_{NT}^{-1}\Pi_{NT}\right)'\bar{X}'_i P_{X_i}\bar{X}_i\left(D_{NT}^{-1}\Pi_{NT}\right) + o_P(1) = \Theta_{2NT} + o_P(1),$$

where ι_{Kk} is a $K \times 1$ vector with 1 in the kth place and zeros elsewhere and $\bar{X}_i \equiv N^{-1}\sum_{l=1}^{N}\alpha_{il}M_{F^0}X_l$ is a $T \times K$ matrix whose kth column is given by $\bar{X}_{i,\cdot k} \equiv \left(\lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\mathbf{X}_k M_{F^0}\right)'$.

To show (ii2), we assume that $K = 1$ for notational simplicity. We write $\mathbf{X}_k$ and $\sum_{k=1}^{K}\left(\beta_k^0 - \hat{\beta}_k\right)M_k^{(0)}$ simply as $\mathbf{X}$ and $(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})M^{(0)}$, respectively, where $M^{(0)} = -M_{F^0}\mathbf{X}'\Phi_1 - \Phi'_1\mathbf{X}M_{F^0}$. Then

$$A_{1NT,22} = \left(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}\right)^2 N^{-1/2}\sum_{i=1}^{N}\left\{\varepsilon'_i M^{(0)}P_{X_i}M^{(0)}\varepsilon_i + c'_i M^{(0)}P_{X_i}M^{(0)}c_i\right.$$
$$+2\varepsilon'_i M^{(0)}P_{X_i}M^{(0)}F^0\lambda_i^0 + 2\varepsilon'_i M^{(0)}P_{X_i}M^{(0)}c_i$$
$$\left.+2\lambda'_i F^{0\prime}M^{(0)}P_{X_i}M^{(0)}c_i\right\}$$
$$\equiv \left(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}\right)^2\left\{A_{1NT,221} + A_{1NT,222} + 2A_{1NT,223} + 2A_{1NT,224}\right.$$
$$\left.+2A_{1NT,225}\right\}, \quad \text{say.}$$

Noting that $||\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}|| = O_P(\gamma_{NT})$, it suffices to prove (ii2) by showing that $\bar{A}_{1NT,22s} \equiv \gamma_{NT}^2 A_{1NT,22s} = o_P(1)$ for $s = 1, 2, \ldots, 5$. Noting that $\left\|P_{X_i}\right\| = 1$ and

$$\left\|M^{(0)}\varepsilon_i\right\| = \left\|\left(M_{F^0}\mathbf{X}'\Phi_1 + \Phi'_1\mathbf{X}M_{F^0}\right)\varepsilon_i\right\|$$
$$= O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\| + O_P\left(N^{-1/2}T^{-1/2}\right)\left\|\mathbf{X}'\varepsilon_i\right\|, \tag{B.11}$$

and by (B.3) and (B.10), we have

$$\left|\bar{A}_{1NT,221}\right| \le \gamma_{NT}^2 N^{-1/2}\sum_{i=1}^{N}\left\|M^{(0)}\varepsilon_i\right\|^2$$
$$\le \gamma_{NT}^2\sum_{i=1}^{N}\left[O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\| + O_P\left(N^{-1/2}T^{-1/2}\right)\left\|\mathbf{X}'\varepsilon_i\right\|\right]^2$$
$$= O_P\left(N^{-1}T^{-1}\right)O_P(N) = O_P\left(T^{-1}\right) = o_P(1),$$
$$\left|\bar{A}_{1NT,222}\right| \le O_P\left(\gamma_{NT}^2\right)(NT)^{-1}\left\|M^{(0)}\right\|^2\sum_{i=1}^{N}\|X_i\|^2 = O_P\left(\gamma_{NT}^2\right)(NT)^{-1}O_P(NT)$$
$$= O_P\left(\gamma_{NT}^2\right) = o_P(1),$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$
\begin{aligned}
\left|\bar{A}_{1NT,223}\right| &\leq \gamma_{NT}^{2} N^{-1/2} \sum_{i=1}^{N} \left\| M^{(0)} \varepsilon_i \right\| \left\| M^{(0)} F^0 \lambda_i^0 \right\| \\
&\leq (NT)^{-1} \left\| M^{(0)} F^0 \right\| \sum_{i=1}^{N} \left[O_P\left(T^{-1/2}\right) \left\| F^{0\prime} \varepsilon_i \right\| \right. \\
&\quad \left. + O_P\left(N^{-1/2} T^{-1/2}\right) \left\| \mathbf{X}' \varepsilon_i \right\| \right] \left\| \lambda_i^0 \right\| \\
&= (NT)^{-1} O_P\left(T^{1/2}\right) O_P(N) = O_P\left(T^{-1/2}\right) = o_P(1), \\
\left|\bar{A}_{1NT,224}\right| &\leq \gamma_{NT}^{2} N^{-1/2} \sum_{i=1}^{N} \left\| M^{(0)} \varepsilon_i \right\| \left\| M^{(0)} c_i \right\| \\
&\leq O_P(\gamma_{NT}) (NT)^{-1} \left\| M^{(0)} \right\| \sum_{i=1}^{N} \left[O_P\left(T^{-1/2}\right) \left\| F^{0\prime} \varepsilon_i \right\| \right. \\
&\quad \left. + O_P\left(N^{-1/2} T^{-1/2}\right) \left\| \mathbf{X}' \varepsilon_i \right\| \right] \| X_i \| \\
&= \gamma_{NT} (NT)^{-1} O_P\left(NT^{1/2}\right) = O_P\left(T^{-1/2} \gamma_{NT}\right) = o_P(1),
\end{aligned}
$$

and $\left|\bar{A}_{1NT,225}\right| \leq \gamma_{NT} (NT)^{-1} \left\| M^{(0)} \right\|^2 \sum_{i=1}^{N} \left\| F^0 \lambda_i^0 \right\| \| X_i \| = \gamma_{NT} (NT)^{-1} O_P(NT) = o_P(1)$.

Step 3. We prove (iii) $A_{1NT,3} = \Theta_{3NT} + o_P(1)$ under $\mathbb{H}_1(\gamma_{NT})$. First, we decompose $A_{1NT,3}$ as follows:

$$
\begin{aligned}
A_{1NT,3} &= \sum_{k=1}^{K} \left(\beta_k^0 - \hat{\beta}_k\right) N^{-1/2} \sum_{i=1}^{N} c_i' M_{F^0} P_{X_i} M_k^{(0)} F^0 \lambda_j^0 \\
&\quad + \sum_{k=1}^{K} \left(\beta_k^0 - \hat{\beta}_k\right) N^{-1/2} \sum_{i=1}^{N} \left\{ \left(\varepsilon_i' + c_i'\right) M_{F^0} P_{X_i} M_k^{(0)} \left(\varepsilon_j + c_j\right) \right. \\
&\qquad \left. + \varepsilon_i' M_{F^0} P_{X_i} M_k^{(0)} F^0 \lambda_j^0 \right\} \\
&\equiv A_{1NT,31} + A_{1NT,32}, \quad \text{say.}
\end{aligned}
$$

We prove (iii) by showing that (iii1) $A_{1NT,31} = \Theta_{3NT} + o_P(1)$ and (iii2) $A_{1NT,32} = o_P(1)$. (iii1) follows because by (A.4)

$$
\begin{aligned}
A_{1NT,31} &= -N^{-1/2} \sum_{k=1}^{K} \left(\beta_k^0 - \hat{\beta}_k\right) \sum_{i=1}^{N} c_i' M_{F^0} P_{X_i} M_{F^0} \mathbf{X}_k' \Phi_1 F^0 \lambda_i^0 \\
&= -N^{-1/2} \sum_{k=1}^{K} \left(\beta_k^0 - \hat{\beta}_k\right) \sum_{i=1}^{N} c_i' M_{F^0} P_{X_i} M_{F^0} \mathbf{X}_k' \lambda^0 \left(\lambda^{0\prime} \lambda^0\right)^{-1} \lambda_i^0 \\
&= \frac{1}{NT} \sum_{k=1}^{K} \imath_{Kk}' D_{NT}^{-1} \Pi_{NT} \\
&\quad \times \sum_{i=1}^{N} \left(\delta_i - D_{NT}^{-1} \Pi_{NT}\right)' X_i' M_{F^0} P_{X_i} M_{F^0} \mathbf{X}_k' \lambda^0 \left(\lambda^{0\prime} \lambda^0\right)^{-1} \lambda_i^0 + o_P(1)
\end{aligned}
$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$= \frac{1}{NT}\sum_{i=1}^{N}\left(\delta_i - D_{NT}^{-1}\Pi_{NT}\right)' X_i' M_{F^0} P_{X_i} \sum_{k=1}^{K} \iota_{Kl}' D_{NT}^{-1}\Pi_{NT}\bar{X}_{k,\cdot i} + o_P(1)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\left(\delta_i - D_{NT}^{-1}\Pi_{NT}\right)' X_i' M_{F^0} P_{X_i} \bar{X}_i \left(D_{NT}^{-1}\Pi_{NT}\right) + o_P(1)$$

$$= \Theta_{3NT} + o_P(1).$$

To show (iii2), again we assume that $K = 1$ for notational simplicity. As before, we write $\mathbf{X}_k$ and $\sum_{k=1}^{K}(\beta_k^0 - \hat{\beta}_k)M_k^{(0)}$ simply as $\mathbf{X}$ and $(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})M^{(0)}$, respectively. Then

$$A_{1NT,32} = (\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})N^{-1/2}\sum_{i=1}^{N}\Big\{\varepsilon_i' M_{F^0} P_{X_i} M^{(0)}\varepsilon_i + \varepsilon_i' M_{F^0} P_{X_i} M^{(0)} c_i + c_i' M_{F^0} P_{X_i} M^{(0)}\varepsilon_i + c_i' M_{F^0} P_{X_i} M^{(0)} c_i + \varepsilon_i' M_{F^0} P_{X_i} M^{(0)} F^0 \lambda_i^0\Big\}$$

$$\equiv (\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})\left(A_{1NT,321} + A_{1NT,322} + A_{1NT,323} + A_{1NT,324} + A_{1NT,325}\right), \quad \text{say.}$$

We prove (iii2) by showing that $\bar{A}_{1NT,32s} \equiv \gamma_{NT} A_{1NT,32s} = o_P(1)$ for $s = 1, 2, \ldots, 5$.

For $\bar{A}_{1NT,321}$, note that $\bar{A}_{1NT,321} = \gamma_{NT} N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' P_{X_i} M^{(0)}\varepsilon_i - \gamma_{NT} N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' P_{F^0} P_{X_i} M^{(0)}\varepsilon_i \equiv \bar{A}_{1NT,321a} - \bar{A}_{1NT,321b}$, say.

$$\left|\bar{A}_{1NT,321a}\right| = \gamma_{NT}\left|\operatorname{tr}\left(M^{(0)} N^{-1/2}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i' P_{X_i}\right)\right|$$

$$\leq \gamma_{NT}\left\|M^{(0)}\right\|_F \left|\operatorname{tr}\left(N^{-1/2}\sum_{i=1}^{N} X_i'\varepsilon_i\varepsilon_i' X_i\left(X_i' X_i\right)^{-1}\right)\right|$$

$$\leq \gamma_{NT}\left\|M^{(0)}\right\|_F \left[\min_{1\leq i\leq N}\mu_{\min}\left(X_i' X_i/T\right)\right]^{-1} \operatorname{tr}\left(T^{-1}N^{-1/2}\sum_{i=1}^{N} X_i'\varepsilon_i\varepsilon_i' X_i\right)$$

$$= \gamma_{NT} O_P\left(N^{1/2}\right) = O_P\left(N^{1/4}T^{-1/2}\right) = o_P(1),$$

as we can readily show that $\operatorname{tr}(T^{-1}N^{-1/2}\sum_{i=1}^{N} X_i'\varepsilon_i\varepsilon_i' X_i) = O_P\left(N^{1/2}\right)$ by Markov's inequality. Similarly, $\bar{A}_{1NT,321b} = O_P\left(N^{1/4}T^{-1/2}\right) = o_P(1)$. It follows that $\bar{A}_{1NT,321} = o_P(1)$. For $\bar{A}_{1NT,322}$, we have

$$\bar{A}_{1NT,322} = \gamma_{NT} N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' P_{X_i} M^{(0)} c_i - \gamma_{NT} N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' P_{F^0} P_{X_i} M^{(0)} c_i$$

$$\equiv \bar{A}_{1NT,322a} - \bar{A}_{1NT,322b}.$$

Using (B.3), (B.10), and the fact that $\left\|P_{X_i}\varepsilon_i\right\|_F = O_P\left(T^{-1/2}\right)\left\|X_i'\varepsilon_i\right\|_F$,

$$\left|\bar{A}_{1NT,322a}\right| = \gamma_{NT}\left|\operatorname{tr}\left(M^{(0)} N^{-1/2}\sum_{i=1}^{N} c_i\varepsilon_i' P_{X_i}\right)\right|$$

$$\leq \gamma_{NT}\left\|M^{(0)}\right\|_F \left\|N^{-1/2}\sum_{i=1}^{N} c_i\varepsilon_i' P_{X_i}\right\|_F$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$\leq T^{-1/2} O_P\left(\gamma_{NT}^2\right) \left\|M^{(0)}\right\|_F N^{-1/2} \sum_{i=1}^N \|X_i\|_F \left\|X_i'\varepsilon_i\right\|_F$$
$$= T^{-1/2} O_P\left(\gamma_{NT}^2\right) O_P\left(N^{1/2}T\right) = O_P\left(T^{-1/2}\right) = o_P(1).$$

Similarly, using $\left\|P_{F^0}\varepsilon_i\right\|_F = O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\|_F$ we have $\left|\bar{A}_{1NT,322b}\right| \leq O_P\left(\gamma_{NT}T^{-1/2}\right)\gamma_{NT}N^{-1/2}\sum_{i=1}^N \left\|F^{0\prime}\varepsilon_i\right\|_F \|X_i\|_F = O_P\left(T^{-1/2}\right) = o_P(1)$. Thus $\bar{A}_{1NT,322} = o_P(1)$. By (B.10), (B.11), and (B.3),

$$\left|\bar{A}_{1NT,323}\right| \leq O_P\left(\gamma_{NT}^2\right) N^{-1/2} \sum_{i=1}^N \|X_i\| \times \left[O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\| + O_P\left(N^{-1/2}T^{-1/2}\right)\left\|\mathbf{X}'\varepsilon_i\right\|\right]$$
$$= \gamma_{NT}^2 O_P\left(N^{1/2}T^{1/2}\right) = O_P\left(T^{-1/2}\right) = o_P(1),$$

and $\left|\bar{A}_{1NT,324}\right| \leq O_P\left(\gamma_{NT}^3\right)\left\|M^{(0)}\right\| N^{-1/2}\sum_{i=1}^N \|X_i\|^2 = \gamma_{NT}^3 O_P\left(N^{1/2}T\right) = O_P\left(N^{-1/4}T^{-1/2}\right) = o_P(1)$.

For $\bar{A}_{1NT,325}$, note that $\bar{A}_{1NT,325} = \gamma_{NT}N^{-1/2}\sum_{i=1}^N \varepsilon_i' P_{X_i} M^{(0)} F^0 \lambda_i^0 - \gamma_{NT}N^{-1/2} \sum_{i=1}^N \varepsilon_i' P_{F^0} P_{X_i} M^{(0)} F^0 \lambda_i^0 \equiv \bar{A}_{1NT,325a} - \bar{A}_{1NT,325b}$, say. Noting that $\left\|M^{(0)}F^0\right\|_F = O_P(1)$ and $\left\|P_{X_i}\varepsilon_i\right\|_F = O_P\left(T^{-1/2}\right)\left\|X_i'\varepsilon_i\right\|_F$,

$$\left|\bar{A}_{1NT,325a}\right| = \gamma_{NT}\left|\operatorname{tr}\left(M^{(0)}F^0 N^{-1/2}\sum_{i=1}^N \lambda_i^0 \varepsilon_i' P_{X_i}\right)\right|$$
$$\leq \gamma_{NT}\left\|M^{(0)}F^0\right\|_F \left\|N^{-1/2}\sum_{i=1}^N \lambda_i^0 \varepsilon_i' P_{X_i}\right\|_F$$
$$\leq O_P\left(T^{-1/2}\gamma_{NT}\right) N^{-1/2}\sum_{i=1}^N \lambda_i^0 \left\|X_i'\varepsilon_i\right\|_F$$
$$= O_P\left(T^{-1/2}\gamma_{NT}\right) O_P\left(N^{1/2}T^{1/2}\right) = o_P(1).$$

Similarly, $\bar{A}_{1NT,325b} = o_P(1)$. Thus $\bar{A}_{1NT,325} = o_P(1)$. This completes the proof. ■

PROPOSITION B.2. *$A_{2NT} = o_P(1)$ under $\mathbb{H}_1(\gamma_{NT})$.*

Proof. Recall from the proof of Theorem 3.3 that $M^{(1)} = -M_{F^0}\varepsilon'\Phi_1 - \Phi_1'\varepsilon M_{F^0}$ and from the discussion following Definition A.1 $\Phi_1 = \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}$. Noting that $\Phi_1 M_{F^0} = 0$ and $\mu_1\left(M_{F^0}\right) = 1$, we have

$$\left\|M^{(1)}\varepsilon_i\right\|_F^2 = \operatorname{tr}\left[\varepsilon_i'\left(M_{F^0}\varepsilon'\Phi_1 + \Phi_1'\varepsilon M_{F^0}\right)\left(M_{F^0}\varepsilon'\Phi_1 + \Phi_1'\varepsilon M_{F^0}\right)\varepsilon_i\right]$$
$$= 2\operatorname{tr}\left(\varepsilon_i'\Phi_1'\varepsilon M_{F^0}\varepsilon'\Phi_1\varepsilon_i\right) \leq 2\operatorname{tr}\left(\varepsilon_i'\Phi_1'\varepsilon\varepsilon'\Phi_1\varepsilon_i\right)$$
$$= 2\operatorname{tr}\left[\varepsilon_i' F^0\left(F^{0\prime}F^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\varepsilon\varepsilon'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\varepsilon_i\right]$$
$$= 2\operatorname{tr}\left[\left(F^{0\prime}F^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\varepsilon\varepsilon'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\varepsilon_i\varepsilon_i'F^0\right]$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$
\begin{aligned}
&\leq 2\mathrm{tr}\left[\left(F^{0\prime}F^{0}\right)^{-1}\left(\lambda^{0\prime}\lambda^{0}\right)^{-1}\lambda^{0\prime}\varepsilon\varepsilon'\lambda^{0}\left(\lambda^{0\prime}\lambda^{0}\right)^{-1}\left(F^{0\prime}F^{0}\right)^{-1}\right]\\
&\quad\times\mathrm{tr}\left(F^{0\prime}\varepsilon_i\varepsilon_i'F^{0}\right)\\
&\leq 2\mathrm{tr}\left[\left(\lambda^{0\prime}\lambda^{0}\right)^{-1}\left(F^{0\prime}F^{0}\right)^{-1}\left(F^{0\prime}F^{0}\right)^{-1}\left(\lambda^{0\prime}\lambda^{0}\right)^{-1}\right]\mathrm{tr}\left(\lambda^{0\prime}\varepsilon\varepsilon'\lambda^{0}\right)\\
&\quad\times\mathrm{tr}\left(F^{0\prime}\varepsilon_i\varepsilon_i'F^{0}\right)\\
&=O_P\left((NT)^{-2}\right)O_P\left(NT\right)\mathrm{tr}\left(F^{0\prime}\varepsilon_i\varepsilon_i'F^{0}\right)=O_P\left((NT)^{-1}\right)\left\|F^{0\prime}\varepsilon_i\right\|_F^2,
\end{aligned}
$$

where we have repeatedly used the rotational property of the trace operator, the fact that

$$\mathrm{tr}\left(AB\right)\leq\mu_1\left(A\right)\mathrm{tr}\left(B\right) \tag{B.12}$$

for any symmetric matrix A and positive semidefinite (p.s.d.) matrix B (see, e.g., Bernstein, 2005, Prop. 8.4.13), and the fact that

$$\mathrm{tr}\left(AB\right)\leq\mathrm{tr}\left(A\right)\mathrm{tr}\left(B\right) \tag{B.13}$$

for any two p.s.d. matrices A and B (see, e.g., Bernstein, 2005, Fact 8.10.7). It follows that

$$\left\|M^{(1)}\varepsilon_i\right\|_F=O_P\left((NT)^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\|_F. \tag{B.14}$$

By the fact that $\left\|M^{(1)}\right\|_F=O_P\left(N^{-1/2}\right)$ and (B.5),

$$\left\|M^{(1)}c_i\right\|_F\leq\left\|M^{(1)}\right\|_F\|c_i\|_F=O_P\left(N^{-1/2}\gamma_{NT}\right)\|X_i\|_F. \tag{B.15}$$

We will use these results frequently.

Now, by Cauchy–Schwarz inequality

$$
\begin{aligned}
A_{2NT}&=N^{-1/2}\sum_{i=1}^{N}\left(\varepsilon_i'+\lambda_i^{0\prime}F^{0\prime}+c_i'\right)M^{(1)}P_{X_i}M^{(1)}\left(\varepsilon_i+F^{0}\lambda_i^{0}+c_i\right)\\
&\leq 3N^{-1/2}\sum_{i=1}^{N}\left\{\varepsilon_i'M^{(1)}P_{X_i}M^{(1)}\varepsilon_i+\lambda_i^{0\prime}F^{0\prime}M^{(1)}P_{X_i}M^{(1)}F^{0}\lambda_i^{0}\right.\\
&\qquad\left.+c_i'M^{(1)}P_{X_i}M^{(1)}c_i\right\}\\
&\equiv 3A_{2NT,1}+3A_{2NT,2}+3A_{2NT,3},\quad\text{say.}
\end{aligned}
$$

We prove the proposition by demonstrating that $A_{2NT,s}=o_P\left(1\right)$ for $s=1,2,3$. By (B.14)–(B.15), (B.3), and the fact that $\mu_1\left(P_{X_i}\right)=1$, we have

$$
\begin{aligned}
\left|A_{2NT,1}\right|&\leq N^{-1/2}\sum_{i=1}^{N}\left\|M^{(1)}\varepsilon_i\right\|_F^2=N^{-1/2}O_P\left((NT)^{-1}\right)\sum_{i=1}^{N}\left\|F^{0\prime}\varepsilon_i\right\|_F^2\\
&=O_P\left(N^{-3/2}T^{-1}\right)O_P\left(NT\right)=O_P\left(N^{-1/2}\right)=o_P\left(1\right)\quad\text{and}\\
\left|A_{2NT,3}\right|&\leq N^{-1/2}\sum_{i=1}^{N}\left\|M^{(1)}c_i\right\|^2=N^{-1/2}O_P\left(N^{-1}\gamma_{NT}^2\right)\sum_{i=1}^{N}\|X_i\|^2\\
&=O_P(N^{-2}T^{-1})O_P\left(NT\right)=O_P\left(N^{-1}\right)=o_P\left(1\right).
\end{aligned}
$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Using $M^{(1)} = -M_{F^0}\boldsymbol{\varepsilon}'\Phi_1 - \Phi_1'\boldsymbol{\varepsilon}M_{F^0}$, $\Phi_1 \equiv \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}$, and $M_{F^0} = I_T - P_{F^0}$, we have

$$\begin{aligned}
\left|A_{2NT,2}\right| &= N^{-1/2}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}M_{F^0}P_{X_i}M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_i^0 \\
&= N^{-1/2}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}P_{X_i}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_i^0 \\
&\quad + N^{-1/2}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}P_{F^0}P_{X_i}P_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_i^0 \\
&\quad - 2N^{-1/2}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}P_{X_i}M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_i^0 \\
&\equiv A_{2NT,21} + A_{2NT,22} - 2A_{2NT,23}, \quad \text{say.}
\end{aligned}$$

Using $\mathbb{E}_{\mathcal{D}}\left\|P_{X_i}\boldsymbol{\varepsilon}'\lambda^0\right\|_F^2 = O_P(N)$ and $\mu_1\left(P_{X_i}\right) = 1$, we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}}\left|A_{2NT,21}\right| &\le T^{-2}N^{-1/2}\left[\mu_{\min}\left(\lambda^{0\prime}\lambda^0/T\right)\right]^{-2}\sum_{i=1}^{N}\left\|P_{X_i}\boldsymbol{\varepsilon}'\lambda^0\right\|^2\left\|\lambda_i^0\right\|^2 \\
&= T^{-2}N^{-1/2}O_P\left(N^2\right) = O_P\left(T^{-2}N^{3/2}\right) = o_P(1),
\end{aligned}$$

and similarly $\mathbb{E}_{\mathcal{D}}\left|A_{2NT,22}\right| \le T^{-2}N^{-1/2}\left[\mu_{\min}\left(\lambda^{0\prime}\lambda^0/T\right)\right]^{-2}\sum_{i=1}^{N}\left\|P_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|^2\left\|\lambda_i^0\right\|^2 = o_P(1)$. By Cauchy–Schwarz inequality, $A_{2NT,23} \le \{A_{2NT,21}\}^{1/2}\{A_{2NT,22}\}^{1/2} = o_P(1)$. It follows that $A_{2NT,2} = o_P(1)$. This completes the proof. ■

PROPOSITION B.3. $A_{3NT} = o_P(1)$ *under* $\mathbb{H}_1(\gamma_{NT})$.

Proof. Noting that $\left\|M^{(2)} + M^{(rem)}\right\| = O_P\left(\delta_{NT}^{-2}\right)$ by (B.3) and (B.2) and $\mu_1\left(P_{X_i}\right) = 1$ we have $|A_{3NT}| \le N^{-1/2}\left\|M^{(2)} + M^{(rem)}\right\|_F^2\sum_{i=1}^{N}\left\|\varepsilon_i + F^0\lambda_i^0 + c_i\right\|_F^2 = N^{-1/2}O_P\left(\delta_{NT}^{-4}\right)O_P(NT) = N^{-1/2}\ O_P\left(N^{-1}T + NT^{-1}\right) = o_P(1)$ by Assumption A.3. ■

PROPOSITION B.4. $A_{4NT} = o_P(1)$ *under* $\mathbb{H}_1(\gamma_{NT})$.

Proof. We decompose A_{4NT} as follows:

$$\begin{aligned}
A_{4NT} &= N^{-1/2}\sum_{i=1}^{N}\left(\varepsilon_i' + c_i'\right)M_{F^0}P_{X_i}M^{(1)}\left(\varepsilon_i + F^0\lambda_i^0 + c_i\right) \\
&\quad + N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\beta_k^0 - \hat{\beta}_k\right)\left(\varepsilon_i' + \lambda_i^{0\prime}F^{0\prime} + c_i'\right)M_k^{(0)}P_{X_i}M^{(1)}\left(\varepsilon_i + F^0\lambda_i^0 + c_i\right) \\
&\equiv A_{4NT,1} + A_{4NT,2}, \quad \text{say.}
\end{aligned}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

First, we study $A_{4NT,1}$ by further decomposing it as follows.

$$\begin{aligned}A_{4NT,1} = N^{-1/2}\sum_{i=1}^{N}\Big\{&\varepsilon_i' M_{F^0} P_{X_i} M^{(1)}\varepsilon_i + \varepsilon_i' M_{F^0} P_{X_i} M^{(1)} F^0\lambda_i^0 + \varepsilon_i' M_{F^0} P_{X_i} M^{(1)} c_i\\ &+ c_i' M_{F^0} P_{X_i} M^{(1)}\varepsilon_i + c_i' M_{F^0} P_{X_i} M^{(1)} F^0\lambda_i^0\\ &+ c_i' M_{F^0} P_{X_i} M^{(1)} c_i\Big\}\\ \equiv\ & A_{4NT,11} + A_{4NT,12} + A_{4NT,13} + A_{4NT,14} + A_{4NT,15} + A_{4NT,16}.\end{aligned}$$

We prove that $A_{4NT,1} = o_P(1)$ by showing that $A_{4NT,1s} = o_P(1)$ for $s = 1, 2, \ldots, 6$. By the triangle inequality, $\|P_{X_i}\| = 1$, $\|M_{F^0}\| = 1$, (B.14), (B.15), and (B.5)

$$\begin{aligned}\left|A_{4NT,14}\right| &\le N^{-1/2}\sum_{i=1}^{N}\|c_i\|\left\|M^{(1)}\varepsilon_i\right\|\\ &\le O_P\left(\gamma_{NT}(NT)^{-1/2}\right)N^{-1/2}\sum_{i=1}^{N}\|X_i\|_F\left\|F^{0\prime}\varepsilon_i\right\|_F\\ &= O_P\left(\gamma_{NT}(NT)^{-1/2}\right)O_P\left(N^{1/2}T\right) = O_P(N^{-1/4}) = o_P(1), \quad \text{and}\end{aligned}$$

$$\begin{aligned}\left|A_{4NT,16}\right| &\le N^{-1/2}\sum_{i=1}^{N}\|c_i\|\left\|M^{(1)}c_i\right\|_F \le O_P\left(N^{-1/2}\gamma_{NT}^2\right)N^{-1/2}\sum_{i=1}^{N}\|X_i\|_F^2\\ &\le O_P\left(N^{-1/2}\gamma_{NT}^2\right)O_P\left(N^{1/2}T\right) = O_P\left(N^{-1/2}\right) = o_P(1).\end{aligned}$$

Next, we show that $A_{4NT,11} = o_P(1)$. Using $M_{F^0} = I_T - P_{F^0}$ and $M^{(1)} = -(M_{F^0}\boldsymbol{\varepsilon}'\Phi_1 + \Phi_1'\boldsymbol{\varepsilon}M_{F^0})$, we first decompose $A_{4NT,11}$ as follows:

$$\begin{aligned}A_{4NT,11} = N^{-1/2}\sum_{i=1}^{N}\Big\{&-\varepsilon_i' P_{X_i} M_{F^0}\boldsymbol{\varepsilon}'\Phi_1\varepsilon_i - \varepsilon_i' P_{X_i}\Phi_1'\boldsymbol{\varepsilon}M_{F^0}\varepsilon_i + \varepsilon_i' P_{F^0} P_{X_i} M_{F^0}\boldsymbol{\varepsilon}'\Phi_1\varepsilon_i\\ &+ \varepsilon_i' P_{F^0} P_{X_i}\Phi_1'\boldsymbol{\varepsilon}M_{F^0}\varepsilon_i\Big\}\\ \equiv\ & -A_{4NT,111} - A_{4NT,112} + A_{4NT,113} + A_{4NT,114}.\end{aligned}$$

Noting that $T^{-2}N^{-3/2}\sum_{i=1}^{N}\left\|\varepsilon_i' X_i\right\|_F^2\left\|F^{0\prime}\varepsilon_i\right\|_F^2 = O_P\left(N^{-1/2}\right)$ and $T^{-2}N^{-3/2}\sum_{i=1}^{N}\left\|X_i' M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F^2 = O_P\left(T^{-1}N^{1/2}\right)$ by Markov's inequality and $\|M_{F^0}\| = 1$, we have

$$\begin{aligned}\left|A_{4NT,111}\right| = \Bigg|&T^{-2}N^{-3/2}\sum_{i=1}^{N}\varepsilon_i' X_i\left(X_i' X_i/T\right)^{-1}X_i' M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\\ &\times\left(F^{0\prime}F^0/T\right)^{-1}F^{0\prime}\varepsilon_i\Bigg|\\ \le\ & c_{1NT}T^{-2}N^{-3/2}\sum_{i=1}^{N}\left\|\varepsilon_i' X_i\right\|_F\left\|F^{0\prime}\varepsilon_i\right\|_F\left\|X_i' M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F\end{aligned}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$\leq c_{1NT}\left\{T^{-2}N^{-3/2}\sum_{i=1}^{N}\left\|\varepsilon_i'X_i\right\|_F^2\left\|F^{0\prime}\varepsilon_i\right\|_F^2\right\}^{1/2}$$
$$\times\left\{T^{-2}N^{-3/2}\sum_{i=1}^{N}\left\|X_i'M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F^2\right\}^{1/2}$$
$$=O_P(1)\,O_P\left(N^{-1/4}\right)O_P\left(T^{-1/2}N^{1/4}\right)=O_P\left(T^{-1/2}\right)=o_P(1),$$

where $c_{1NT}\equiv\max_{1\leq i\leq N}\left\|\left(X_i'X_i/T\right)^{-1}\right\|_F\left\|\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\left(F^{0\prime}F^0/T\right)^{-1}\right\|_F=O_P(1)$. Similarly, we can show that $A_{4NT,11s}=o_P(1)$ for $s=2,3,4$. It follows that $A_{4NT,11}=o_P(1)$.

For $A_{4NT,12}$, we write $A_{4NT,12}=-N^{-1/2}\sum_{i=1}^{N}\varepsilon_i'P_{X_i}M_{F^0}\boldsymbol{\varepsilon}'\Phi_1F^0\lambda_i^0-N^{-1/2}\sum_{i=1}^{N}\varepsilon_i'P_{F^0}P_{X_i}M_{F^0}\boldsymbol{\varepsilon}'\Phi_1F^0\lambda_i^0\equiv-A_{4NT,121}-A_{4NT,122}$. We further decompose $A_{4NT,121}$ as follows:

$$A_{4NT,121}=T^{-1}N^{-3/2}\sum_{i=1}^{N}\varepsilon_i'X_i\left(X_i'X_i/T\right)^{-1}X_i'M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\lambda_i^0$$
$$=T^{-1}N^{-3/2}\sum_{i=1}^{N}\varepsilon_i'X_i\Omega_i^{-1}X_i'M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\lambda_i^0$$
$$+T^{-1}N^{-3/2}\sum_{i=1}^{N}\varepsilon_i'X_i\left[\left(X_i'X_i/T\right)^{-1}-\Omega_i^{-1}\right]X_i'M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\lambda_i^0$$
$$\equiv A_{4NT,121a}+A_{4NT,121b}.$$

By straightforward moment calculations $\mathbb{E}_{\mathcal{D}}\left(A_{4NT,121a}^2\right)=O_P\left(N^{-1}\right)$, implying that $A_{4NT,121a}=O_P\left(N^{-1/2}\right)=o_P(1)$. As in the study of $A_{4NT,111}$,

$$A_{4NT,121b}\leq c_{2NT}T^{-1}N^{-3/2}\sum_{i=1}^{N}\left\|\varepsilon_i'X_i\right\|_F\left\|\lambda_i^0\right\|_F\left\|X_i'M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F$$
$$\leq c_{2NT}\left\{T^{-1}N^{-3/2}\sum_{i=1}^{N}\left\|\varepsilon_i'X_i\right\|_F^2\left\|\lambda_i^0\right\|_F^2\right\}^{1/2}$$
$$\times\left\{T^{-1}N^{-3/2}\sum_{i=1}^{N}\left\|X_i'M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F^2\right\}^{1/2}$$
$$=O_P(a_{NT})\,O_P\left(N^{-1/4}\right)O_P\left(N^{1/4}\right)=O_P(a_{NT})=o_P(1),$$

where $c_{2NT}\equiv\max_{1\leq i\leq N}\left\|\left(X_i'X_i/T\right)^{-1}-\Omega_i^{-1}\right\|_F\left\|\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\right\|_F=O_P(a_{NT})$. It follows that $A_{4NT,121}=o_P(1)$. Analogously, we can show that $A_{4NT,122}=o_P(1)$.

For $A_{4NT,13}$, we write $A_{4NT,13}=N^{-1/2}\sum_{i=1}^{N}\varepsilon_i'P_{X_i}M^{(1)}c_i-N^{-1/2}\sum_{i=1}^{N}\varepsilon_i'P_{F^0}P_{X_i}M^{(1)}c_i\equiv A_{4NT,131}-A_{4NT,132}$.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$A_{4NT,131} = T^{-1}N^{-1/2}\sum_{i=1}^{N}\varepsilon_i' X_i \left(X_i' X_i/T\right)^{-1} X_i' M^{(1)} c_i$$

$$\leq c_{3NT}\left\{T^{-1}N^{-1/2}\sum_{i=1}^{N}\left\|\varepsilon_i' X_i\right\|_F \left\|X_i'\right\|_F \|X_i\|_F\right\}$$

$$= O_P\left(N^{-1/2}\gamma_{NT}\right) O_P\left(T^{1/2}N^{1/2}\right) = O_P\left(N^{-1/4}\right) = o_P(1),$$

where $c_{3NT} \equiv \max_{1\leq i\leq N} \|\left(X_i' X_i/T\right)^{-1}\|_F \max_{1\leq i\leq N} \|\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}} + \gamma_{NT}\delta_i\|_F \left\|M^{(1)}\right\|_F = O_P\left(N^{-1/2}\gamma_{NT}\right)$. By the same token $A_{4NT,132} = o_P(1)$. Thus $A_{4NT,13} = o_P(1)$.

Next, noting that $\left\|P_{X_i} M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F^2 = O(N)$ uniformly in i, we have

$$\left|A_{4NT,15}\right| = \left|N^{-3/2}\sum_{i=1}^{N} c_i' M_{F^0} P_{X_i} M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\lambda_i^0\right|$$

$$\leq O_P(\gamma_{NT})N^{-3/2}\sum_{i=1}^{N}\|X_i\|_F\left\|\lambda_i^0\right\|_F\left\|P_{X_i} M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F$$

$$\leq O_P(\gamma_{NT})\left\{N^{-3/2}\sum_{i=1}^{N}\|X_i\|_F^2\left\|\lambda_i^0\right\|_F^2\right\}^{1/2}$$

$$\times\left\{N^{-3/2}\sum_{i=1}^{N}\left\|P_{X_i} M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F^2\right\}^{1/2}$$

$$= O_P(\gamma_{NT}) O_P\left(N^{-1/4}T^{1/2}\right) O_P\left(N^{1/4}\right) = O_P\left(N^{-1/4}\right) = o_P(1).$$

In sum, we have shown that $A_{4NT,1} = o_P(1)$ under $\mathbb{H}_1(\gamma_{NT})$.

Now, we study $A_{4NT,2}$. Observe that

$$A_{4NT,2} = \sum_{k=1}^{K}\left(\beta_k^0 - \hat{\beta}_k\right)N^{-1/2}$$

$$\times\sum_{i=1}^{N}\left\{\left(\varepsilon_i' + c_i'\right)M^{(1)} P_{X_i} M_k^{(0)}\left(\varepsilon_i + F^0\lambda_i^0 + c_i\right) + \lambda_i^{0\prime} F^{0\prime} M^{(1)} P_{X_i} M_k^{(0)}\varepsilon_i\right.$$

$$\left. + \lambda_i^{0\prime} F^{0\prime} M^{(1)} P_{X_i} M_k^{(0)} F^0\lambda_i^0 + \lambda_i^{0\prime} F^{0\prime} M^{(1)} P_{X_i} M_k^{(0)} c_i\right\}$$

$$\equiv \sum_{k=1}^{K}\left(\beta_k^0 - \hat{\beta}_k\right)\left[A_{4NT,21}(k) + A_{4NT,22}(k) + A_{4NT,23}(k)\right.$$

$$\left. + A_{4NT,24}(k)\right], \quad \text{say.}$$

We prove $A_{4NT,2} = o_P(1)$ by showing that $\bar{A}_{4NT,2s}(k) \equiv \gamma_{NT} A_{4NT,2s}(k) = o_P(1)$ for $s = 1, 2, 3, 4$ and $k = 1, \ldots, K$. By (B.14), (B.15), and (B.10), we can readily show that

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$\begin{aligned}
\left|\bar{A}_{4NT,21}(k)\right| &\le K^{1/2}\gamma_{NT}\left\|M_k^{(0)}\right\| N^{-1/2}\sum_{i=1}^{N}\left\|\varepsilon_i+F^0\lambda_i^0+c_i\right\|\left\|M^{(1)}(\varepsilon_i+c_i)\right\| \\
&= O_P(\gamma_{NT})N^{-1/2}\sum_{i=1}^{N}\left\|\varepsilon_i+F^0\lambda_i^0+c_i\right\| \\
&\quad\times\left\{O_P((NT)^{-1/2})\left\|F^{0\prime}\varepsilon_i\right\|+O_P(N^{-1/2}\gamma_{NT})\|X_i\|\right\} \\
&= O_P(\gamma_{NT})O_P\left(T^{1/2}\right)=O_P\left(N^{-1/4}\right)=o_P(1), \quad \text{and} \\
\left|\bar{A}_{4NT,24}(k)\right| &\le K^{1/2}\gamma_{NT}\left\|M^{(1)}F^0\right\| N^{-1/2}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|\|c_i\| = O_P\left(\gamma_{NT}^2N^{-1/2}T^{1/2}\right) \\
&\quad\times N^{-1/2}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|\|X_i\| \\
&= O_P\left(N^{-1}T^{-1/2}\right)O_P\left(N^{1/2}T^{1/2}\right)=O_P\left(N^{-1/2}\right)=o_P(1),
\end{aligned}$$

For $A_{4NT,22}(k)$ we have

$$\begin{aligned}
\left|A_{4NT,22}(k)\right| &= \left|\gamma_{NT}N^{-1/2}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}M_{F^0}P_{X_i}M_{F^0}\mathbf{X}_k'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\right. \\
&\quad\left.\times\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\varepsilon_i\right| \\
&\le K^{1/2}c_{4NT}\gamma_{NT}T^{-1}N^{-5/2}\left\|\mathbf{X}_k'\lambda^0\right\| \\
&\quad\times\left\{\sum_{i=1}^{N}\left\|\lambda_i^0\right\|_F^2\left\|F^{0\prime}\varepsilon_i\right\|_F\left\|P_{X_i}M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F\right\} \\
&\le K^{1/2}c_{4NT}\gamma_{NT}T^{-1}N^{-5/2}\left\|\mathbf{X}_k'\lambda^0\right\|\left\{\sum_{i=1}^{N}\left\|\lambda_i^0\right\|_F^4\left\|F^{0\prime}\varepsilon_i\right\|_F^2\right\}^{1/2} \\
&\quad\times\left\{\sum_{i=1}^{N}\left\|P_{X_i}M_{F^0}\boldsymbol{\varepsilon}'\lambda^0\right\|_F^2\right\}^{1/2} \\
&= \gamma_{NT}T^{-1}N^{-5/2}O_P\left(N^{1/2}T^{1/2}\right)O_P\left(N^{1/2}T^{1/2}\right)O_P(N) \\
&= O_P\left(T^{-1/2}N^{-3/4}\right)=o_P(1),
\end{aligned}$$

where $c_{4NT}\equiv\left\|\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\right\|_F^2\left\|\left(F^{0\prime}F^0/T\right)^{-1}\right\|_F=O_P(1)$. Similarly,

$$\begin{aligned}
\left|A_{4NT,23}(k)\right| &= \left|\gamma_{NT}N^{-5/2}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}M_{F^0}P_{X_i}M_{F^0}\mathbf{X}_k'\lambda^0\right. \\
&\quad\left.\times\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\lambda_i^0\right|
\end{aligned}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$
\begin{aligned}
&\leq K^{1/2} c_{5NT} \gamma_{NT} \left\| \mathbf{X}_k' \lambda^0 \right\| \left\{ N^{-5/2} \sum_{i=1}^{N} \left\| \lambda_i^0 \right\|_F^2 \left\| P_{X_i} M_{F^0} \boldsymbol{\varepsilon}' \lambda^0 \right\|_F \right\} \\
&\leq K^{1/2} c_{5NT} \gamma_{NT} \left\| \mathbf{X}_k' \lambda^0 \right\| \left\{ N^{-5/2} \sum_{i=1}^{N} \left\| \lambda_i^0 \right\|_F^4 \right\}^{1/2} \\
&\quad \times \left\{ N^{-5/2} \sum_{i=1}^{N} \left\| P_{X_i} M_{F^0} \boldsymbol{\varepsilon}' \lambda^0 \right\|_F^2 \right\}^{1/2} \\
&= \gamma_{NT} O_P \left(N^{1/2} T^{1/2} \right) O_P \left(N^{-3/4} \right) O_P \left(N^{-1/4} \right) \\
&= O_P \left(N^{-3/4} \right) = o_P (1),
\end{aligned}
$$

where $c_{5NT} \equiv \left\| \left(\lambda^{0\prime} \lambda^0 / N \right)^{-1} \right\|_F^2 = O_P (1)$. This completes the proof. ■

PROPOSITION B.5. *$A_{5NT} = o_P (1)$ under $\mathbb{H}_1 (\gamma_{NT})$.*

Proof. We decompose A_{5NT} as follows:

$$
\begin{aligned}
A_{5NT} &= N^{-1/2} \sum_{i=1}^{N} \left(\varepsilon_i' + c_i' \right) M_{F^0} P_{X_i} \left(M^{(2)} + M^{(rem)} \right) \left(\varepsilon_i + F^0 \lambda_i^0 + c_i \right) \\
&\quad + N^{-1/2} \sum_{i=1}^{N} \sum_{k=1}^{K} \left(\beta_k^0 - \hat{\beta}_k \right) \left(\varepsilon_i' + \lambda_i^{0\prime} F^{0\prime} + c_i' \right) M_k^{(0)} P_{X_i} \left(M^{(2)} + M^{(rem)} \right) \\
&\quad \times \left(\varepsilon_i + F^0 \lambda_i^0 + c_i \right) \\
&\equiv A_{5NT,1} + A_{5NT,2}, \quad \text{say.}
\end{aligned}
$$

We dispense with the term that is easy to analyze first. By the triangle inequality, the fact that $||\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}|| = O_P (\gamma_{NT})$, and $\left\| P_{X_i} \right\| = 1$, the submultiplicative property of $\|\cdot\|$, (B.2), and (B.3), we have

$$
\begin{aligned}
\left| A_{5NT,2} \right| &\leq \sum_{k=1}^{K} \left| \beta_k^0 - \hat{\beta}_k \right| \left\| M_k^{(0)} \right\| \left\| M^{(2)} + M^{(rem)} \right\| N^{-1/2} \sum_{i=1}^{N} \left\| \varepsilon_i + F^0 \lambda_i^0 + c_i \right\|^2 \\
&= O_P (\gamma_{NT}) O_P (1) O_P \left(\delta_{NT}^{-2} \right) O_P \left(N^{1/2} T \right) \\
&= O_P \left(N^{1/4} T^{-1/2} + N^{-3/4} T^{1/2} \right) = o_P (1).
\end{aligned}
$$

Now, we analyze $A_{5NT,1}$ by further decomposing it as follows:

$$
\begin{aligned}
A_{5NT,1} = N^{-1/2} \sum_{i=1}^{N} \Big\{ & c_i' M_{F^0} P_{X_i} \left(M^{(2)} + M^{(rem)} \right) \left(\varepsilon_i + F^0 \lambda_i^0 + c_i \right) + \varepsilon_i' M_{F^0} P_{X_i} M^{(2)} \varepsilon_i \\
& + \varepsilon_i' M_{F^0} P_{X_i} M^{(2)} F^0 \lambda_i^0 + \varepsilon_i' M_{F^0} P_{X_i} M^{(2)} c_i \\
& + \varepsilon_i' M_{F^0} P_{X_i} M^{(rem)} \varepsilon_i + \varepsilon_i' M_{F^0} P_{X_i} M^{(rem)} F^0 \lambda_i^0 \\
& + \varepsilon_i' M_{F^0} P_{X_i} M^{(rem)} c_i \Big\} \\
\equiv\ & A_{5NT,11} + A_{5NT,12} + A_{5NT,13} + A_{5NT,14} + A_{5NT,15} \\
& + A_{5NT,16} + A_{5NT,17}.
\end{aligned}
$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

By the triangle inequality, the submultiplicative property of $\|\cdot\|$, (B.10), (B.2), and (B.3),

$$\begin{aligned}
\left|A_{5NT,11}\right| &\leq O_P\left(\gamma_{NT}\right)\left\|M^{(2)}+M^{(rem)}\right\| N^{-1/2}\sum_{i=1}^{N}\|X_i\|\left(\|\varepsilon_i\|+\left\|F^0\lambda_i^0\right\|+\|c_i\|\right)\\
&= O_P\left(\gamma_{NT}\delta_{NT}^{-2}\right)O_P\left(N^{1/2}T\right)=O_P\left(N^{-3/4}T^{1/2}+N^{1/4}T^{-1/2}\right)=o_P\left(1\right),\\
\left|A_{5NT,14}\right| &\leq O_P\left(\gamma_{NT}\right)\left\|M^{(2)}\right\| N^{-1/2}\sum_{i=1}^{N}\|\varepsilon_i\|\,\|X_i\|\\
&= O_P\left(\gamma_{NT}\delta_{NT}^{-2}\right)O_P\left(N^{1/2}T\right)\\
&= O_P\left(N^{-3/4}T^{1/2}+N^{1/4}T^{-1/2}\right)=o_P\left(1\right),\quad \text{and}\\
\left|A_{5NT,17}\right| &\leq O_P\left(\gamma_{NT}\right)\left\|M^{(rem)}\right\| N^{-1/2}\sum_{i=1}^{N}\|\varepsilon_i\|\,\|X_i\|\\
&= O_P\left(\delta_{NT}^{-1}\gamma_{NT}^{2}\right)O_P\left(N^{1/2}T\right)=O_P\left(\delta_{NT}^{-1}\right)=o_P\left(1\right).
\end{aligned}$$

For $A_{5NT,15}$, we have $A_{5NT,15}=N^{-1/2}\sum_{i=1}^{N}\varepsilon_i'P_{X_i}M^{(rem)}\varepsilon_i-N^{-1/2}\sum_{i=1}^{N}\varepsilon_i'P_{F^0}P_{X_i}M^{(rem)}\varepsilon_i\equiv A_{5NT,15a}-A_{5NT,15b}$. Noting that $\left\|P_{F^0}\varepsilon_i\right\|=O_P(T^{-1/2})\left\|F^{0\prime}\varepsilon_i\right\|_F$, we have by (B.2),

$$\begin{aligned}
\left|A_{5NT,15b}\right| &\leq O_P\left(T^{-1/2}\right)O_P\left(\delta_{NT}^{-1}\gamma_{NT}\right)N^{-1/2}\sum_{i=1}^{N}\left\|F^{0\prime}\varepsilon_i\right\|_F\|\varepsilon_i\|_F\\
&= O_P\left(\delta_{NT}^{-1}N^{-1/4}T^{-1}\right)O_P\left(N^{1/2}T\right)=O_P\left(N^{1/4}\delta_{NT}^{-1}\right)=o_P\left(1\right).
\end{aligned}$$

By (B.2), the fact that

$$\begin{aligned}
\mathbb{E}\left\|N^{-1/2}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'P_{X_i}\right\|_F^2 &= N^{-1}\mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\operatorname{tr}\left(P_{X_i}P_{X_j}\varepsilon_j\varepsilon_j'\varepsilon_i\varepsilon_i'\right)\right]\\
&\leq N^{-1}\mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\operatorname{tr}\left[\left(P_{X_i}P_{X_j}\right)^2\right]\right\}^{1/2}\right.\\
&\quad\left.\times\left\{\operatorname{tr}\left[\left(\varepsilon_j\varepsilon_j'\varepsilon_i\varepsilon_i'\right)^2\right]\right\}^{1/2}\right]\\
&\leq K^{1/2}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left(\varepsilon_i'\varepsilon_j\right)^2=O\left(NT\right),
\end{aligned}$$

and Chebyshev's inequality, we have

$$\begin{aligned}
\left|A_{5NT,15a}\right| &= \left|\operatorname{tr}\left(M^{(rem)}N^{-1/2}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'P_{X_i}\right)\right|\leq\left\|M^{(rem)}\right\|_F\left\|N^{-1/2}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'P_{X_i}\right\|_F\\
&= O_P\left(\delta_{NT}^{-1}\gamma_{NT}\right)O_P\left(N^{1/2}T^{1/2}\right)=O_P(\delta_{NT}^{-1}N^{1/4})=o_P\left(1\right).
\end{aligned}$$

It follows that $A_{5NT,15}=o_P\left(1\right)$.

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

Now, we write $A_{5NT,16}$ as follows: $A_{5NT,16} = N^{-1/2}\sum_{i=1}^N \varepsilon_i' P_{X_i} M^{(rem)} F^0 \lambda_i^0 - N^{-1/2}\sum_{i=1}^N \varepsilon_i' P_{F^0} P_{X_i} \ M^{(rem)} F^0 \lambda_i^0 \equiv A_{5NT,16a} - A_{5NT,16b}$. As in the study of $A_{5NT,15b}$, we can bound $A_{5NT,16b}$ by $o_P(1)$. Similarly, as in the study of $A_{5NT,15a}$, we have by (B.2) and Chebyshev's inequality

$$\begin{aligned}
|A_{5NT,16a}| &= \left|\operatorname{tr}\left(M^{(rem)} F^0 N^{-1/2} \sum_{i=1}^N \lambda_i^0 \varepsilon_i' P_{X_i}\right)\right| \\
&\le \left\|M^{(rem)} F^0\right\|_F \left\|N^{-1/2} \sum_{i=1}^N \lambda_i^0 \varepsilon_i' P_{X_i}\right\|_F \\
&= O_P\left(\delta_{NT}^{-1} \gamma_{NT} T^{1/2}\right) O_P\left(T^{1/2}\right) = O_P\left(N^{-1/4} + N^{-3/4} T^{1/2}\right) = o_P(1).
\end{aligned}$$

It follows that $A_{5NT,16} = o_P(1)$.

For $A_{5NT,12}$, we have $A_{5NT,12} = N^{-1/2}\sum_{i=1}^N \varepsilon_i' P_{X_i} M^{(2)} \varepsilon_i - N^{-1/2}\sum_{i=1}^N \varepsilon_i' P_{F^0} P_{X_i} M^{(2)} \varepsilon_i \equiv A_{5NT,12a} - A_{5NT,12b}$. We bound $\left\|N^{-1/2}\sum_{i=1}^N \varepsilon_i \varepsilon_i' P_{X_i}\right\|_F$ first. Observe that

$$\begin{aligned}
\left\|N^{-1/2} \sum_{i=1}^N \varepsilon_i \varepsilon_i' P_{X_i}\right\|_F^2 &= T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \operatorname{tr}\left[\varepsilon_i \varepsilon_i' X_i \hat{\Omega}_i^{-1} X_i' X_j \hat{\Omega}_j^{-1} X_j' \varepsilon_j \varepsilon_j'\right] \\
&= T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \operatorname{tr}\left[\varepsilon_i \varepsilon_i' X_i \Omega_i^{-1} X_i' X_j \Omega_j^{-1} X_j' \varepsilon_j \varepsilon_j'\right] \\
&\quad + T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \operatorname{tr}\left[\varepsilon_i \varepsilon_i' X_i (\hat{\Omega}_i^{-1} - \Omega_i^{-1}) X_i' X_j \Omega_j^{-1} X_j' \varepsilon_j \varepsilon_j'\right] \\
&\quad + T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \operatorname{tr}\left[\varepsilon_i \varepsilon_i' X_i \Omega_i^{-1} X_i' X_j (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) X_j' \varepsilon_j \varepsilon_j'\right] \\
&\quad + T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \operatorname{tr}\left[\varepsilon_i \varepsilon_i' X_i (\hat{\Omega}_i^{-1} - \Omega_i^{-1}) X_i' X_j^{-1} \right. \\
&\qquad\qquad \left. \times (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) X_j' \varepsilon_j \varepsilon_j'\right] \\
&\equiv D_{1NT} + D_{2NT} + D_{3NT} + D_{4NT}, \quad \text{say.}
\end{aligned}$$

Noting that $\mathbb{E}|D_{1NT}| = T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}\left\{\operatorname{tr}\left[\varepsilon_i \varepsilon_i' X_i \Omega_i^{-1} X_i' X_j \Omega_j^{-1} X_j' \varepsilon_j \varepsilon_j'\right]\right\} = O(N)$, $D_{1NT} = O_P(N)$ by Markov's inequality. For D_{2NT}, using $|\operatorname{tr}(AB)| \le \|A\|_F \|B\|_F$, $\operatorname{tr}(ABA') \le \mu_1(B)\operatorname{tr}(AA')$ for p.s.d. B, $|\mu_1(A)| \le \|A\|_F$ for symmetric matrix A, and Lemma A.7, we have

$$\begin{aligned}
|D_{2NT}| &= \left|T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \operatorname{tr}\left[\varepsilon_i \varepsilon_i' X_i (\hat{\Omega}_i^{-1} - \Omega_i^{-1}) X_i' \ X_j \Omega_j^{-1} X_j' \varepsilon_j \varepsilon_j'\right]\right| \\
&\le T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \left\{\operatorname{tr}\left[\varepsilon_i \varepsilon_i' X_i (\hat{\Omega}_i^{-1} - \Omega_i^{-1}) X_i' X_i (\hat{\Omega}_i^{-1} - \Omega_i^{-1}) X_i' \varepsilon_i \varepsilon_i'\right]\right\}^{1/2} \\
&\quad \times \left\{\operatorname{tr}\left[X_j \Omega_j^{-1} X_j' \varepsilon_j \varepsilon_j' \varepsilon_j \varepsilon_j' X_j \Omega_j^{-1} X_j'\right]\right\}^{1/2}
\end{aligned}$$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$$
\begin{aligned}
&= \max_{1\le i\le N}\left\{\mu_1\left[(\hat{\Omega}_i^{-1}-\Omega_i^{-1})X_i'X_i\left(\hat{\Omega}_i^{-1}-\Omega_i^{-1}\right)\right]\right\}^{1/2}\\
&\quad\times \max_{1\le j\le N}\left\{\mu_1\left(\Omega_j^{-1}X_j'X_j\Omega_j^{-1}\right)\right\}^{1/2}\\
&\quad\times T^{-2}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\mathrm{tr}\left[\varepsilon_i\varepsilon_i'X_iX_i'\varepsilon_i\varepsilon_i'\right]\right\}^{1/2}\left\{\mathrm{tr}\left[X_j'\varepsilon_j\varepsilon_j'\varepsilon_j\varepsilon_j'X_j\right]\right\}^{1/2}\\
&= O_P\left(T^{1/2}a_{NT}\right)O_P\left(T^{1/2}\right)O_P(N) = O_P(NTa_{NT}).
\end{aligned}
$$

Analogously, we can show that $D_{3NT} = O_P(NTa_{NT})$ and $D_{4NT} = O_P\left(NTa_{NT}^2\right)$. It follows that $\left\|N^{-1/2}\sum_{i=1}^N \varepsilon_i\varepsilon_i'P_{X_i}\right\|_F = O_P(N^{1/2}T^{1/2}a_{NT}^{1/2})$ and

$$
\begin{aligned}
A_{5NT,12a} &= \left|N^{-1/2}\mathrm{tr}\left(M^{(2)}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'P_{X_i}\right)\right| \le \left\|M^{(2)}\right\|_F\left\|N^{-1/2}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'P_{X_i}\right\|_F\\
&= O_P\left(\delta_{NT}^{-2}\right)O_P\left(N^{1/2}T^{1/2}a_{NT}^{1/2}\right) = a_{NT}^{1/2}O_P\left(N^{-1/2}T^{1/2}+N^{1/2}T^{-1/2}\right)\\
&= o_P(1).
\end{aligned}
$$

Similarly, we can show that $\left\|N^{-1/2}\sum_{i=1}^N \varepsilon_i\varepsilon_i'P_{F^0}P_{X_i}\right\|_F = O_P\left(N^{1/2}T^{1/2}a_{NT}^{1/2}\right)$ and thus

$$
\begin{aligned}
A_{5NT,12b} &= \left|N^{-1/2}\mathrm{tr}\left(M^{(2)}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'P_{F^0}P_{X_i}\right)\right| \le \left\|M^{(2)}\right\|_F\left\|N^{-1/2}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'P_{F^0}P_{X_i}\right\|_F\\
&= O_P\left(\delta_{NT}^{-2}\right)O_P\left(N^{1/2}T^{1/2}a_{NT}^{1/2}\right) = a_{NT}^{1/2}O_P\left(N^{-1/2}T^{1/2}+N^{1/2}T^{-1/2}\right)\\
&= o_P(1).
\end{aligned}
$$

Analogously, we can show that $A_{5NT,13} = o_P(1)$. This completes the proof of the proposition. ■

PROPOSITION B.6. $A_{6NT} = o_P(1)$ *under* $\mathbb{H}_1(\gamma_{NT})$.

Proof. By Cauchy–Schwarz inequality and Propositions B.2 and B.3, $A_{6NT} = N^{-1/2}\sum_{i=1}^N d_{2i}'P_{X_i}d_{3i} \le \{A_{2NT}\}^{1/2}\{A_{3NT}\}^{1/2} = o_P(1)o_P(1) = o_P(1)$. ■

Proof of Theorem 3.4. Let P^* denote the probability measure induced by the WB conditional on the original sample $\mathcal{W}_{NT} \equiv \{(Y_i, X_i), i = 1, \ldots, N\}$ and $\mathbb{E}^*$ and Var* denote the expectation and variance with respect to P^*. Let $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$ denote the probability order under P^*, e.g., $b_{NT} = o_{P^*}(1)$ if for any $\epsilon > 0$, $P^*\left(\|b_{NT}\|_F > \epsilon\right) = o_P(1)$. Note that $b_{NT} = o_P(1)$ implies that $b_{NT} = o_{P^*}(1)$.

Observing that $Y_{it}^* = \hat{\boldsymbol{\beta}}'X_{it} + \hat{\lambda}_i'\hat{F}_t + \varepsilon_{it}^*$, the null hypothesis is maintained in the bootstrap world. Given $\mathcal{W}_{NT}$, ε_{it}^* are independent across i, and are independent of X_{js}, $\hat{\lambda}_j$, and $\hat{F}_s$ for all i, t, j, s, because the latter objects are fixed in the fixed-design bootstrap world. Let $\mathcal{F}_{NT,t}^*$ denote the σ-field generated by $\{\varepsilon_{it}^*, \ldots, \varepsilon_{i1}^*\}_{i=1}^N$. For each i, $\left\{\varepsilon_{it}^*, \mathcal{F}_{NT,t}^*\right\}$ is an m.d.s. such that $\mathbb{E}^*\left(\varepsilon_{it}^*|\mathcal{F}_{NT,t-1}^*\right) = \hat{\varepsilon}_{it}\mathbb{E}(\varsigma_{it}) = 0$ and $\mathbb{E}^*[\left(\varepsilon_{it}^*\right)^2|\mathcal{F}_{NT,t-1}^*] =$

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press

$\hat{\varepsilon}_{it}^2 \mathbb{E}^*\left(\varsigma_{it}^2\right) = \hat{\varepsilon}_{it}^2$. These observations greatly simplify the proofs in the bootstrap world. In particular, we can show that $\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}} = O_{P^*}\left(N^{-1/2}T^{-1/2}\right)$.

Let LM_{NT}^*, J_{NT}^*, B_{NT}^*, V_{NT}^*, $\hat{B}_{NT}^*$, and $\hat{V}_{NT}^*$ denote the bootstrap analogue of LM_{NT}, J_{NT}, B_{NT}, V_{NT}, $\hat{B}_{NT}$, and $\hat{V}_{NT}$, respectively. Then $J_{NT}^* \equiv \left(N^{-1/2}LM_{NT}^* - B_{NT}^*\right)/\sqrt{V_{NT}^*}$ and $\hat{J}_{NT}^* \equiv \left(N^{-1/2}LM_{NT}^* - \hat{B}_{NT}^*\right)/\sqrt{\hat{V}_{NT}^*}$. Let $\hat{\varepsilon}_i^*$, d_{li}^*, and A_{sNT}^* denote the bootstrap analogue of $\hat{\varepsilon}_i$, d_{li}, and A_{sNT}, respectively, for $i = 1, 2, \ldots, N$, $l = 1, 2, 3$, and $s = 1, 2, \ldots, 6$. As in the proof of Theorem 3.3, we have

$$\begin{aligned}\sqrt{V_{NT}^*}J_{NT}^* &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\varepsilon}_i^{*\prime}P_{X_i}\hat{\varepsilon}_i^* - B_{NT}^* \\ &= \left(A_{1NT}^* - B_{NT}^*\right) + A_{2NT}^* + A_{3NT}^* + 2A_{4NT}^* + 2A_{5NT}^* + 2A_{6NT}^*.\end{aligned}$$

We prove the theorem by showing that (i) $\left(A_{1NT}^* - B_{NT}^*\right)/\sqrt{V_{NT}^*} \xrightarrow{D^*} N(0,1)$, (ii) $A_{sNT}^* = o_{P^*}(1)$ for $s = 2, \ldots, 6$, (iii) $\hat{B}_{NT}^* = B_{NT}^* + o_{P^*}(1)$, and (iv) $\hat{V}_{NT}^* = V_{NT}^* + o_{P^*}(1)$.

We only outline the proof of (i) as those of other parts are analogous to the corresponding parts in the proofs of Theorems 3.3 and 3.2. Analogously to the proof of Proposition B.1, we can show that $A_{1NT}^* - B_{NT}^* = \sum_{t=2}^{T} Z_{NT,t}^* + o_{P^*}(1)$ where $Z_{NT,t}^* \equiv 2T^{-1}N^{-1/2}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\varepsilon_{it}^*\varepsilon_{is}^*\hat{b}_{it}'\hat{b}_{is}$ and $\hat{b}_{it}$ denotes the tth row of $M_{\hat{F}}X_i\hat{\Omega}_i^{-1/2}$. Noting that $\{Z_{NT,t}^*, \mathcal{F}_{NT,t}^*\}$ is an m.d.s., we can continue to apply the martingale CLT in Pollard (1984, p. 171) by showing that

$$\mathcal{Z}^* \equiv \sum_{t=2}^{T}\mathbb{E}_{\mathcal{F}_{NT,t-1}^*}^*\left|Z_{NT,t}^*\right|^4 = o_{P^*}(1) \quad \text{and} \quad \sum_{t=2}^{T}Z_{NT,t}^{*2} - V_{NT}^* = o_{P^*}(1). \tag{B.16}$$

Using the i.i.d. property of ς_{it} and the fact that $\mathbb{E}^*\left(\varepsilon_{it}^{*2}\right) = \hat{\varepsilon}_{it}^2$ and $\mathbb{E}^*\left(\varepsilon_{it}^{*4}\right) = 3\hat{\varepsilon}_{it}^4$, we can readily show that

$$\begin{aligned}\mathbb{E}^*\left(\mathcal{Z}^*\right) = {} & \frac{48}{T^4N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{s=1}^{t-1}\sum_{q=1}^{t-1}\hat{b}_{it}'\hat{b}_{is}\hat{b}_{it}'\hat{b}_{is}\hat{b}_{jt}'\hat{b}_{jq}\hat{b}_{jt}'\hat{b}_{jq}\hat{\varepsilon}_{is}^2\hat{\varepsilon}_{jq}^2\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{jt}^2 \\ & + \frac{48}{T^4N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{1\leq r,s,q,v\leq t-1}\hat{b}_{it}'\hat{b}_{is}\hat{b}_{it}'\hat{b}_{ir}\hat{b}_{it}'\hat{b}_{iq}\hat{b}_{it}'\hat{b}_{iv}\mathbb{E}^*\left(\varepsilon_{is}^*\varepsilon_{ir}^*\varepsilon_{iq}^*\varepsilon_{iv}^*\right)\hat{\varepsilon}_{it}^4.\end{aligned}$$

One can readily show that the first term is $o_P(1)$ by noting that the total number of terms in the summation is of order $O\left(N^2T^3\right)$. Similarly, noting that $\mathbb{E}^*\left(\varepsilon_{is}^*\varepsilon_{ir}^*\varepsilon_{iq}^*\varepsilon_{iv}^*\right) = 0$ if $\#\{r,s,q,v\} = 3$ or 4, we can show that the second term is $o_P(1)$. Then $\mathcal{Z}^* = o_{P^*}(1)$ by the conditional Markov inequality. Now $\sum_{t=2}^{T}\mathbb{E}^*(Z_{NT,t}^{*2}) = 4T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\mathbb{E}^*[\varepsilon_{it}^*\hat{b}_{it}'\sum_{s=1}^{t-1}\hat{b}_{is}\varepsilon_{is}^*]^2 = V_{NT}^*$. (Apparently one can simplify the expression for V_{NT}^* by using the i.i.d. property of ς_{it} used in generating ε_{it}^*.) In addition, straightforward moment calculations yield that $\mathbb{E}^*(\sum_{t=2}^{T}Z_{NT,t}^{*2})^2 = V_{NT}^{*2} + o_P(1)$. Thus $\mathrm{Var}^*(\sum_{t=2}^{T}Z_{NT,t}^{*2}) = o_P(1)$ and $\sum_{t=2}^{T}Z_{NT,t}^{*2} - V_{NT}^* = o_{P^*}(1)$. ■

https://doi.org/10.1017/S0266466613000017 Published online by Cambridge University Press