Policy Analysis Using Panel and Multilevel Regression Models with Group Interactive Fixed Effects*

Min Seong Kim
University of Connecticut

January, 2018

Abstract

The use of panel and multilevel regression models is prevalent in policy analysis to estimate the effect of group level policies on individual outcomes. In order to allow for the time varying effect of group heterogeneity and the group specific impact of time effects, we propose a group interactive fixed effects approach that employs interactive terms of group fixed effects and common time effects in this model. For this approach, we consider the least squares estimator and associated inference procedure. We examine their properties under the large n, T asymptotics such that $T/n \to 0$. The number of groups, G, is allowed to grow but at a slower rate. The group structure of this model helps relieve the incidental parameters problem, and we provide conditions under which asymptotic unbiasedness and normality are achieved. We also develop a test for the appropriate level of grouping to specify group fixed effects. Finally, we study a GMM approach to address policy endogeneity with respect to the idiosyncratic error. Monte Carlo simulations show that the proposed estimators and tests perform well in finite samples. For empirical illustration, we revisit Buccirossi et al. (2013) and apply the proposed method to understand the effect of country level competition policy on the country-industry level productivity growth.

Keywords: endogeneity, GMM estimation, group interactive fixed effects, group level test, least squares estimation, panel and multilevel model, policy analysis

JEL Classification Number: C12, C13, C23, C54

---

*Email: minseong_kim@uconn.edu. The author is grateful for comments from Victor Aguirregabiria, Tim Conley, Chihwa Kao, Yixiao Sun, and seminar and conference participants at the University of Montreal, the University of Connecticut, the 33rd Meeting of the Canadian Econometrics Study Group, the 26th Annual Meeting of the Midwest Econometrics Group and NY Camp Econometrics XII.

1 Introduction

In this paper, we consider policy analysis using a panel and multilevel linear regression model that includes interactive terms of group fixed effects and common time effects. Multilevel regression models are prevalent in policy analysis to examine the effect of group level policies on individual level outcomes. See, for example, county level food stamp programs and household-head employment status (Hoynes and Schanzenbach, 2012), and country level competition policy and country-industry specific TFP growth (Buccirossi et al., 2013). In policy analysis using panel data, a researcher usually employs the additive fixed effects regression model to control for endogeneity due to correlations between the policy variable and unobserved group heterogeneity/time effects. The difference in differences (DID) is a typical example of this approach. A shortcoming of this method is that its validity crucially depends on the condition that the effect of unobserved group heterogeneity is time invariant and the impact of time effects is homogeneous across groups. This condition may be implausible in empirical applications. For example, in the evaluation of training programs, a training participant tends to have a temporary dip in earning that influences the participation decision (Ashenfelter and Card, 1985). Differential trends can also appear if groups based on different regions, markets or ages show heterogenous responses to common time effects such as cyclical fluctuations (Blundell and Dias, 2009).

To address this problem, we propose an interactive fixed effects model in the panel and multilevel regression setting. The proposed model is given by

$$
\begin{aligned}
Y_{it} &= X_{it}'\beta^0 + \lambda_{g_i}^{0\prime} F_t^0 + \varepsilon_{it}, \quad (i = 1, ..., n;\ t = 1, ..., T)\,, \qquad (1)\\
X_{it} &= \left[Z_{g_i t}', W_{it}'\right]',\ \beta^0 = \left[\beta_Z^{0\prime}, \beta_W^{0\prime}\right]',
\end{aligned}
$$

where a $(d_F \times 1)$ vector of unobserved time effects F_t^0 interacts with unobserved group fixed effects $\lambda_{g_i}^0$. X_{it} is a $(d_x \times 1)$ vector of regressors, β^0 is a vector of unknown coefficients, and ε_{it} is an idiosyncratic error. X_{it} contains $Z_{g_i t}$, a $(d_z \times 1)$ vector of group level regressors, and W_{it}, a vector of individual covariates. The former includes a policy variable of interest that is common within a group $g \in \{1, ..., G\}$ but may vary across different groups. Each individual that belongs to a group g shares λ_g^0. We assume the group membership $\{g_i\}_{i=1}^n$ to be known to researchers. As F_t^0 is multiplicative of $\lambda_{g_i}^0$, this model accounts for the effect of unobserved group heterogeneity that is time varying as well as the impact of unobserved time effects that is group specific. Our model nests the conventional additive model as a special case with $\lambda_{g_i}^0 = \left[\alpha_{g_i}^0, 1\right]'$ and $F_t^0 = \left[1, f_t^0\right]'$, where $\alpha_{g_i}^0$ and f_t^0 denote scalar group effects and time effects respectively.

We examine estimation and inference procedures for this model. We employ the least squares (LS) estimator, which we henceforth refer to as the "group interactive fixed effects estimator." Here, we assume X_{it} to be strictly exogenous with respect to ε_{it} but allow it to be arbitrarily correlated with $\lambda_{g_i}^{0\prime} F_t^0$. We investigate the asymptotic properties of this estimator under the asymptotics that $(G, n, T) \to \infty$ jointly. Our LS approach builds on Bai (2009), who proposes the LS estimator for

$$
Y_{it} = X_{it}'\beta^0 + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \qquad (2)
$$

which we will refer to as the "standard interactive fixed effects estimator" to distinguish it from the proposed estimator in this paper. Bai (2009) establishes the properties under the large n, T asymptotics. Different estimation methods for (2) have been studied in the literature. See, for example, the quasi-differencing method (Holtz-Eakin, Newey and Rosen, 1988), the GMM method (Ahn, Lee and Schmidt, 2001), the common correlated effects method (Pesaran, 2006), the QMLE (Bai and Li, 2014), and the method of Lasso (Lu and Su, 2016). Bai (2009) shows that when n grows at a faster rate than T, such that $T/n \to 0$, the standard interactive estimator for β^0 is $\sqrt{nT}$ consistent in the absence of heteroskedasticity and serial correlation in ε_{it}. When heteroskedasticity or serial correlation exist, the estimator is asymptotically biased due to the so-called incidental parameters problem (Neyman and Scott, 1948). As shown in Bertrand, Duflo, and Mullainathan (2004), researchers often confront positive serial correlation in policy analysis using panel data. Serial correlation can be removed by introducing a lagged dependent variable as a regressor, but, as is well known in the literature, this causes a different type of asymptotic bias. See Moon and Weidner (2017) for this issue. Bai (2009) provides a bias corrected estimator for (2). In our setting, since λ_g is common within a group, the source of this bias is of order $G/\sqrt{nT}$, and the estimator is asymptotically unbiased in the presence of heteroskedasticity and serial and cross sectional correlations given $G/\sqrt{nT} \to 0$ and $T/n \to 0$ as $(G, n, T) \to \infty$. These rate conditions are relevant in many empirical applications where a policy variable is group specific (e.g., state, industry or country) and many individuals belong to each group. An example of considering group structures that address the incidental parameters problem appears in the literature. Bester and Hansen (2016) study nonlinear panel data models with time invariant group effects under the asymptotics that $(G, n, T) \to \infty$.

Another contribution of this paper is to propose a Hausman type test for the level of grouping for the group fixed effects. While the asymptotics is established under the assumption that the group membership is known, it can be challenging, in a practical situation, to decide on the appropriate level of grouping to specify the group fixed effects. For example, when a policy is country specific and the outcome is at firm level, a researcher may specify the group fixed effects at country level. However, a finer level of grouping should be considered if he/she suspects that, within each country, the impact of unobserved time effects varies across, for example, different industry sectors. We suppose that two different grouping schemes, $\mathbb{A}_0$ and $\mathbb{A}_a$, of which the latter demonstrates a finer level of grouping and nests the former as a special case, are available. The null hypothesis of this test is that $\mathbb{A}_0$ is correctly specified against the alternative that $\mathbb{A}_0$ is misspecified and only $\mathbb{A}_a$ is correctly specified. Utilizing the fact that, under the null, both $\mathbb{A}_0$ and $\mathbb{A}_a$ yield consistent estimators while only $\mathbb{A}_a$ leads to consistency under the alternative, our test compares the group fixed effects estimator based on $\mathbb{A}_0$ with the one based on $\mathbb{A}_a$.

This paper also extends our approach to the case that a policy variable is endogenous with respect to the idiosyncratic error ε_{it}. Some sources of endogeneity, such as simultaneity and measurement error, may remain even when the group interactive terms are introduced. To address this, we consider a moment condition based GMM approach which we call the "interactive fixed effects GMM" (IFE-GMM) estimator. The idea of this approach based on the standard interactive fixed effects model is discussed by Moon, Shum and Weidner (2017) in the random coefficients logit demand models, but they do not provide the asymptotics. Thus, our contribution to this method is to provide the estimation procedure in the linear model setting and to establish consistency and asymptotic normality of the estimator under the asymptotics that $(G, n, T) \to \infty$. The IFE-GMM estimation needs instruments, Ψ_{it}, and has a larger set of potential instruments than the one without interactive fixed effects terms. For the latter, the exogeneity condition should hold not only for ε_{it} but also for $\lambda_{g_i}^{0\prime} F_t^0$, while the former requires exogeneity only with respect to ε_{it}. This can be empirically important, since usual instruments in the policy analysis literature tend to be correlated with group fixed effects. See Besley and Case (2000) for a detailed discussion of policy endogeneity in the panel regression framework. The IFE-GMM method includes the LS estimator as a special case with $\Psi_{it} = X_{it}$. Moon, Shum and Weidner (2017) first consider this endogeneity problem and propose the "least-squares minimum distance (LS-MD)" estimation method in the random coefficients logit demand models. Moon and Weidner (2017) and Lee, Moon and Weidner (2012) extend this method to the linear regression model. Recently, Lu (2017) proposes QMLE and iterative generalized principal components (IGPC) method to estimate spatial interactive fixed effects models in the presence of simultaneity.

We apply the proposed approach to estimate the effect of country level competition policy on country-industry level productivity growth. In this regard, Buccirossi, Ciari, Duso, Spagnolo and Vitale (2013, BCDSV hereafter) estimate the panel and multilevel regression model using the additive fixed effects approach to provide empirical evidence that the effect of competition policy on the total factor productivity (TFP) growth is positive and significant. We revisit their evidence with our group interactive fixed effects method. Our estimation is based on BCDSV (2013)'s 1995-2005 balanced subsample. We also conduct the test on the appropriate level to specify the group fixed effects and estimate the model using the IFE-GMM estimator. Our analysis finds that the magnitude of the coefficient for the competition policy variable and the degree of its significance are substantially decreased when we employ the proposed approach compared to the additive fixed effects method.

The proposed model is related to

$$Y_{it} = X_{it}'\beta^0 + \alpha_{g_i t}^0 + \varepsilon_{it},$$

where $\alpha_{g_i t}^0$ denotes a scalar group specific time effects term. This model accommodates time varying group heterogeneity in accordance with our model and can be estimated with the standard fixed effects approach. In contrast to our model, however, this model is not useful for estimating the effect of a group level policy due to multicollinearity of the dummies for $\alpha_{g_i t}^0$ and group level regressors. Bertrand, Duflo, and Mullainathan (2004) and Hansen (2007) consider this model within the multilevel regression framework. They assume $\alpha_{g_i t}^0$ to be uncorrelated with regressors and study inference issues caused by the serial correlation in $\alpha_{g_i t}^0$. The former discuss cluster robust inference and the latter studies GLS estimation of this model.

Recently, increasing attention has been paid to grouped panel data models in which the group membership is unknown. See, for example, Sun (2005), Hahn and Moon (2010), Bonhomme and Manresa (2012), Ando and Bai (2015), and Su, Shi and Phillips (2016). Among them, Ando and Bai (2015) consider the interactive fixed effects approach in the group structure when group membership is unobserved. In their paper, time effects are group specific and interact with individual specific effects, which is different from our setting. In addition, they assume the number of groups, G, to be fixed as $(n, T) \to \infty$.

The outline of this paper is as follows. Section 2 introduces our proposed model and its LS estimator. Section 3 examines the asymptotic properties of the estimator and associated test statistics. We assume that the number of factors is known and the panel data is balanced. Section 4 provides a test on the appropriate level of grouping for the group fixed effects. In Section 5, we study GMM estimation to address policy endogeneity with respect to the idiosyncratic error. Section 6 reports simulation evidence. In Section 7, we revisit BCDSV (2013) and apply the proposed method. The last section concludes.

Certain notation is used throughout the paper. The paper defines projection matrices $P_F = F(F'F)^{-1}F'$ and $M_F = I_T - P_F$. For a column vector x, the Euclidean norm is defined by $\|x\| = \sqrt{x'x}$. For an $(a \times b)$ matrix A, the Frobenius norm is $\|A\| = \sqrt{tr\,(A'A)}$.

2 Model and estimation

Model (1) can be rewritten as

$$Y_i = X_i\beta^0 + F^0\lambda^0_{g_i} + \varepsilon_i, \tag{3}$$

where $Y_i = (Y_{i1}, ..., Y_{iT})', X_i = (X_{i1}, ..., X_{iT})', \varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$ and $F^0 = \left(F^0_1, ..., F^0_T\right)'$. In matrix notation, it can also be expressed as

$$Y = \beta^0_1 X^{(1)} + \cdots + \beta^0_{d_x} X^{(d_x)} + F^0\Lambda^{0\prime} + \varepsilon,$$

where $\Lambda^0 = \left(\lambda^0_{g_1}, ..., \lambda^0_{g_n}\right)', Y = (Y_1, ..., Y_n), \varepsilon = (\varepsilon_1, ..., \varepsilon_n)$, β^0_ℓ is the ℓ-th component of β^0, and $X^{(\ell)} = \left(X^{(\ell)}_1, ..., X^{(\ell)}_n\right)$ is the $(T \times n)$ matrix of the ℓ-th regressor.

We follow Bai (2009) in considering LS estimation for this model. As λ_g is common within each group, the LS objective function is written as

$$\mathcal{Q}(\beta, F, \Lambda) = \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}(Y_i - X_i\beta - F\lambda_g)'(Y_i - X_i\beta - F\lambda_g), \tag{4}$$

where $\mathcal{A}_g = \{i : g_i = g\}$ is the set of individuals in group g. The estimators $\left(\hat{\beta}, \hat{F}, \hat{\Lambda}\right)$ minimize (4). F^0 and Λ^0 are not separately identifiable because they are multiplicative in the model. To address this, the following normalization

$$F'F/T = I_{d_F} \text{ and } \Lambda'\Lambda = \text{diagonal} \tag{5}$$

is employed in the literature. Under (5), Λ and F are uniquely determined given that the product $F\Lambda'$ and the LS estimator $\left(\hat{F}, \hat{\Lambda}\right)$ satisfy these restrictions.

Let $n_g = \sum_{i=1}^n 1\{i \in \mathcal{A}_g\}$, and $\bar{w}_g = n_g^{-1}\sum_{i\in\mathcal{A}_g} w_i$ for a random variable w. Concentrating

$$\hat{\lambda}_g(\beta, F) = \left(F'F\right)^{-1} F'\left(\frac{1}{n_g}\sum_{i\in\mathcal{A}_g}(Y_i - X_i\beta)\right) = \frac{F'\left(\bar{Y}_g - \bar{X}_g\beta\right)}{T} \tag{6}$$

out of (4), we have

$$\begin{aligned}
&\mathcal{Q}(\beta, F) \\
&= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(Y_i - P_F\bar{Y}_g - \left(X_i - P_F\bar{X}_g\right)\beta\right)'\left(Y_i - P_F\bar{Y}_g - \left(X_i - P_F\bar{X}_g\right)\beta\right) &(7)\\
&= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}(Y_i - X_i\beta)'(Y_i - X_i\beta) - \frac{1}{nT}\sum_{g=1}^{G} n_g\left(\bar{Y}_g - \bar{X}_g\beta\right)'\left(\frac{FF'}{T}\right)\left(\bar{Y}_g - \bar{X}_g\beta\right). &(8)
\end{aligned}$$

From (7), $\hat{\beta}(F)$ that minimizes $\mathcal{Q}(\beta, F)$ given F is

$$\hat{\beta}(F) = \left[\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_i - P_F\bar{X}_g\right)'\left(X_i - P_F\bar{X}_g\right)\right]^{-1}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_i - P_F\bar{X}_g\right)'\left(Y_i - P_F\bar{Y}_g\right).$$

We also obtain $\hat{F}(\beta)$ given β. Since the first term in (8) does not depend on F, minimization of (8) with respect to F reduces to maximization of

$$\begin{aligned}\sum_{g=1}^{G} n_g \left(\bar{Y}_g - \bar{X}_g\beta\right)' \frac{FF'}{T} \left(\bar{Y}_g - \bar{X}_g\beta\right) &= \frac{1}{T} tr \left[F' \sum_{g=1}^{G} n_g \left(\bar{Y}_g - \bar{X}_g\beta\right) \left(\bar{Y}_g - \bar{X}_g\beta\right)' F \right] \\ &= \frac{1}{T} tr \left[F' \bar{\mathcal{R}}(\beta) \bar{\mathcal{R}}(\beta)' F \right], \end{aligned} \tag{9}$$

where $\bar{\mathcal{R}}(\beta) = \left[\sqrt{n_1}\left(\bar{Y}_1 - \bar{X}_1\beta\right), ..., \sqrt{n_g}\left(\bar{Y}_g - \bar{X}_g\beta\right), ...\sqrt{n_G}\left(\bar{Y}_G - \bar{X}_G\beta\right)\right]$. It is known in the principal component analysis that the solution to maximization of (11) is the $(T \times d_F)$ matrix whose columns are the eigenvectors multiplied by $\sqrt{T}$ associated with the d_F largest eigenvalues of $\bar{\mathcal{R}}(\beta)\bar{\mathcal{R}}(\beta)'$. Therefore, we obtain $\left(\hat{\beta}, \hat{F}\right)$ based on

$$\hat{\beta} = \left[\sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(X_i - P_{\hat{F}}\bar{X}_g\right)' \left(X_i - P_{\hat{F}}\bar{X}_g\right) \right]^{-1} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(X_i - P_{\hat{F}}\bar{X}_g\right)' \left(Y_i - P_{\hat{F}}\bar{Y}_g\right), \tag{10}$$

and

$$\frac{1}{nT} \bar{\mathcal{R}}\left(\hat{\beta}\right) \bar{\mathcal{R}}\left(\hat{\beta}\right)' \hat{F} = \hat{F}\hat{\Gamma}, \tag{11}$$

where $\hat{\Gamma}$ is a diagonal matrix that includes the d_F largest eigenvalues of $(nT)^{-1} \bar{\mathcal{R}}\left(\hat{\beta}\right) \bar{\mathcal{R}}\left(\hat{\beta}\right)'$. For implementation, we plug an initial value of β in (11) or an initial value of F in (10) and iterate (10) and (11) to convergence. As discussed in Bai (2009) and Su and Chen (2013), this procedure can lead to a local minimum of the objective function (7) depending on the initial value we use. Thus, we should conduct iteration with several initial values and choose the one that produces the smallest value of (7). Applying $\left(\hat{\beta}, \hat{F}\right)$ to (6), we have

$$\hat{\lambda}_g = \frac{\hat{F}' \left(\bar{Y}_g - \bar{X}_g\hat{\beta}\right)}{T} \text{ and } \hat{\Lambda} = [\hat{\lambda}_{g_1}, ..., \hat{\lambda}_{g_n}]'. \tag{12}$$

Note that the rank condition requires $d_F \leq T - 1$. If $d_F = T$, we have $P_{\hat{F}} = I_{d_F}$ because each column of $\hat{F}$ is orthonormal. Thus,

$$X_i - P_{\hat{F}}\bar{X}_{g_i} = [Z_{g_i}, W_i] - \left[Z_{g_i}, \bar{W}_{g_i}\right] = \left[O, W_i - \bar{W}_{g_i}\right],$$

where O denotes a $(T \times d_z)$ zero matrix. It is obvious that $\sum_{i=1}^{n}\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)'\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)$ is not of full rank in this case. This implies, for example, that when $T = 5$, 4 is the maximum number of interactive terms we can employ in the model.

3 Asymptotic theory and inference

In this section, we examine the asymptotic properties of $\hat{\beta}$ and the associated Wald test statistics. Define

$$Q_{nT}^{vw}(F) = \frac{1}{nT} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(v_i - P_F\bar{v}_g\right)' \left(w_i - P_F\bar{w}_g\right),$$

for random variables v and w, and

$$\mathcal{X}_i^X(F) = \left(X_i - P_F \bar{X}_{g_i}\right) - \frac{1}{n}\sum_{j=1}^{n} a_{ij}^0 M_F X_j, \tag{13}$$

$$\begin{aligned} B_{nT}^{XX}(F) &= \frac{1}{nT}\sum_{i=1}^{n} \mathcal{X}_i^X(F)' \mathcal{X}_i^X(F) \\ &= Q_{nT}^{XX}(F) - \frac{1}{T}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^0 X_i' M_F X_j\right\} \end{aligned} \tag{14}$$

with $a_{ij}^0 = \lambda_{g_j}^{0\prime}\left(\Lambda^{0\prime}\Lambda^0/n\right)^{-1}\lambda_{g_i}^0$. To understand these two expressions, we can look at

$$\sqrt{nT}\left(\hat{\beta} - \beta^0\right) = Q_{nT}^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[\bar{X}_g' M_{\hat{F}} F^0 \lambda_g^0 + \left(X_i - P_{\hat{F}}\bar{X}_g\right)'\varepsilon_i\right] \tag{15}$$

which is directly obtained from (10). We can see that the first part of (15) comes from the estimation error in $\hat{F}$ because $M_{F^0}F^0 = 0$. For this term, Proposition A2 in the appendix indicates that, under the regularity assumptions presented below,

$$\begin{aligned} &\frac{1}{\sqrt{nT}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\bar{X}_g' M_{\hat{F}} F^0 \lambda_g^0 \\ &= \frac{1}{T}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^0 X_i' M_{\hat{F}} X_j\right\}\sqrt{nT}\left(\hat{\beta} - \beta\right) - \frac{1}{\sqrt{nT}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^0 X_i' M_{\hat{F}}\varepsilon_j \\ &+ O_p\left(\frac{G}{\sqrt{nT}}\right) + O_p\left(\frac{\sqrt{T}}{n}\right) + o_p(1) \end{aligned}$$

Combining this expression with (15), we obtain

$$\begin{aligned} &\sqrt{nT}\left(\hat{\beta} - \beta^0\right) \\ &= B_{nT}^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{j=1}^{n}\mathcal{X}_i^X\left(\hat{F}\right)'\varepsilon_i + O_p\left(\frac{G}{\sqrt{nT}}\right) + O_p\left(\frac{\sqrt{T}}{n}\right) + o_p(1). \end{aligned}$$

Thus, both (13) and (14) involve the effect of estimation error in $\hat{F}$.

To further investigate the asymptotics of $\hat{\beta}$, we make the following assumptions.

Assumption 1 *(i)* $E\|X_{it}\|^4 \leq M$. *(ii) Let* $\mathcal{F} = \{F : F'F/T = I_{d_F}\}$.

$$\inf_{F\in\mathcal{F}} B_{nT}^{XX}(F) > 0.$$

Assumption 2 *(i)* $E\|F_t\|^4 \leq M$ *and* $T^{-1}\sum_{t=1}^{T} F_t F_t' \to^p \sum_F > 0$ *as* $T \to \infty$; *(ii)* $E\|\lambda_g\|^4 \leq M$ *and* $\Lambda'\Lambda/n \to^p \sum_\Lambda > 0$ *as* $n \to \infty$.

Assumption 3 *(i) For all* i *and* t, $E(\varepsilon_{it}) = 0$ *and* $E\left(\varepsilon_{it}^8\right) \leq M$.
(ii) For all (t,s),

$$\lim_{n\to\infty} \frac{1}{G} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} \sqrt{n_g n_{\tilde{g}}} \left| E\left(\bar{\varepsilon}_{gt} \bar{\varepsilon}_{\tilde{g}s}\right)\right| < M.$$

(iii) For all $(g, \tilde{g})$

$$\lim_{n\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \sqrt{n_g n_{\tilde{g}}} \left| E\left(\bar{\varepsilon}_{gt} \bar{\varepsilon}_{\tilde{g}s}\right)\right| < M.$$

(iv)

$$\lim_{n,T\to\infty} \frac{1}{GT} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} \sum_{t=1}^{T} \sum_{s=1}^{T} \sqrt{n_g n_{\tilde{g}}} \left| E\left(\bar{\varepsilon}_{gt} \bar{\varepsilon}_{\tilde{g}s}\right)\right| < M.$$

(v) Let $\rho_{\max}(A)$ *be the largest eigenvalue of* A.

$$\left|\rho_{\max}\left(n_g E\left(\bar{\varepsilon}_g \bar{\varepsilon}_g'\right)\right)\right| < M$$

uniformly in g *and* T.
(vi) For all (t,s),

$$E\left(\frac{1}{\sqrt{G}} \sum_{g=1}^{G} n_g \left\{\bar{\varepsilon}_{gt}\bar{\varepsilon}_{gs} - E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{gs}\right)\right\}\right)^4 < M.$$

(vii) We have

$$\lim_{n,T\to\infty} \frac{1}{GT^2} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{\tilde{t}=1}^{T} \sum_{\tilde{s}=1}^{T} n_g n_{\tilde{g}} \left| Cov\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{gs}, \bar{\varepsilon}_{\tilde{g}\tilde{t}}\bar{\varepsilon}_{\tilde{g}\tilde{s}}\right)\right| < M,$$

$$\lim_{n,T\to\infty} \frac{1}{G^2 T} \sum_{g=1}^{G} \sum_{\tilde{g}_1=1}^{G} \sum_{\tilde{g}_2=1}^{G} \sum_{t=1}^{T} \sum_{s=1}^{T} n_g \sqrt{n_{\tilde{g}_1} n_{\tilde{g}_2}} \left| Cov\left(\bar{\varepsilon}_{\tilde{g}_1 t}\bar{\varepsilon}_{gt}, \bar{\varepsilon}_{\tilde{g}_2 s}\bar{\varepsilon}_{gs}\right)\right| < M.$$

Assumption 4 *(i)* ε_{it} *is independent of* λ_g *and* F_s *for all* i, t, s *and* g; *(ii)* $E(\varepsilon_{it}|X_1, ..., X_n) = 0$ *for all* i *and* t.

Assumption 5 *For all* g, $n_g/n^{\alpha} \to c_g$ *where* $c_g > 0$ *and* $0 < \alpha < 1$.

Assumption 1 is the identification condition for the proposed LS estimator. Assumption 2 provides the moment conditions for $\{F_t\}$ and $\{\lambda_g\}$ and ensures that there exist d_F distinct time effects. Assumption 3 states the moment conditions and weak dependence conditions for $\{\varepsilon_{it}\}$. This assumption is adapted from Bai (2009), and weak dependence conditions are given based on (scaled) mean, $\sqrt{n_g}\bar{\varepsilon}_{gt}$, due to the group structure of our model. Weak dependence of $\left\{\sqrt{n_g}\bar{\varepsilon}_{gt}\right\}$ across groups is not restrictive, as strong dependence is captured by interactive terms.

In Assumption 4, we assume independence of $\{\varepsilon_{it}\}$ with $\{\lambda_g\}$ and $\{F_s\}$ which are standard in the literature. We introduce the strict exogeneity condition for $\{X_{it}\}$ which is also standard for the additive fixed effects model. Assumption 5 implies that group sizes are comparable to each other and that, for all g, n_g grows as n increases but at a slower rate.

Theorem 1 *Suppose that Assumptions 1-5 hold. Then, we have*

$$\hat{\beta} - \beta^0 \to^p 0$$

as (i) $(G, n, T) \to \infty$*, or (ii)* $(G, n) \to \infty$ *such that* $G/n \to 0$ *for fixed* T.

The proof of Theorem 1 is in the appendix. This result is analogous to consistency for the standard interactive fixed effects model in Bai (2009, Proposition 1(i)). We can compare the rate conditions between these two estimators. If λ_i^0 were known in Bai's model, it would be easy to show that consistency is achieved as $n \to \infty$ regardless of T. Hence, the rate condition $(n, T) \to \infty$ in Bai (2009) reflects the fact that λ_i^0 is unknown and estimated. In our model, λ_g^0 is common within each group. Due to this group structure, $\hat{\beta}$ is consistent not only when $(G, n, T) \to \infty$, but they also remain consistent for a fixed T, if the number of group members in each group grows such that $G/n \to 0$ as $n \to \infty$. In the rest of this paper, we only consider the asymptotics that $(G, n, T) \to \infty$. As for $\hat{F}$, its average norm consistency is provided in Proposition A1 in the appendix.

To obtain the asymptotic normality, we introduce the following assumptions.

Assumption 6 $(G, n, T) \to \infty$ *such that* $T/n \to 0$ *and* $G/\sqrt{nT} \to 0$.

Assumption 7

$$\lim_{n,T\to\infty} \frac{1}{GnT} \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{s=1}^{T} n_g \left| E\left(X_{it}' \bar{\varepsilon}_{gt} X_{is} \bar{\varepsilon}_{gs} \right) \right| < M,$$

$$\lim_{n,T\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \left\| E\left(X_{it} \varepsilon_{it} X_{js}' \varepsilon_{js} \right) \right\| < M,$$

$$\lim_{n,T\to\infty} \frac{1}{GT} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} \sum_{t=1}^{T} \sum_{s=1}^{T} \sqrt{n_g n_{\tilde{g}}} \left\| E\left(\bar{X}_{gt} \bar{\varepsilon}_{gt} \bar{X}_{\tilde{g}s}' \bar{\varepsilon}_{\tilde{g}s} \right) \right\| < M.$$

The rate condition, $T/n \to 0$ and $G/\sqrt{nT} \to 0$, in Assumption 6 is necessary to address the incidental parameters problem when $(G, n, T) \to \infty$. The weak dependence conditions in Assumption 7 are devised to obtain the asymptotic distribution under strict exogeneity of $\{X_{it}\}$.

Define

$$V_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[\mathcal{X}_i^X \left(F^0\right)' \varepsilon_i \varepsilon_j' \mathcal{X}_j^X \left(F^0\right) \right]. \tag{16}$$

Assumption 8 *Let* $\mathcal{X}_i^X = \mathcal{X}_i^X \left(F^0\right)$. *We have*

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left(\mathcal{X}_i^X\right)' \varepsilon_i \to^d N\left(0, V\right),$$

where $V = \lim_{n,T\to\infty} V_{nT}$ *is positive definite.*

A similar assumption is made in Bai (2009) and Lu and Su (2016). Note that this assumption allows heteroskedasticity and serial and cross sectional correlation in the idiosyncratic errors. Theorem 2 states the asymptotic normality of $\hat{\beta}$.

Theorem 2 *Suppose that Assumptions 1-8 hold. We then have*

$$\sqrt{nT}\left(\hat{\beta}-\beta^0\right) \to^d N\left(0, B_{XX}^{-1} V B_{XX}^{-1}\right),$$

where $B_{XX} = \text{plim}_{n,T\to\infty} B_{nT}^{XX}\left(F^0\right)$.

The proof is in the appendix. To obtain this result, we consider the following expansion which appears in the proof of this theorem.

$$\begin{aligned}\sqrt{nT}\left(\hat{\beta}-\beta^0\right) &= B_{XX}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i^X\left(F^0\right)'\varepsilon_i + \frac{G}{\sqrt{nT}}B_{XX}^{-1}A_{nT}^{(1)} + \sqrt{\frac{T}{n}}B_{XX}^{-1}A_{nT}^{(2)} \\ &\quad + o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta^0\right\|\right) + o_p\left(\frac{G}{\sqrt{nT}}\right) + o_p\left(\sqrt{\frac{T}{n}}\right),\end{aligned} \tag{17}$$

where

$$\begin{aligned}A_{nT}^{(1)} &= \frac{1}{nT}\sum_{i=1}^{n}X_i' M_{F^0}\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)F^0 H\Upsilon\lambda_{g_i}, \\ A_{nT}^{(2)} &= \frac{1}{GT}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\frac{\left(\bar{X}_g-K_g\right)F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1}\lambda_{\tilde{g}}^0\sqrt{n_{\tilde{g}}n_g}\bar{\varepsilon}_{\tilde{g}}'\bar{\varepsilon}_g, \\ \text{with } \Upsilon &= \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1} \text{ and } K_g = \lambda_g^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1}\left(\frac{1}{n}\sum_{g=1}^{G}n_g\lambda_g^0\bar{X}_g\right).\end{aligned}$$

When G grows at the same rate as $\sqrt{nT}$, $A_{nT}^{(1)}$ may become a source of asymptotic bias that should be corrected for valid inference. An extreme case is when $G = n$ and $g_i = i$ for all i, so that $\hat{\beta}$ reduces to the standard interactive fixed effects estimator. As shown by Bai (2009), in this case, the interactive estimator is asymptotically biased as $T/n \to c > 0$ unless the error term is homoskedastic and serially and cross sectionally uncorrelated. For Bai's model, if n grows faster than T ($T/n \to 0$), then the third term vanishes. However, the second term becomes divergent in the presence of heterogeneity or serial correlation and dominates the distribution. In our model, we assume that G grows, but slowly, such that $G/\sqrt{nT} \to 0$ as $T/n \to 0$. This rate condition is empirically relevant in policy analysis in which each group often includes a large number of individuals, for example, when we have a state or county level policy and individual outcome and when we have a country level policy and country-industry or firm level outcome. Under this condition, $\hat{\beta}$ is still centered at β^0 and is asymptotically normal when normalized by the sample size.

We consider inference on β^0 based on Theorem 2. Suppose that we are interested in the following null and alternative hypotheses

$$\mathcal{H}_0 : R\beta = \mathfrak{r}^0 \text{ vs } \mathcal{H}_1 : R\beta \neq \mathfrak{r}^0, \tag{18}$$

where R is a $(d_R \times d_x)$ matrix and $\mathfrak{r}^0$ is a $(d_R \times 1)$ vector. To do this, we first need to estimate B_{nT}^{XX} and V_{nT} in the variance term. Estimation of B_{nT}^{XX} is straightforward because we can use the sample analogue.

$$\hat{B}_{nT}^{XX} = \frac{1}{nT}\sum_{i=1}^{n}\left(\hat{\mathcal{X}}_i^X\right)'\hat{\mathcal{X}}_i^X$$
$$= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_i - P_{\hat{F}}\bar{X}_g\right)'\left(X_i - P_{\hat{F}}\bar{X}_g\right) - \frac{1}{T}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{a}_{ij}X_i'M_{\hat{F}}X_j\right\} \tag{19}$$

where

$$\hat{\mathcal{X}}_i^X = \left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right) - \frac{1}{n}\sum_{j=1}^{n}\hat{a}_{ij}M_{\hat{F}}X_j \text{ with } \hat{a}_{ij} = \hat{\lambda}_i'\left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1}\hat{\lambda}_j, \tag{20}$$

and $\hat{\lambda}_g$ and $\hat{\Lambda}$ are given in (12). For estimation of V_{nT}, we introduce the following conditions.

Assumption 9 *(i) $\mathcal{X}_{it}^X\varepsilon_{it}$ and $\mathcal{X}_{js}^X\varepsilon_{js}$ are independent if $g_i \neq g_j$, (ii) $\mathcal{X}_{it}^X\varepsilon_{it}$ and $\mathcal{X}_{js}^X\varepsilon_{js}$ are independent if $i \neq j$ or $t \neq s$ and $Var\left(\varepsilon_{it}|X, F^0, \Lambda^0\right) = \sigma^2$, and $B_{nT}^{XX}\left(F^0\right) - (nT)^{-1}\sum_{i=1}^{n}E\left[\left(\mathcal{X}_i^X\right)'\mathcal{X}_i^X\right] \to^p 0$.*

The cluster covariance structure in Assumption 9(i) is commonly employed in the panel and multilevel regression, e.g., Moulton (1990), Donald and Lang (2001), Bertrand, et al. (2004) and Hansen (2007). While this assumption characterizes the cluster dependence of $\left\{\mathcal{X}_{it}^X\varepsilon_{it}\right\}$ based on $\{\mathcal{A}_{g_i}, i = 1, ..., n\}$ for notational simplicity, it can be easily generalized to any level of clustering as long as the independence condition holds across clusters. A researcher selects an appropriate level of clustering by using his/her prior information about data or by conducting a test for this choice. See, for example, Ibragimov and Mueller (2014) who develop a test about the level of clustering. We can also relax this assumption to allow weak dependence across clusters by assuming the number of observations located on the boundaries to be negligible. See Bester, Conley and Hansen (2010) for details. The independence and homoskedasticity condition in Assumption 9(ii) is made to develop a test about the appropriate level of grouping to specify group fixed effects in Section 4.

Under Assumption 9(i) and (ii), V_{nT} reduces to

$$V_{nT}^c = \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}E\left[\left(\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\mathcal{X}_j^X\right], \text{ and}$$
$$V_{nT}^s = \frac{\sigma^2}{nT}\sum_{i=1}^{n}E\left[\left(\mathcal{X}_i^X\right)'\mathcal{X}_i^X\right]$$

respectively and we estimate them with

$$\hat{V}_{nT}^c = \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_i^X\right)'\hat{\varepsilon}_i\hat{\varepsilon}_j'\hat{\mathcal{X}}_j^X, \text{ and}$$
$$\hat{V}_{nT}^s = \hat{\sigma}^2\hat{B}_{nT}^{XX}, \text{ where } \hat{\sigma}^2 = \frac{1}{nT}\sum_{i=1}^{n}\hat{\varepsilon}_i'\hat{\varepsilon}_i \text{ with } \hat{\varepsilon}_i = Y_i - X_i\hat{\beta} - \hat{F}\hat{\lambda}_{g_i}.$$

Theorem 3 *Suppose that Assumptions 1-8 hold. If $\mathcal{H}_0$ is true, then we have*

$$\left(\hat{B}_{nT}^{XX}\right)^{-1}\hat{V}_{nT}^{c}\left(\hat{B}_{nT}^{XX}\right)^{-1}-B_{XX}^{-1}V_{nT}^{c}B_{XX}^{-1}\rightarrow^{p}0,\ \textit{and}$$

$$\hat{\sigma}^{2}\left(\hat{B}_{nT}^{XX}\right)^{-1}-\sigma^{2}B_{XX}^{-1}\rightarrow^{p}0$$

under Assumption 9(i) and Assumptions 9(ii) respectively.

The Wald statistics are given by

$$\mathbb{W}^{c}=\sqrt{nT}\left(R\hat{\beta}-\mathfrak{r}^{0}\right)'\left(R\left(\hat{B}_{nT}^{XX}\right)^{-1}\hat{V}_{nT}^{c}\left(\hat{B}_{nT}^{XX}\right)^{-1}R'\right)^{-1}\sqrt{nT}\left(R\hat{\beta}-\mathfrak{r}^{0}\right),$$
$$\mathbb{W}^{s}=\sqrt{nT}\left(R\hat{\beta}-\mathfrak{r}^{0}\right)'\left(\hat{\sigma}^{2}R\left(\hat{B}_{nT}^{XX}\right)^{-1}R'\right)^{-1}\sqrt{nT}\left(R\hat{\beta}-\mathfrak{r}^{0}\right).$$

The corollary below follows from Theorems 2 and 3.

Corollary 1 *Suppose that Assumptions 1-8 hold. If $\mathcal{H}_0$ is true, then we have*

$$\mathbb{W}^{c}\rightarrow^{d}\chi^{2}(d_R)\ \textit{and}\ \mathbb{W}^{s}\rightarrow^{d}\chi^{2}(d_R)$$

under Assumption 9(i) and Assumption 9(ii) respectively.

4 Testing the level of grouping for group fixed effects

To establish the asymptotics of $\hat{\beta}$, we have assumed the group membership to be known. In empirical applications, however, we may have to decide on the appropriate level of grouping from among several alternatives. For example, when we estimate the effect of a country level policy on a firm level outcome, it may be natural to introduce the interactive terms using the country level group effects. However, if we suspect that the sensitivity to the common time effects varies across different industries within a same country, then we should consider a finer (country-industry) level of grouping. To address this practical problem, we develop a Hausman type test for the appropriate level of grouping to specify group fixed effects.

Suppose that two different levels of grouping are available:

$$\mathbb{A}_0=\{\mathcal{A}_1,...,\mathcal{A}_g,...,\mathcal{A}_{G_0}\}\text{ and}$$
$$\mathbb{A}_a=\{\mathcal{A}_1^{(1)},...,\mathcal{A}_1^{(\kappa_1)},...,\mathcal{A}_g^{(1)}...,\mathcal{A}_g^{(\kappa_g)},...,\mathcal{A}_G^{(1)}...,\mathcal{A}_G^{(\kappa_G)}\},$$

between which $\mathbb{A}_a$ is the finer level of grouping in that $\mathcal{A}_g=\cup_{\ell=1}^{\kappa_g}\mathcal{A}_g^{(\ell)}$. Let $\hat{\beta}_0$ and $\hat{\beta}_a$ denote the group interactive fixed effects estimators based on $\mathbb{A}_0$ and $\mathbb{A}_a$ respectively and $\lambda_g^{(\ell)}$ denotes the group fixed effects for $\mathcal{A}_g^{(\ell)}$. We also define G_0 and $G_a=\sum_{g=1}^{G_0}\kappa_g$ as the numbers of groups under $\mathbb{A}_0$ and $\mathbb{A}_a$ respectively. We assume that the rate conditions in Assumptions 5 and 6 hold for both $\mathbb{A}_0$ and $\mathbb{A}_a$. We consider the following null and alternative hypotheses.

$\mathcal{H}_0$: $\mathbb{A}_0$ is correctly specified.
$\mathcal{H}_a$: $\mathbb{A}_0$ is misspecified and $\mathbb{A}_a$ is correctly specified.

We develop the test based on the facts that (i) $\hat{\beta}_0$ and $\hat{\beta}_a$ are consistent under $\mathcal{H}_0$ because $\mathbb{A}_0$ is nested in $\mathbb{A}_a$ with $\lambda_{g_i} = \lambda_g$ for all $i \in \cup_{\ell=1}^{\kappa_g} \mathcal{A}_g^{(\ell)}$, and that (ii) only $\hat{\beta}_a$ is consistent under $\mathcal{H}_a$, since $\lambda_g^{(\ell)} \neq \lambda_g^{(k)}$ for $\ell \neq k$ within the group $\mathcal{A}_g$, .

Let

$$\mathcal{X}_{0,i}^X (F) = \left(X_i - P_F \bar{X}_{0,g_i}\right) - \frac{1}{n} \sum_{j=1}^{n} a_{ij}^0 M_F X_j,\ \mathcal{X}_{0,i}^X = \mathcal{X}_{0,i}^X \left(F^0\right)$$

$$B_{0,nT}^{XX} = \frac{1}{nT} \sum_{i=1}^{n} \left(\mathcal{X}_{0,i}^X\right)' \mathcal{X}_{0,i}^X \text{ and } B_{0,XX} = \underset{n,T\to\infty}{\text{plim}}\ B_{0,nT}^{XX}$$

where $\bar{X}_{0,g_i}$ denotes the group average of X_i based on $\mathbb{A}_0$. We define $\mathcal{X}_{a,i}^X (F)$, $\mathcal{X}_{a,i}^X$, $B_{a,nT}^{XX}$ and $B_{a,XX}$ for the model based on $\mathbb{A}_a$ in the same manner. The following result is a direct consequence of Theorem 2.

Corollary 2 *Suppose that Assumptions 1-8 hold. Then, under $\mathcal{H}_0$, we have*

$$\begin{aligned}\sqrt{nT}\left(\hat{\beta}_0 - \hat{\beta}_a\right) &= \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left(\left(B_{0,nT}^{XX}\right)^{-1} \left(\mathcal{X}_{0,i}^X\right)' - \left(B_{a,nT}^{XX}\right)^{-1} \left(\mathcal{X}_{a,i}^X\right)'\right) \varepsilon_i + o_p(1) \\ &\to^d N\left(0, V_{\mathcal{T}}\right)\end{aligned}$$

where

$$V_{\mathcal{T}} = \lim_{n,T\to\infty} Var\left(\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left(\left(B_{0,nT}^{XX}\right)^{-1} \left(\mathcal{X}_{0,i}^X\right)' - \left(B_{a,nT}^{XX}\right)^{-1} \left(\mathcal{X}_{a,i}^X\right)'\right) \varepsilon_i\right). \tag{21}$$

This result allows us to test the level of grouping. Under Assumption 9(ii), $V_{\mathcal{T}}$ is written as

$$\begin{aligned}V_{\mathcal{T}} &= \lim_{n,T\to\infty} \frac{\sigma^2}{nT} \sum_{i=1}^{n} E\left[\left(\left(B_{0,nT}^{XX}\right)^{-1} \left(\mathcal{X}_{0,i}^X\right)' - \left(B_{a,nT}^{XX}\right)^{-1} \left(\mathcal{X}_{a,i}^X\right)'\right)\right. \\ &\quad \left.\times \left(\left(B_{0,nT}^{XX}\right)^{-1} \left(\mathcal{X}_{0,i}^X\right)' - \left(B_{a,nT}^{XX}\right)^{-1} \left(\mathcal{X}_{a,i}^X\right)'\right)'\right] \\ &= \sigma^2 \left(B_{0,XX}^{-1} + B_{a,XX}^{-1} - B_{0,XX}^{-1} C_{0a,XX} B_{a,XX}^{-1} - B_{a,XX}^{-1} C_{a0,XX} B_{0,XX}^{-1}\right),\end{aligned} \tag{22}$$

where $C_{0a,XX} = \lim_{n,T\to\infty} C_{0a,nT}^{XX}$, $C_{0a,nT}^{XX} = (nT)^{-1} \sum_{i=1}^{n} E\left[\left(\mathcal{X}_{0,i}^X\right)' \mathcal{X}_{a,i}^X\right]$ and $C_{a0,XX} = C_{0a,XX}'$.

Let

$$\hat{B}_{0,nT}^{XX} = \frac{1}{nT} \sum_{i=1}^{n} \left(\hat{\mathcal{X}}_{0,i}^X\right)' \hat{\mathcal{X}}_{0,i}^X,\ \hat{B}_{a,nT}^{XX} = \frac{1}{nT} \sum_{i=1}^{n} \left(\hat{\mathcal{X}}_{a,i}^X\right)' \hat{\mathcal{X}}_{a,i}^X,$$

$$\hat{C}_{0a,nT}^{XX} = \frac{1}{nT} \sum_{i=1}^{n} \left(\hat{\mathcal{X}}_{0,i}^X\right)' \hat{\mathcal{X}}_{a,i}^X,$$

where $\hat{\mathcal{X}}_{0,i}^X$ and $\hat{\mathcal{X}}_{a,i}^X$ are defined in the same way as $\hat{\mathcal{X}}_i^X$ in (20) based on $\mathbb{A}_0$ and $\mathbb{A}_a$ respectively. We introduce the test statistic $\mathcal{T}$ given by

$$\mathcal{T} = nT \left(\hat{\beta}_0 - \hat{\beta}_a\right)' \hat{V}_{\mathcal{T}}^{-1} \left(\hat{\beta}_0 - \hat{\beta}_a\right),$$

where

$$\hat{V}_{\mathcal{T}} = \hat{\sigma}_a^2 \left[\left(\hat{B}_{0,nT}^{XX} \right)^{-1} + \left(\hat{B}_{a,nT}^{XX} \right)^{-1} - \left(\hat{B}_{0,nT}^{XX} \right)^{-1} \hat{C}_{0a,nT}^{XX} \left(\hat{B}_{a,nT}^{XX} \right)^{-1} \right.$$
$$\left. - \left(\hat{B}_{a,nT}^{XX} \right)^{-1} \hat{C}_{a0,nT}^{XX} \left(\hat{B}_{0,nT}^{XX} \right)^{-1} \right]. \tag{23}$$

We estimate σ^2 based on $\mathbb{A}_a$ because it is consistent under both the null and alternative hypotheses. We make an additional assumption.

Assumption 10 *Let* $\tilde{C}_{0a,nT}^{XX} = (nT)^{-1} \sum_{i=1}^{n} \left(\mathcal{X}_{0,i}^X \right)' \mathcal{X}_{a,i}^X$. $\hat{C}_{0a,nT}^{XX} - C_{0a,nT}^{XX} \to^p 0$.

It is worth noting that we may simplify the test statistic further. Using the fact that a finer group $\mathcal{A}_g^{(1)} ..., \mathcal{A}_g^{(\kappa_g)}$ are the subsets of $\mathcal{A}_g$, we can show that

$$\frac{1}{nT} \sum_{i=1}^{n} \bar{X}_{0,g_i}' P_F \bar{X}_{a,g_i} = \frac{1}{nT} \sum_{i=1}^{n} \bar{X}_{0,g_i}' P_F X_i. \tag{24}$$

If we apply (24) to (22), then, under the null hypothesis, the covariance of $\hat{\beta}_0$ and $\hat{\beta}_a$ equals the variance of $\hat{\beta}_0$. That is,

$$\lim_{n,T \to \infty} \frac{\sigma^2}{nT} \sum_{i=1}^{n} E \left[\left(\mathcal{X}_{0,i}^X \right)' \mathcal{X}_{a,i}^X \right] = \sigma^2 B_{0,XX}$$

and $V_{\mathcal{T}}$ reduces to $\sigma^2 \left(B_{0,XX}^{-1} - B_{a,XX}^{-1} \right)$. This yields another candidate variance estimator

$$\tilde{V}_{\mathcal{T}} = \hat{\sigma}_a^2 \left(\left(\hat{B}_{0,nT}^{XX} \right)^{-1} - \left(\hat{B}_{a,nT}^{XX} \right)^{-1} \right).$$

It may be tempting to construct the test statistic based on $\tilde{V}_{\mathcal{T}}$, say $\tilde{\mathcal{T}}$, which is analogous to the standard Hausman (1978) test statistic. Though both $\mathcal{T}$ and $\tilde{\mathcal{T}}$ are valid in the asymptotic sense, we suggest using the former. Let $\hat{F}_0$ and $\hat{F}_a$ denote the estimators of $F^0 H$ based on $\mathbb{A}_0$ and $\mathbb{A}_a$ respectively. As implied by Proposition A2 in the appendix, the crucial condition we need to rely on for $\tilde{\mathcal{T}}$ is

$$\frac{1}{nT} \sum_{i=1}^{n} \mathcal{X}_{0,i}^X \left(\hat{F}_0 \right)' \mathcal{X}_{a,i}^X \left(\hat{F}_a \right) \approx \frac{1}{nT} \sum_{i=1}^{n} \mathcal{X}_{0,i}^X \left(\hat{F}_0 \right)' \mathcal{X}_{0,i}^X \left(\hat{F}_0 \right)$$

which requires $\hat{F}_0 \approx \hat{F}_a$. However, even when the null is true, this approximation tends to be poor due to the estimation errors in $\hat{F}_0$ and $\hat{F}_a$. As shown in Proposition A1, the estimator $\hat{F}$ exhibits a slower convergence rate than $\hat{\beta}$, so it is too optimistic to rely on this approximation. When this approximation does not work well, $\tilde{\mathcal{T}}$ will suffer from poor finite sample properties. In contrast, $\mathcal{T}$ does not impose such a restriction in $\hat{V}_{\mathcal{T}}$ to accommodate estimation uncertainty in $\hat{F}_0$ and $\hat{F}_a$. Another advantage of using $\hat{V}_{\mathcal{T}}$ over $\tilde{V}_{\mathcal{T}}$ is that the former is positive semi-definite by construction, which is an important property for the practical use of variance estimators. $\tilde{V}_{\mathcal{T}}$ may not yields a positive semi-definite estimate.

The asymptotics of $\mathcal{T}$ under the null and alternative hypotheses are characterized as follows.

Theorem 4 *Suppose that Assumptions 1-8, 9(ii) and 10 hold. If $\mathcal{H}_0$ is true, then*

$$\mathcal{T} \to^d \chi^2 (d_\beta).$$

If $\mathcal{H}_a$ is true and $\hat{\beta}_0 \to^p \beta^0 + \Delta$ with $\Delta \neq 0$, then for any $C > 0$

$$P(\mathcal{T} > C) \to 1.$$

5 Policy endogeneity with respect to ε_{it}: IFE-GMM approach

The validity of the LS estimation discussed thus far crucially relies on the assumption that the regressors are exogenous with respect to ε_{it}. This condition may not be met in empirical applications. For example, simultaneity often appears in policy analysis between the policy variable and outcome, and in this case endogeneity persists even when the group interactive fixed effects are introduced. To address this, we suppose that a vector of instruments, Ψ_i, is available and consider a moment condition based GMM approach. We refer to this as the "interactive fixed effects GMM (IFE-GMM)" estimation. This approach includes the LS estimator in Section 2 as a special case with $\Psi_i = X_i$.

We first assume that Ψ_i satisfies the following conditions.

Assumption 11 *(i) $E(\varepsilon_{it}|\Psi_1, ..., \Psi_n) = 0$ for all i and t. (ii) rank$\left(Q_{nT}^{X\Psi}(F)\right) = d_x$ and rank$\left(B_{nT}^{X\Psi}(F)\right) = d_x$ for any $F \in \mathcal{F}$ and $d_\Psi \geq d_x$.*

Assumption 11 states strict exogeneity of Ψ_i and rank conditions. The IFE-GMM approach has a larger set of potential instruments than the one based on the model without interactive terms. In the latter case, the exogeneity condition is required to hold not only for ε_{it} but also for $\lambda_{g_i}^{0\prime} F_t^0$. This can be restrictive in empirical studies, because usual instruments in the policy analysis literature tend to be correlated with group characteristics. See Besley and Case (2000) for further discussion. They use the fraction of female legislators in state lower and upper houses as an instrument for the manual rate to study the impact of state workers' compensation benefits on the employment and earnings of construction workers. It is natural to expect such a political variable to be correlated with unobserved state characteristics.

The IFE-GMM estimator is given by

$$\left(\hat{F}(\beta), \hat{\Lambda}(\beta)\right) = \underset{(F,\Lambda)}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} (Y_i - X_i\beta - F\lambda_g)' (Y_i - X_i\beta - F\lambda_g) \tag{25}$$

and

$$\hat{\beta}_{gmm}(F, \Lambda) = \underset{\beta}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left[\Psi_i' (Y_i - X_i\beta - F\lambda_g)\right]' \Omega_{nT}^{-1} \times \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left[\Psi_i' (Y_i - X_i\beta - F\lambda_g)\right], \tag{26}$$

where Ω_{nT} is a positive definite $(d_\Psi \times d_\Psi)$ weighting matrix. Note that while $\hat{\beta}_{gmm}$ is obtained via the GMM criterion based on the moment conditions in Assumption 11, the fixed effects F^0

and Λ^0 are estimated via the LS criterion and principal component method. We have $\hat{\lambda}_g(\beta, F) = F'\left(\bar{Y}_g - \bar{X}_g\beta\right)/T$ from (25). Plugging this into (26), we have the GMM estimator given by

$$\begin{aligned}\hat{\beta}_{gmm} &= \underset{\beta}{\text{argmin}} \sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[\left(\Psi_i - P_{\hat{F}_{gmm}}\bar{\Psi}_g\right)'\left(Y_i - P_{\hat{F}_{gmm}}\bar{Y}_g - \left(X_i - P_{\hat{F}_{gmm}}\bar{X}_g\right)\beta\right)\right]'\Omega_{nT}^{-1} \\ &\quad\times \sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_i - P_{\hat{F}_{gmm}}\bar{\Psi}_g\right)'\left(Y_i - P_{\hat{F}_{gmm}}\bar{Y}_g - \left(X_i - P_{\hat{F}_{gmm}}\bar{X}_g\right)\beta\right) \\ &= \left[Q_{nT}^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_{nT}^{-1}Q_{nT}^{\Psi X}\left(\hat{F}_{gmm}\right)\right]^{-1}Q_{nT}^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_{nT}^{-1}Q_{nT}^{\Psi Y}\left(\hat{F}_{gmm}\right), \end{aligned} \tag{27}$$

where $\hat{F}_{gmm}$ satisfies

$$\frac{1}{nT}\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)'\hat{F}_{gmm} = \hat{F}_{gmm}\hat{\Gamma}_{gmm}. \tag{28}$$

$\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)$ is defined in (10) and $\hat{\Gamma}_{gmm}$ is a diagonal matrix of the d_F largest eigenvalues of $(nT)^{-1}\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)'$. As in LS estimation, we can obtain $\hat{\beta}_{gmm}$ by iterating (27) and (28) to convergence. If $d_\Psi = d_x$, we have

$$\hat{\beta}_{gmm} = Q_{nT}^{\Psi X}\left(\hat{F}_{gmm}\right)^{-1}Q_{nT}^{\Psi Y}\left(\hat{F}_{gmm}\right)$$

and it is easy to see that $\hat{\beta}_{gmm} = \hat{\beta}$ with $\Psi_i = X_i$.

As pointed out by Moon, Shum and Weidner (2017), a drawback of the GMM approach is that it can lead to a local minimum, and this is in contrast to their LS-MD estimator. To minimize the risk of falsely choosing a local minimum as the solution to (27), we should conduct the iterations of (27) and (28) using several initial values. The potential local minima problem is not the unique issue for this method. The LS estimator has the same problem.

We introduce additional assumptions to establish the asymptotics for $\hat{\beta}_{gmm}$.

Assumption 12 *(i)* $E\|\Psi_{it}\|^4 \le M$. *(ii)* $\Omega_{nT} \to^p \Omega$ *and* Ω *is positive definite.*

The theorem below states the consistency of the IFE-GMM estimator.

Theorem 5 *Under Assumptions 1-4(i), 5 and 11-12, we have*

$$\hat{\beta}_{gmm} - \beta^0 \to^p 0 \text{ and } \frac{1}{\sqrt{T}}\left\|\hat{F}_{gmm} - F^0 H\right\| \to^p 0,$$

where $H_{gmm} = \left(\Lambda^{0\prime}\Lambda^0/n\right)\left(F^{0\prime}\hat{F}_{gmm}/T\right)\hat{\Gamma}_{gmm}^{-1}$ *as* $(G, n, T) \to \infty$.

The following high level assumptions are made to obtain the asymptotic normality of $\hat{\beta}_{gmm}$

Assumption 13

$$\lim_{n,T\to\infty}\frac{1}{GnT}\sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{s=1}^{T}n_g\left|E\left(\Psi_{it}'\bar{\varepsilon}_{gt}\Psi_{is}\bar{\varepsilon}_{gs}\right)\right| < M,$$

$$\lim_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|E\left(\Psi_{it}\varepsilon_{it}\Psi_{js}'\varepsilon_{js}\right)\right\|<M,$$

$$\lim_{n,T\to\infty}\frac{1}{GT}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\sum_{t=1}^{T}\sum_{s=1}^{T}\sqrt{n_g n_{\tilde{g}}}\left\|E\left(\bar{\Psi}_{gt}\bar{\varepsilon}_{gt}\bar{\Psi}_{\tilde{g}s}'\bar{\varepsilon}_{\tilde{g}s}\right)\right\|<M.$$

Assumption 14 *Let* $\mathcal{X}_i^{\Psi}=\mathcal{X}_i^{\Psi}\left(F^0\right)$. *We have*

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\mathcal{X}_i^{\Psi}\right)'\varepsilon_i\to^d N\left(0,V_{gmm}\right),$$

where $V_{gmm}=\text{plim}_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[\mathcal{X}_i^{\Psi}\left(F^0\right)'\varepsilon_i\varepsilon_j'\mathcal{X}_j^{\Psi}\left(F^0\right)\right]$ *is positive definite.*

The asymptotic normality of $\hat{\beta}_{gmm}$ is presented as follows.

Theorem 6 *Under Assumptions 1-6 and 11-14, we have*

$$\sqrt{nT}\left(\hat{\beta}_{gmm}-\beta^0\right)\to^d N\left(0,\left(Q_{X\Psi}\Omega^{-1}B_{\Psi X}\right)^{-1}Q_{X\Psi}\Omega^{-1}V_{gmm}\Omega^{-1}Q_{\Psi X}\left(B_{X\Psi}\Omega^{-1}Q_{\Psi X}\right)^{-1}\right),$$

where

$$Q_{X\Psi}=\underset{n,T\to\infty}{\text{plim}}\,Q_{nT}^{X\Psi}\left(F^0\right),\ Q_{\Psi X}=Q_{X\Psi}',\ B_{\Psi X}=\underset{n,T\to\infty}{\text{plim}}\,B_{nT}^{\Psi X}.$$

The proof is in the appendix. We can make inference on β^0 based on Theorem 6. This procedure is analogous to that for the LS estimator in Section 3 and is here omitted to save space.

Regarding the choice of Ω_{nT}, it is well known in the literature that the asymptotic variance of the sample moments is optimal in a standard GMM framework (Hansen , 1982) that minimizes the asymptotic variance of the GMM estimator. This optimality scheme, however, does not apply to our method. The estimation error in $\hat{F}_{gmm}$ affects the asymptotic variance of $\hat{\beta}_{gmm}$ via $B_{\Psi X}$, so the choice of $\Omega_{nT}=V_{gmm}$ does not yield the usual variance form of the efficient GMM estimator. It will be interesting to study the optimal Ω_{nT} in IFE-GMM estimation and we leave this to future research. In Section 7, we set $\Omega_{nT}=Q_{nT}^{\Psi\Psi}\left(\hat{F}_{gmm}\right)$ for empirical illustration, which is motivated from 2SLS estimation.

6 Monte Carlo simulation

This section reports simulation results on the finite sample properties of the proposed estimation and tests in the panel and multilevel regression model:

$$Y_{it}=\beta_Z^0 Z_{g_i t}+\beta_W^0 W_{it}+u_{it},\ i=1,...,n,\ t=1,...,T,\ g=1,...,G, \tag{29}$$

where $Z_{g_i t}$ is a scalar group level regressor and W_{it} is a scalar individual regressor.

We first examine the performance of our method when u_{it} includes multiplicative terms of group fixed effects and time effects as well as individual fixed effects. We generate u_{it} in (29) based on the following DGP:

$$u_{it} = \lambda_{g_i}^{0\prime} F_t^0 + \gamma_i^0 + \varepsilon_{it}, \tag{30}$$

$$\lambda_g^0 \sim^{iid} U(0,1), \ \gamma_i^0 \sim^{iid} U(0,1), \tag{31}$$

$$F_t^0 = \rho_F F_{t-1}^0 + \sqrt{1-\rho_F^2} v_t^F, \text{ with } F_1^0, v_t^F \sim^{iid} N(0, I_{d_F}), \tag{32}$$

$$\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it-1} + \sqrt{1-\rho_\varepsilon^2} v_{it}^\varepsilon, \text{ with } \varepsilon_{i1}, v_{it}^\varepsilon \sim^{iid} N(0,1). \tag{33}$$

We set $d_F = 2$, and Z_{g_it} and W_{it} are generated from the following processes:

$$Z_{gt} = \lambda_g^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \lambda_{1,g}^0 + \lambda_{2,g}^0 + v_{gt}^Z, \text{ with } v_{gt}^Z \sim^{iid} N(0,1), \tag{34}$$

$$W_{it} = \lambda_{g_i}^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \lambda_{1,g_i}^0 + \lambda_{2,g_i}^0 + v_{it}^W, \text{ with } v_{it}^W \sim^{iid} N(0,1) \tag{35}$$

Both the regressors are correlated with F_t^0, λ_g^0 and $\lambda_g^{0\prime} F_t^0$. We address the individual fixed effects in the procedure. Let $\ddot{w}_i = w_{it} - T^{-1}\sum_{t=1}^T w_{it}$ for some random variable w. Using within transformation, we have

$$\ddot{Y}_{it} = \beta_Z^0 \ddot{Z}_{g_it} + \beta_W^0 \ddot{W}_{it} + \lambda_{g_i}^{0\prime} \ddot{F}_t^0 + \ddot{\varepsilon}_{it}$$

and can estimate $\left(\beta_Z^0, \beta_W^0\right)$ using the estimation procedure in Section 2. We set $\beta_Z^0 = \beta_W^0 = 0$, and the number of replications is 5000. Through this section, the number of individuals is the same in each group. i.e., $n_g = n/G$ for $g = 1, ..., G$.

Table 1 presents the bias, standard deviation (SD) and empirical rejection probability (ERP) at the 5% level with $(\rho_F, \rho_\varepsilon) = (0.5, 0.5)$. From the table, we first observe that, when the number of interactive terms used in estimation is the same or larger than the number of interactive terms in the DGP, our method performs very well. In contrast, when our regression model includes only one interactive term $\left(d_{\tilde{F}} = 1\right)$, it does not yield valid estimation and inference results. Moon and Weidner (2015) show that the limiting distribution of the standard interactive fixed effects estimator is independent from the number of interactive terms in the regression model as long as this number is not smaller than the true number of interactive terms. Our simulation result implies that our estimator may have the same property. Table 1 also indicates that our procedure produces a relatively large bias and poor size property as G becomes larger given n and T. For example, when $G = 250$ with $(n, T) = (1000, 5)$ and $\tilde{d}_F = 2$, the bias and ERP of our estimator of β_Z^0 are 0.017 and 0.229, while they are 0.002 and 0.078 when $G = 50$ with $(n, T) = (1000, 5)$ and $\tilde{d}_F = 2$. This is expected, since, as G becomes larger given n, our group interactive estimator gets close to the standard interactive estimator, which is asymptotically biased when $\rho_\varepsilon \neq 0$ and $T/n \to 0$ (Bai, 2009). Our estimation and inference become more accurate as n and T increase.

Table 2 compares our method with the standard interactive fixed effects approach. For this simulation, we generate data based on (29)-(35) as the first simulation experiment, but we set $\rho_\varepsilon = 0$, under which the standard interactive fixed effects estimator is asymptotically unbiased. Group based clustering variance estimation is used to calculate the standard errors. Though both estimators are asymptotically valid, the difference in their finite sample performances is not trivial. We observe that the standard interactive fixed effects estimator tends to be much less accurate than our estimator when T is short. For example, when $(n, T) = (1000, 5)$, the bias and SD of this estimator are 0.017 and 0.037 respectively, which is substantially larger than those of the group interactive effects estimator at 0.001 and 0.024. The ERP is 0.186 for the former, while it is 0.085 for the latter.

Table 3 reports the finite sample properties of the proposed test on the level of grouping. We have two candidate group structures:

$$\mathbb{A}_0 = \{\mathcal{A}_1, ..., \mathcal{A}_g, ..., \mathcal{A}_{G_0}\},$$
$$\mathbb{A}_a = \{\mathcal{A}_1^{(1)}, \mathcal{A}_1^{(2)}, ..., \mathcal{A}_g^{(1)}, \mathcal{A}_g^{(2)}, ..., \mathcal{A}_{G_0}^{(1)}, \mathcal{A}_{G_0}^{(2)}\}.$$

We first examine the size property. In this simulation, we generate data based on (29)-(35) using $\mathbb{A}_0$. The group structure for the group fixed effects is set to be identical to the one for the group level regressor. The table presents ERPs at a 5% nominal level and we can see that the size is well controlled.

We also investigate the power of the test. To simulate the power, the group fixed effects are generated based on the finer level of grouping $\mathbb{A}_a$, while the group level regressor is still based on $\mathbb{A}_0$. We generate the correlation of the regressors with the fixed effects based on

$$Z_{gt} = \sum_{\ell=1}^{2} \lambda_{g(\ell)}^{(\ell)\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \sum_{\ell=1}^{2} \left(\lambda_{1,g}^{(\ell)} + \lambda_{2,g}^{(\ell)}\right) + v_{gt}^Z,$$
$$W_{it} = \sum_{\ell=1}^{2} \lambda_g^{(\ell)\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \sum_{\ell=1}^{2} \left(\lambda_{1,g_i}^{(\ell)} + \lambda_{2,g_i}^{(\ell)}\right) + v_{it}^W,$$

where $v_{gt}^Z, v_{it}^W \sim^{iid} N(0,1)$. $\lambda_g^{(\ell)} = \left[\lambda_{1,g}^{(\ell)}, \lambda_{2,g}^{(\ell)}\right]'$ denote the group fixed effects for $\mathcal{A}_g^{(\ell)}$ and are generated by

$$\lambda_{1,g}^{(1)}, \lambda_{2,g}^{(1)} \sim^{iid} \delta U(0,1), \text{ and } \lambda_{1,g}^{(2)} = -\lambda_{1,g}^{(1)}, \lambda_{2,g}^{(2)} = -\lambda_{2,g}^{(1)}. \quad (36)$$

The DGP (36) implies that $\mathcal{A}_g^{(1)}$ and $\mathcal{A}_g^{(2)}$ respond to the time effects in the opposite direction and δ represents the degree of heterogeneity of the group fixed effects between these two subgroups. Table 3 shows nontrivial power against the alternatives based on (36). The power monotonically rises as the degree of heterogeneity between $\mathcal{A}_g^{(1)}$ and $\mathcal{A}_g^{(2)}$ grows.

Table 4 compares the finite sample performances of the LS and IFE-GMM estimator in the presence of endogeneity with respect to ε_{it}. We generate data from (29)–(33) and employ

$$\Psi_{gt}^Z = \lambda_g^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \lambda_{1,g}^0 + \lambda_{2,g}^0 + v_{gt}^{\Psi 1}, \text{ with } v_{gt}^{\Psi 1} \sim^{iid} N(0,1),$$
$$\Psi_{it}^W = \lambda_{g_i}^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \lambda_{1,g_i}^0 + \lambda_{2,g_i}^0 + v_{it}^{\Psi 2}, \text{ with } v_{it}^{\Psi 2} \sim^{iid} N(0,1),$$

and

$$Z_{gt} = \Psi_{gt}^Z + 0.5\bar{\varepsilon}_{gt} + v_{gt}^Z, \text{ with } v_{gt}^Z \sim^{iid} N(0,1), \quad (37)$$
$$W_{it} = \Psi_{it}^W + 0.5\varepsilon_{it} + v_{it}^W, \text{ with } v_{it}^W \sim^{iid} N(0,1), \quad (38)$$

to generate instruments $\Psi_{g_i t}^Z$ and Ψ_{it}^W and regressor endogeneity in Z_{gt} and W_{it} with respect to ε_{it}. The table shows that the LS estimator suffers from a serious bias and size distortion problem in the presence of endogeneity. The IFE-GMM addresses this substantially by producing accurate estimates and ERPs.

7 Empirical illustration: competition policy and productivity growth

There is broad consensus in economics that competition tends to enhance economic efficiency, but there is no such agreement on the effectiveness of competition policy. For example, Baker (2003) and Werden (2003) argue that the benefit of antitrust enforcement for the economy outweighs the cost, while Crandall and Winston (2003) claim that antitrust law has been ineffective in the US. In this regard, BCDSV (2013) examine the importance of competition policy for improving productivity growth. Using the panel and multilevel data, they provide empirical evidence that the effect of competition policy on the total factor productivity (TFP) growth is positive and significant. As a measure of competition policy, they construct the Competition Policy Indicator (CPI) that summarizes the key features of a country's competition policy.

We revisit the evidence provided by BCDSV (2013) using the group interactive fixed effects approach. The regression model is given by

$$\Delta TFP_{it} = \beta_0 CPI_{g_i t-1} + \beta_1 \Delta TFP_{L(i),t} + \beta_2 \frac{TFP_{L(i),t}}{TFP_{it}} + \beta_3 W_{it-1} + \beta_4 Z_{g_i t-1} + \lambda'_{g_i} F_t + \alpha_i + \varepsilon_{it}, \quad (39)$$

where ΔTFP_{it} is the TFP growth of industry i in country g_i at time t, CPI_{gt} is the CPI of country g at time t. Thus, β_0 is the parameter of interest that represents the effect of country level competition policy on country-industry specific TFP growth. $\Delta TFP_{L(i),t}$ and $\left(TFP_{L(i),t}/TFP_{it}\right)$ denote the technology transfer from a technological frontier country and the productivity gap to the technological frontier respectively. W_{it} is a vector of country-industry specific covariates including trade openness and country-industry specific trends, and $Z_{g_i t}$ denotes country specific product market regulation (prm). Including individual fixed effects, α_i, (39) also controls for unobserved country-industry level heterogeneity.

The model is estimated based on the 1995-2005 balanced panel data. It consists of 22 industries in 7 countries (Czech Republic, Germany, Italy, Japan, Sweden, UK and US).[1] BCDSV (2013) include 5 more countries (Canada, France, Hungary, Netherlands and Spain), but we exclude them to obtain the balanced panel. Note that the omission of these 5 countries yields only a mild change in additive fixed effects estimates.[2] We follow ISIC Rev.3 for industry classification.

We first conduct the test for the appropriate level of grouping to specify the group fixed effects. The grouping scheme under the null, $\mathbb{A}_0$, is at country level, so resulting in 7 groups containing 22 industries each. The finer grouping under the alternative, $\mathbb{A}_a$, results in two subgroups in each country divided between the manufacturing and non-manufacturing sectors. According to the ISIC Rev 3, 12 out of 22 industries belong to the manufacturing sector, while the other 10 belong to the non-manufacturing sector. The result is presented in the table below.

<Test on the level of grouping>

$\mathcal{H}_0$: Country level of grouping is correctly specified.				
$\mathcal{H}_a$: A finer level of grouping based on manufacturing in each country is correctly specified.				
Number of interactive terms (d_F)	1	2	3	4
$\mathcal{T}$	12.03	11.89	6.48	3.82
critical value	$\chi^2_{0.95}(6) = 12.59$			

[1] The data is available at https://dataverse.harvard.edu/dataverse/restat

[2] The additive fixed effects estimates in Tables 5 and 6 are based on the same model specifications used for column 4 of Table 2 and column 1 of Table 4 in BCDSV (2013).

As presented in the table, the test does not reject the null hypothesis with various choice of d_F at 5% level. According to this result, we set the group fixed effects at country level. Thus, in this application, we have $n = 154$, $T = 10$, and $G = 7$.

We estimate (39) using the group interactive fixed effects LS estimation. For comparison, we also consider the following additive fixed effects regression model as employed by BCDSV (2013):

$$\Delta TFP_{it} = \beta_0 CPI_{g_i t-1} + \beta_1 \Delta TFP_{L(i),t} + \beta_2 \frac{TFP_{L(i),t}}{TFP_{it}} + \beta_3 W_{it-1} + \beta_4 Z_{g_i t-1} + \alpha_i + f_t + \varepsilon_{it}.$$

Table 5 reports the coefficients and their t-statistics based on the country based clustering standard errors. We observe that the magnitude of coefficients for CPI obtained from our LS estimation are substantially smaller than that from the additive fixed effects method. The former are between 0.029 and 0.044 when $d_F = 2{\sim}4$, while the latter is 0.095. Comparison of t-statistics shows us a similar result. Thus, we may conclude that the significance of the coefficient for CPI in the additive fixed effects model is due to endogeneity associated with the time varying effect of country heterogeneity and country specific impact of time effects. Table 6 reports the coefficients and their t-statistics using IFE-GMM and 2SLS estimation based on the additive fixed effects model. We use the political variables developed by Cusack and Fuchs (2002) as instruments. They include Market regulation (per403), Economic planning (per404), Welfare state limitations planning (per505) and European Community (per108). BCDSV (2013) also use them as instruments. The qualitative results are the same as the LS estimation case. Compared to the additive fixed effects estimation, the magnitude of the coefficient for CPI and the degree of its significance are substantially reduced when the interactive effects are employed with various choices of d_F.

8 Conclusion

The panel and multilevel regression model is very useful for studying the effects of group level policies on individual outcomes. In this setting, researchers often employ the additive fixed effects regression to allow for correlation between the policy variable and unobserved group heterogeneity/time effects. A shortcoming of this approach is that its validity crucially depends on the assumption that the group heterogeneity is time invariant and that the time effects are common across groups. However, this assumption may not be met in many applications. This paper proposes the group interactive fixed effects model in the multilevel setting and establishes the asymptotics. This approach accounts for group specific impact of time effects as well as time varying effect of group heterogeneity. The group structure helps address the incidental parameters problem. To decide on the appropriate group structure, we propose a Hausman type test that compares two different levels of grouping. We also consider the GMM method to address the endogeneity with respect to the idiosyncratic error.

Table 1: Mean, Standard Deviation and ERP of group interactive fixed effects estimators

			$\beta_W^0 = 0$			$\beta_Z^0 = 0$		
			$(\rho_F, \rho_\varepsilon) = (0.5, 0.5), d_F = 2$					
	n	T	Bias	SD	ERP	Bias	SD	ERP
	1000	5	0.043	0.038	0.595	0.058	0.050	0.617
$G = 50$	1000	10	0.061	0.038	0.904	0.069	0.044	0.897
$\tilde{d}_F = 1$	2000	5	0.041	0.035	0.694	0.055	0.047	0.663
	2000	10	0.058	0.037	0.952	0.066	0.042	0.935
	1000	5	0.001	0.013	0.055	0.002	0.018	0.078
$G = 50$	1000	10	0.001	0.010	0.061	0.001	0.011	0.075
$\tilde{d}_F = 2$	2000	5	0.000	0.009	0.063	-0.000	0.013	0.078
	2000	10	0.000	0.007	0.054	-0.000	0.008	0.070
	1000	5	0.000	0.013	0.054	0.001	0.024	0.106
$G = 50$	1000	10	0.000	0.010	0.060	0.000	0.011	0.082
$\tilde{d}_F = 3$	2000	5	0.000	0.009	0.059	0.000	0.017	0.108
	2000	10	0.000	0.007	0.052	0.000	0.008	0.079
	1000	5	0.010	0.019	0.164	0.017	0.029	0.229
$G = 250$	1000	10	0.009	0.013	0.191	0.011	0.015	0.228
$\tilde{d}_F = 2$	2000	5	0.003	0.011	0.088	0.006	0.015	0.120
	2000	10	0.002	0.007	0.078	0.003	0.008	0.089
	1000	5	0.002	0.015	0.070	0.006	0.024	0.087
$G = 250$	1000	10	0.004	0.011	0.089	0.005	0.012	0.123
$\tilde{d}_F = 3$	2000	5	0.001	0.010	0.054	0.002	0.016	0.072
	2000	10	0.001	0.007	0.063	0.002	0.007	0.068

Table 2: Comparison between the group interactive fixed effects estimator and individual interactive fixed effects estimator

		$\beta_W^0 = 0$			$\beta_Z^0 = 0$		
		$(\rho_f, \rho_\varepsilon) = (0.5, 0.0), G = 50, d_F = 2$					
n	T	Bias	SD	ERP	Bias	SD	ERP
		Group Interactive FE ($\tilde{d}_F = 2$)					
1000	5	0.001	0.016	0.059	0.001	0.024	0.085
1000	10	0.000	0.011	0.060	0.000	0.012	0.070
1000	20	0.000	0.007	0.053	0.000	0.008	0.065
2000	5	0.000	0.011	0.061	-0.000	0.017	0.075
2000	10	0.000	0.008	0.055	0.000	0.009	0.067
2000	20	0.000	0.005	0.060	0.000	0.006	0.068
		Standard Interactive FE ($\tilde{d}_F = 2$)					
1000	5	0.017	0.033	0.189	0.017	0.037	0.186
1000	10	0.004	0.013	0.098	0.004	0.014	0.103
1000	20	0.001	0.008	0.065	0.001	0.008	0.076
2000	5	0.008	0.020	0.140	0.008	0.023	0.135
2000	10	0.002	0.009	0.072	0.002	0.009	0.086
2000	20	0.001	0.006	0.060	0.001	0.006	0.075

Table 3: Test about the level of grouping for group fixed effects

n	T	Empirical Size	
		$G_0 = 50, G_a = 200, d_{\tilde{F}} = 2$	$G_0 = 50, G_a = 200, d_{\tilde{F}} = 3$
1000	5	0.098	0.081
1000	10	0.038	0.051
2000	5	0.058	0.048
2000	10	0.021	0.040
		$G_0 = 50, G_a = 100, d_{\tilde{F}} = 2$	$G_0 = 50, G_a = 100, d_{\tilde{F}} = 3$
1000	5	0.055	0.055
1000	10	0.023	0.039
2000	5	0.038	0.033
2000	10	0.020	0.037
		Power	
		$G_0 = 50, G_a = 100, d_{\tilde{F}} = 2$	
		$\delta = 1$	$\delta = 10$
1000	5	0.145	0.394
1000	10	0.154	0.315
2000	5	0.154	0.420
2000	10	0.149	0.321
		$\delta = 20$	$\delta = 30$
1000	5	0.526	0.603
1000	10	0.475	0.589
2000	5	0.553	0.637
2000	10	0.478	0.609

Table 4: Comparison between the LS estimator and IFE-GMM estimator in the presence of endogeneity with respect to idiosyncratic errors

		$\beta_W^0 = 0$			$\beta_Z^0 = 0$		
		$(\rho_f, \rho_\varepsilon) = (0.5, 0.5)\,, G = 50, d_F = 2$					
n	T	Bias	SD	ERP	Bias	SD	ERP
		Group Interactive FE ($\tilde{d}_F = 2$)					
1000	5	0.008	0.012	0.153	0.011	0.014	0.220
1000	10	0.010	0.008	0.290	0.011	0.008	0.361
2000	5	0.004	0.008	0.105	0.005	0.009	0.132
2000	10	0.005	0.005	0.179	0.006	0.006	0.215
		Group Interactive FE ($\tilde{d}_F = 3$)					
1000	5	0.004	0.012	0.082	0.007	0.013	0.156
1000	10	0.007	0.007	0.188	0.009	0.008	0.285
2000	5	0.002	0.008	0.069	0.004	0.009	0.111
2000	10	0.004	0.005	0.123	0.005	0.005	0.188
		IFE-GMM ($\tilde{d}_F = 2$)					
1000	5	0.003	0.021	0.131	0.003	0.022	0.137
1000	10	0.001	0.012	0.090	0.001	0.012	0.094
2000	5	0.001	0.014	0.096	0.001	0.014	0.101
2000	10	0.000	0.008	0.078	0.000	0.008	0.073
		IFE-GMM ($\tilde{d}_F = 3$)					
1000	5	0.001	0.019	0.101	0.001	0.019	0.098
1000	10	0.000	0.011	0.082	0.000	0.011	0.091
2000	5	0.000	0.013	0.096	0.000	0.013	0.092
2000	10	0.000	0.008	0.084	0.000	0.008	0.080

Table 5: The effect of competition policy on TFP growth: LS estimation

Dependent Variable	ΔTFP_{it}				
	Group Interactive FE				Additive FE
d_F	1	2	3	4	
$CPI_{g_i t-1}$	0.010	0.035	0.029	0.044	0.095
	(0.377)	(1.356)	(1.421)	(1.876)	(2.616)
$\Delta TFP_{L(i),t}$	0.077	0.078	0.076	0.075	0.072
	(2.042)	(2.055)	(2.015)	(1.932)	(2.786)
$(TFP_{L(i),t}/TFP_{it})$	0.014	0.014	0.014	0.014	0.013
	(4.557)	(4.578)	(4.413)	(4.274)	(5.804)
Industry trend	0.104	0.097	0.099	0.099	0.097
	(9.728)	(9.044)	(8.278)	(8.417)	(3.146)
Import penetration	0.014	0.015	0.015	0.016	0.016
	(1.755)	(1.832)	(1.803)	(1.907)	(2.882)
pmr	0.010	0.012	0.013	0.010	−0.033
	(1.982)	(2.373)	(2.799)	(1.167)	(−1.483)

The numbers in parentheses represent t statistics based on country level clustered standard errors.

Table 6: The effect of competition policy on TFP growth: IV estimation

Dep Var	ΔTFP_{it}				
	IFE-GMM				Additive FE 2SLS
d_F	1	2	3	4	
$CPI_{g_i t-1}$	0.203	0.178	0.122	0.174	0.472
	(1.064)	(1.180)	(1.163)	(1.547)	(1.913)
$\Delta TFP_{L,it}$	0.067	0.069	0.069	0.065	0.069
	(1.780)	(1.767)	(1.806)	(1.702)	(1.874)
$(TFP_{L(i),t}/TFP_{it})$	0.014	0.014	0.014	0.014	0.014
	(4.499)	(4.505)	(4.263)	(4.274)	(4.334)
Industry trend	0.108	0.096	0.096	0.097	0.108
	(8.962)	(9.555)	(9.436)	(8.417)	(6.457)
Import penetration	0.013	0.014	0.014	0.015	0.016
	(1.579)	(1.705)	(1.702)	(1.907)	(1.590)
pmr	0.020	0.023	0.021	0.024	−0.045
	(1.800)	(2.398)	(3.469)	(1.167)	(−1.568)

The numbers in parentheses represent t statistics based on country level clustered standard errors.

9 Appendix

Proof of Theorem 1. We first consider the rate condition that $(G,n,T)\to\infty$. Suppose that $\beta^0=0$ without loss of generality. $\left(\hat{\beta},\hat{F}\right)$ minimizes

$$\tilde{\mathcal{Q}}(\beta,F)=\mathcal{Q}(\beta,F)-\frac{1}{nT}\sum_{i=1}^{n}\left[\left(\varepsilon_i-P_{F^0}\bar{\varepsilon}_{g_i}\right)'\left(\varepsilon_i-P_{F^0}\bar{\varepsilon}_{g_i}\right)\right]$$

as the second term of the right hand side does not depend on (β,F). Let

$$\begin{aligned}\mathcal{Q}^*(\beta,F)&=\frac{1}{nT}\sum_{i=1}^{n}\lambda_{g_i}^{0\prime}F^{0\prime}M_F F^0\lambda_{g_i}^0-\frac{2}{nT}\sum_{i=1}^{n}\lambda_{g_i}^{0\prime}F^{0\prime}M_F\left(X_i-P_F\bar{X}_{g_i}\right)\beta\\&+\beta'\frac{1}{nT}\sum_{i=1}^{n}\left(X_i-P_F\bar{X}_{g_i}\right)'\left(X_i-P_F\bar{X}_{g_i}\right)\beta\end{aligned}$$

Note that given H, $M_{F^0H}=M_{F^0}$ and

$$\mathcal{Q}^*(\beta^0,F^0H)=0.$$

The first step is to show that

$$\begin{aligned}&\tilde{\mathcal{Q}}(\beta,F)-\mathcal{Q}^*(\beta,F)\\&=\frac{1}{nT}\sum_{i=1}^{n}\left[2\lambda_{g_i}^{0\prime}F^{0\prime}M_F\left(\varepsilon_i-P_F\bar{\varepsilon}_{g_i}\right)-2\left(\varepsilon_i'X_i-\bar{\varepsilon}_{g_i}'P_F\bar{X}_{g_i}\right)\beta-\bar{\varepsilon}_{g_i}'\left(P_F-P_{F^0}\right)\bar{\varepsilon}_{g_i}\right]\\&=o_p(1)\end{aligned}\tag{A.1}$$

for all bounded β and $F\in\mathcal{F}=\{F:F'F/T=I_{d_F}\}$ as $(G,n,T)\to\infty$. For the first term, note that $\sum_{i=1}^{n}\lambda_{g_i}^{0\prime}F^{0\prime}M_F\varepsilon_i=\sum_{g=1}^{G}n_g\lambda_g^{0\prime}F^{0\prime}M_F\bar{\varepsilon}_g$. We have

$$\begin{aligned}&\left|\frac{1}{nT}\sum_{i=1}^{n}\lambda_{g_i}^{0\prime}F^{0\prime}M_F\left(\varepsilon_i-P_F\bar{\varepsilon}_{g_i}\right)\right|\\&\leq\left|\frac{1}{nT}\sum_{g=1}^{G}\sum_{t=1}^{T}n_g\lambda_g^{0\prime}F_t^0\bar{\varepsilon}_{gt}\right|+\left|\frac{1}{n}\sum_{g=1}^{G}n_g\left(\frac{1}{T}\sum_{t=1}^{T}\lambda_g^{0\prime}F_t^0F_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T}F_t\bar{\varepsilon}_{gt}\right)\right|\\&=a1+a2.\end{aligned}$$

It is easy to show that $a1=O_p\left((nT)^{-1/2}\right)$ under Assumptions 2, 3(ii), 4(i) and 5. For $a2$,

$$\begin{aligned}a2&\leq\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\|\lambda_{g_i}^{0\prime}F_t^0F_t'\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{g=1}^{G}n_g\left\|\frac{1}{T}\sum_{t=1}^{T}F_t\bar{\varepsilon}_{gt}\right\|^2\right)^{1/2}\\&=O_p(1)\left(\frac{1}{n}\sum_{g=1}^{G}n_g\left\|\frac{1}{T}\sum_{t=1}^{T}F_t\bar{\varepsilon}_{gt}\right\|^2\right)^{1/2},\end{aligned}$$

and

$$P\left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\frac{1}{T}\sum_{t=1}^{T} F_t \bar{\varepsilon}_{gt}\right\|^2 > \Delta\right) \leq \frac{1}{\Delta}\frac{G}{nT}\sum_{\ell=1}^{d_F}\left(\frac{1}{GT}\sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{s=1}^{T} n_g E\left(F_{t\ell}F_{s\ell}\right)E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{gs}\right)\right)$$
$$= O\left(\frac{G}{nT}\right).$$

Thus, $a1$ and $a2$ are $o_p(1)$ as $(G, n, T) \to \infty$. Using the same procedure, we can show that the second and third terms in (A.1) are also $o_p(1)$. Therefore, (A.1) holds.

The second step is to show

$$\mathcal{Q}^*(\beta, F) > 0 \tag{A.2}$$

for any $(\beta, F) \neq (\beta^0, F^0 H)$, and the proof of consistency in Bai (2009, Proposition 1) can directly apply here. Therefore, $\mathcal{Q}^*(\beta, F) \geq 0$ and $\mathcal{Q}^*(\beta, F) > 0$ if $(\beta, F) \neq (\beta^0, F^0 H)$, which completes the proof of Part (i).

As for consistency under the asymptotics that $(G, n) \to \infty$ such that $G/n \to 0$ for fixed T, we can follow the same steps above and only need to show (A.1), which is straightforward. ■

Proposition A1 *Suppose that Assumptions 1-5 hold. Then, we have*

$$\frac{1}{\sqrt{T}}\left\|\hat{F} - F^0 H\right\| = O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n\sqrt{T}}\right)$$

where $H = \left(\Lambda^{0\prime}\Lambda^0/n\right)\left(F^{0\prime}\hat{F}/T\right)\hat{\Gamma}^{-1}$ *as* $(G, n, T) \to \infty$.

Proof. From (11), we have

$$\begin{aligned}
&\hat{F}\hat{\Gamma} - F^0\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right) \\
&= \frac{1}{nT}\sum_{g=1}^{G} n_g \bar{X}_g\left(\beta^0 - \hat{\beta}\right)\left(\beta^0 - \hat{\beta}\right)'\bar{X}_g'\hat{F} \\
&+ \frac{1}{nT}\sum_{g=1}^{G} n_g \bar{X}_g\left(\beta^0 - \hat{\beta}\right)\lambda_g' F^{0\prime}\hat{F} + \frac{1}{nT}\sum_{g=1}^{G} n_g \bar{X}_g\left(\beta^0 - \hat{\beta}\right)\bar{\varepsilon}_g'\hat{F} \\
&+ \frac{1}{nT}\sum_{g=1}^{G} n_g F^0\lambda_g\left(\beta^0 - \hat{\beta}\right)'\bar{X}_g'\hat{F} + \frac{1}{nT}\sum_{g=1}^{G} n_g \bar{\varepsilon}_g\left(\beta^0 - \hat{\beta}\right)'\bar{X}_g'\hat{F} \\
&+ \frac{1}{nT}\sum_{g=1}^{G} n_g F^0\lambda_g^0\bar{\varepsilon}_g'\hat{F} + \frac{1}{nT}\sum_{g=1}^{G} n_g\bar{\varepsilon}_g\lambda_g^{0\prime}F^{0\prime}\hat{F} + \frac{1}{nT}\sum_{g=1}^{G} n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\hat{F} \\
&= I1 + I2 + \cdots + I8
\end{aligned} \tag{A.3}$$

We multiply $\left(F^{0\prime}\hat{F}/T\right)^{-1}\left(\Lambda^{0\prime}\Lambda^{0}/n\right)^{-1}$ to obtain

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|\hat{F}\hat{\Gamma}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{n}\right)^{-1}-F^{0}\right\| \\
&\leq\frac{1}{\sqrt{T}}\left(\|I1\|+\cdots+\|I8\|\right)\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{n}\right)^{-1}\right\|.
\end{aligned}$$

For $I1$,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\|I1\| &\leq \frac{1}{nT}\sum_{g=1}^{G}\sum_{t=1}^{T}n_g\left\|\bar{X}_{gt}\right\|^2\left\|\beta^0-\hat{\beta}\right\|^2\sqrt{d_F} \\
&= O_p\left(\left\|\hat{\beta}-\beta^0\right\|^2\right).
\end{aligned}$$

Using similar procedures, we can show that $\|I2\|=\cdots=\|I5\|=O_p\left(\left\|\beta^0-\hat{\beta}\right\|\right).$

For $I6$, we have

$$\begin{aligned}
\frac{1}{\sqrt{T}}\|I6\| &\leq \frac{1}{\sqrt{n}}\left\|\frac{1}{\sqrt{GT}}\sum_{g=1}^{G}\sum_{t=1}^{T}\sqrt{n_g}\lambda_g^0\bar{\varepsilon}_{gt}\right\|\left(\frac{\|F^0\|}{\sqrt{T}}\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)O(1) \\
&= O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

In the same way, we can show that $T^{-1/2}\|I7\|=O_p\left(n^{-1/2}\right).$

For $I8$, we have

$$\begin{aligned}
\frac{1}{\sqrt{T}}\|I8\| &\leq \frac{1}{\sqrt{T}}\left\|\frac{1}{nT}\sum_{g=1}^{G}n_g\left[\bar{\varepsilon}_g\bar{\varepsilon}_g'-E\left(\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\right]\hat{F}\right\|+\frac{1}{\sqrt{T}}\left\|\frac{1}{nT}\sum_{g=1}^{G}n_gE\left(\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\hat{F}\right\| \\
&= \frac{1}{\sqrt{T}}I81+\frac{1}{\sqrt{T}}I82.
\end{aligned}$$

For the first term,

$$\begin{aligned}
\frac{1}{T}\|I81\|^2 &= \frac{1}{n^2T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\sum_{\tau=1}^{T}n_gn_{\tilde{g}}\left[\bar{\varepsilon}_{g\tau}\bar{\varepsilon}_{gt}-E\left(\bar{\varepsilon}_{g\tau}\bar{\varepsilon}_{gt}\right)\right]\left[\bar{\varepsilon}_{\tilde{g}\tau}\bar{\varepsilon}_{\tilde{g}s}-E\left(\bar{\varepsilon}_{\tilde{g}\tau}\bar{\varepsilon}_{\tilde{g}s}\right)\right]\sum_{\ell=1}^{d_F}\hat{F}_t^{(\ell)}\hat{F}_s^{(\ell)}\right] \\
&\leq O_p\left(\frac{G}{n^2}\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{1}{GT}\sum_{\tau=1}^{T}\sum_{g=1}^{G}n_g\left[\bar{\varepsilon}_{g\tau}\bar{\varepsilon}_{gt}-E\left(\bar{\varepsilon}_{g\tau}\bar{\varepsilon}_{gt}\right)\right]\sum_{\tilde{g}=1}^{G}n_{\tilde{g}}\left[\bar{\varepsilon}_{\tilde{g}\tau}\bar{\varepsilon}_{\tilde{g}s}-E\left(\bar{\varepsilon}_{\tilde{g}\tau}\bar{\varepsilon}_{\tilde{g}s}\right)\right]\right)^2\right)^{1/2}.
\end{aligned}$$

Let $\varsigma_{g,ts} = n_g \left[\bar{\varepsilon}_{gt} \bar{\varepsilon}_{gs} - E\left(\bar{\varepsilon}_{gt} \bar{\varepsilon}_{gs} \right) \right]$. Since

$$
\begin{aligned}
& P\left[\left(\frac{1}{GT} \sum_{\tau=1}^{T} \sum_{g=1}^{G} n_g \left[\bar{\varepsilon}_{g\tau} \bar{\varepsilon}_{gt} - E\left(\bar{\varepsilon}_{g\tau} \bar{\varepsilon}_{gt} \right) \right] \sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \left[\bar{\varepsilon}_{\tilde{g}\tau} \bar{\varepsilon}_{\tilde{g}s} - E\left(\bar{\varepsilon}_{\tilde{g}\tau} \bar{\varepsilon}_{\tilde{g}s} \right) \right] \right)^2 > \Delta \right] \\
& \leq E\left[\frac{1}{G^2T^2} \sum_{\tau_1=1}^{T} \left(\sum_{g_1=1}^{G} n_{g_1} \left[\bar{\varepsilon}_{g_1\tau_1} \bar{\varepsilon}_{g_1t_1} - E\left(\bar{\varepsilon}_{g_1\tau_1} \bar{\varepsilon}_{g_1t_1} \right) \right] \right) \left(\sum_{g_2=1}^{G} n_{g_2} \left[\bar{\varepsilon}_{g_2\tau_1} \varepsilon_{g_2s_1} - E\left(\bar{\varepsilon}_{g_2\tau_1} \varepsilon_{g_2s_1} \right) \right] \right) \right. \\
& \left. \times \sum_{\tau_2=1}^{T} \left(\sum_{g_3=1}^{G} n_{g_3} \left[\bar{\varepsilon}_{g_3\tau_2} \bar{\varepsilon}_{g_3t_2} - E\left(\bar{\varepsilon}_{g_3\tau_2} \bar{\varepsilon}_{g_3t_2} \right) \right] \right) \left(\sum_{g_4=1}^{G} n_{g_4} \left\{ \bar{\varepsilon}_{g_4\tau_2} \bar{\varepsilon}_{g_4s_2} - E\left(\bar{\varepsilon}_{g_4\tau_2} \bar{\varepsilon}_{g_4s_2} \right) \right\} \right) \right] \\
& = E\left[\frac{1}{G^2T^2} \sum_{\tau_1=1}^{T} \sum_{\tau_2=1}^{T} \left(\sum_{g_1=1}^{G} \varsigma_{g_1,\tau_1t_1} \right) \left(\sum_{g_2=1}^{G} \varsigma_{g_2,\tau_1s_1} \right) \left(\sum_{g_3=1}^{G} \varsigma_{g_3,\tau_2t_2} \right) \left(\sum_{g_4=1}^{G} \varsigma_{g_4,\tau_2s_2} \right) \right] \\
& \leq \max_{t,s} E\left(\frac{1}{\sqrt{G}} \sum_{g=1}^{G} \varsigma_{g,ts} \right)^4,
\end{aligned}
$$

we have $T^{-1/2} \|I81\| = O_p\left(\sqrt{G}/n \right)$ under Assumption 3(vi). For $I82$,

$$
\begin{aligned}
\frac{1}{\sqrt{T}} \|I82\| &\leq \frac{G}{n\sqrt{T}} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[\frac{1}{G} \sum_{g=1}^{G} n_g E\left(\bar{\varepsilon}_{gt} \bar{\varepsilon}_{gs} \right) \right]^2} \frac{\left\| \hat{F} \right\|}{\sqrt{T}} \\
&= O\left(\frac{G}{n\sqrt{T}} \right).
\end{aligned}
$$

Therefore, $T^{-1/2} \|I82\| = O\left(G/\left(n\sqrt{T} \right) \right)$, and

$$
\frac{1}{\sqrt{T}} \|I8\| = O_p\left(\frac{\sqrt{G}}{n} + \frac{G}{nT} \right). \tag{A.4}
$$

Combining $I1$-$I8$, we obtain

$$
\begin{aligned}
& \frac{1}{\sqrt{T}} \left(\hat{F} \hat{\Gamma} \left(\frac{F^{0\prime} \hat{F}}{T} \right)^{-1} \left(\frac{\Lambda^{0\prime} \Lambda^0}{n} \right)^{-1} - F^0 \right) \\
& = O_p\left(\left\| \hat{\beta} - \beta^0 \right\| \right) + O_p\left(\frac{1}{\sqrt{n}} \right) + O_p\left(\frac{G}{n\sqrt{T}} \right).
\end{aligned} \tag{A.5}
$$

Premultiplying $\hat{F}'/\sqrt{T}$ in (A.5), we obtain

$$
\hat{\Gamma} = \left(\frac{\hat{F}' F^0}{T} \right) \left(\frac{\Lambda^{0\prime} \Lambda^0}{n} \right) \left(\frac{F^{0\prime} \hat{F}}{T} \right) + o_p(1). \tag{A.6}
$$

As shown in Bai (2009), $F^{0\prime} \hat{F}/T$ is invertible, so $\hat{\Gamma}$ is invertible. Thus, from (A.5) we have

$$
\frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| = O_p\left(\left\| \hat{\beta} - \beta^0 \right\| \right) + O_p\left(\frac{1}{\sqrt{n}} \right) + O_p\left(\frac{G}{n\sqrt{T}} \right).
$$

■

Lemma A1 *Under Assumptions 1-5, for each g*

$$\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}=o_p\left(\left\|\beta^0-\hat{\beta}\right\|\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{1}{\sqrt{nG}}\right)+O_p\left(\frac{1}{T}\right)$$

as $(G,n,T)\to\infty$.

The proof is in the supplementary appendix.

Lemma A2 *Under Assumptions 1-5,*

$$HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}=O_p\left(\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{G}{nT}\right).$$

The proof is in the supplementary appendix.

Proposition A2 *Let*

$$A_{nT}^{(1)}=\frac{1}{nT}\sum_{g=1}^{G}n_g\bar{X}_g'M_{F^0}\left(\frac{1}{G}\sum_{\tilde{g}=1}^{G}n_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}'\right)F^0H\Upsilon\lambda_g^0$$

with $\Upsilon=\left(F^{0\prime}\hat{F}/T\right)^{-1}\left(\Lambda^{0\prime}\Lambda^0/n\right)^{-1}$. *Under Assumptions 1-7,*

$$\begin{aligned}\sqrt{nT}\left(\hat{\beta}-\beta^0\right)=&\,B_{nT}^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\left(X_i-P_{\hat{F}}\bar{X}_{g_i}\right)-\frac{1}{n}\sum_{j=1}^{n}a_{ij}^0M_{\hat{F}}X_j\right)'\varepsilon_i\\&+\frac{G}{\sqrt{nT}}B_{nT}^{XX}\left(\hat{F}\right)^{-1}A_{nT}^{(1)}+O_p\left(\frac{\sqrt{T}}{n}\right)+o_p\left(\frac{G}{\sqrt{nT}}\right)+O_p\left(\frac{\sqrt{G}}{n}\right)\\&+o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta\right\|\right).\end{aligned}$$

Proof. Note that

$$\begin{aligned}\sqrt{nT}\left(\hat{\beta}-\beta^0\right)=&\left[\frac{1}{nT}\sum_{i=1}^{n}\left(X_i-P_{\hat{F}}\bar{X}_{g_i}\right)'\left(X_i-P_{\hat{F}}\bar{X}_{g_i}\right)\right]^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\Bigg(X_i'M_{\hat{F}}F^0\lambda_{g_i}^0\\&+\left(X_i-P_{\hat{F}}\bar{X}_{g_i}\right)'\varepsilon_i\Bigg)\end{aligned}\tag{A.7}$$

Let $\Upsilon=\left(F^{0\prime}\hat{F}/T\right)^{-1}\left(\Lambda^{0\prime}\Lambda^0/n\right)^{-1}$. For the first term of (A.7), using $M_{\hat{F}}\hat{F}=0$, we have

$$\begin{aligned}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}X_i'M_{\hat{F}}F^0\lambda_{g_i}^0&=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}X_i'M_{\hat{F}}\left[F^0-\hat{F}\hat{\Gamma}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1}\right]\lambda_{g_i}^0\\&=-\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}X_i'M_{\hat{F}}\left(I1+\cdots+I8\right)\Upsilon\lambda_{g_i}^0\\&:=J1+\cdots+J8\end{aligned}$$

For $J1$,

$$\|J1\| \leq \sqrt{nT}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\|X_i'\|^2}{T}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_{g_i}\|^2\right)^{1/2}\frac{1}{\sqrt{T}}\|I1\|\|\Upsilon\|$$
$$= o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta\right\|\right) \tag{A.8}$$

in which we use

$$\left\|X_i'M_{\hat{F}}\right\|^2 = tr\left(X_i'X_i\right) - tr\left(X_i'\hat{F}\left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'X_i\right)$$
$$= \|X_i\|^2 - \frac{1}{T}\left\|\hat{F}'X_i\right\|^2 \leq \left\|X_i'\right\|^2. \tag{A.9}$$

For $J2$, we have

$$J2 = \frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}X_i'M_{\hat{F}}X_j\left\{\lambda_{g_j}^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1}\lambda_{g_i}^0\right\}\sqrt{nT}\left(\hat{\beta}-\beta^0\right)$$
$$= \frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^0X_i'M_{\hat{F}}X_j\sqrt{nT}\left(\hat{\beta}-\beta^0\right), \tag{A.10}$$

where

$$\left\|\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}X_iM_{\hat{F}}X_j\left\{\lambda_{g_j}^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1}\lambda_{g_i}^0\right\}\right\|$$
$$\leq \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\|X_i'\|}{\sqrt{T}}\left\|\lambda_{g_i}^0\right\|\right)^2\left\|\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1}\right\| = O_p(1).$$

For $J3$,

$$\|J3\| \leq \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\|X_i'\|^2}{T}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_{g_i}\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{g=1}^{G}\left\|\frac{\sqrt{n_g}\bar{\varepsilon}_g'\hat{F}}{T}\right\|^2\right)^{1/2}$$
$$\times\left\|\sqrt{nT}\left(\beta-\hat{\beta}\right)\right\|\|\Upsilon\|,$$

where the equality holds because $\left(\bar{\varepsilon}_g'\hat{F}/T\right)\Upsilon\lambda_{g_i}$ is a scalar. Note that

$$\frac{\bar{\varepsilon}_g'\hat{F}}{T} = \frac{1}{\sqrt{T}}\frac{\bar{\varepsilon}_g'F^0H}{\sqrt{T}} + \frac{\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}$$

and

$$P\left(\frac{1}{n}\sum_{g=1}^{G}\left\|\frac{\sqrt{n_g}\bar{\varepsilon}_g'F^0H}{\sqrt{T}}\right\|^2 > \Delta\right) \leq \frac{1}{nT}\sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{s=1}^{T}n_g\left|E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{gs}\right)\right|\left|E\left(F_t^{0\prime}F_s^0\right)\right|\|H\|^2$$
$$= o(1).$$

Using a similar way, we can show that

$$\frac{1}{n}\sum_{g=1}^{G}\left\|\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\right\| = o_p\left(\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{1}{\sqrt{nG}}\right)+O_p\left(\frac{1}{T}\right). \quad (A.11)$$

Thus, $\|J3\| = o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta^0\right\|\right)$.

For $J4$,

$$\begin{aligned}\|J4\| &\le \left(\frac{1}{nT}\sum_{i=1}^{n}\left\|X_i'\right\|^2\right)^{1/2}\frac{\left\|F^0-\hat{F}H^{-1}\right\|}{\sqrt{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\lambda_{g_i}^0\right\|^2\right)\left(\frac{1}{n}\sum_{j=1}^{n}\left\|\frac{X_j'\hat{F}}{T}\right\|^2\right)^{1/2}\left\|\sqrt{nT}\left(\hat{\beta}-\beta^0\right)\right\|\|\Upsilon\| \\ &= o_p\left(\left\|\sqrt{nT}\left(\beta^0-\hat{\beta}\right)\right\|\right).\end{aligned}$$

For $J5$,

$$\begin{aligned}\|J5\| &\le \left\|\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}X_{it}\bar{\varepsilon}_{g_jt}\sqrt{nT}\left(\beta^0-\hat{\beta}\right)'\frac{1}{T}\sum_{s=1}^{T}\bar{X}_{g_js}\hat{F}_s'\Upsilon\lambda_{g_i}^0\right\| \\ &+\left\|\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}X_i'\frac{\hat{F}\hat{F}'}{T}\bar{\varepsilon}_{g_j}\sqrt{nT}\left(\beta^0-\hat{\beta}\right)'\frac{1}{T}\sum_{s=1}^{T}\bar{X}_{g_js}\hat{F}_s'\Upsilon\lambda_{g_i}^0\right\| \\ &= J51+J52.\end{aligned}$$

$$\begin{aligned}J51 &\le \frac{G}{n\sqrt{T}}\left(\frac{1}{nG}\sum_{i=1}^{n}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_g}X_{it}\bar{\varepsilon}_{gt}\right\|^2\right)^{1/2}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\frac{1}{T}\sum_{s=1}^{T}\bar{X}_{g_js}\hat{F}_s'\Upsilon\lambda_{g_i}^0\right\|^2\right)^{1/2} \\ &\times\left\|\sqrt{nT}\left(\beta^0-\hat{\beta}\right)\right\| \\ &= o_p\left(\left\|\sqrt{nT}\left(\beta^0-\hat{\beta}\right)\right\|\right),\end{aligned}$$

because

$$\begin{aligned}P\left(\frac{1}{nG}\sum_{i=1}^{n}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_g}X_{it}\bar{\varepsilon}_{gt}\right\|^2>\Delta\right) &\le \frac{1}{nGT}\sum_{\ell=1}^{d_x}\sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{s=1}^{T}n_g\left|E\left(X_{it}^{(\ell)}\bar{\varepsilon}_{gt}X_{is}^{(\ell)}\bar{\varepsilon}_{gs}\right)\right| \\ &= O(1).\end{aligned}$$

by Assumption 7. Using a similar way, we can show that $J52 = o_p\left(\left\|\sqrt{nT}\left(\beta^0-\hat{\beta}\right)\right\|\right)$.

For $J6$, since $M_{\hat{F}}\hat{F}=0$,

$$\begin{aligned}\|J6\| &\le \sqrt{nT}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\left\|X_i'\right\|^2}{T}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\lambda_{g_i}^0\right\|^2\right)^{1/2}\left\|\frac{1}{\sqrt{T}}\left(F^0-\hat{F}H^{-1}\right)\right\| \\ &\times\left\|\frac{1}{nT}\sum_{j=1}^{n}\lambda_{g_j}^0\varepsilon_j'\hat{F}\right\|\|\Upsilon\|\end{aligned}$$

and

$$\frac{1}{nT}\sum_{j=1}^{n}\lambda_{g_j}^{0}\varepsilon_j'\hat{F}=\frac{1}{\sqrt{nT}}\left(\frac{1}{\sqrt{GT}}\sum_{g=1}^{G}\sum_{t=1}^{T}\sqrt{n_g}\lambda_g^0\bar{\varepsilon}_{gt}F_t^0H\right)+\frac{1}{nT}\sum_{g=1}^{G}n_g\lambda_g^0\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)$$

$$=O_p\left(\frac{1}{\sqrt{n}}\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{\sqrt{nT}}\right) \tag{A.12}$$

because

$$\left\|\frac{1}{nT}\sum_{g=1}^{G}n_g\lambda_g^0\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)\right\|\leq O\left(\frac{1}{\sqrt{n}}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\sqrt{n_g}\lambda_g^0\bar{\varepsilon}_{gt}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t-H'F_t^0\right\|^2\right)^{1/2}$$

$$=O_p\left(\frac{1}{\sqrt{n}}\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{n}\right)+O_p\left(\frac{G}{n\sqrt{nT}}\right).$$

Therefore,

$$\|J6\|=O_p\left(\sqrt{T}\left\|\hat{\beta}-\beta^0\right\|^2\right)+O_p\left(\frac{\sqrt{T}}{n}\right)+O_p\left(\frac{G}{n\sqrt{T}}\right)+O_p\left(\frac{G}{n\sqrt{n}}\right).$$

For $J7$, since a_{ij}^0 is a scalar,

$$J7=-\frac{1}{n\sqrt{nT}}\sum_{i=1}^{n}\sum_{j=1}^{n}X_i'M_{\hat{F}}\varepsilon_j\lambda_{g_j}^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1}\lambda_{g_i}^0$$

$$=-\frac{1}{n\sqrt{nT}}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^0X_i'M_{\hat{F}}\varepsilon_j.$$

Let

$$A_{nT}^{(1)}=\frac{1}{nT}\sum_{i=1}^{n}X_i'M_{F^0}\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)F^0H\Upsilon\lambda_{g_i}^0.$$

Then, $J8$ can be rewritten as

$$\begin{aligned}J8=&-\frac{G}{\sqrt{nT}}A_{nT}^{(1)}\\&-\frac{G}{nT\sqrt{nT}}\sum_{i=1}^{n}X_i'\left(M_{\hat{F}}-M_{F^0}\right)\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\left(\hat{F}-F^0H\right)\Upsilon\lambda_{g_i}^0\\&-\frac{G}{nT\sqrt{nT}}\sum_{i=1}^{n}X_i'\left(M_{\hat{F}}-M_{F^0}\right)\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)F^0H\Upsilon\lambda_{g_i}^0\\&-\frac{G}{nT\sqrt{nT}}\sum_{i=1}^{n}X_i'M_{F^0}\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\left(\hat{F}-F^0H\right)\Upsilon\lambda_{g_i}^0\\&=J81+J82+J83+J84.\end{aligned}$$

For $J81$,

$$\begin{aligned}\|J81\| &\le \frac{G}{\sqrt{nT}}\left\|\frac{1}{nT}\sum_{i=1}^{n}X_i'\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)F^0H\Upsilon\lambda_{g_i}^0\right\| \\ &+\frac{G}{\sqrt{nT}}\left\|\frac{1}{nT}\sum_{i=1}^{n}X_i'F^0\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)F^0H\Upsilon\lambda_{g_i}^0\right\| \\ &= a11+a12,\end{aligned}$$

where

$$\begin{aligned}a11 &\le \frac{G}{\sqrt{nT}}\left(\frac{1}{nG}\sum_{i=1}^{n}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_g}X_{it}\bar{\varepsilon}_{gt}\right\|^2\right)^{1/2} \\ &\times\left(\frac{1}{G}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\sqrt{n_g}\bar{\varepsilon}_{gs}F_s^{0\prime}\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\lambda_{g_i}^0\right\|^2\right)^{1/2}\|H\|\,\|\Upsilon\| \\ &= O_p\left(\frac{G}{\sqrt{nT}}\right).\end{aligned}$$

We can show that $a12$ is also $O_p\left(G/\sqrt{nT}\right)$ using the same steps.

$$\begin{aligned}J82 &= \left(\frac{G}{\sqrt{n}}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{X_i'}{\sqrt{T}}\left(\frac{\hat{F}-F^0H}{\sqrt{T}}\right)H'\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0\sqrt{n_g}\bar{\varepsilon}_g\right)\left(\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\right)\Upsilon\lambda_{g_i}^0 \\ &+\left(\frac{G\sqrt{T}}{\sqrt{n}}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{X_i'}{\sqrt{T}}\frac{\left(\hat{F}-F^0H\right)}{\sqrt{T}}\frac{1}{G}\sum_{g=1}^{G}\frac{\left(\hat{F}-F^0H\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T}\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\Upsilon\lambda_{g_i}^0 \\ &+\left(\frac{G\sqrt{T}}{\sqrt{n}}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{X_i'F^0H}{T}\frac{1}{G}\sum_{g=1}^{G}\frac{\left(\hat{F}-F^0H\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T}\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\Upsilon\lambda_{g_i}^0 \\ &+\left(\frac{G}{\sqrt{n}}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{X_i'F^0}{T}\left[HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0\sqrt{n_g}\bar{\varepsilon}_{gt}\right)\left(\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\right)\Upsilon\lambda_{g_i}^0 \\ &= o_p\left(\sqrt{nT}\left\|\beta^0-\hat{\beta}\right\|\right)+O_p\left(\frac{G}{T\sqrt{nT}}\right)+O_p\left(\frac{\sqrt{T}}{n\sqrt{n}}\right)+O_p\left(\frac{\sqrt{G}}{n\sqrt{n}}\right)+O_p\left(\frac{G}{nT}\right)+O_p\left(\frac{\sqrt{G}}{n\sqrt{T}}\right)\end{aligned}$$

For $J83$,

$$\|J83\| = \left\| \frac{G}{nT\sqrt{nT}} \sum_{i=1}^{n} X_i' \hat{F} \left(\frac{\hat{F}' - \left(F^0 H\right)'}{T} \right) \left(\frac{1}{G} \sum_{g=1}^{G} n_g \bar{\varepsilon}_g \bar{\varepsilon}_g' \right) F^0 H \Upsilon \lambda_{g_i}^0 \right\|$$

$$+ \left\| \frac{G}{nT\sqrt{nT}} \sum_{i=1}^{n} X_i' \left(\hat{F} - F^0 H \right) \frac{\left(F^0 H\right)'}{T} \left(\frac{1}{G} \sum_{g=1}^{G} n_g \bar{\varepsilon}_g \bar{\varepsilon}_g' \right) F^0 H \Upsilon \lambda_{g_i}^0 \right\|$$

$$= o_p \left(\sqrt{nT} \left\| \beta^0 - \hat{\beta} \right\| \right) + O_p \left(\frac{G}{n\sqrt{T}} \right) + O_p \left(\frac{\sqrt{G}}{n} \right) + O_p \left(\frac{G}{T\sqrt{n}} \right)$$

For $J84$

$$J84 = \frac{G}{nT\sqrt{nT}} \sum_{i=1}^{n} X_i' \left(\frac{1}{G} \sum_{g=1}^{G} n_g \bar{\varepsilon}_g \bar{\varepsilon}_g' \right) \left(\hat{F} - F^0 H \right) \Upsilon \lambda_{g_i}^0$$

$$+ \frac{G}{nT\sqrt{nT}} \sum_{i=1}^{n} X_i' \frac{F^0 H \left(F^0 H\right)'}{T} \left(\frac{1}{G} \sum_{g=1}^{G} n_g \bar{\varepsilon}_g \bar{\varepsilon}_g' \right) \left(\hat{F} - F^0 H \right) \Upsilon \lambda_{g_i}^0$$

$$= a41 + a42,$$

$$\|a41\| \leq \frac{G}{\sqrt{n}} \left(\frac{1}{nG} \sum_{i=1}^{n} \sum_{g=1}^{G} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sqrt{n_g} X_{it} \bar{\varepsilon}_{gt} \right\|^2 \left(\frac{1}{G} \sum_{g=1}^{G} \left\| \frac{\sqrt{n_g} \bar{\varepsilon}_g' \left(\hat{F} - F^0 H \right)}{T} \right\|^2 \right) \right)^{1/2}$$

$$\times \|\Upsilon\| \left(\frac{1}{n} \sum_{i=1}^{n} \left\| \lambda_{g_i}^0 \right\|^2 \right)^{1/2}$$

$$= o_p \left(\sqrt{nT} \left\| \beta^0 - \hat{\beta} \right\| \right) + O_p \left(\frac{G}{n\sqrt{T}} \right) + O_p \left(\frac{\sqrt{G}}{n} \right) + O_p \left(\frac{G}{T\sqrt{n}} \right)$$

Using the same steps, we can show the same result for $a42$. Thus,

$$J8 = J81 + o_p \left(\sqrt{nT} \left\| \beta^0 - \hat{\beta} \right\| \right) + O_p \left(\frac{\sqrt{T}}{n\sqrt{n}} \right) + O_p \left(\frac{G}{n\sqrt{T}} \right) + O_p \left(\frac{\sqrt{G}}{n} \right) + O_p \left(\frac{G}{T\sqrt{n}} \right)$$

Combining $J1$-$J8$, we have

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} X_i' M_{\hat{F}} F^0 \lambda_{g_i}^0$$

$$= \left\{ \frac{1}{n^2 T} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^0 X_i' M_{\hat{F}} X_j \right\} \sqrt{nT} \left(\hat{\beta} - \beta^0 \right) - \frac{1}{n\sqrt{nT}} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^0 X_i' M_{\hat{F}} \varepsilon_j + \left(\frac{G}{\sqrt{nT}} \right) A_{nT}^{(1)}$$

$$+ o_p \left(\sqrt{nT} \left\| \hat{\beta} - \beta \right\| \right) + O_p \left(\frac{\sqrt{T}}{n} \right) + O_p \left(\frac{\sqrt{G}}{n} \right) + o_p \left(\frac{G}{\sqrt{nT}} \right) \tag{A.13}$$

Combining this result with (A.7), under the rate conditions in Assumption 5, we have

$$\left(\frac{1}{nT}\sum_{i=1}^{n}\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)'\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right) - \frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^0 X_i' M_{\hat{F}} X_j\right)\sqrt{nT}\left(\hat{\beta}-\beta^0\right)$$
$$= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)' - \frac{1}{n}\sum_{i=1}^{n}a_{ij}^0 X_j' M_{\hat{F}}\right)\varepsilon_i + \frac{G}{\sqrt{nT}}A_{nT}^{(1)}$$
$$+ O_p\left(\frac{\sqrt{T}}{n}\right) + o_p\left(\frac{G}{\sqrt{nT}}\right) + O_p\left(\frac{\sqrt{G}}{n}\right) + o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta\right\|\right)$$

and by premultiplying $B_{nT}^{XX}\left(\hat{F}\right)^{-1}$ we have

$$\sqrt{nT}\left(\hat{\beta}-\beta^0\right) = B_{nT}^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right) - \frac{1}{n}\sum_{j=1}^{n}a_{ij}^0 M_{\hat{F}} X_j\right)'\varepsilon_i$$
$$+ \frac{G}{\sqrt{nT}}B_{nT}^{XX}\left(\hat{F}\right)^{-1}A_{nT}^{(1)} + O_p\left(\frac{\sqrt{T}}{n}\right) + o_p\left(\frac{G}{\sqrt{nT}}\right)$$
$$+ O_p\left(\frac{\sqrt{G}}{n}\right) + o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta\right\|\right).$$

■

Proof of Theorem 2. Let $\mathcal{X}_i = \mathcal{X}_i\left(F^0\right)$ for notational simplicity. From Proposition A2, we have

$$\sqrt{nT}\left(\hat{\beta}-\beta^0\right) = B_{nT}^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i\left(\hat{F}\right)'\varepsilon_i + \frac{G}{\sqrt{nT}}B_{nT}^{XX}\left(\hat{F}\right)^{-1}A_{nT}^{(1)}$$
$$+ O_p\left(\frac{\sqrt{T}}{n}\right) + O_p\left(\frac{\sqrt{G}}{n}\right) + o_p\left(\frac{G}{\sqrt{nT}}\right) + o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta^0\right\|\right).$$

Thus, we need to show

$$\text{(i) } \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i\left(\hat{F}\right)'\varepsilon_i - \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i'\varepsilon_i = o_p(1),$$
$$\text{(ii) } B_{nT}^{XX}\left(\hat{F}\right) - B_{nT}^{XX}\left(F^0\right) = o_p(1)$$

to complete the proof.

Part (i) We have

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i\left(\hat{F}\right)'\varepsilon_i - \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i'\varepsilon_i$$
$$= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\bar{X}_{g_i}'\left(\left(\frac{\hat{F}\hat{F}'}{T}\right) - P_{F^0}\right)\varepsilon_i - \frac{1}{n\sqrt{nT}}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^0 X_j'\left(M_{F^0} - M_{\hat{F}}\right)\varepsilon_i$$
$$= \mathcal{H}1 + \mathcal{H}2,$$

and

$$\mathcal{H}1=\frac{1}{\sqrt{nT}}\sum_{g=1}^{G}\frac{n_g\bar{X}_g'\left(\hat{F}-F^0H\right)}{T}H'F^{0\prime}\bar{\varepsilon}_g+\frac{1}{\sqrt{nT}}\sum_{g=1}^{G}\frac{n_g\bar{X}_g'\left(\hat{F}-F^0H\right)}{T}\left(\hat{F}-F^0H\right)'\bar{\varepsilon}_g$$
$$+\frac{1}{\sqrt{nT}}\sum_{g=1}^{G}\frac{n_g\bar{X}_g'F^0H}{T}\left(\hat{F}-F^0H\right)'\bar{\varepsilon}_g+\frac{1}{\sqrt{nT}}\sum_{g=1}^{G}\frac{n_g\bar{X}_g'F^0}{T}\left[HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]F^{0\prime}\bar{\varepsilon}_g$$
$$=h1+h2+h3+h4.$$

For $h1$,

$$\|h1\|\le\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{g=1}^{G}\bar{X}_{gs}\sqrt{n_g}\bar{\varepsilon}_{gt}F_t^{0\prime}\right\|^2\right)^{1/2}\|H\|\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s-HF_s^0\right\|^2\right)^{1/2}$$
$$=O_p\left(\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\frac{G}{n\sqrt{T}}\right)\tag{A.14}$$

For $h2$,

$$\|h2\|\le\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t-H'F_t^0\right)\right\|^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\sqrt{n_g}\bar{X}_{gt}\bar{\varepsilon}_{gs}\right\|^2\right)^{1/2}$$
$$=O_p\left(\sqrt{T}\left\|\hat{\beta}-\beta^0\right\|^2\right)+O_p\left(\frac{\sqrt{T}}{n}\right)+O_p\left(\frac{G^2}{n^2\sqrt{T}}\right).\tag{A.15}$$

For $h3$

$$\|h3\|=\frac{1}{\sqrt{nT}}\sum_{g=1}^{G}\frac{\bar{X}_g'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\hat{F}H^{-1}-F^0\right)'n_g\bar{\varepsilon}_g$$
$$+\frac{1}{\sqrt{nT}}\sum_{g=1}^{G}\frac{\bar{X}_g'F^0}{T}\left[HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]\left(\hat{F}H^{-1}-F^0\right)'n_g\bar{\varepsilon}_g$$
$$=h31+h32$$

For $h32$, we have

$$h32=\sqrt{GT}\frac{1}{G}\sum_{g=1}^{G}\frac{\bar{X}_g'F^0}{T}\left[HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]\frac{\left(\hat{F}H^{-1}-F^0\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T}O(1)$$
$$=o_p\left(\sqrt{nT}\left\|\beta^0-\hat{\beta}\right\|\right)+O_p\left(\frac{\sqrt{T}}{n\sqrt{n}}\right)+O_p\left(\frac{\sqrt{G}}{n\sqrt{T}}\right)+O_p\left(\frac{1}{n}\right)$$
$$+O_p\left(\frac{\sqrt{G}}{\sqrt{n}T}\right)+O_p\left(\frac{G\sqrt{G}}{nT\sqrt{T}}\right)\tag{A.16}$$

For $h31$

$$
\begin{aligned}
h31 &= \frac{1}{\sqrt{nT}} \sum_{g=1}^{G} \frac{\bar{X}_g' F^0}{T} \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \Upsilon' \left(I1 + \cdots + I8 \right)' n_g \bar{\varepsilon}_g \\
&= \mathcal{K}1 + \cdots + \mathcal{K}8,
\end{aligned}
$$

It is easy to show that $\mathcal{K}1$-$\mathcal{K}5$ are $o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right)$. For $\mathcal{K}6$,

$$
\begin{aligned}
\|\mathcal{K}6\| &= O\left(\sqrt{\frac{G}{n}}\right) \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \frac{1}{\sqrt{GT}} \sum_{g=1}^{G} \sum_{t=1}^{T} \bar{X}_{gs} F_t^{0\prime} \sqrt{n_g} \bar{\varepsilon}_{gt} \right\|^2 \right)^{1/2} \left(\frac{1}{G} \sum_{\tilde{g}=1}^{G} \left\| \frac{\left(\hat{F} - F^0 H\right)' \sqrt{n_{\tilde{g}}} \bar{\varepsilon}_{\tilde{g}}}{T} \right\|^2 \right)^{1/2} \\
&\quad \times \left(\frac{1}{G} \sum_{\tilde{g}=1}^{G} \left\|\lambda_{\tilde{g}}^{0\prime}\right\|^2 \right)^{1/2} \|\Upsilon\| \left\| \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \right\| \left(\frac{1}{T} \sum_{s=1}^{T} \left\|F_s^0\right\|^2 \right)^{1/2} \\
&\quad + O\left(\frac{1}{\sqrt{nT}}\right) \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \frac{1}{\sqrt{GT}} \sum_{g=1}^{G} \sum_{t=1}^{T} \bar{X}_{gs} F_t^{0\prime} \sqrt{n_g} \bar{\varepsilon}_{gt} \right\|^2 \right)^{1/2} \left\| \frac{1}{\sqrt{GT}} \sum_{\tilde{g}=1}^{G} \sum_{\tau=1}^{T} H' F_\tau^0 \sqrt{n_{\tilde{g}}} \bar{\varepsilon}_{\tilde{g}\tau} \lambda_{\tilde{g}}^{0\prime} \right\| \\
&\quad \times \|\Upsilon\| \left\| \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \right\| \left(\frac{1}{T} \sum_{s=1}^{T} \left\|F_s^0\right\|^2 \right)^{1/2} \\
&= o_p\left(\sqrt{nT}\left\|\beta^0 - \hat{\beta}\right\|\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right).
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{K}7 &= \sqrt{\frac{T}{n}} \frac{1}{GT} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} \frac{\bar{X}_g' F^0}{T} \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \left(\frac{\Lambda^{0\prime} \Lambda^0}{n} \right)^{-1} \lambda_{\tilde{g}}^0 \sqrt{n_{\tilde{g}} n_g} \bar{\varepsilon}_{\tilde{g}}' \bar{\varepsilon}_g \\
&\equiv \sqrt{\frac{T}{n}} A_{nT}^{(21)}
\end{aligned}
$$

where

$$
\begin{aligned}
A_{nT}^{(21)} &= \left(\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left\| \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \sqrt{n_g} \bar{\varepsilon}_{gt} \bar{X}_{gs} \right\|^2 \left(\frac{1}{T} \sum_{s=1}^{T} \left\|F_s^{0\prime}\right\|^2 \right) \right)^{1/2} \\
&\quad \times \left\| \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \right\| \left\| \left(\frac{\Lambda' \Lambda}{n} \right)^{-1} \right\| \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{\sqrt{G}} \sum_{\tilde{g}=1}^{G} \lambda_{\tilde{g}} \sqrt{n_{\tilde{g}}} \bar{\varepsilon}_{\tilde{g}t} \right\|^2 \right)^{1/2} \\
&= O_p(1).
\end{aligned}
$$

$$\mathcal{K}8 = O\left(\frac{1}{nT}\right)\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{g=1}^{G}\sum_{t=1}^{T}\sqrt{n_{\tilde{g}}n_g}\left[\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt} - E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt}\right)\right]\frac{\bar{X}_g' F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\Upsilon'\left(\hat{F} - F^0 H\right)'\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}$$
$$+ O\left(\frac{1}{nT}\right)\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{g=1}^{G}\sum_{t=1}^{T}\sqrt{n_{\tilde{g}}n_g}E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt}\right)\frac{\bar{X}_g' F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\Upsilon'\left(\hat{F} - F^0 H\right)'\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}$$
$$+ O\left(\frac{1}{nT}\right)\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{g=1}^{G}\sum_{t=1}^{T}\sqrt{n_g}\bar{\varepsilon}_{gt}\frac{\bar{X}_g' F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\Upsilon' H'\sum_{\tau=1}^{T}F_\tau^0 n_{\tilde{g}}\left[\bar{\varepsilon}_{\tilde{g}\tau}\bar{\varepsilon}_{\tilde{g}t} - E\left(\bar{\varepsilon}_{\tilde{g}\tau}\bar{\varepsilon}_{\tilde{g}t}\right)\right]$$
$$+ O\left(\frac{1}{nT}\right)\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{g=1}^{G}\sum_{t=1}^{T}\sqrt{n_g}\bar{\varepsilon}_{gt}\frac{\bar{X}_g' F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\Upsilon' H'\sum_{\tau=1}^{T}F_\tau^0 n_{\tilde{g}}E\left(\bar{\varepsilon}_{\tilde{g}\tau}\bar{\varepsilon}_{\tilde{g}t}\right)$$
$$= \mathcal{K}81 + \mathcal{K}82 + \mathcal{K}83 + \mathcal{K}84.$$

$$\|\mathcal{K}81\| \le O\left(\frac{G\sqrt{G}}{n}\right)\left(\left(\frac{1}{G^2}\sum_{\tilde{g}=1}^{G}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_{\tilde{g}}n_g}\left[\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt} - E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt}\right)\right]\right\|^2\left(\frac{1}{G}\sum_{g=1}^{G}\left\|\frac{1}{T}\sum_{s=1}^{T}\bar{X}_{gs}F_s^{0\prime}\right\|^2\right)\right)\right)^{1/2}$$
$$\times\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|\left\|\Upsilon'\right\|\left(\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left\|\frac{\left(\hat{F} - F^0 H\right)'\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}}{T}\right\|^2\right)^{1/2}$$
$$= o_p\left(\sqrt{nT}\left\|\beta - \hat{\beta}\right\|\right) + O_p\left(\frac{G\sqrt{G}}{n\sqrt{nT}}\right) + O_p\left(\frac{G}{n\sqrt{n}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right)$$

$$\|\mathcal{K}82\| = O\left(\frac{1}{nT}\right)\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{g=1}^{G}\sum_{t=1}^{T}\sqrt{n_{\tilde{g}}n_g}E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt}\right)\frac{\bar{X}_g' F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\Upsilon'\left(\hat{F} - F^0 H\right)'\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}$$
$$\le O\left(\frac{\sqrt{G}}{n}\right)\left(\frac{1}{GT}\sum_{\tilde{g}=1}^{G}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{g=1}^{G}\sqrt{n_{\tilde{g}}n_g}E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt}\right)\bar{X}_{gs}\right\|^2\right)^{1/2}\left(\sum_{s=1}^{T}\left\|F_s^{0\prime}\right\|^2\right)^{1/2}$$
$$\times\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|\left\|\Upsilon'\right\|\left(\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left\|\frac{\left(\hat{F} - F^0 H\right)'\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}}{T}\right\|^2\right)^{1/2}$$
$$= o_p\left(\sqrt{nT}\left\|\beta - \hat{\beta}\right\|\right) + O_p\left(\frac{\sqrt{G}}{n\sqrt{nT}}\right) + O_p\left(\frac{1}{n\sqrt{n}}\right) + O_p\left(\frac{\sqrt{G}}{nT}\right)$$

$$\|\mathcal{K}83\| \leq O\left(\frac{\sqrt{G}}{n}\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\sqrt{n_g}\bar{\varepsilon}_{gt}\bar{X}_{gs}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|F_s^{0\prime}\right\|^2\right)^{1/2}\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|$$
$$\times\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{\tau=1}^{T}F_\tau^0 n_{\tilde{g}}\left[\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}\tau}-E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}\tau}\right)\right]\right\|^2\right)^{1/2}\left\|\Upsilon'\right\|\left\|H'\right\|$$
$$= O_p\left(\frac{\sqrt{G}}{n}\right)$$

because

$$P\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{\tau=1}^{T}F_\tau^0 n_{\tilde{g}}\left[\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}\tau}-E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}\tau}\right)\right]\right\|^2>\Delta\right)$$
$$\leq \frac{M}{\Delta GT^2}\sum_{\tilde{g}_1=1}^{G}\sum_{\tilde{g}_2=1}^{G}\sum_{t=1}^{T}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T}n_{\tilde{g}_1}n_{\tilde{g}_2}Cov\left(\bar{\varepsilon}_{\tilde{g}_1t}\bar{\varepsilon}_{\tilde{g}_1\tau_1},\bar{\varepsilon}_{\tilde{g}_2t}\bar{\varepsilon}_{\tilde{g}_2\tau_2}\right)$$
$$= O(1)$$

under Assumption 3(vii).

$$\mathcal{K}84 = O\left(\frac{G}{n\sqrt{T}}\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\sqrt{n_g}\bar{\varepsilon}_{gt}\bar{X}_{gs}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|F_s^{0\prime}\right\|^2\right)^{1/2}\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|$$
$$\times\left\|\Upsilon'\right\|\left\|H'\right\|\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{G}\sum_{\tilde{g}=1}^{G}\sum_{\tau=1}^{T}F_\tau^0 n_{\tilde{g}}E\left(\bar{\varepsilon}_{\tilde{g}\tau}\bar{\varepsilon}_{\tilde{g}t}\right)\right\|^2\right)^{1/2}$$
$$= O_p\left(\frac{G}{n\sqrt{T}}\right).$$

Thus,

$$\mathcal{K}8 = o_p\left(\sqrt{nT}\left\|\beta^0-\hat{\beta}\right\|\right)+O_p\left(\frac{G\sqrt{G}}{nT}\right)+O_p\left(\frac{\sqrt{G}}{n}\right)+O_p\left(\frac{G}{n\sqrt{T}}\right).$$

It follows from $\mathcal{K}1$-$\mathcal{K}8$ that

$$h31 = \sqrt{\frac{T}{n}}A_{nT}^{(21)}+o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{\sqrt{G}}{n}\right)$$
$$+O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{G\sqrt{G}}{nT}\right)+O_p\left(\frac{G}{n\sqrt{T}}\right).$$

Combining this with (A.16), we have

$$h3 = \sqrt{\frac{T}{n}} A_{nT}^{(21)} + o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{\sqrt{G}}{n}\right)$$
$$+ O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right) + O_p\left(\frac{G}{n\sqrt{T}}\right). \quad (A.17)$$

For $h4$, we apply Lemma A2 to have

$$\|h4\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{GT}}\sum_{g=1}^{G}\sum_{t=1}^{T}\bar{X}_{gs}F_t^{0\prime}\sqrt{n_g}\bar{\varepsilon}_{gt}\right\|^2\right)^{1/2}\left\|HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|\left(\frac{1}{T}\sum_{s=1}^{T}\left\|F_s^0\right\|^2\right)^{1/2}$$
$$= O_p\left(\left\|\hat{\beta} - \beta\right\|\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{G}{nT}\right)$$

Therefore, we have

$$\mathcal{H}1 = \sqrt{\frac{T}{n}} A_{nT}^{(21)} + o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right) + o_p\left(\sqrt{\frac{T}{n}}\right)$$
$$+ O_p\left(\frac{G}{n\sqrt{T}}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right) \quad (A.18)$$

Let $K_g = \lambda_g' \left(\Lambda'\Lambda/n\right)^{-1}\left(n^{-1}\sum_{j=1}^{n}\lambda_{g_j}X_j\right)$. Replacing $\bar{X}_{g_i}$ with K_g, we can use the same procedure for $\mathcal{H}2$, and we have

$$\mathcal{H}2 = \sqrt{\frac{T}{n}} A_{nT}^{(22)} + o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right) + o_p\left(\sqrt{\frac{T}{n}}\right)$$
$$+ O_p\left(\frac{G}{n\sqrt{T}}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right) \quad (A.19)$$

with

$$A_{nT}^{(22)} = \frac{1}{GT}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\frac{K_g'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{n}\right)^{-1}\lambda_{\tilde{g}}\sqrt{n_{\tilde{g}}n_g}\bar{\varepsilon}_{\tilde{g}}'\bar{\varepsilon}_g$$

From (A.18) and (A.19), we have

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\left(X_i - \left(\frac{\hat{F}\hat{F}'}{T}\right)\bar{X}_{g_i}\right)' - \frac{1}{n}\sum_{j=1}^{n}a_{ij}^0 X_j' M_{\hat{F}}\right)\varepsilon_i$$
$$= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i'\varepsilon_i + \sqrt{\frac{T}{n}}A_{nT}^{(2)} + o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right)$$
$$+ o_p\left(\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{G}{n\sqrt{T}}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right) \quad (A.20)$$

where

$$A_{nT}^{(2)}=\frac{1}{GT}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\frac{\left(\bar{X}_g-K_g\right)F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{n}\right)^{-1}\lambda_{\tilde{g}}\sqrt{n_{\tilde{g}}n_g}\bar{\varepsilon}_{\tilde{g}}'\bar{\varepsilon}_g$$

as $(G,n,T)\to\infty$.
Part (ii) We have

$$\begin{aligned}
&B_{nT}^{XX}\left(\hat{F}\right)-B_{nT}^{XX}\left(F^0\right)\\
&=\frac{1}{nT}\sum_{i=1}^{n}\bar{X}_{g_i}'\left(P_{F^0}-\frac{\hat{F}\hat{F}'}{T}\right)\bar{X}_{g_i}\\
&-\left(\frac{1}{T}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^0X_i'M_{\hat{F}}X_j\right\}-\frac{1}{T}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^0X_i'M_{F^0}X_j\right\}\right)\\
&=\mathcal{G}1+\mathcal{G}2,
\end{aligned}$$

$$\begin{aligned}
\mathcal{G}1=&-\frac{1}{nT}\sum_{i=1}^{n}\frac{\bar{X}_{g_i}'\left(\hat{F}-F^0H\right)}{T}H'F^{0\prime}\bar{X}_{g_i}-\frac{1}{nT}\sum_{i=1}^{n}\frac{\bar{X}_{g_i}'\left(\hat{F}-F^0H\right)}{T}\left(\hat{F}-F^0H\right)'\bar{X}_{g_i}\\
&-\frac{1}{nT}\sum_{i=1}^{n}\frac{\bar{X}_{g_i}'F^0H}{T}\left(\hat{F}-F^0H\right)'\bar{X}_{g_i}-\frac{1}{nT}\sum_{i=1}^{n}\frac{\bar{X}_{g_i}'F^0}{T}\left[HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]F^{0\prime}\bar{X}_{g_i}\\
&=o_p\left(1\right).
\end{aligned}$$

Similarly, we have $\mathcal{G}2=o_p\left(1\right)$. Combining (i) and (ii), we have

$$\begin{aligned}
\sqrt{nT}\left(\hat{\beta}-\beta\right)&=\left(B_{nT}^{XX}\left(F^0\right)+o_p\left(1\right)\right)^{-1}\left[\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i'\varepsilon_i+\frac{G}{\sqrt{nT}}A_{nT}^{(1)}+\sqrt{\frac{T}{n}}A_{nT}^{(2)}\right.\\
&\left.+o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta\right\|\right)+o_p\left(\sqrt{\frac{T}{n}}\right)+o_p\left(\frac{G}{\sqrt{nT}}\right)\right]\\
&=B_{nT}^{XX}\left(F^0\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i'\varepsilon_i+o_p\left(1\right)
\end{aligned}$$

under Assumption 6. ■

Lemma A3 *Under Assumptions 1-5,*

$$\frac{1}{\sqrt{n}}\left(\hat{\Lambda}'-H^{-1}\Lambda'\right)=O_p\left(\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\sqrt{\frac{G}{nT}}\right).$$

The proof is in the supplementary appendix.
Proof of Theorem 3. Part (i) Due to Proposition A2 and Theorem 2, we need to prove

$$\text{(a) }\hat{B}_{nT}^{XX}-B_{nT}^{XX}\left(\hat{F}\right)=o_p\left(1\right),$$

$$\text{(b) }\hat{V}_{nT}^{c}-V_{nT}^{c}=o_p\left(1\right).$$

For (a), we have

$$\begin{aligned}\hat{B}_{nT}^{XX}-B_{nT}^{XX}\left(\hat{F}\right) &= \frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\lambda_{g_i}^{0\prime}\left(H^{-1}\right)'\left(H^{-1}\Lambda^{0\prime}\Lambda^{0}\left(H^{-1}\right)'/n\right)^{-1}\left(H^{-1}\lambda_{g_i}^{0}-\hat{\lambda}_{g_j}\right)\right]X_i'M_{\hat{F}}X_j \\ &+\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\lambda_{g_i}^{0\prime}\left(H^{-1}\right)'\left(\left(H^{-1}\Lambda^{0\prime}\Lambda^{0}\left(H^{-1}\right)'/n\right)^{-1}-\left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1}\right)\hat{\lambda}_{g_j}\right]X_i'M_{\hat{F}}X_j \\ &+\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(\lambda_{g_i}^{0\prime}\left(H^{-1}\right)'-\hat{\lambda}_{g_i}'\right)\left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1}\hat{\lambda}_{g_j}\right]X_i'M_{\hat{F}}X_j \\ &= o_p\left(1\right),\end{aligned}$$

due to $\left(n^2T\right)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|X_i'M_{\hat{F}}X_j\right\|^2=O_p\left(1\right)$ and Lemma A3.

For (b), let

$$\tilde{V}_{nT}^{c}=\frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\mathcal{X}_j^X \text{ and } \ddot{V}_{nT}^{c}=\frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_i^X\right)'\varepsilon_i\varepsilon_j'\hat{\mathcal{X}}_j^X.$$

We have

$$\hat{V}_{nT}^{c}-V_{nT}^{c}=\left(\hat{V}_{nT}^{c}-\ddot{V}_{nT}^{c}\right)+\left(\ddot{V}_{nT}^{c}-\tilde{V}_{nT}^{c}\right)+\left(\tilde{V}_{nT}^{c}-V_{nT}^{c}\right) \tag{A.21}$$

and need to show that each term in the rhs is $o_p\left(1\right)$.

$$\begin{aligned}\hat{V}_{nT}^{c}-\ddot{V}_{nT}^{c} &= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_i^X\right)'\left[X_i\left(\beta^0-\hat{\beta}\right)\left(\beta^0-\hat{\beta}\right)'X_j'\right. \\ &+\left(F^0\lambda_{g_i}^0-\hat{F}\hat{\lambda}_{g_i}\right)\left(\beta^0-\hat{\beta}\right)'X_j'+\varepsilon_i\left(\beta^0-\hat{\beta}\right)'X_j' \\ &\left.+X_i\left(\beta^0-\hat{\beta}\right)\left(F^0\lambda_{g_j}^0-\hat{F}\hat{\lambda}_{g_j}\right)'+\varepsilon_i\left(F^0\lambda_{g_j}^0-\hat{F}\hat{\lambda}_{g_j}\right)'\right]\hat{\mathcal{X}}_j^X \\ &= o_p\left(1\right).\end{aligned} \tag{A.22}$$

For the second term in (A.21),

$$\begin{aligned}\ddot{V}_{nT}^{c}-\tilde{V}_{nT}^{c} &= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_i^X-\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\left(\hat{\mathcal{X}}_j^X-\mathcal{X}_i^X\right) \\ &+\frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_i^X-\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\mathcal{X}_j^X \\ &+\frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\left(\hat{\mathcal{X}}_j^X-\mathcal{X}_j^X\right) \\ &= R1+R2+R3\end{aligned} \tag{A.23}$$

First, we have

$$
\begin{aligned}
R1 &= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\bar{X}_g'\left(P_{F^0}-P_{\hat{F}}\right)\varepsilon_i\varepsilon_j'\left(P_{F^0}-P_{\hat{F}}\right)\bar{X}_g \\
&+ \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\frac{1}{n}\sum_{c=1}^{n}X_c'\left(a_{ic}M_{F^0}-\hat{a}_{ic}M_{\hat{F}}\right)\varepsilon_i\varepsilon_j'\left(P_{F^0}-P_{\hat{F}}\right)\bar{X}_g \\
&+ \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\frac{1}{n}\sum_{c=1}^{n}\bar{X}_g'\left(P_{F^0}-P_{\hat{F}}\right)\varepsilon_i\varepsilon_j'\frac{1}{n}\sum_{c=1}^{n}\left(a_{jc}M_{F^0}-\hat{a}_{jc}M_{\hat{F}}\right)X_c \\
&+ \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\frac{1}{n}\sum_{c=1}^{n}X_c'\left(a_{ic}M_{F^0}-\hat{a}_{ic}M_{\hat{F}}\right)\varepsilon_i\varepsilon_j'\frac{1}{n}\sum_{c=1}^{n}\left(a_{jc}M_{F^0}-\hat{a}_{jc}M_{\hat{F}}\right)X_c \\
&= R11+R12+R13+R14.
\end{aligned}
$$

$R11 = o_p(1)$ because

$$
R11 = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{T}}\bar{X}_g'\left(P_{F^0}-P_{\hat{F}}\right)\sqrt{n_g}\bar{\varepsilon}_g\right)\left(\frac{1}{\sqrt{T}}\sqrt{n_g}\bar{\varepsilon}_j'\left(P_{F^0}-P_{\hat{F}}\right)\bar{X}_g\right)
$$

and for each g

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\bar{X}_g'\left(P_{\hat{F}}-P_{F^0}\right)\sqrt{n_g}\bar{\varepsilon}_g \\
&= \frac{\bar{X}_g'\left(\hat{F}-F^0H\right)}{T}H'\frac{F^{0\prime}\sqrt{n_g}\bar{\varepsilon}_g}{\sqrt{T}} + \sqrt{T}\frac{\bar{X}_g'\left(\hat{F}-F^0H\right)}{T}\frac{\left(\hat{F}-F^0H\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T} \\
&+ \sqrt{T}\frac{\bar{X}_g'F^0H}{T}\frac{\left(\hat{F}-F^0H\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T} + \frac{\bar{X}_g'F^0}{T}\left(HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right)\frac{F^{0\prime}\sqrt{n_g}\bar{\varepsilon}_g}{\sqrt{T}} \\
&= o_p(1) \qquad (A.24)
\end{aligned}
$$

as $(G,n,T)\to\infty$.

For $R12$, let $a_{g\tilde{g}} = a_{ic} = \lambda_{g_i}^{0\prime}\left(\Lambda^{0\prime}\Lambda^0/n\right)^{-1}\lambda_{g_c}^0$ and $\hat{a}_{g\tilde{g}} = \hat{a}_{ic} = \hat{\lambda}_{g_i}'\left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1}\hat{\lambda}_{g_c}$ respectively if $i\in\mathcal{A}_g$ and $c\in\mathcal{A}_{\tilde{g}}$. Then,

$$
\begin{aligned}
R12 &= \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{T}}\left(\frac{1}{n}\sum_{\tilde{g}=1}^{G}a_{g\tilde{g}}n_g\bar{X}_{\tilde{g}}\right)'\left(P_{\hat{F}}-P_{F^0}\right)\sqrt{n_g}\bar{\varepsilon}_g\right)\left(\frac{1}{\sqrt{T}}\sqrt{n_g}\varepsilon_g'\left(P_{F^0}-P_{\hat{F}}\right)\bar{X}_g\right) \\
&+ \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{n\sqrt{T}}\sum_{\tilde{g}=1}^{G}\left(a_{g\tilde{g}}-\hat{a}_{g\tilde{g}}\right)n_{\tilde{g}}\bar{X}_{\tilde{g}}'M_{\hat{F}}\sqrt{n_g}\bar{\varepsilon}_g\right)\left(\frac{1}{\sqrt{T}}\sqrt{n_g}\bar{\varepsilon}_g'\left(P_{F^0}-P_{\hat{F}}\right)\bar{X}_g\right) \\
&= \mathcal{L}1+\mathcal{L}2
\end{aligned}
$$

and we can show that $\mathcal{L}1 = o_p(1)$ using the same procedure to show $R11 = o_p(1)$. For $\mathcal{L}2$, note that

$$\begin{aligned}
&\frac{1}{n\sqrt{T}}\sum_{\tilde{g}=1}^{G}\left(a_{g\tilde{g}} - \hat{a}_{g\tilde{g}}\right) n_{\tilde{g}} \bar{X}_{\tilde{g}}' M_{\hat{F}} \sqrt{n_g}\bar{\varepsilon}_g \\
&= \left(H^{-1}\lambda_g^0\right)' \left(H^{-1}\Lambda^{0\prime}\Lambda^0\left(H^{-1}\right)'/n\right)^{-1} \frac{1}{n}\sum_{j=1}^{n}\left(H^{-1}\lambda_{g_j}^0 - \hat{\lambda}_{g_j}\right) \frac{X_j' M_{\hat{F}}\sqrt{n_g}\bar{\varepsilon}_g}{\sqrt{T}} \\
&+ \left(H^{-1}\lambda_g^0\right)' \left[\left(H^{-1}\Lambda^{0\prime}\Lambda^0\left(H^{-1}\right)'/n\right)^{-1} - \left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1}\right] \frac{1}{n}\sum_{j=1}^{n}\hat{\lambda}_{g_j}\frac{X_j' M_{\hat{F}}\sqrt{n_g}\bar{\varepsilon}_g}{\sqrt{T}} \\
&+ \left(H^{-1}\lambda_g^0 - \hat{\lambda}_g\right)' \left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1} \frac{1}{n}\sum_{j=1}^{n}\hat{\lambda}_{g_j}\frac{X_j' M_{\hat{F}}\sqrt{n_g}\bar{\varepsilon}_g}{\sqrt{T}} \\
&= o_p(1). \qquad \text{(A.25)}
\end{aligned}$$

Thus, $\mathcal{L}2 = o_p(1)$ and $R12 = o_p(1)$. Using similar procedures to show $R12 = o_p(1)$, we can prove that $R13 = o_p(1)$ and $R14 = o_p(1)$. Thus, $R1 = o_p(1)$.

For $R2$,

$$\begin{aligned}
R2 &= \frac{1}{G}\sum_{g=1}^{G}\left[\left(\frac{1}{\sqrt{T}}\bar{X}_g'\left(P_{F^0} - P_{\hat{F}}\right)\sqrt{n_g}\bar{\varepsilon}_g\right) + \left(\frac{1}{n\sqrt{T}}\sum_{c=1}^{n} X_c'\left(a_{ic}M_{F^0} - \hat{a}_{ic}M_{\hat{F}}\right)\sqrt{n_g}\bar{\varepsilon}_g\right)\right] \\
&\times\left(\frac{1}{\sqrt{n_g T}}\sum_{j\in\mathcal{A}_g}\varepsilon_j'\mathcal{X}_j^X\right) \\
&= o_p(1)
\end{aligned}$$

The detail of proof is omitted because it is the same as the proofs of $R11 = o_p(1)$ and $R12 = o_p(1)$. We also have $R3 = o_p(1)$ since $R3 = R2'$.

For the last term in (A.21), since the convergence is elementwise, let's assume that $\tilde{V}_{nT}^c - V_{nT}^c$ is a scalar. Then, under independence among groups in Assumption 9(i), we have

$$\begin{aligned}
&P\left(\left|\tilde{V}_{nT}^c - V_{nT}^c\right| > \Delta\right) \\
&\leq \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g} E\left(\left(\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\mathcal{X}_j^X - E\left[\left(\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\mathcal{X}_j^X\right]\right)^2 \\
&= o(1)
\end{aligned}$$

as $(G, n, T) \to \infty$.

Part (ii) Due to the proofs of Theorem 2 and Theorem 3 Part (i), we only need to show that

$$\hat{\sigma}^2 - \sigma^2 \to^p 0. \qquad \text{(A.26)}$$

We have

$$\hat{\sigma}^2 = \frac{1}{nT}\sum_{i=1}^{n}\left[\left(\beta^0-\hat{\beta}\right)' X_i' X_i \left(\beta^0-\hat{\beta}\right) + 2\left(\beta^0-\hat{\beta}\right)' X_i'\left(F^0\lambda^0_{g_i} - \hat{F}\hat{\lambda}_{g_i}\right) + 2\left(\beta^0-\hat{\beta}\right)' X_i'\varepsilon_i \right.$$
$$\left. + \left(F^0\lambda^0_{g_i} - \hat{F}\hat{\lambda}_{g_i}\right)'\left(F^0\lambda^0_{g_i} - \hat{F}\hat{\lambda}_{g_i}\right) + 2\left(F^0\lambda^0_{g_i} - \hat{F}\hat{\lambda}_{g_i}\right)'\varepsilon_i + \varepsilon_i'\varepsilon_i\right]$$
$$= T1 + ... + T6.$$

It is straightforward to show that $T1 = ... = T5 = o_p(1)$.

For $T6$, we have

$$T6 = \frac{1}{nT}\sum_{i=1}^{n}\varepsilon_i'\varepsilon_i \to^p \sigma^2,$$

which completes the proof of (A.26). ■

Proof of Theorem 4. Part (i) Due to the proofs of Theorem 2 and Theorem 3, it is sufficient to show that $\hat{C}^{XX}_{0a,nT} - C^{XX}_{0a,nT} \to^p 0$.

We have

$$\hat{C}^{XX}_{0a,nT} - C^{XX}_{0a,nT} = \left(\hat{C}^{XX}_{0a,nT} - \tilde{C}^{XX}_{0a,nT}\right) + \left(\tilde{C}^{XX}_{0a,nT} - C^{XX}_{0a,nT}\right). \tag{A.27}$$

The second term is $o_p(1)$ by Assumption 10. For the first term,

$$\hat{C}^{XX}_{0a,nT} - \tilde{C}^{XX}_{0a,nT}$$
$$= \frac{1}{nT}\sum_{i=1}^{n}\left[\left(X_i - P_{\hat{F}_0}\bar{X}_{0,g_i}\right)'\left(X_i - P_{\hat{F}_a}\bar{X}_{a,g_i}\right) - \left(X_i - P_{F^0}\bar{X}_{0,g_i}\right)'\left(X_i - P_{F^0}\bar{X}_{a,g_i}\right)\right]$$
$$+ \frac{1}{nT}\sum_{i=1}^{n}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\hat{a}_{0,ij}M_{\hat{F}_0}X_j\right)'\left(\frac{1}{n}\sum_{j=1}^{n}\hat{a}_{a,ij}M_{\hat{F}_a}X_j\right) - \left(\frac{1}{n}\sum_{j=1}^{n}a^0_{ij}M_{F^0}X_j\right)'\left(\frac{1}{n}\sum_{j=1}^{n}a^0_{ij}M_{F^0}X_j\right)\right]$$
$$- \frac{1}{nT}\sum_{i=1}^{n}\left[\left(X_i - P_{\hat{F}_0}\bar{X}_{0,g_i}\right)'\left(\frac{1}{n}\sum_{j=1}^{n}\hat{a}_{a,ij}M_{\hat{F}_a}X_j\right) - \left(X_i - P_{F^0}\bar{X}_{0,g_i}\right)'\left(\frac{1}{n}\sum_{j=1}^{n}a^0_{ij}M_{F^0}X_j\right)\right]$$
$$- \frac{1}{nT}\sum_{i=1}^{n}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\hat{a}_{0,ij}M_{\hat{F}_0}X_j\right)'\left(X_i - P_{\hat{F}_a}\bar{X}_{a,g_i}\right) - \left(\frac{1}{n}\sum_{j=1}^{n}a^0_{ij}M_{F^0}X_j\right)'\left(X_i - P_{F^0}\bar{X}_{a,g_i}\right)\right]$$
$$= o_p(1).$$

Part (ii) $\hat{V}_{\mathcal{T}}$ is positive semi-definite under Assumption 1(ii), and $\hat{\beta}_0 - \beta^0 \to^p \Delta$. Thus, it is straightforward that $\mathcal{T}$ diverges as $n, T \to \infty$. ■

Proof of Theorem 5. Part (i) Without loss of generality, we set $\beta^0 = 0$. Let

$$\tilde{\mathcal{Q}}_{gmm}(\beta, F) = \frac{1}{nT}\sum_{i=1}^{n}\left[M_F F^0\lambda^0_{g_i} + \left(\varepsilon_i - P_F\bar{\varepsilon}_{g_i}\right) - \left(X_i - P_F\bar{X}_{g_i}\right)\beta\right]'\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)\Omega^{-1}_{nT}$$
$$\times \frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)'\left[M_F F^0\lambda^0_{g_i} + \left(\varepsilon_i - P_F\bar{\varepsilon}_{g_i}\right) - \left(X_i - P_F\bar{X}_{g_i}\right)\beta\right]$$
$$- \frac{1}{nT}\sum_{i=1}^{n}\left(\varepsilon_i - P_{F^0}\bar{\varepsilon}_{g_i}\right)'\left(\Psi_i - P_{F^0}\bar{\Psi}_{g_i}\right)\Omega^{-1}_{nT}\frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i - P_{F^0}\bar{\Psi}_{g_i}\right)'\left(\varepsilon_i - P_{F^0}\bar{\varepsilon}_{g_i}\right),$$

$$
\begin{aligned}
\mathcal{Q}_{gmm}^{*}(\beta,F) &= \frac{1}{nT}\sum_{i=1}^{n}\lambda_{g_i}^{0\prime}F^{0\prime}M_F'\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)\Omega_{nT}^{-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)' M_F F^0\lambda_{g_i}^0 \\
&\quad - \frac{2}{nT}\sum_{i=1}^{n}\lambda_{g_i}^{0\prime}F^{0\prime}M_F'\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)\Omega_{nT}^{-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)'\left(X_i - P_F\bar{X}_{g_i}\right)\beta \\
&\quad + \frac{1}{nT}\sum_{i=1}^{n}\beta'\left(X_i - P_F\bar{X}_{g_i}\right)'\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)\Omega_{nT}^{-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)'\left(X_i - P_F\bar{X}_{g_i}\right)\beta,
\end{aligned}
$$

$$
\tilde{\mathcal{Q}}_{LS}(F) = \tilde{\mathcal{Q}}(\hat{\beta}_{gmm}(F),F),
$$

and

$$
\begin{aligned}
\mathcal{Q}_{LS}^{*}(F) &= \frac{1}{nT}\sum_{i=1}^{n}\lambda_{g_i}'F^{0\prime}M_F F^0\lambda_{g_i} - \frac{2}{nT}\sum_{i=1}^{n}\lambda_{g_i}'F^{0\prime}M_F\left(X_i - P_F\bar{X}_{g_i}\right)\mathbb{C}_{nT}(F) \\
&\quad + \mathbb{C}_{nT}(F)'\left(\frac{1}{nT}\sum_{i=1}^{n}\left(X_i - P_F\bar{X}_{g_i}\right)'\left(X_i - P_F\bar{X}_{g_i}\right)\right)\mathbb{C}_{nT}(F),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{C}_{nT}(F) &= \mathbb{D}_{nT}(F)\left[\frac{1}{nT}\sum_{j=1}^{n}\left(\Psi_j - P_F\bar{\Psi}_{g_j}\right)M_F F^0\lambda_{g_j}\right], \\
\mathbb{D}_{nT}(F) &= \left[Q_{nT}^{X\Psi}(F)\Omega_{nT}^{-1}Q_{nT}^{\Psi X}(F)\right]^{-1}Q_{nT}^{X\Psi}(F)\Omega_{nT}^{-1}
\end{aligned}
$$

The proof consists of two steps. The first step is to show that $\left(\beta^0, F^0H\right)$ is the unique minimizer of $\mathcal{Q}_{gmm}^{*}(\beta,F)$ for all bounded β and $F\in\mathcal{F}$ and $\tilde{\mathcal{Q}}_{gmm}(\beta,F) - \mathcal{Q}_{gmm}^{*}(\beta,F) = o_p(1)$. The second step is to show F^0H is the unique minimizer of $\mathcal{Q}_{LS}^{*}(F)$ and $\tilde{\mathcal{Q}}_{LS}(F) - \mathcal{Q}_{LS}^{*}(F) = o_p(1)$ for $F\in\mathcal{F}$.

For the first step, it is easy to see that $\mathcal{Q}_{gmm}^{*}\left(\beta^0, F^0H\right) = 0$ and since Ω_{nT}^{-1} is positive definite

$$
\begin{aligned}
\mathcal{Q}_{gmm}^{*}(\beta,F) &= \frac{1}{nT}\sum_{i=1}^{n}\left[\lambda_{g_i}^{0\prime}F^{0\prime}M_F' - \beta'\left(X_i - P_F\bar{X}_{g_i}\right)'\right]\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)\Omega_{nT}^{-1} \\
&\quad \times\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)'\left[M_F F^0\lambda_{g_i}^0 - \left(X_i - P_F\bar{X}_{g_i}\right)\beta\right] \\
&> 0
\end{aligned}
$$

if $(\beta,F) \neq \left(\beta^0, F^0H\right)$. Thus, $\left(\beta^0, F^0H\right)$ is the unique minimizer of $\mathcal{Q}_{gmm}^{*}(\beta,F)$. We also have

$$
\begin{aligned}
&\tilde{\mathcal{Q}}_{gmm}(\beta,F) - \mathcal{Q}_{gmm}^{*}(\beta,F) \\
&= \frac{2}{nT}\sum_{i=1}^{n}\lambda_{g_i}^{0\prime}F^{0\prime}M_F\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)\Omega_{nT}^{-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)'\left(\varepsilon_i - P_F\bar{\varepsilon}_{g_i}\right) \\
&+ \frac{1}{nT}\sum_{i=1}^{n}\left(\varepsilon_i - P_F\bar{\varepsilon}_{g_i}\right)'\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)\Omega_{nT}^{-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)'\left(\varepsilon_i - P_F\bar{\varepsilon}_{g_i}\right) \\
&- \frac{2}{nT}\sum_{i=1}^{n}\left(\varepsilon_i - P_F\bar{\varepsilon}_{g_i}\right)'\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)\Omega_{nT}^{-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i - P_F\bar{\Psi}_{g_i}\right)'\left(X_i - P_F\bar{X}_{g_i}\right)\beta \\
&= o_p(1).
\end{aligned}
$$

For the second step, we have $\mathcal{Q}_{LS}^{*}(F^0 H) = 0$, and it is easy to prove that $\tilde{\mathcal{Q}}_{LS}(F) - \mathcal{Q}_{LS}^{*}(F) = o_p(1)$. We can show that $\mathcal{Q}_{LS}^{*}(F) > 0$ if $F \neq F^0 H$ using the proof of consistency in Bai (2009, Proposition 1) by replacing β with $\mathbb{C}_{nT}(F)$.

Part (ii) The proof is omitted because it is the same as the proof of Proposition A1. ■

Proof of Theorem 6. Note that

$$\begin{aligned}
&\left[Q_{nT}^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_{nT}^{-1}Q_{nT}^{\Psi X}\left(\hat{F}_{gmm}\right)\right]\sqrt{nT}\left(\hat{\beta}_{gmm}-\beta^0\right)\\
&= Q_{nT}^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_{nT}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\Psi_i' M_{\hat{F}_{gmm}}F^0\lambda_{g_i} - \left(\Psi_i - P_{F^0}\bar{\Psi}_{g_i}\right)'\varepsilon_i\right). \qquad \text{(A.28)}
\end{aligned}$$

Using a similar procedure in the proof of Proposition A2, we can have

$$\begin{aligned}
&\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\Psi_i' M_{\hat{F}_{gmm}}F^0\lambda_{g_i}\\
&= \frac{1}{n^2 T}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^0\Psi_i' M_{\hat{F}_{gmm}}X_j\sqrt{nT}\left(\hat{\beta}_{gmm}-\beta\right) - \frac{1}{n\sqrt{nT}}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^0\Psi_i' M_{\hat{F}_{gmm}}\varepsilon_j\\
&+ o_p\left(\sqrt{nT}\left\|\hat{\beta}_{gmm}-\beta\right\|\right) + O_p\left(\frac{\sqrt{T}}{n}\right) + O_p\left(\frac{\sqrt{G}}{n}\right) + O_p\left(\frac{G}{\sqrt{nT}}\right).
\end{aligned}$$

Applying this result to (A.28), under the rate condition in Assumption 5, we have

$$Q_{nT}^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_{nT}^{-1}B_{nT}^{\Psi X}\left(\hat{F}_{gmm}\right)\sqrt{nT}\left(\hat{\beta}_{gmm}-\beta\right) = Q_{nT}^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_{nT}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i^{\Psi}\left(\hat{F}_{gmm}\right)'\varepsilon_i + o_p(1).$$

Then, using a similar procedure in the proof of Theorem 2, we have

$$\sqrt{nT}\left(\hat{\beta}_{gmm}-\beta\right) = \left(Q_{X\Psi}\Omega_{nT}^{-1}B_{\Psi X}\right)^{-1}Q_{X\Psi}\Omega_{nT}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\mathcal{X}_i^{\Psi}\right)'\varepsilon_i + o_p(1).$$

Applying Assumptions 12(ii) and 14 to this result completes the proof. ■

References

[1] Ahn, S., Y. Lee, and P. Schmidt (2001). GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects, *Journal of Econometrics*, 101(2), 219-255.

[2] Ando, T., and J. Bai (2016). Panel Data Models with Grouped Factor Structure Under Unknown Group Membership, *Journal of Applied Econometrics*, 31(1), 163–191.

[3] Ashenfelter, O., and C. Card (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs, *Review of Economics and Statistics*, 67(4), 648-660.

[4] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions, *Econometrica*, 71(1), 135–171.

[5] Bai, J. (2009). Panel Data Models with Interactive Fixed Effects, *Econometrica*, 77(4), 1229–1279.

[6] Baker, B. (2003). The Case for Antitrust Enforcement, *Journal of Economic Perspectives*, 17, 27–55.

[7] Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-In-Differences Estimates?, *Quarterly Journal of Economics*, 119(1), 249-275.

[8] Besley, T., and A. Case (2000) Unnatural Experiments? Estimating the Incidence of Endogenous Policies, *Economic Journal*, 110(467), 672-694.

[9] Bester, A., and C. Hansen (2016). Grouped Effects Estimators in Fixed Effects Models, *Journal of Econometrics*, 190(1), 197-208.

[10] Bester, A., T. Conley, and C. Hansen (2011). Inference with Dependent Data Using Cluster Covariance Estimators, *Journal of Econometrics*, 165(2), 137-151.

[11] Blundell, R., and M. Dias (2009). Alternative Approaches to Evaluation in Empirical Microeconomics, *Journal of Human Resources*, 44(3), 565-640.

[12] Bonhomme, S., and E. Manresa (2015). Grouped Patterns of Heterogeneity in Panel Data (2015), *Econometrica*, 83, 1147-1184.

[13] Buccirossi, P., L. Ciari, T. Duso, G. Spagnolo, and C. Vitale (2013). Competition Policy and Productivity Growth: An Empirical Assessment, *Review of Economics and Statistics*, 95(4), 1324-1336.

[14] Crandall, R., and C. Winston (2003). Does Antitrust Policy Improve Consumer Welfare? Assessing the Evidence, *Journal of Economic Perspectives*, 17, 3–26.

[15] Cusack, T. and S. Fuchs (2002). Documentation Notes for Parties, Governments, and Legislatures Data Set, mimeo, Wissenschaftszentrum Berlin für Sozialforschung.

[16] Donald, S., and K. Lang (2001), Inference with Difference in Differences and Other Panel Data, unpublished manuscript, Boston University.

[17] Hahn, J., and H. Moon (2010). Panel Data Models with Finite Number of Multiple Equilibria, *Econometric Theory*, 26, 863-881.

[18] Hansen, C. (2007). Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects, *Journal of Econometrics*, 140(2), 670-694.

[19] Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50(4), 1029-1054.

[20] Hausman, J. (1978). Specification Tests in Econometrics, *Econometrica*, 46(6), 1251-1271.

[21] Holtz-Eakin, D., W. Newey, and H. Rosen (1988). Estimating Vector Autoregressions with Panel Data, *Econometrica*, 56(6),1371-1395.

[22] Honyes, H., and D. Schanzenbach (2012). Work Incentives and the Food Stamp Program, *Journal of Public Economics*, 96(1-2), 151-162.

[23] Ibragimov, R., and U. Mueller (2016). Inference with Few Heterogeneous Clusters, *Review of Economics and Statistics*, 98, 83-96.

[24] Lee, N., H. Moon, and M. Weidner (2012) Analysis of Interactive Fixed Effects Dynamic Linear Panel Regression with Measurement Error, *Economics Letters*, 117(1), 239-242

[25] Lu, Lina (2017) Simultaneous Spatial Panel Data Models with Common Shocks, *FRB of Boston Working paper series.*

[26] Lu, X., and L. Su (2016). Shrinkage Estimation of Dynamic Panel Data Models with Interactive Fixed Effects, *Journal of Econometrics*, 190, 148-175.

[27] Moon, H., M. Shum, and M. Weidner (2017). Interactive Fixed effects in the BLP Random Coefficients Demand Model, *Journal of Econometrics*, forthcoming.

[28] Moon, H., and M. Weidner (2017). Dynamic Linear Panel Regression Models with Interactive Fixed Effects, *Econometric Theory*, 33(1), 158-195.

[29] Moon, H., and M. Weidner (2015). Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects, *Econometrica*, 83(4), 1543-1579.

[30] Moulton, B. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units, *Review of Economics and Statistics*, 72, 335-338.

[31] Neyman, J., and E. Scott (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica,* 16(1),1-32.

[32] Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, *Econometrica*, 74(4), 967-1012.

[33] Su, L., and Q. Chen (2013). Testing Homogeneity in Panel Data Models with Interactive Fixed Effects, *Econometric Theory*, 29, 1079-1135.

[34] Su, L., Z. Shi, and P. C. B. Phillips (2016). Identifying Latent Structures in Panel Data, *Econometrica*, forthcoming.

[35] Sun, Y. (2005). Estimation and Inference in Panel Structure Models, *UCSD working paper series.*

[36] Werden, G. (2003). The 1982 Merger Guidelines and the Ascent of the Hypothetical Monopolist Paradigm, Antitrust Law Review, 71, 253–269.