Panel and Multilevel Models with Interactive Terms of Group Fixed Effects and Common Time Effects

Min Seong Kim*
Department of Economics
Ryerson University

November, 2016

Abstract

In this paper, we consider a linear panel and multilevel model with interactive terms of group fixed effects and common time effects. The panel and multilevel regression model is particularly prevalent in policy analysis to accommodate a group level policy variable in the regression with an individual outcome dependent variable. For this model, we first study the least squares estimator and associated inference procedure. Secondly, we develop a test for the appropriate level of grouping to specify group fixed effects. Finally, we propose a new control function approach to address policy endogeneity with respect to the idiosyncratic error. A GMM approach is also studied in this regard. We establish the asymptotics under the large n, T asymptotics such that $T/n \to 0$. We allow the number of group, G, to grow but at a slower rate. Monte Carlo simulations show that the proposed estimators and tests perform well in finite samples.

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

JEL Classification Number: C23

*Email: minseong.kim@ryerson.ca. Any comments and suggestions are appreciated.

1 Introduction

In this paper, we propose a panel and multilevel linear regression model that includes interactive terms of group fixed effects and common time effects. A multilevel regression model is prevalent in policy analysis to accommodate a group level policy variable in the regression with an individual outcome dependent variable. See, for example, county level food stamp program and household head's 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 often 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 an example of this approach. A shortcoming of the additive fixed effects model is that its validity crucially depends on the assumption that the effect of unobserved group heterogeneity is time invariant and the impact of time effects is the same across groups. This assumption is implausible in some 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 may also appear if different groups based on regions, markets or ages show heterogenous responses to common time effects such as cyclical fluctuations. See Blundell and Dias (2009) for details.

To address this, we consider the interactive fixed effects model in the panel and multilevel regression setting. Our model is given by

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

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 $((d_x - d_z) \times 1)$ vector of individual level regressors. 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. We assume the group membership $\{g_i\}_{i=1}^n$ to be known to researchers. As F_t^0 is multiplicative of $\lambda_{g_i}^0$, our 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 $F_t^0 = [1, f_t^0]'$, where $\alpha_{g_i}^0$ and f_t^0 denote scalar group fixed effects and time fixed effects respectively.

We examine estimation and inference procedures for (1). First, we consider a least squares (LS) estimator, which we will 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 estimation is built on Bai (2009) who proposes the LS estimator for

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

which we will refer to as the "standard interactive fixed effects estimator" in this paper to distinguish it from our estimator. He establishes the properties under the large n, T asymptotics. (2) has also been studied by Kiefer (1980), Holtz-Eakin, Newey, and Rosen (1988), Ahn, Lee, and Schmidt (2001), Pesaran (2006), Moon and Weidner (2014, 2015), and Lu and Su (2016) among others. Bai (2009) shows that when n grows faster 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 there exists serial correlation or heteroskedasticity, the estimator is asymptotically biased due to 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 well known in the literature, this causes a different type of asymptotic bias. See Moon and Weidner (2015) for this issue. Bai (2009) provides a bias corrected estimator for (2). In our setting, since λ_g is common within a group, this source of bias is of order $G/\sqrt{nT}$, and our estimator is asymptotically unbiased in the presence of correlation and heteroskedasticity given $G/\sqrt{nT} \to 0$ and $T/n \to 0$ as $n, T \to \infty$. These rate conditions are relevant in policy analysis where a policy variable is group (e.g., state and county) specific and many individuals belong to each group. The idea of considering a group structure to 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 new test about 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 practical situation, to decide on the appropriate level of grouping to specify the group fixed effects. For example, when a policy is state specific, the group fixed effects can be set at state level, but a finer level of grouping should also be considered if a researcher suspects within each state the impact of unobserved time effects to vary, depending on employment status or income level. We suppose that two different grouping schemes, $\mathbb{A}_0$ and $\mathbb{A}_a$, between which the latter is a finer level of grouping, 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 the estimator of β^0 based on $\mathbb{A}_a$ is consistent. Utilizing the fact that under the null not only $\mathbb{A}_0$ but $\mathbb{A}_a$ also yields a consistent estimator which is less efficient, we conduct the test by comparing the group fixed effects estimator based on $\mathbb{A}_0$ with the one based on $\mathbb{A}_a$.

We extend our approach to the case that a policy variable is endogenous with respect to ε_{it}. Some sources of endogeneity, such as simultaneity and measurement error, may remain even if the group interactive terms are introduced. To address this, we propose a new control function approach which we will call “interactive fixed effects control function" (IFE-CF) estimator. This approach needs instruments, Ψ_i, that yield mean independence of ε_{it} from the policy variable, $\{\Psi_i\}$, and $\{\lambda_{g_i}^{0\prime} F_t\}$ conditional on the error term in the reduced from regression. The CF approach in the interactive fixed effects model may have a larger set of potential instruments than the one without interactive fixed effects terms. For the latter, the mean independence assumption should holds not only for ε_{it} but also for $\lambda_{g_i}' F_t$. While we develop this method based on (1), it can be extended to (2). In this setting, Moon, Shum and Weidner (2014) propose the “least-squares minimum distance (LS-MD)" estimation method in the random coefficient multinomial logit demand model context, and it is extended to a linear regression model by Moon and Weidner (2014) and Lee, Moon and Weidner (2012). In the paper, we also consider a moment condition based GMM approach and establish the consistency and asymptotic normality under the asymptotics that $(G, n, T) \to \infty$.[1]

The proposed model is related to

$$Y_{it} = X_{it}'\beta^0 + \alpha_{g_i t}^0 + \varepsilon_{it}, \tag{3}$$

where $\alpha_{g_i t}^0$ represent unobserved group specific time effects. This model accommodates time varying group heterogeneity as our model does and can be estimated with the standard fixed effects approach. In contrast to our model, however, this model is not useful to estimate the effect of a group specific policy due to multicollinearity of the dummies for $\alpha_{g_i t}^0$ and group level regressors. Hansen (2007) and Bertrand, Duflo, and Mullainathan (2004) consider (3) in the multilevel regression framework. But, they assume $\alpha_{g_i t}^0$ to be uncorrelated with regressors and focus on 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), Su, Shi and Phillips (2016). Among them, Ando and Bai (2015) study an interactive fixed effects approach in the group structure when the group membership is unobserved. In their paper, time effects are group specific and they are interacted with individual specific effects, which is different from our setting. In addition, they assume the number of group G to be fixed as $(n, T) \to \infty$.

The outline of the paper is as follows. Section 2 introduces our model and LS estimator. Section 3 examines the asymptotic properties of the estimator and associated test statistic. We follow Bai (2009) to assume that the number of factors is known and the panel data is balanced. Section 4 provides a Hausman type test about the appropriate level of grouping. In Section 5, we study the control function and GMM approaches to address the endogeneity of a policy variable with respect to the idiosyncratic error. Section 6 reports simulation evidence. In the Section 7, we use our procedures to Buccirossi et al., (2013) and reexamine their main results. The last section concludes. Proofs are given in the appendix.

Some notation is used throughout the paper. We define projection matrices $P_F = F(F'F)^{-1}F' = FF'/T$ 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

Our model (1) can be rewritten as

$$Y_i = X_i\beta^0 + F^0\lambda_{g_i}^0 + \varepsilon_i, \tag{4}$$

where $Y_i = (Y_{i1}, ..., Y_{iT})'$, $X_i = (X_{i1}, ..., X_{iT})'$ and $F^0 = (F_1^0, ..., F_T^0)'$. Using matrix notation, we also have

$$Y = \beta_1^0 X^{(1)} + \cdots + \beta_{d_x}^0 X^{(d_x)} + F^0\Lambda^{0\prime} + \varepsilon,$$

[1]The procedure of this approach based on the standard interactive fixed effects model is discussed by Moon, Shum and Weidner (2014) but they do not provide the asymptotics.

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. Note that F^0 and Λ^0 are not separately identifiable because they are multiplicative in the model. We follow Bai (2003) to impose the following normalization in estimation:

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

Under (5), $P_F = FF'/T$ and Λ and F are uniquely determined given the product $F\Lambda'$.

We consider LS estimation for this model. Since λ^0_g is common within each group, the objective function is written as

$$\mathcal{Q}(\beta, F, \Lambda) = \frac{1}{nT} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} (Y_i - X_i\beta - F\lambda_g)' (Y_i - X_i\beta - F\lambda_g), \tag{6}$$

where $\mathcal{A}_g = \{i : g_i = g\}$ is the set of individuals that belong to the group g. The LS estimators $\left(\hat{\beta}, \hat{F}, \hat{\Lambda}\right)$ minimize (6). 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

$$\hat{\lambda}_g(\beta, F) = (F'F)^{-1} F' \left(\frac{1}{n_g} \sum_{i \in \mathcal{A}_g} Y_i - X_i\beta \right) = \frac{F'\left(\bar{Y}_g - \bar{X}_g\beta\right)}{T} \tag{7}$$

out of (6), we have

$$\begin{aligned}
&\mathcal{Q}(\beta, F) \\
&= \frac{1}{nT} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(Y_i - P_F\bar{Y}_g - \left(X_i - P_F\bar{X}_g\right)\beta\right)' \left(Y_i - P_F\bar{Y}_g - \left(X_i - P_F\bar{X}_g\right)\beta\right) &(8)\\
&= \frac{1}{nT} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} (Y_i - X_i\beta)' (Y_i - X_i\beta) - \frac{1}{nT} \sum_{g=1}^{G} n_g \left(\bar{Y}_g - \bar{X}_g\beta\right)' \left(\frac{FF'}{T}\right) \left(\bar{Y}_g - \bar{X}_g\beta\right). &(9)
\end{aligned}$$

Based on (8), we have $\hat{\beta}(F)$ that minimizes $\mathcal{Q}(\beta, F)$ given F as follows:

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

We also derive $\hat{F}(\beta)$ given β. Since the first term in (9) does not depend on F, minimization of (9) with respect to F reduces to maximization of

$$\sum_{g=1}^{G} n_g \left(\bar{Y}_g - \bar{X}_g\beta\right)' \left(\frac{FF'}{T}\right) \left(\bar{Y}_g - \bar{X}_g\beta\right) = \frac{1}{T} tr\left[F'\bar{\mathcal{R}}(\beta)\bar{\mathcal{R}}(\beta)'F\right], \tag{10}$$

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 this maximization 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

$$\hat{\beta} = \left[\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_i - P_{\hat{F}}\bar{X}_g\right)'\left(X_i - P_{\hat{F}}\bar{X}_g\right)\right]^{-1}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_i - P_{\hat{F}}\bar{X}_g\right)'\left(Y_i - P_{\hat{F}}\bar{Y}_g\right), \tag{11}$$

and

$$\frac{1}{nT}\bar{\mathcal{R}}\left(\hat{\beta}\right)\bar{\mathcal{R}}\left(\hat{\beta}\right)'\hat{F} = \hat{F}\hat{\Gamma}, \tag{12}$$

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 (12) or an initial value of F in (11) to iterate (11) and (12) to convergence. As discussed in Bai (2009) and Su and Chen (2013), this procedure can yield a local minimum of the objective function (8) depending on the initial value we use. We may try the iteration with several initial values and choose the one that produces the smallest value of (8). Applying $\left(\hat{\beta}, \hat{F}\right)$ to (7), we have

$$\hat{\lambda}_g = \frac{\hat{F}'\left(\bar{Y}_g - \bar{X}_g\hat{\beta}\right)}{T} \text{ and } \hat{\Lambda} = [\hat{\lambda}_{g_1}, ..., \hat{\lambda}_{g_n}]'. \tag{13}$$

Note that the LS estimators $\left(\hat{F}, \hat{\Lambda}\right)$ satisfy the restrictions in (5).

The rank condition requires $T \geq d_F + 1$. Each column of $\hat{F}$ is orthogonal, so $P_{\hat{F}} = I_{d_F}$ when $T = d_F$. 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, when $T = 5$, four is the maximum number of interactive terms we can employ in the model.

3 Asymptotic theory and inference

In this section, we examine the properties of $\hat{\beta}$ and the associated inference under the asymptotics that $(n, T, G) \rightarrow \infty$. Let

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

for random variables v and w. From (11), we have

$$\hat{\beta} - \beta^0 = Q_{nT}^{XX}\left(\hat{F}\right)^{-1}\frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[\bar{X}_g'M_{\hat{F}}F^0\lambda_g + \left(X_i - P_{\hat{F}}\bar{X}_g\right)'\varepsilon_i\right]. \tag{14}$$

We show in the Proposition A1 in the appendix that, under regularity assumptions presented below,

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

with $a_{ij} = \lambda_{g_i}' \left(\Lambda'\Lambda/n\right)^{-1} \lambda_{g_j}$. Combining this expression with (14), we obtain

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

where

$$\mathcal{X}_i^X(F) = \left(X_i - P_F \bar{X}_{g_i}\right) - \frac{1}{n}\sum_{j=1}^{n} a_{ij} M_F X_j \tag{15}$$

and

$$\begin{aligned} B_{nT}^{XX}(F) &= \frac{1}{nT}\sum_{i=1}^{n}\mathcal{X}_i^X(F)'\,\mathcal{X}_i^X(F) \\ &= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_i - P_F\bar{X}_g\right)'\left(X_i - P_F\bar{X}_g\right) - \frac{1}{T}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} X_i' M_F X_j\right\}. \end{aligned} \tag{16}$$

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

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

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

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

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

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

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

(iii) 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.
(iv) For all (t,s),

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

(v) We have

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

$$\lim_{n,T\to\infty}\frac{1}{G^2T}\sum_{g=1}^{G}\sum_{\tilde{g}_1=1}^{G}\sum_{\tilde{g}_2=1}^{G}\sum_{t=1}^{T}\sum_{s=1}^{T} n_g\sqrt{n_{\tilde{g}_1}n_{\tilde{g}_2}}\left|Cov\left(\bar{\varepsilon}_{\tilde{g}_1t}\bar{\varepsilon}_{gt},\bar{\varepsilon}_{\tilde{g}_2s}\bar{\varepsilon}_{gs}\right)\right| < M.$$

Assumption 4

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

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

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

Assumption 5 *(i)* ε_{it} *is independent of* λ_g *and* F_s *for all* i,t,g *and* s; *(ii)* $E\left(\varepsilon_{it}|X_1,...,X_n\right)=0$ *for all* i *and* t.

Assumption 6 *For all* g, $n_g = cn^{\alpha}$ *for some* $c>0$ *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, $\sqrt{n_g}\bar{\varepsilon}_{gt}$ here, due to the group structure of our model. Weak dependence of $\left\{\sqrt{n_g}\bar{\varepsilon}_{gt}\right\}$ across groups should not be restrictive, since strong dependence is captured by interactive terms.

In Assumption 5, we assume independence of $\{\varepsilon_{it}\}$ with $\{\lambda_g\}$ and $\{F_s\}$ which are standard in the literature. Regarding $\{X_{it}\}$, we introduce the strict exogeneity condition instead of independence. This assumption is particularly useful in Section 5.1 to develop the control function method to address endogeneity with respect to $\{\varepsilon_{it}\}$. Assumption 6 implies the size of each group is comparable and that, for all g, n_g grows as n increases.

Assumption 7 *(i) $(G, n, T) \to \infty$; (ii) $(G, n) \to \infty$ such that $G/n \to 0$ for fixed T.*

Assumption 7 provides the rate conditions that we employ to derive consistency of the LS estimator $\left(\hat{\beta}, \hat{F}\right)$ in Theorem 1 below.

Theorem 1 *Suppose that Assumptions 1-5 and 6(ii) hold. Then, we have*

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

where $H = (\Lambda'\Lambda/n) \left(F'\hat{F}/T\right) \hat{\Gamma}^{-1}$ as (i) $(G, n, T) \to \infty$ or (ii) $(G, n) \to \infty$ such that $G/n \to 0$.

The proofs of Theorem 1 are in the appendix. The first part presents consistency of $\hat{\beta}$, and the second part states the average norm consistency of $\hat{F}$ for F^0 with a rotation. These results are analogous to the consistency results for the standard interactive fixed effects model in Bai (2009, Proposition 1(i) and Proposition A.1(i)). We can compare the rate conditions between these two models. If λ_i were known in Bai's model, it would be easy to show that consistency is achieved as $n \to \infty$ regardless of T. Thus, the rate condition $(n, T) \to \infty$ in Bai (2009) reflects the fact that λ_i are unknown and estimated. In our model, λ_g is common within each group. Due to this group structure, $\left(\hat{\beta}, \hat{F}\right)$ is consistent not only as $(G, n, T) \to \infty$, but they also remain consistent for a fixed T, if $G/n \to 0$ as $n \to \infty$ so that each group has a large number of group members.

From now on, we focus on the asymptotics that $(G, n, T) \to \infty$.

Assumption 8 *$(G, n, T) \to \infty$ such that $T/n \to 0$ and $G/\sqrt{nT} \to 0$.*

Let

$$V_{nT}(F) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[\mathcal{X}_i^X(F)' \varepsilon_i \varepsilon_j' \mathcal{X}_j^X(F)\right]. \tag{17}$$

We make the following high level assumption to derive the asymptotic distribution for $\hat{\beta}$.

Assumption 9 *Let $\mathcal{X}_i^X = \mathcal{X}_i^X(F^0)$. We have*

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

where $V = \operatorname{plim}_{n,T\to\infty} V_{nT}(F^0)$ is positive definite.

Similar assumptions are introduced in Bai (2009) and Lu and Su (2016). Note that this assumption allows heteroskedasticity and serial and cross sectional correlation in the idiosyncratic errors. We now state the asymptotic distribution of $\hat{\beta}$.

Theorem 2 *Suppose that Assumptions 1-6 and 8-9 hold. Then, we have*

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

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

The proof is in the appendix. To understand this result, we can consider the following expansion derived in the proof of this theorem.

$$\begin{aligned}\sqrt{nT}\left(\hat{\beta}-\beta^0\right) &= B^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i^X\left(F^0\right)'\varepsilon_i + \frac{G}{\sqrt{nT}}B^{-1}A_{nT}^{(1)} + \sqrt{\frac{T}{n}}B^{-1}A_{nT}^{(2)} \\ &\quad + o_p\left(\sqrt{nT}\left\|\hat{\beta}-\beta^0\right\|\right) + o_p\left(\frac{G}{\sqrt{nT}}\right) + o_p\left(\sqrt{\frac{T}{n}}\right),\end{aligned} \tag{18}$$

where

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

When G grows at the same rate as $\sqrt{nT}$, $A_{nT}^{(1)}$ may become a source of 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 this model, if n grows faster than T ($T/n \to 0$), then the third term becomes negligible. However, the second term becomes divergent (in the presence of heterogeneity or serial correlation) and dominates the distribution. In our model, we assume G to grow but slowly such that $G/\sqrt{nT} \to 0$ as $T/n \to 0$. This rate condition allows $\hat{\beta}$ to remain centered at β^0 and to be asymptotically normal when normalized by the sample size. It is also empirically relevant in policy analysis in which each group often includes a large number of individuals.

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

$$\mathcal{H}_0 : R\beta = \mathfrak{r}^0 \text{ vs } \mathcal{H}_1 : R\beta \neq \mathfrak{r}^0, \tag{19}$$

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 following sample analogue

$$\begin{aligned}\hat{B}_{nT}^{XX} &= \frac{1}{nT}\sum_{i=1}^{n}\left(\hat{\mathcal{X}}_i^X\right)'\hat{\mathcal{X}}_i^X \\ &= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)'\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right) - \frac{1}{T}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{a}_{ij}X_i'M_{\hat{F}}X_j\right\} \qquad (20)\end{aligned}$$

where

$$\hat{\mathcal{X}}_i^X = \left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right) - \frac{1}{n}\sum_{j=1}^{n}\hat{a}_{ij}M_{\hat{F}}X_j \text{ with } \hat{a}_{ij} = \hat{\lambda}_i'\left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1}\hat{\lambda}_j, \qquad (21)$$

and $\hat{\lambda}_g$ and $\hat{\Lambda}$ are given in (13). For estimation of V_{nT}, we introduce the following conditions.

Assumption 10 *(i) $E(\varepsilon_{it}\varepsilon_{js}|X) = 0$ for $t, s = 1, ..., T$ if $g_i \neq g_j$; (ii) $Var(\varepsilon_{it}|X) = \sigma^2$ and $E(\varepsilon_{it}\varepsilon_{js}) = 0$ if $i \neq j$ or $t \neq s$ and .*

Group based covariance structure in Assumption 10(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). This assumption can be relaxed to allow weak dependence among groups 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 conditions in (ii) is likely to be restrictive in many applications, but is useful to develop a test about the appropriate level of grouping in Section 4.

Under Assumption 10(i) and (ii), V_{nT} reduces to

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

respectively, and we can estimate them with

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

Theorem 3 *Suppose that Assumptions 1-9 hold. When $\mathcal{H}_0$ is true, we have*

$$\begin{aligned}\left(\hat{B}_{nT}^{XX}\right)^{-1}\hat{V}_{nT}^c\left(\hat{B}_{nT}^{XX}\right)^{-1} - \left(B_{nT}^{XX}\right)^{-1}V_{nT}^c\left(B_{nT}^{XX}\right)^{-1} &\to^p 0, \text{ and} \\ \hat{\sigma}^2\left(\hat{B}_{nT}^{XX}\right)^{-1} - \sigma^2\left(B_{nT}^{XX}\right)^{-1} &\to^p 0\end{aligned}$$

under Assumption 10(i) and Assumption 10(ii) respectively.

We consider the Wald statistics

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

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

The corollary below follows from Theorems 2 and 3.

Corollary 1 *Suppose that Assumptions 1-9 hold. When $\mathcal{H}_0$ is true, we have*

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

under Assumption 10(i) and Assumption 10(ii) respectively.

4 Testing the level of grouping for group fixed effects

To establish the asymptotics of $\hat{\beta}$, we assume that the group membership to be known. In empirical applications, however, we may have to decide on the appropriate level of grouping among a few of alternatives. For example, when we estimate the effect of a state level policy on the invidual level outcome, it seems natural to introduce the interactive terms using the state level group effects. However, if we suspect the sensitivity to the common time effects to vary within a group, depending on, for example, the employment status or income level, then we should consider a finer level of grouping. To address this practical problem, we develop a Hausman type test about the appropriate level of grouping to specify group fixed effects.

Suppose that two different levels of grouping are available:

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

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. Let G_0 and $G_a = \sum_{g=1}^{G_0}\kappa_g$ denote the numbers of groups under $\mathbb{A}_0$ and $\mathbb{A}_a$ respectively. We assume that the rate conditions in Assumptions 6 and 8 hold for both $\mathbb{A}_0$ and $\mathbb{A}_a$. The null hypothesis in our test is that $\mathbb{A}_0$ is correctly specified against the alternative hypothesis that $\mathbb{A}_0$ is misspecified and only $\hat{\beta}_a$ is consistent. Since $\mathbb{A}_0$ is nested in $\mathbb{A}_a$ with $\lambda_{g_i} = \lambda_g$ for all $i \in \mathcal{A}_g^{(\ell)}$, $\ell = 1, ..., \kappa_g$, $\hat{\beta}_0$ and $\hat{\beta}_a$ are consistent under the null, but the former is less efficient.

Let

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

$$B^{XX}_{a,nT} = \frac{1}{nT}\sum_{i=1}^{n}\left(\mathcal{X}^X_{a,i}\right)'\mathcal{X}^X_{a,i} \text{ and } B^{XX}_a = \operatorname*{plim}_{n,T\to\infty} B^{XX}_{a,nT}$$

where $\bar{X}_{g_i}^{(\ell)}$ denote the average of X_i for $i \in \mathcal{A}_g^{(\ell)}$. We also define $\mathcal{X}_{0,i}^{X}(F)$, $\mathcal{X}_{0,i}^{X}$, $B_{0,nT}^{XX}$ and B_0^{XX} for the model based on $\mathbb{A}_0$ in the same manner. The following result is a direct consequence of Theorem 2.

Corollary 2 *Suppose that Assumptions 1-6, 8-9 hold. Then, under the null hypothesis, we have*

$$\begin{aligned}\sqrt{nT}\left(\hat{\beta}_0 - \hat{\beta}_a\right) &= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\left(B_0^{XX}\right)^{-1}\left(\mathcal{X}_{0,i}^{X}\right)' - \left(B_a^{XX}\right)^{-1}\left(\mathcal{X}_{a,i}^{X}\right)'\right)\varepsilon_i + o_p(1) \\ &\to^d N\left(0, V_{\mathcal{T}}\right)\end{aligned}$$

where

$$V_{\mathcal{T}} = \lim_{n,T\to\infty} Var\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\left(B_0^{XX}\right)^{-1}\left(\mathcal{X}_{0,i}^{X}\right)' - \left(B_a^{XX}\right)^{-1}\left(\mathcal{X}_{a,i}^{X}\right)'\right)\varepsilon_i\right). \tag{22}$$

This result allows us to test the level of grouping. Under Assumption 10(ii), $V_{\mathcal{T}}$ is written as

$$V_{\mathcal{T}} = \lim_{n,T\to\infty}\frac{\sigma^2}{nT}\sum_{i=1}^{n}E\left[\left(\left(B_0^{XX}\right)^{-1}\left(\mathcal{X}_{0,i}^{X}\right)' - \left(B_a^{XX}\right)^{-1}\left(\mathcal{X}_{a,i}^{X}\right)'\right)\left(\left(B_0^{XX}\right)^{-1}\left(\mathcal{X}_{0,i}^{X}\right)' - \left(B_a^{XX}\right)^{-1}\left(\mathcal{X}_{a,i}^{X}\right)'\right)'\right]. \tag{23}$$

Let

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

where $\hat{\mathcal{X}}_{0,i}^{X}$ and $\hat{\mathcal{X}}_{a,i}^{X}$ are defined in the same manner as $\hat{\mathcal{X}}_{i}^{X}$ in (21) except that the former are based on $\mathbb{A}_0$ and $\mathbb{A}_a$ respectively. Based on (23), we introduce the test statistic $\mathcal{T}$ as follows

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

where

$$\begin{aligned}\hat{V}_{\mathcal{T}} &= \frac{\hat{\sigma}_a^2}{nT}\sum_{i=1}^{n}\left(\left(\hat{B}_{0,nT}^{XX}\right)^{-1}\left(\hat{\mathcal{X}}_{0,i}^{X}\right)' - \left(\hat{B}_{a,nT}^{XX}\right)^{-1}\left(\hat{\mathcal{X}}_{a,i}^{X}\right)'\right)\left(\left(\hat{B}_{0,nT}^{XX}\right)^{-1}\left(\hat{\mathcal{X}}_{0,i}^{X}\right)' - \left(\hat{B}_{a,nT}^{XX}\right)^{-1}\left(\hat{\mathcal{X}}_{a,i}^{X}\right)'\right)' \\ &= \hat{\sigma}_a^2\left[\left(\hat{B}_{0,nT}^{XX}\right)^{-1} + \left(\hat{B}_{a,nT}^{XX}\right)^{-1} - \left(\hat{B}_{0,nT}^{XX}\right)^{-1}\hat{C}_{0a,nT}^{XX}\left(\hat{B}_{a,nT}^{XX}\right)^{-1} - \left(\hat{B}_{a,nT}^{XX}\right)^{-1}\hat{C}_{a0,nT}^{XX}\left(\hat{B}_{0,nT}^{XX}\right)^{-1}\right].\end{aligned} \tag{24}$$

We estimate σ^2 using the finer model because it is consistent under the null and alternative hypotheses.

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

$$\frac{1}{nT}\sum_{i=1}^{n}\bar{X}_{g_i}'P_F\bar{X}_{g_i}^{(\ell)} = \frac{1}{nT}\sum_{i=1}^{n}\bar{X}_{g_i}'P_F X_i.$$

If we apply this equality to (23), 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 $\sigma^2 \left(\left(B_0^{XX}\right)^{-1} - \left(B_a^{XX}\right)^{-1} \right)$. 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 based on $\mathbb{A}_0$ and $\mathbb{A}_a$ respectively. The crucial condition to justify $\tilde{\mathcal{T}}$ is that $\hat{F}_0 \approx \hat{F}_a$ from which we can have

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

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 the proof of Theorem 1 Part (ii) in the appendix, the estimator $\hat{F}$ exhibts a slower convergence rate than $\hat{\beta}$, so it is too optimistic to rely on this approximation. In case that this approximation does not work well, $\tilde{\mathcal{T}}$ will suffer from poor finite sample properties. In contrast, $\mathcal{T}$ does not impose such restrictions 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 practical use of variance estimator. $\tilde{V}_{\mathcal{T}}$ may not yields a positive semidefinite estimate. The asymptotic distributions of $\mathcal{T}$ under the null hypothesis and alternative hypothesis are provided as follows:

Theorem 4 *Suppose that Assumptions 1-6, 8-9 and 10(ii) hold. Under the null hypothesis., we have*

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

Under the alternative, for any $C > 0$

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

5 Policy endogeneity with respect to ε_{it}

The validity of LS estimation discussed so 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. In particular, simultaneity often appears between a policy regressor and policy outcome, in which case endogeneity still persists even when the group interactive fixed effects are introduced. To address this, we first propose a control function approach, which we refer to as "interactive fixed effects CF (IFE-CF)" estimation. We provide its implementation procedure and show its consistency. While this method is established based on the group structure, it can be extended to the case where the interactive terms are between individual fixed effects and time effects. In the latter setting, regressor endogeneity with respect to the idiosyncratic error is first considered by Moon, Shum and Weidner (2014) in BLP demand model estimation. They propose an IV approach which they refer to as the "Least Squares-Minimum Distance (LS-MD)" method, and it is extended to the linear model by Moon and Weidner (2010) and Lee, Moon and Weidner (2012).

In addition to the control function method, we also examine the properties of a GMM approach, which we refer to as "interactive fixed effects GMM (IFE-GMM)" estimation. This approach is discussed in Moon, Shum and Weidner (2014) in the BLP demand model estimation, but they do not provide the asymptotics.

In this section, we suppose that $X_{it} = (\mathcal{P}_{g_it}, X_{it}^{ex\prime})'$ where $\mathcal{P}_{g_it}$ denotes a scalar endogenous policy regressor and X_{it}^{ex} is a $((d_x - 1) \times 1)$ vector of exogenous regressors. X_{it}^{ex} is a subvector of instrumental variables Ψ_{g_it} with $\dim(\Psi_{g_it}) = d_\Psi$.

5.1 IFE-CF approach

A crucial issue to develop the control function approach is how to write the reduced form to define the control variable. We assume endogeneity to come from simultaneity of $\mathcal{P}_g$ with the policy outcome. Consider following system of simultaneous equations:

$$\begin{aligned} Y_i &= \mathcal{P}_{g_i}\beta_p^0 + X_i^{ex}\beta_{ex}^0 + F^0\lambda_{g_i}^0 + \varepsilon_i \\ \mathcal{P}_{g_i} &= \bar{Y}_{g_i}\theta_Y^0 + \Psi_{g_i}^{(1)}\theta_1^0 + v_{g_i}. \end{aligned}$$

Combining these two equations, we can have the reduced form as

$$\mathcal{P}_{g_i} = \Psi_{g_i}\Pi^0 + \left(F^0\lambda_{g_i}^0\right)\delta^0 + \eta_{g_i}, \tag{25}$$

where

$$\Pi^0 = \frac{1}{1-\beta_p^0\theta_Y^0}\begin{bmatrix} \beta_{ex}^0\theta_Y^0 \\ \theta_1^0 \end{bmatrix},\ \delta^0 = \frac{\theta_Y^0}{1-\beta_p^0\theta_Y^0},\ \eta_{g_i} = \frac{\varepsilon_{g_i}\theta_Y^0 + v_{g_i}}{1-\beta_p^0\theta_Y^0},$$

and $\Psi_{g_i} = \left[\bar{X}_{g_i}^{ex}, \Psi_{g_i}\right]$. The reduce form in (25) includes the interactive term in (1) as a regressor.

The IFE-CF approach is established based on the following assumption.

Assumption 11 $E\left(\varepsilon_{it}|\mathcal{P}_1, ..., \mathcal{P}_G, \Psi_1^{(1)}, ..., \Psi_G^{(1)}, X_1^{ex}, ..., X_n^{ex}, F^0\Lambda^{0\prime}\right) = E\left(\varepsilon_{it}|\eta_1, ..., \eta_G\right) = \eta_{g_it}'\phi.$

This assumption states mean independence of ε_{it} from the instruments, endogenous policy variable and the interactive terms conditional on η. The IFE-CF approach should have a larger set of candidate instruments than the one based on the model without interactive terms. In the latter case, the error term is the sum of ε_{it} and $\lambda_{g_i}^{0\prime}F_t^0$, so the mean independence 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, because usual instruments used in policy analysis are likely to be correlated with group fixed effects. For example, Besley and Case (2000) examine the effect of state based workers' compensation benefits and use the fraction of women legislators in state lower and upper houses as an instrument for the manual rate. It seems natural to expect such a political variable to be potentially correlated with unobserved state characteristics.

We consider the following estimating equation

$$Y_i = X_i\beta^0 + F^0\lambda^0_{g_i} + \eta_{g_i}\phi + e_i. \tag{26}$$

Since the conditioning variables in Assumption 11 include other individuals' instruments and covariates, $(X_i^{ex}, \mathcal{P}_{g_i}, \eta_{g_i})$ satisfy the strict exogeneity condition with respect to e_{it}. The IEF-CF estimator is given by

$$\begin{aligned}\begin{pmatrix} \hat{\beta}_{cf} \\ \hat{\phi} \end{pmatrix} &= \underset{(\beta',\phi)'}{\operatorname{argmin}} \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[Y_i - P_{\hat{F}_{cf}}\bar{Y}_g - \left(X_i - P_{\hat{F}_{cf}}\bar{X}_g\right)\beta - \left(\tilde{\eta}_i - P_{\hat{F}_{cf}}\tilde{\eta}_g\right)\phi\right]' \\ &\quad\times\left[Y_i - P_{\hat{F}_{cf}}\bar{Y}_g - \left(X_i - P_{\hat{F}_{cf}}\bar{X}_g\right)\beta - \left(\tilde{\eta}_i - P_{\hat{F}_{cf}}\tilde{\eta}_g\right)\phi\right] \\ &= \begin{pmatrix} Q^{XX}_{nT}\left(\hat{F}_{cf}\right) & Q^{X\tilde{\eta}}_{nT}\left(\hat{F}_{cf}\right) \\ Q^{\tilde{\eta}X}_{nT}\left(\hat{F}_{cf}\right) & Q^{\tilde{\eta}\tilde{\eta}}_{nT}\left(\hat{F}_{cf}\right)\end{pmatrix}^{-1}\begin{pmatrix} Q^{XY}_{nT}\left(\hat{F}_{cf}\right) \\ Q^{\tilde{\eta}Y}_{nT}\left(\hat{F}_{cf}\right)\end{pmatrix}\end{aligned} \tag{27}$$

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

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

and

$$\tilde{\eta}_g = \mathcal{P}_g - \Psi_g\tilde{\Pi} - \left(\hat{F}_{cf}\hat{\lambda}_g\right)\tilde{\delta} \tag{29}$$

$\bar{\mathcal{R}}\left(\hat{\beta}_{cf}\right)$ and $\hat{\Gamma}_{cf}$ are defined in the same manner as $\bar{\mathcal{R}}\left(\hat{\beta}\right)$ and $\hat{\Gamma}$ in Section 2.

This estimator can be implemented based on the following steps.

1. Choose an initial estimator $\hat{\beta}^{(1)}$. For example we can use the OLS without including $\tilde{\eta}_i$ and $F\lambda_{g_i}$.

2. Plugging $\hat{\beta}^{(1)}$ in (28) and using (13), we compute $\hat{F}^{(1)}$ and $\hat{\lambda}^{(1)}_{g_i}$.

3. We obtain $\tilde{\eta}^{(1)}_{g_i}$ from (29) by regressing $\mathcal{P}_{g_i}$ on Ψ_{g_i} and $\hat{F}^{(1)}\hat{\lambda}^{(1)}_{g_i}$.

4. We use $\tilde{\eta}^{(1)}_{g_i}$ and $\hat{F}^{(1)}$ to obtain the group interactive fixed effects estimates $\left(\hat{\beta}^{(2)}, \hat{\phi}^{(2)}\right)$ from (27)

5. We iterate steps 2–4 to convergence.

Define $\mathcal{X}_i^v(F)$ and $B_{nT}^{vw}(F)$ for some random variables v and w in the same manner as (15) and (16). Since $\eta_g(F,\Pi) = \mathcal{P}_g - \Psi_g\Pi - (F\lambda_g)\delta$, we can show that

$$\begin{aligned} B_{nT}^{X\eta}(F) &= Q_{nT}^{X\mathcal{P}}(F) - Q_{nT}^{X\Psi}(F)\Pi - \frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_i'M_F\left(\mathcal{P}_{g_j} - \Psi_{g_j}\Pi\right), \\ B_{nT}^{\eta\eta}(F) &= Q_{nT}^{\mathcal{P}\mathcal{P}}(F) - Q_{nT}^{\mathcal{P}\Psi}(F)\Pi - \Pi' Q_{nT}^{\Psi\mathcal{P}}(F) + \Pi' Q_{nT}^{\Psi\Psi}(F)\Pi \\ &\quad - \frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left[\mathcal{P}_{g_i}'M_F\mathcal{P}_{g_j} + \Pi'\Psi_{g_i}'M_F\Psi_{g_j}\Pi\right] \\ &:= D_{nT}^{P\Psi}(F,\Pi). \end{aligned}$$

Let

$$\Xi_{nT}^{X\eta\eta X}(F,\Pi) = B_{nT}^{XX}(F) - B_{nT}^{X\eta}(F)\left[B_{nT}^{\eta\eta}(F)\right]^{-1}B_{nT}^{\eta X}(F).$$

We introduce the following additional assumption.

Assumption 12 *(i)* $E\|\Psi_{it}\|^4 \leq M$. *(ii)* $(nT)^{-1}\sum_{i=1}^{n}\Psi_i'\Psi_i \to_p \Sigma_\Psi > 0$ *and* $d_\Psi \geq d_x$. *(iii) For all bounded* Π *and* $F \in \mathcal{F}$, *we have* $\inf_{(\Pi,F)}\Xi_{nT}^{X\eta\eta X}(F,\Pi) > 0$ *and* $\inf_{(\Pi,F)}E_{nT}^{P\Psi}(F,\Pi) > 0$.

The following theorem states the consistency of the IFE-CF estimator.

Theorem 5 *Suppose that Assumptions 1-5(i), 6, 7 and 11-12 hold. Then, we have*

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

where $H_{cf} = (\Lambda'\Lambda/n)\left(F'\hat{F}_{cf}/T\right)\hat{\Gamma}_{cf}$.

5.2 IFE-GMM approach

We also examine the asymptotics of the GMM method. For this, we first assume that we have a vector of instruments, Ψ_{it}, that satisfy the following conditions.

Assumption 13 *(i)* $E(\varepsilon_{it}|\Psi_1,...,\Psi_n) = 0$ *for all* i *and* t. *(ii) rank*$\left(Q_{nT}^{X\Psi}(F)\right) = d_x$ *for any* $F \in \mathcal{F}$ *and* $d_\Psi \geq d_x$.

The IFE-GMM estimator is given by

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

and

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

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 from Assumption 13, the fixed effects F^0 and Λ^0 are estimated via the LS criterion and principal component method. By concentrating $\hat{\lambda}_g(\beta)$ out, we can rewrite the estimator as

$$\hat{\beta}_{gmm} = \left[Q_{nT}^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_{nT}^{-1}Q_{nT}^{\Psi X}\left(\hat{F}_{gmm}\right)\right]^{-1} Q_{nT}^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_{nT}^{-1}Q_{nT}^{\Psi Y}\left(\hat{F}_{gmm}\right), \tag{30}$$

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

$$\frac{1}{nT}\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)'\hat{F}_{gmm} = \hat{F}_{gmm}\hat{\Gamma}_{gmm}. \tag{31}$$

$\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)$ and $\hat{\Gamma}_{gmm}$ are defined based on $\hat{\beta}_{gmm}$. As in LS estimation, we can obtain $\hat{\beta}_{gmm}$ by iterating (30) and (31) to convergence.

Moon, Shum and Weidner (2014) discuss this approach in the BLP demand model context. They point out that a shortcoming of this estimator is a local minimum problem, which their LS-MD estimator does not suffer from. The potential local minimum is not the unique issue for this method. The LS estimator has the same issue. The LS estimator in Section 2 is a special case of the IFE-GMM method with $\Psi_i = X_i$. We can reduce the chance of reaching a local minimum, by conducting iteration with several initial values of F (when we start from (30)) and β (when we start from (31)).

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

Assumption 14 *(i)* $E\|\Psi_{it}\|^4 \leq M$. *(ii)* $\Omega_{nT} \to^p \Omega$ *and* Ω *is positive definite.*

Theorem 6 *Under Assumptions 1-5(i), 6, 7 and 13-14, 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 $H_{gmm} = (\Lambda'\Lambda/n)\left(F'\hat{F}_{gmm}/T\right)\hat{\Gamma}_{gmm}$.

The proof are in the appendix. Let

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

We make the following high level assumption:

Assumption 15 *Let* $\mathcal{X}_i^{\Psi} = \mathcal{X}_i^{\Psi}(F^0)$. *We have*

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

where $V_{gmm} = \text{plim}_{n,T\to\infty} V_{nT}^{gmm}(F^0)$ *is positive definite.*

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

Theorem 7 *Under Assumptions 1-15, 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

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

6 Monte Carlo simulation

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

$$Y_{it} = \beta_Z^0 Z_{g_i t} + \beta_W^0 W_{it} + u_{it},\ i=1,...,n,\ t=1,...,T,\ g=1,...,G; \tag{32}$$

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

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 (32) based on the following DGP:

$$u_{it} = \lambda_{g_i}^{0\prime} F_t^0 + \gamma_i^0 + \varepsilon_{it}, \tag{33}$$

$$\lambda_g^0 \sim^{iid} U(0,1),\ \gamma_i^0 \sim^{iid} U(0,1), \tag{34}$$

$$F_t^0 = \rho_F F_{t-1}^0 + \sqrt{1-\rho_F^2} v_t^F,\ \text{with } F_1^0, v_t^F \sim^{iid} N(0, I_{d_F}), \tag{35}$$

$$\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it-1} + \sqrt{1-\rho_\varepsilon^2} v_{it}^\varepsilon,\ \text{with } \varepsilon_{i1}, v_{it}^\varepsilon \sim^{iid} N(0,1). \tag{36}$$

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

$$Z_{gt} = \lambda_g^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \lambda_{1,g}^0 + \lambda_{2,g}^0 + v_{gt}^Z,\ \text{with } v_{gt}^Z \sim^{iid} N(0,1), \tag{37}$$

$$W_{it} = \lambda_{g_i}^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \lambda_{1,g_i}^0 + \lambda_{2,g_i}^0 + v_{it}^W,\ \text{with } v_{it}^W \sim^{iid} N(0,1) \tag{38}$$

Both the regressors are correlated with F_t^0, λ_g^0 and $\lambda_g^{0\prime} F_t^0$. We accommodate 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

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

and can estimate (β_Z^0, β_W^0) 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 identical in each group. i.e., $n_g = n/G$ for $g = 1, ..., G$.

Table 1 reports the bias, standard deviation (SD) and empirical rejection probability (ERP) at 5% level. 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 procedure performs very well. In contrast, when our regression model includes only one interactive term

$(d_{\tilde{F}} = 1)$, it does not yield valid estimation and inference results. Moon and Weidner (2015) show that the limiting distribution of standard interactive fixed effects estimator is independent of 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 shows that our estimator may have the same property, though it is not theoretically proved. Table 1 also shows that our procedure shows a relatively large bias and poor size property when G gets 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 well expected, because as G becomes large given n, our group interactive estimator gets close to the standard interactive estimator and the latter 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 increases.

Table 2 compares our method with the standard interactive fixed effects approach. For this simulation, we generate data based on (33)-(38) as the first simulation experiment, but we set $\rho_\varepsilon = 0$ for the standard interactive fixed effects estimator to be asymptotically unbiased. Group based clustering variance estimation is used to calculate the standard errors. Though both estimators are asymtotically valid, difference in their finite sample performance is not trivial. We observe that the standard interactive fixed effects estimator is not accurate 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, and this are much larger than those of the group interactive fixed effects estimator, 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 (32)–(38) 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 the ERP at 5% nominal level. 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 $\mathbb{A}_a$, while the group level regressor is still based on $\mathbb{A}_0$. We generate the regressors based on

$$
\begin{aligned}
Z_{gt} &= \sum_{\ell=1}^{2} \lambda_{g(\ell)}^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \sum_{\ell=1}^{2} \left(\lambda_{1,g(\ell)}^0 + \lambda_{2,g(\ell)}^0 \right) + v_{gt}^Z, \\
W_{it} &= \sum_{\ell=1}^{2} \lambda_{g(\ell)}^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \sum_{\ell=1}^{\kappa_g} \left(\lambda_{1,g_i(\ell)}^0 + \lambda_{2,g_i(\ell)}^0 \right) + v_{it}^W, \\
v_{gt}^Z, v_{it}^W &\sim^{iid} N(0,1),
\end{aligned}
$$

where $\lambda_{g(\ell)}^0 = \left[\lambda_{1,g(\ell)}^0, \lambda_{2,g(\ell)}^0 \right]'$ denote the group fixed effects for $\mathcal{A}_g^{(\ell)}$ and they are generated by

$$\lambda_{1,g(1)}^0, \lambda_{2,g(1)}^0 \sim^{iid} \delta U(0,1), \text{ and } \lambda_{1,g(2)}^0 = -\lambda_{1,g(1)}^0, \lambda_{2,g(2)}^0 = -\delta \lambda_{2,g(1)}^0. \tag{39}$$

The DGP (39) implies that $\mathcal{A}_g^{(1)}$ and $\mathcal{A}_g^{(2)}$ respond to the time effects in the opposite way. δ represent the degree of heterogeneity of the group fixed effects between these two subgroups. Table 3 shows nontrivial power against the alternatives based on (39). The power improves as the degree of heterogeneity between $\mathcal{A}_g^{(1)}$ and $\mathcal{A}_g^{(2)}$ grows.

Table 4 compares finite sample performance of the LS estimator and IFE-GMM estimator in the presence of endogeneity with respect to ε_{it}. For the IFE-GMM estimator, we follow the idea of the 2SLS estimation to use $Q_{nT}^{\Psi\Psi}\left(\hat{F}_{gmm}\right)$ as the weighting matrix (i.e., $\Omega_{nT} = Q_{nT}^{\Psi\Psi}\left(\hat{F}_{gmm}\right)$). We generate data from (33)–(36) 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

$$Z_{gt} = \Psi_{gt}^{Z} + 0.5\bar{\varepsilon}_{gt} + v_{gt}^{Z}, \text{ with } v_{gt}^{Z} \sim^{iid} N(0,1), \tag{40}$$
$$W_{it} = \Psi_{it}^{W} + 0.5\varepsilon_{it} + v_{it}^{W}, \text{ with } v_{it}^{W} \sim^{iid} N(0,1), \tag{41}$$

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 IFE-GMM produces accurate estimates and ERPs in the presence of endogeneity while the LS estimator suffers from relatively large bias and serious size distortion.

Table 5 compares the IFE-CF estimator and IFE-GMM estimator in the presence of endogeneity with respect to ε_{it}. We also generate the group level regressor and instrument based on

$$Z_{gt} = \Psi_{gt}^{Z} + \bar{Y}_{gt} + v_{gt}^{Z} \text{ with } v_{gt}^{Z} \sim^{iid} N(0,1),$$
$$\Psi_{gt}^{Z} = \lambda_g^{0\prime} F_t^0 + F_{1,t}^0 + F_{2,t}^0 + \lambda_{1,g}^0 + \lambda_{2,g}^0 + v_{gt}^{\Psi}, \text{ with } v_{gt}^{\Psi} \sim^{iid} N(0,1).$$

We can see that endogeneity is generated by simultaneity between the group level regressor and dependent variable. We also employ (32)-(36), (38) not including γ_i to generate the other variables. From this table, we can see that both the IFE-CF and IFE-GMM address the endogeneity well. Compared to the LS estimator, these methods yield smaller bias and RMSE, and become more accurate as n and T increase. Between the IFE-CF and IFE-GMM estimators, they are comparable in terms of RMSE.

7 Empirical illustration: competition policy and productivity growth

There is broad consensus in economics that competition tends to enhance economic efficiency, but no such agreement is on the effectiveness of competition policy. See, for example, Baker (2003) and Werden (2003) who argue that the benefit of antitrust enforcement for the economy is larger than the cost, while Crandall and Winston (2003) claim that antitrust law has been ineffective in the US. In this regard, Buccirossi, et al. (2013) examine the importance of competition policy to improve 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 all the key features of competition policy of a country.

We revisit the evidence by Buccirossi, et al. (2013) using the group interactive fixed effects estimation approach. The regression model considered in this section is

$$\Delta TFP_{it} = \beta_0 CPI_{g_i t-1} + \beta_1 \Delta TFP_{L(i),t} + \beta_2 \frac{TFP_{L(i),t}}{TFP_{it}} + \beta_3 W_{it-1} + \beta_4 Z_{g_i t-1} + \lambda_{g_i}' F_t + \alpha_i + \varepsilon_{it}, \quad (42)$$

where ΔTFP_{it} is 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 the country on the technological frontier and productivity gap to the technological frontier respectively. W_{it} is a vector of country-industry specific covariates (trade openness and country-industry specific trend) and $Z_{g_i t}$ denotes country specific control (product market regulation). Including individual fixed effects, α_i, (42) also control for unobserved industry level heterogeneity.

The model is estimated based on the 1995-2005 balanced subsample of Buccirossi, et al. (2013). It consists of 22 industries in 7 countries (Czech Republic, Germany, Italy, Japan, Sweden, UK and US),[2] and the industry classification is based on ISIC Rev.3. For more details on data description, refer to on Buccirossi, et al. (2013).

We first conduct the test on 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 there are 7 groups each of which contains 22 industries. The finer grouping under the alternative, $\mathbb{A}_a$, make two subgroups in each country based on the manufacturing sector and non-manufacturing sector. According to the ISIC Rev 3., 12 out of 22 industries belong to the manufacturing sector, and the others to the non-manufacturing sector. The result is presented in the table below.

<Test on the level of grouping>

$\mathbb{H}_0$: Country level of grouping is correctly specified.				
$\mathbb{H}_a$: 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	$\chi^2_{0.95}(6) = 12.59$			

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

We first estimate (42) using our group interactive fixed effects LS estimation. For comparison, we also consider the following additive fixed effects regression model

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

[2] Data is available at https://dataverse.harvard.edu/dataverse/restat

that Beuccirossi, et al. (2013) use. Table 6 report the the coefficients and their t-statistics based on the country based clustering standard errors. We find that the magnitude of coefficients for CPI with our LS estimation are clearly smaller than the one with the additive fixed effects estimation. While the latter is significant at 10% level, the former are not. Thus, we may conclude that significance of the coefficient for CPI in the additive fixed effects model may be due to endogeneity associated with time varying country heterogeneity. Table 7 reports the coefficients and their t-statistics using IFE-GMM and IFE-CF methods as well as 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). Beuccirossi, et al. (2013) use the same instruments as we do. The qualitative results are the same for the LS estimation case. Compared to the additive fixed effects approach, the magnitudes of the coefficients for CPI and the degrees of their significances are reduced when the interactive effects are considered.

8 Conclusion

The panel and multilevel regression model is very useful to study the effect of a group level policy on individual outcome. 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 the time effects are common across groups. However, this assumption may not be met in many applications. This paper proposes the interactive fixed effects model in the multilevel setting. This approach accounts for group specific time effects as well as time varying group heterogeneity. To decide on the appropriate group structure, we propose a Hausman type test that compares two alternative levels of grouping. We also develop the control function estimator and examine the properties of the GMM method to address the endogeneity of a policy variable 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 ($\tilde{d}_F = 2$)					
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
		Individual Interactive ($\tilde{d}_F = 2$)					
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

To compute the power, we generate the data based on (39).

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 ($\tilde{d}_F = 2$)					
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 ($\tilde{d}_F = 3$)					
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: Comparison between the LS estimator, IFE-CF estimator and IFE-GMM estimator in the presence of endogeneity with respect to idiosyncratic errors

$\beta_Z = 0$, $G = 50$, $d_F = 2$							
n	T	Bias	SD	RMSE	Bias	SD	RMSE
Group Interactive							
		$\tilde{d}_F = 2$			$\tilde{d}_F = 3$		
1000	5	0.023	0.018	0.029	0.015	0.014	0.020
1000	10	0.023	0.010	0.025	0.018	0.009	0.020
1000	20	0.023	0.006	0.024	0.020	0.006	0.021
2000	5	0.011	0.011	0.016	0.007	0.010	0.012
2000	10	0.011	0.006	0.013	0.009	0.006	0.011
2000	20	0.011	0.004	0.012	0.010	0.004	0.011
IFE-CF							
		$\tilde{d}_F = 2$			$\tilde{d}_F = 3$		
1000	5	0.009	0.020	0.022	0.007	0.016	0.018
1000	10	0.006	0.012	0.013	0.008	0.010	0.013
1000	20	0.005	0.007	0.009	0.007	0.007	0.010
2000	5	0.003	0.013	0.013	0.004	0.011	0.012
2000	10	0.003	0.007	0.008	0.004	0.007	0.008
2000	20	0.002	0.005	0.005	0.003	0.005	0.006
IFE-GMM							
		$\tilde{d}_F = 2$			$\tilde{d}_F = 3$		
1000	5	0.003	0.022	0.022	0.001	0.019	0.019
1000	10	0.001	0.013	0.013	0.001	0.012	0.012
1000	20	0.000	0.008	0.008	0.000	0.008	0.008
2000	5	0.001	0.014	0.014	0.000	0.014	0.014
2000	10	0.000	0.008	0.008	−0.000	0.008	0.008
2000	20	0.000	0.005	0.005	−0.000	0.005	0.005

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

Dep Var	ΔTFP_{it}				
	Interactive				Additive
d_F	1	2	3	4	
CPI_{g_it-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)
$\left(TFP_{L(i),t}/TFP_{it}\right)$	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 7: The effect of competition policy on TFP growth: IV estimation

Dep Var	ΔTFP_{it}								
	Interactive								Additive
	GMM				CF				
d_F	1	2	3	4	1	2	3	4	
CPI_{g_it-1}	0.203	0.178	0.122	0.174	0.116	0.052	0.020	0.065	0.472
	(1.064)	(1.180)	(1.163)	(1.547)					(1.913)
$\Delta TFP_{L,it}$	0.067	0.069	0.069	0.065	0.072	0.078	0.076	0.074	0.069
	(1.780)	(1.767)	(1.806)	(1.702)					(1.874)
$\left(TFP_{L(i),t}/TFP_{it}\right)$	0.014	0.014	0.014	0.014	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	0.106	0.097	0.099	0.100	0.108
	(8.962)	(9.555)	(9.436)	(8.417)					(6.457)
Import penetration	0.013	0.014	0.014	0.015	0.014	0.015	0.015	0.016	0.016
	(1.579)	(1.705)	(1.702)	(1.907)					(1.590)
pmr	0.020	0.023	0.021	0.024	0.016	0.013	0.013	0.011	−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. Part (i) Suppose that $\beta^0 = 0$ without loss of generality. $\left(\hat{\beta}, \hat{F}\right)$ minimizes

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

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

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

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

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

We first show that

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

for all bounded β and $F \in \mathcal{F} = \{F : F'F/T = I_{d_F}\}$ under the rate conditions in Theorem 1. Note that $\sum_{i=1}^{n}\lambda_{g_i}\bar{\varepsilon}_{g_i}' = \sum_{g=1}^{G} n_g \lambda_g \bar{\varepsilon}_g'$. For the first term, we have

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

Assumption 4(ii) implies $a1 = O_p\left((nT)^{-1/2}\right)$. For $a2$,

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

and

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

Thus, $a1$ and $a2$ are $o_p(1)$ under the rate conditions in Theorem 1. Using the same procedure, we can show that the second term and third term in (A.1) are also $o_p(1)$. Therefore, (A.1) holds.

Second, we should show

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

for any $(\beta,F)\neq(\beta^0,F^0H)$. Let

$$\theta=vec\left(M_FF^0\right)-L^{-1}C\beta,$$
$$L=\left(\frac{\Lambda\Lambda'}{n}\otimes I_T\right),\ C=\frac{1}{nT}\sum_{i=1}^{n}\left(\lambda_{g_i}'\otimes M_FX_i\right).$$

Since

$$\frac{1}{nT}\sum_{i=1}^{n}\lambda_{g_i}'F^{0\prime}\left(X_i-P_F\bar{X}_{g_i}\right)\beta=\frac{1}{nT}\sum_{i=1}^{n}\lambda_{g_i}'F^{0\prime}M_FX_i\beta,$$

(A.2) can be rewritten as

$$\begin{aligned}\mathcal{Q}^*(\beta,F)&=\frac{1}{nT}\sum_{i=1}^{n}\lambda_{g_i}'F^{0\prime}M_FF^0\lambda_{g_i}-\frac{2}{nT}\sum_{i=1}^{n}\lambda_{g_i}'F^{0\prime}M_FX_i\beta\\&+\beta'\frac{1}{nT}\sum_{i=1}^{n}\left(X_i-P_F\bar{X}_{g_i}\right)'\left(X_i-P_F\bar{X}_{g_i}\right)\beta\\&=\beta'B_{nT}^{XX}(F)\beta+\theta'L\theta,\end{aligned} \tag{A.3}$$

where the second equality is shown in Bai (2009, pp1264-1265). Assumption 1 states $B_{nT}^{XX}(F)$ is positive definite, and L is also positive definite. Therefore, $\mathcal{Q}^*(\beta,F)\geq 0$ and $\mathcal{Q}^*(\beta,F)>0$ if $(\beta,F)\neq(\beta^0,F^0H)$, which completes the proof of Part (i).

Part (ii) In this proof, we show

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

where $H=(\Lambda'\Lambda/n)\left(F'\hat{F}/T\right)\hat{\Gamma}^{-1}$.

From (12), we have

$$
\begin{aligned}
&\hat{F}\hat{\Gamma} - F^0\left(\frac{\Lambda'\Lambda}{n}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right) \\
&= \frac{1}{nT}\sum_{i=1}^{n}\bar{X}_{g_i}\left(\beta^0-\hat{\beta}\right)\left(\beta^0-\hat{\beta}\right)'\bar{X}_{g_i}'\hat{F} \\
&+ \frac{1}{nT}\sum_{i=1}^{n}\bar{X}_{g_i}\left(\beta^0-\hat{\beta}\right)\lambda_{g_i}'F^{0\prime}\hat{F} + \frac{1}{nT}\sum_{i=1}^{n}\bar{X}_{g_i}\left(\beta^0-\hat{\beta}\right)\bar{\varepsilon}_{g_i}'\hat{F} \\
&+ \frac{1}{nT}\sum_{i=1}^{n}F^0\lambda_{g_i}\left(\beta^0-\hat{\beta}\right)'\bar{X}_{g_i}'\hat{F} + \frac{1}{nT}\sum_{i=1}^{n}\bar{\varepsilon}_{g_i}\left(\beta^0-\hat{\beta}\right)'\bar{X}_{g_i}'\hat{F} \\
&+ \frac{1}{nT}\sum_{i=1}^{n}F^0\lambda_{g_i}\bar{\varepsilon}_{g_i}'\hat{F} + \frac{1}{nT}\sum_{i=1}^{n}\bar{\varepsilon}_{g_i}\lambda_{g_i}'F^{0\prime}\hat{F} + \frac{1}{nT}\sum_{i=1}^{n}\bar{\varepsilon}_{g_i}\bar{\varepsilon}_{g_i}'\hat{F} \\
&= I1 + I2 + \cdots + I8 \qquad \text{(A.4)}
\end{aligned}
$$

We multiply $\left(F^{0\prime}\hat{F}/T\right)^{-1}(\Lambda'\Lambda/n)^{-1}$ to obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|\hat{F}\hat{\Gamma}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{n}\right)^{-1} - F^0\right\| \\
&\leq \frac{1}{\sqrt{T}}\left(\|I1\| + \cdots + \|I8\|\right)\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{n}\right)^{-1}\right\|.
\end{aligned}
$$

For $I1$,

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

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

For $I6$, we have

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

under Assumption 4(1). In the same way, we can show that $T^{1/2}\|I7\| = O_p\left(n^{-1/2}\right)$.

For $I8$, we have

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

For the first term,

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

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

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

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

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

$$\frac{1}{\sqrt{T}}\|I82\| \leq \left\|\frac{G}{nT}\frac{1}{G}\sum_{g=1}^{G}n_g E\left(\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\right\|\frac{\left\|\hat{F}\right\|}{\sqrt{T}}$$
$$\leq \frac{G}{n\sqrt{T}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\frac{1}{G}\sum_{g=1}^{G}n_g E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{gs}\right)\right]^2}\frac{\left\|\hat{F}\right\|}{\sqrt{T}}$$
$$= O\left(\frac{G}{n\sqrt{T}}\right).$$

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

$$\frac{1}{\sqrt{T}}\|I8\| = O_p\left(\frac{\sqrt{G}}{n} + \frac{G}{nT}\right). \tag{A.5}$$

Combining $I1$-$I8$, we obtain

$$\frac{1}{\sqrt{T}}\left(\hat{F}\hat{\Gamma}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{n}\right)^{-1} - F^0\right)$$
$$= O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n\sqrt{T}}\right). \tag{A.6}$$

Premultiplying $\hat{F}'/\sqrt{T}$ in (A.6), we obtain

$$\hat{\Gamma} = \left(\frac{\hat{F}'F^0}{T}\right)\left(\frac{\Lambda'\Lambda}{n}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right) + o_p(1). \tag{A.7}$$

As shown in Bai (2009), $F^{0\prime}\hat{F}/T$ is invertible, so $\hat{\Gamma}$ is invertible. Thus, from (A.6) 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, for each g*

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

Proof. We have

$$\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}=\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(I1+\cdots+I8\right)\hat{\Gamma}^{-1}}{T}$$
$$=\mathcal{I}1+\cdots+\mathcal{I}8.$$

It is easy to show that $\|\mathcal{I}1\|=\cdots=\|\mathcal{I}4\|=O_p\left(T^{-1/2}\left\|\beta-\hat{\beta}\right\|\right)$. For $\mathcal{I}5$,

$$\|\mathcal{I}5\|\leq\sqrt{\frac{G}{nT}}\left(\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_gn_{\tilde{g}}}\left[\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}-E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}\right)\right]\right\|^2\right)^{1/2}\left\|\beta^0-\hat{\beta}\right\|\left(\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left\|\frac{\bar{X}_{\tilde{g}}'\hat{F}}{T}\right\|^2\right)^{1/2}\left\|\hat{\Gamma}^{-1}\right\|$$
$$+\frac{1}{\sqrt{nG}}\left\|\beta-\hat{\beta}\right\|\left\|\frac{1}{T}\sum_{\tilde{g}=1}^{G}\sum_{t=1}^{T}\sqrt{n_gn_{\tilde{g}}}E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}\right)\frac{\bar{X}_{\tilde{g}}'\hat{F}}{T}\right\|\left\|\hat{\Gamma}^{-1}\right\|$$
$$=O_p\left(\left(\sqrt{\frac{G}{nT}}+\frac{1}{\sqrt{nG}}\right)\left\|\beta^0-\hat{\beta}\right\|\right).$$

For $\mathcal{I}6$,

$$\|\mathcal{I}6\|\leq\left\|\frac{1}{T}\sqrt{n_g}\bar{\varepsilon}_g'\left(\frac{1}{nT}\sum_{i=1}^{n}F^0\lambda_{g_i}\bar{\varepsilon}_{g_i}'F^0H\right)\hat{\Gamma}^{-1}\right\|$$
$$+\left\|\frac{1}{T}\sqrt{n_g}\bar{\varepsilon}_g'\left(\frac{1}{nT}\sum_{i=1}^{n}F^0\lambda_{g_i}\bar{\varepsilon}_{g_i}'\left(\hat{F}-F^0H\right)\right)\hat{\Gamma}^{-1}\right\|$$
$$=\mathcal{I}61+\mathcal{I}62,$$

$$\mathcal{I}61\leq\frac{1}{\sqrt{nT}}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_g}\bar{\varepsilon}_{gt}F_t^0\right\|\left\|\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{s=1}^{T}\sqrt{n_{\tilde{g}}}\lambda_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}s}F_s^{0\prime}\right\|\|H\|\left\|\hat{\Gamma}^{-1}\right\|O(1)$$
$$=O_p\left(\frac{1}{\sqrt{nT}}\right),$$

$$\mathcal{I}62\leq\frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_g}\bar{\varepsilon}_{gt}F_t^{0\prime}\right\|\left(\left(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_{g_i}\|^2\right)^{1/2}\left(\frac{1}{G}\sum_{\tilde{g}=1}^{G}\frac{n_{\tilde{g}}\left\|\bar{\varepsilon}_{\tilde{g}}'\right\|^2}{T}\right)^{1/2}\right)\frac{\left\|\hat{F}-F^0H\right\|}{\sqrt{T}}\left\|\hat{\Gamma}^{-1}\right\|O(1)$$
$$=O_p\left(\frac{1}{\sqrt{T}}\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{G}{nT}\right).$$

Thus,

$$\mathcal{I}6 = O_p\left(\frac{1}{\sqrt{T}}\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{G}{nT}\right).$$

For $\mathcal{I}7$,

$$\begin{aligned}\|\mathcal{I}7\| &= \frac{1}{\sqrt{nT}}\left(\frac{1}{\sqrt{GT}}\sum_{\tilde{g}=1}^{G}\sum_{t=1}^{T}\sqrt{n_g n_{\tilde{g}}}\left[\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}-E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}\right)\right]\lambda_{\tilde{g}}'\frac{F^{0\prime}\hat{F}}{T}\right)\hat{\Gamma}^{-1}\\
&+\frac{1}{\sqrt{nG}}\left(\frac{1}{T}\sum_{\tilde{g}=1}^{G}\sum_{t=1}^{T}\sqrt{n_g n_{\tilde{g}}}E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}\right)\lambda_{\tilde{g}}'\frac{F^{0\prime}\hat{F}}{T}\right)\hat{\Gamma}^{-1}\\
&= O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{1}{\sqrt{nG}}\right).\end{aligned}$$

For $\mathcal{I}8$,

$$\begin{aligned}\mathcal{I}8 &= \frac{1}{T}\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_g n_{\tilde{g}}}\left[\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}-E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}\right)\right]\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}s}F_s^{0\prime}\right)H\hat{\Gamma}^{-1}\\
&+\frac{\sqrt{G}}{n\sqrt{T}}\frac{1}{\sqrt{G}}\sum_{\tilde{g}=1}^{G}\left(\frac{1}{T}\sum_{t=1}^{T}\sqrt{n_g n_{\tilde{g}}}E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}\right)\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}s}F_s^{0\prime}\right)H\hat{\Gamma}^{-1}\\
&+\frac{G}{n\sqrt{T}}\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_g n_{\tilde{g}}}\left[\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}-E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}\right)\right]\right)\frac{\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}'}{\sqrt{T}}\frac{\left(\hat{F}-FH^0\right)}{\sqrt{T}}\hat{\Gamma}^{-1}\\
&+\frac{\sqrt{G}}{n}\frac{1}{\sqrt{G}}\sum_{\tilde{g}=1}^{G}\left(\frac{1}{T}\sum_{t=1}^{T}\sqrt{n_g n_{\tilde{g}}}E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}t}\right)\right)\frac{\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}'}{\sqrt{T}}\frac{\left(\hat{F}-FH^0\right)}{\sqrt{T}}\hat{\Gamma}^{-1}\\
&= O_p\left(\left(\frac{G}{n\sqrt{T}}+\frac{\sqrt{G}}{n}\right)\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{T}\right)+O_p\left(\frac{\sqrt{G}}{n\sqrt{T}}\right)+O_p\left(\frac{\sqrt{G}}{n\sqrt{n}}\right).\end{aligned}$$

Combining $\mathcal{I}1-\mathcal{I}8$, we have

$$\begin{aligned}\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T} &= O_p\left(\frac{1}{\sqrt{T}}\left\|\beta^0-\hat{\beta}\right\|\right)+O_p\left(\frac{1}{\sqrt{nG}}\left\|\beta^0-\hat{\beta}\right\|\right)+O_p\left(\frac{\sqrt{G}}{n}\left\|\hat{\beta}-\beta^0\right\|\right)\\
&+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{1}{\sqrt{nG}}\right)+O_p\left(\frac{1}{T}\right).\end{aligned}$$

■

Proposition A1 *Let*

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

with $\Upsilon = \left(F^{0\prime}\hat{F}/T\right)^{-1}(\Lambda'\Lambda/n)^{-1}$. *Under Assumptions 1-6,*

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

Proof. Note that

$$\begin{aligned}
\hat{\beta}-\beta^0 &= \left[\frac{1}{nT}\sum_{i=1}^{n}\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)'\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)\right]^{-1} \\
&\quad \times \frac{1}{nT}\sum_{i=1}^{n}\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)'\left(M_{\hat{F}}F^0\lambda_{g_i} + \varepsilon_i - P_{\hat{F}}\bar{\varepsilon}_{g_i}\right) \\
&= \left[\frac{1}{nT}\sum_{i=1}^{n}\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)'\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)\right]^{-1}\frac{1}{nT}\sum_{i=1}^{n}\Bigg(X_i'M_{\hat{F}}F^0\lambda_{g_i} \\
&\quad + \left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)'\varepsilon_i\Bigg) \qquad \text{(A.8)}
\end{aligned}$$

Let $\Upsilon = \left(F^{0\prime}\hat{F}/T\right)^{-1}(\Lambda'\Lambda/n)^{-1}$. For the first term of (A.8), since $M_{\hat{F}}\hat{F}=0$, we have

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

For $J1$,

$$\begin{aligned}
\|J1\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\|X_i'\|^2}{T}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_{g_i}\|^2\right)^{1/2}\frac{1}{\sqrt{T}}\|I1\|\,\|\Upsilon\| \\
&= o_p\left(\left\|\hat{\beta}-\beta\right\|\right) \qquad \text{(A.9)}
\end{aligned}$$

in which we use

$$\|X_i' M_{\hat{F}}\|^2 = tr\left(X_i' X_i\right) - tr\left(X_i' \hat{F}\left(\hat{F}'\hat{F}\right)^{-1} \hat{F}' X_i\right)$$
$$= \|X_i\|^2 - \frac{1}{T}\left\|\hat{F}' X_i\right\|^2 \leq \|X_i'\|^2. \qquad \text{(A.10)}$$

For $J2$, we have

$$J2 = \frac{1}{n^2 T}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i' M_{\hat{F}} X_j \left\{\lambda_{g_j}'\left(\frac{\Lambda'\Lambda}{n}\right)^{-1}\lambda_{g_i}\right\}\left(\hat{\beta} - \beta^0\right), \qquad \text{(A.11)}$$

where

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

For $J3$,

$$\|J3\| = \left\|\frac{1}{n^2 T}\sum_{i=1}^{n} X_i' M_{\hat{F}} \lambda_{g_i}' \Upsilon' \sum_{j=1}^{n}\left(\frac{\hat{F}'\bar{\varepsilon}_{g_j}}{T}\right) X_j \left(\beta^0 - \hat{\beta}\right)\right\|$$
$$\leq \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\|X_i'\|^2}{T}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_{g_i}\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{g=1}^{G}\left\|\frac{\sqrt{n_g}\bar{\varepsilon}_g'\hat{F}}{T}\right\|^2\right)^{1/2}$$
$$\times \left\|\beta - \hat{\beta}\right\| \|\Upsilon\|,$$

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

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

and

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

by Assumption 4(i). Using a similar way, we can show that

$$\frac{1}{n}\sum_{g=1}^{G}\left\|\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F} - F^0 H\right)}{T}\right\| = o_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{1}{\sqrt{nG}}\right) + O_p\left(\frac{1}{T}\right). \qquad \text{(A.12)}$$

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

For $J4$,

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

$$= o_p\left(\left\|\beta^0 - \hat{\beta}\right\|\right).$$

For $J5$,

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

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

$$= o_p\left(\left\|\beta^0 - \hat{\beta}\right\|\right),$$

because

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

by Assumption 4(ii). Using a similar way, we can show that $J52 = o_p\left(\left\|\beta^0 - \hat{\beta}\right\|\right)$.

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

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

and

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

$$= O_p\left(\frac{1}{\sqrt{n}}\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) \tag{A.13}$$

because

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

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

Therefore,

$$\|J6\| = O_p\left(\frac{1}{\sqrt{n}}\left\|\hat{\beta} - \beta^0\right\|^2\right) + O_p\left(\frac{1}{n\sqrt{n}}\right) + O_p\left(\frac{G}{nT\sqrt{n}}\right) + O_p\left(\frac{G}{n^2\sqrt{T}}\right).$$

For $J7$, since $a_{ij} = \lambda_{g_j}'\left(\Lambda'\Lambda/n\right)^{-1}\lambda_{g_i}$ is a scalar,

$$J7 = -\frac{1}{n^2 T}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_i' M_{\hat{F}}\varepsilon_j.$$

Let

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

Then, $J8$ can be rewritten as

$$\begin{aligned}
J8 &= -\frac{G}{(nT)^2}\sum_{i=1}^{n} X_i' M_{\hat{F}}\frac{1}{G}\sum_{g=1}^{G} n_g\left(\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\hat{F}\Upsilon\lambda_{g_i} \\
&= -\left(\frac{G}{nT}\right)A_{nT}^{(1)} \\
&- \frac{G}{(nT)^2}\sum_{i=1}^{n} X_i'\left(M_{\hat{F}} - M_{F^0}\right)\left(\frac{1}{G}\sum_{g=1}^{G} n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\left(\hat{F} - F^0 H\right)\Upsilon\lambda_{g_i} \\
&- \frac{G}{(nT)^2}\sum_{i=1}^{n} X_i'\left(M_{\hat{F}} - M_{F^0}\right)\left(\frac{1}{G}\sum_{g=1}^{G} n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)F^0 H\Upsilon\lambda_{g_i} \\
&- \frac{G}{(nT)^2}\sum_{i=1}^{n} X_i' M_{F^0}\left(\frac{1}{G}\sum_{g=1}^{G} n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\left(\hat{F} - F^0 H\right)\Upsilon\lambda_{g_i} \\
&= J81 + J82 + J83 + J84.
\end{aligned}$$

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

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

Using the same steps, we can show that $a12$ is also $O_p\left(G/\left(nT\right)\right)$.

$$
\begin{aligned}
J82 &= \left(\frac{G}{nT}\right)\frac{1}{nT}\sum_{i=1}^{n}X_i'\left(\frac{\hat{F}\hat{F}'}{T}-F^0\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\right)\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)\Upsilon\lambda_{g_i} \\
&= \left(\frac{G}{n\sqrt{T}}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{X_i'}{\sqrt{T}}\left(\frac{\hat{F}-F^0H}{\sqrt{T}}\right)H'\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0\sqrt{n_g}\bar{\varepsilon}_{gt}\right)\left(\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\right)\Upsilon\lambda_{g_i} \\
&+\left(\frac{G}{n}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{X_i'}{\sqrt{T}}\frac{\left(\hat{F}-F^0H\right)}{\sqrt{T}}\frac{1}{G}\sum_{g=1}^{G}\frac{\left(\hat{F}-F^0H\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T}\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\Upsilon\lambda_{g_i} \\
&+\left(\frac{G}{n}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{X_i'F^0H}{T}\frac{1}{G}\sum_{g=1}^{G}\frac{\left(\hat{F}-F^0H\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T}\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\Upsilon\lambda_{g_i} \\
&+\left(\frac{G}{n\sqrt{T}}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{X_i'F^0}{T}\left[HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0\sqrt{n_g}\bar{\varepsilon}_{gt}\right) \\
&\times\left(\frac{\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F}-F^0H\right)}{T}\right)\Upsilon\lambda_{g_i} \\
&= o_p\left(\left\|\beta^0-\hat{\beta}\right\|\right)+O_p\left(\frac{G}{n^2T}\right)+O_p\left(\frac{1}{n^2}\right)+O_p\left(\frac{G}{nT^2}\right)+O_p\left(\frac{\sqrt{G}}{n^2\sqrt{T}}\right)
\end{aligned}
$$

For $J83$,

$$\|J83\| \leq \left(\frac{G}{n\sqrt{T}}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{X_i'\hat{F}}{T}\right\|^2\right)^{1/2}\left(\frac{1}{G}\sum_{g=1}^{G}\left\|\frac{\left(F^0H-\hat{F}\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T}\right\|^2\right)^{1/2}$$

$$\times\left(\frac{1}{G}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\sqrt{n_g}\bar{\varepsilon}_{gs}F_s^{0\prime}\right\|^2\right)^{1/2}\|H\|\,\|\Upsilon\|\left(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_{g_i}\|^2\right)^{1/2}$$

$$+\left(\frac{G}{nT}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\|X_i'\|^2\right)^{1/2}\left\|F^0H\left(HF^{0\prime}F^0H\right)^{-1}-\frac{\hat{F}}{T}\right\|\|H'\|\left(\frac{1}{G}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0\sqrt{n_g}\bar{\varepsilon}_{gt}\right\|^2\right)^{1/2}$$

$$\times\left(\frac{1}{G}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\sqrt{n_g}\bar{\varepsilon}_{gs}F_s^{0\prime}\right\|^2\right)^{1/2}\|H\|\,\|\Upsilon\|\left(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_{g_i}\|^2\right)^{1/2}$$

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

For $J84$

$$J84 = \frac{G}{(nT)^2}\sum_{i=1}^{n}X_i'\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\left(\hat{F}-F^0H\right)\Upsilon\lambda_{g_i}$$

$$+\frac{G}{(nT)^2}\sum_{i=1}^{n}X_i'F^0\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\left(\frac{1}{G}\sum_{g=1}^{G}n_g\bar{\varepsilon}_g\bar{\varepsilon}_g'\right)\left(\hat{F}-F^0H\right)\Upsilon\lambda_{g_i}$$

$$= a41 + a42,$$

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

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

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

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

$$J8 = J81 + o_p\left(\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{n^2}\right)+O_p\left(\frac{G}{n\sqrt{nT}}\right)+O_p\left(\frac{\sqrt{G}}{n\sqrt{nT}}\right)+O_p\left(\frac{G}{nT\sqrt{T}}\right)$$

Combining $J1$-$J8$, we have

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

Combining this result with (A.8), under the rate conditions in Assumption 6, we have

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

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

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

■

Lemma A2 *Under Assumptions 1-5 and 6(ii),*

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

Proof of Lemma A2.

$$HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} = H\left\{\frac{H'F^{0\prime}F^0H}{T} - I\right\}\left(\frac{H'F^{0\prime}F^0H}{T}\right)^{-1}H'$$
$$= H\left\{\frac{1}{T}H'F^{0\prime}\left(F^0H-\hat F\right) + \frac{1}{T}\left(F^0H-\hat F\right)'\left(\hat F - F^0H\right)\right.$$
$$\left.+\frac{1}{T}\left(F^0H-\hat F\right)'F^0H\right\}\left(\frac{H'F^{0\prime}F^0H}{T}\right)^{-1}H', \tag{A.14}$$

$$
\begin{aligned}
\frac{1}{T} F^{0\prime}\left(F^{0} H-\hat{F}\right) &= \frac{1}{T} F^{0\prime}\left(F^{0} H \hat{\Gamma}-\hat{F} \hat{\Gamma}\right) \hat{\Gamma}^{-1} \\
&= \frac{1}{T} F^{0\prime}\left(\frac{1}{n T} \sum_{i=1}^{n} \bar{X}_{g_i}\left(\beta^{0}-\hat{\beta}\right)\left(\beta^{0}-\hat{\beta}\right)^{\prime} \bar{X}_{g_i}^{\prime} \hat{F}\right. \\
&+\frac{1}{n T} \sum_{i=1}^{n} \bar{X}_{g_i}\left(\beta^{0}-\hat{\beta}\right) \lambda_{g_i}^{\prime} F^{0\prime} \hat{F}+\frac{1}{n T} \sum_{i=1}^{n} \bar{X}_{g_i}\left(\beta^{0}-\hat{\beta}\right) \bar{\varepsilon}_{g_i}^{\prime} \hat{F} \\
&+\frac{1}{n T} \sum_{i=1}^{n} F^{0} \lambda_{g_i}\left(\beta^{0}-\hat{\beta}\right)^{\prime} \bar{X}_{g_i}^{\prime} \hat{F}+\frac{1}{n T} \sum_{i=1}^{n} \bar{\varepsilon}_{g_i}\left(\beta^{0}-\hat{\beta}\right)^{\prime} \bar{X}_{g_i}^{\prime} \hat{F} \\
&\left.+\frac{1}{n T} \sum_{i=1}^{n} F^{0} \lambda_{g_i} \bar{\varepsilon}_{g_i}^{\prime} \hat{F}+\frac{1}{n T} \sum_{i=1}^{n} \bar{\varepsilon}_{g_i} \lambda_{g_i}^{\prime} F^{0\prime} \hat{F}+\frac{1}{n T} \sum_{i=1}^{n} \bar{\varepsilon}_{g_i} \bar{\varepsilon}_{g_i}^{\prime} \hat{F}\right) \hat{\Gamma}^{-1} \\
&= P1+P2+\cdots+P8 .
\end{aligned}
$$

Using the same arguments for $\|I1\|=\cdots=\|I5\|$, we can show that $\|P1\|=\cdots=\|P5\|=O_p\left(\left\|\hat{\beta}-\beta^{0}\right\|\right)$. For $P6$, using the result in (A.13), we have

$$
\begin{aligned}
P6 &= \left(\frac{1}{T} \sum_{t=1}^{T} F_t^{0} F_t^{0\prime}\right) \frac{1}{n T} \sum_{g=1}^{G} n_g \lambda_g \varepsilon_g^{\prime} \hat{F} \hat{\Gamma}^{-1} \\
&= \frac{1}{\sqrt{n}} O_p\left(\left\|\beta^{0}-\hat{\beta}\right\|\right)+O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{\sqrt{n T}}\right) .
\end{aligned}
$$

For $P7$,

$$
\begin{aligned}
P7 &= \frac{1}{n T} \sum_{g=1}^{G} \sum_{t=1}^{T} F_t^{0} n_g \varepsilon_{gt} \lambda_g^{\prime}\left(\frac{F^{0\prime} \hat{F}}{T}\right) \hat{\Gamma}^{-1} \\
&= O_p\left(\frac{1}{\sqrt{n T}}\right) .
\end{aligned}
$$

For $P8$, using the Lemma A1, we have

$$
\begin{aligned}
P8 &= \frac{G}{n \sqrt{T}} \frac{1}{G} \sum_{g=1}^{G}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t^{0} \sqrt{n_g} \bar{\varepsilon}_{gt}\right)\left[\frac{\sqrt{n_g} \bar{\varepsilon}_g^{\prime}\left(\hat{F}-F^{0} H\right)}{T}+\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \sqrt{n_g} \bar{\varepsilon}_{gs} F_s^{0} H\right)\right] \hat{\Gamma}^{-1} \\
&= o_p\left(\left\|\beta^{0}-\hat{\beta}\right\|\right)+O_p\left(\frac{G}{n T \sqrt{n}}\right)+O_p\left(\frac{\sqrt{G}}{n \sqrt{n T}}\right)+O_p\left(\frac{G}{n T}\right)
\end{aligned}
$$

From $P1$-$P8$,

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

For the second term in (A.14), it follows from the Theorem 1 Part (ii) that

$$\frac{1}{T}\left(F^0H-\hat{F}\right)'\left(\hat{F}-F^0H\right)=O_p\left(\left\|\hat{\beta}-\beta^0\right\|^2\right)+O_p\left(\frac{1}{n}\right)+O_p\left(\frac{G^2}{n^2T}\right)$$

Therefore,

$$HH'-\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}=O_p\left(\left\|\hat{\beta}-\beta^0\right\|\right)+O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{G}{nT}\right). \tag{A.15}$$

■

Proof of Theorem 2. Let $\mathcal{X}_i=\mathcal{X}_i\left(F^0\right)$ for notational simplicity. From Proposition A1, we have

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

Thus, we need to show

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

to complete the proof.

Part (i) We have

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

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 $h1$,

$$\|h1\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{GT}}\sum_{t=1}^{T}\sum_{g=1}^{G}\bar{X}_{gs}\sqrt{n_g}\bar{\varepsilon}_{gt}F_t^{0\prime}\right\|^2\right)^{1/2}\|H\|\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - HF_s^0\right\|^2\right)^{1/2}$$

$$= O_p\left(\left\|\hat{\beta}-\beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n\sqrt{T}}\right) \tag{A.16}$$

For $h2$,

$$\|h2\| \leq \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t - H'F_t^0\right)\right\|^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\sqrt{n_g}\bar{X}_{gt}\bar{\varepsilon}_{gs}\right\|^2\right)^{1/2}$$

$$= O_p\left(\sqrt{T}\left\|\hat{\beta}-\beta^0\right\|^2\right) + O_p\left(\frac{\sqrt{T}}{n}\right) + O_p\left(\frac{G^2}{n^2\sqrt{T}}\right). \tag{A.17}$$

For $h3$

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

For $h32$, we have

$$\begin{aligned} h32 &= \sqrt{GT}\frac{1}{G}\sum_{g=1}^{G}\frac{\bar{X}_g'F^0}{T}\left[HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]\frac{\left(\hat{F}H^{-1}-F^0\right)'\sqrt{n_g}\bar{\varepsilon}_g}{T} \\ &= o_p\left(\sqrt{nT}\left\|\beta^0-\hat{\beta}\right\|\right) + O_p\left(\frac{\sqrt{T}}{n\sqrt{n}}\right) + O_p\left(\frac{\sqrt{G}}{n\sqrt{T}}\right) \\ &+ O_p\left(\frac{1}{n}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT\sqrt{T}}\right) \end{aligned} \tag{A.18}$$

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 $\mathcal{K}6$,

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

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

where

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

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

$$
\begin{aligned}
\|\mathcal{K}81\| &\leq \frac{G\sqrt{G}}{n}\left(\frac{1}{G^{2}}\sum_{\tilde{g}=1}^{G}\sum_{g=1}^{G}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{n_{\tilde{g}}n_{g}}\left[\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt}-E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt}\right)\right]\right\|^{2}\left(\frac{1}{G}\sum_{g=1}^{G}\left\|\frac{1}{T}\sum_{s=1}^{T}\bar{X}_{gs}F_{s}^{0\prime}\right\|^{2}\right)\right)^{1/2} \\
&\times\left\|\left(\frac{F^{0\prime}F^{0}}{T}\right)^{-1}\right\|\|\Upsilon'\|\left(\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left\|\frac{\left(\hat{F}-F^{0}H\right)'\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}}{T}\right\|^{2}\right)^{1/2} \\
&=o_{p}\left(\sqrt{nT}\left\|\beta-\hat{\beta}\right\|\right)+O_{p}\left(\frac{G\sqrt{G}}{n\sqrt{nT}}\right)+O_{p}\left(\frac{G}{n\sqrt{n}}\right)+O_{p}\left(\frac{G\sqrt{G}}{nT}\right)
\end{aligned}
$$

$$
\begin{aligned}
\|\mathcal{K}82\| &\leq \frac{\sqrt{G}}{n}\left(\frac{1}{GT}\sum_{\tilde{g}=1}^{G}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{g=1}^{G}\sqrt{n_{\tilde{g}}n_{g}}E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{gt}\right)\bar{X}_{gs}\right\|^{2}\right)^{1/2}\left(\sum_{s=1}^{T}\left\|F_{s}^{0\prime}\right\|^{2}\right)^{1/2} \\
&\times\left\|\left(\frac{F^{0\prime}F^{0}}{T}\right)^{-1}\right\|\|\Upsilon'\|\left(\frac{1}{G}\sum_{\tilde{g}=1}^{G}\left\|\frac{\left(\hat{F}-F^{0}H\right)'\sqrt{n_{\tilde{g}}}\bar{\varepsilon}_{\tilde{g}}}{T}\right\|^{2}\right)^{1/2} \\
&=o_{p}\left(\sqrt{nT}\left\|\beta-\hat{\beta}\right\|\right)+O_{p}\left(\frac{\sqrt{G}}{n\sqrt{nT}}\right)+O_{p}\left(\frac{1}{n\sqrt{n}}\right)+O_{p}\left(\frac{\sqrt{G}}{nT}\right)
\end{aligned}
$$

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

because

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

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

Thus,

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

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

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

Combining this with (A.18), we have

$$h3 = \sqrt{\frac{T}{n}} A_{nT}^{(21)} + o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{G}{n\sqrt{n}}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right)$$
$$+ O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{n\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right) + O_p\left(\frac{\sqrt{T}}{n\sqrt{n}}\right). \tag{A.19}$$

For $h4$, we apply (A.15) 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

$$\mathcal{H}1 = \sqrt{\frac{T}{n}} A_{nT}^{(21)} + o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right) + o_p\left(\sqrt{\frac{T}{n}}\right)$$
$$+ O_p\left(\frac{G}{n\sqrt{T}}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right) \tag{A.20}$$

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

$$\mathcal{H}2 = \sqrt{\frac{T}{n}} A_{nT}^{(22)} + o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right) + o_p\left(\sqrt{\frac{T}{n}}\right)$$
$$+ O_p\left(\frac{G}{n\sqrt{T}}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right) \tag{A.21}$$

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.20) and (A.21), we have

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\left(X_i - \left(\frac{\hat{F}\hat{F}'}{T}\right)\bar{X}_{g_i}\right)' - \frac{1}{n}\sum_{j=1}^{n}a_{ij}X_j'M_{\hat{F}}\right)\varepsilon_i$$
$$= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathcal{X}_i'\varepsilon_i + \sqrt{\frac{T}{n}}A_{nT}^{(2)} + o_p\left(\sqrt{nT}\left\|\hat{\beta} - \beta^0\right\|\right)$$
$$+ o_p\left(\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{G}{n\sqrt{T}}\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{nT}}\right) + O_p\left(\frac{G\sqrt{G}}{nT}\right) \tag{A.22}$$

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 $(G, n, T) \to \infty$ and $T/n^2 \to 0$.

Part (ii) We have

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

$$\begin{aligned}
\mathcal{G}1 &= \frac{1}{nT} \sum_{i=1}^{n} \bar{X}_{g_i}' \left(\left(\frac{\hat{F}\hat{F}'}{T}\right) - P_{F^0}\right) \bar{X}_{g_i} \\
&= \frac{1}{nT} \sum_{i=1}^{n} \frac{\bar{X}_{g_i}'\left(\hat{F} - F^0 H\right)}{T} H' F^{0\prime} \bar{X}_{g_i} + \frac{1}{nT} \sum_{i=1}^{n} \frac{\bar{X}_{g_i}'\left(\hat{F} - F^0 H\right)}{T} \left(\hat{F} - F^0 H\right)' \bar{X}_{g_i} \\
&+ \frac{1}{nT} \sum_{i=1}^{n} \frac{\bar{X}_{g_i}' F^0 H}{T} \left(\hat{F} - F^0 H\right)' \bar{X}_{g_i} + \frac{1}{nT} \sum_{i=1}^{n} \frac{\bar{X}_{g_i}' F^0}{T} \left[H H' - \left(\frac{F^{0\prime} F^0}{T}\right)^{-1}\right] F^{0\prime} \bar{X}_{g_i} \\
&= \mathcal{G}11 + \mathcal{G}12 + \mathcal{G}13 + \mathcal{G}14,
\end{aligned}$$

$$\begin{aligned}
\|\mathcal{G}11\| &\leq \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\left\|\bar{X}_{g_i}'\right\|^2}{T}\right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} \left\|\frac{1}{T} \sum_{s=1}^{T} F_s^0 \bar{X}_{g_i s}'\right\|^2\right)^{1/2} \frac{\left\|\hat{F} - F^0 H\right\|}{\sqrt{T}} \|H\| = o_p(1), \\
\|\mathcal{G}12\| &\leq \frac{1}{nT} \sum_{i=1}^{n} \left\|\bar{X}_{g_i}'\right\|^2 \frac{\left\|\hat{F} - F^0 H\right\|^2}{T} = o_p(1), \\
\|\mathcal{G}14\| &\leq \frac{1}{n} \sum_{i=1}^{n} \left\|\frac{\bar{X}_{g_i}' F^0}{T}\right\|^2 \left\|H H' - \left(\frac{F^{0\prime} F^0}{T}\right)^{-1}\right\| = o_p(1),
\end{aligned}$$

and $\mathcal{G}13$ is also $o_p(1)$ because $\mathcal{G}13 = \mathcal{G}11'$. Therefore, $\mathcal{G}1 = o_p(1)$. Using the same procedure, we can show that $\mathcal{G}2 = o_p(1)$.

Combining (i) and (ii), we have

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

under the rate conditions in Assumption 6. ■

Lemma A3 *Under Assumptions 1-6,*

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

Proof of Lemma A3. Note that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\hat{\Lambda}'-H^{-1}\Lambda'\right)\\
&= \frac{1}{\sqrt{n}}\hat{F}'\left[\frac{1}{T}\left(\bar{Y}_{g_1}-\bar{X}_{g_1}\hat{\beta}\right),\cdots,\frac{1}{T}\left(\bar{Y}_{g_n}-\bar{X}_{g_n}\hat{\beta}\right)\right] - \frac{1}{\sqrt{n}}H^{-1}\Lambda'\\
&= \frac{\hat{F}'}{T\sqrt{n}}\left(F^0-\hat{F}H^{-1}\right)\Lambda' + \frac{\hat{F}'}{T\sqrt{n}}\left[\bar{\varepsilon}_{g_1},\cdots,\bar{\varepsilon}_{g_n}\right] + \frac{\hat{F}'}{T\sqrt{n}}\left[\bar{X}_{g_1},\cdots,\bar{X}_{g_n}\right]\left(\beta-\hat{\beta}\right)\\
&= S1+S2+S3,
\end{aligned}$$

where

$$\begin{aligned}
\|S1\| &\le \frac{\left\|\hat{F}'\right\|}{\sqrt{T}}\frac{\left(F^0-\hat{F}H^{-1}\right)}{\sqrt{T}}\frac{\|\Lambda\|}{\sqrt{n}}\\
&= O_p\left(\left\|\beta^0-\hat{\beta}\right\|\right) + O_p\left(\frac{1}{\bar{\bar{\sqrt{n}}}}\right) + O_p\left(\frac{G}{n\sqrt{T}}\right),\\
\|S2\| &\le \frac{\left\|\hat{F}-F^0H\right\|}{\sqrt{T}}\frac{1}{\sqrt{nT}}\left\|\bar{\varepsilon}_{g_1},\cdots,\bar{\varepsilon}_{g_n}\right\| + \frac{\|H'\|}{T\sqrt{n}}\left\|F^{0\prime}\left[\bar{\varepsilon}_{g_1},\cdots,\bar{\varepsilon}_{g_n}\right]\right\|\\
&= O_p\left(\left\|\beta^0-\hat{\beta}\right\|\right) + O_p\left(\frac{1}{\bar{\bar{\sqrt{n}}}}\right) + O_p\left(\sqrt{\frac{G}{nT}}\right),\\
\|S3\| &\le \frac{\left\|\hat{F}'\right\|}{\sqrt{T}}\frac{1}{\sqrt{nT}}\left\|\left[\bar{X}_{g_1},\cdots,\bar{X}_{g_n}\right]\right\|O_p\left(\left\|\beta^0-\hat{\beta}\right\|\right)\\
&= O_p\left(\left\|\beta^0-\hat{\beta}\right\|\right)
\end{aligned}$$

Therefore,

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

■

Proof of Theorem 3. Due to Proposition A1 and Theorem 2, we need to prove

$$\text{(i) } \hat{B}_{nT}^{XX}-B_{nT}^{XX}\left(\hat{F}\right)=o_p(1),$$
$$\text{(ii) } \hat{V}_{nT}-V_{nT}^{c}=o_p(1).$$

Part (i) We have

$$\begin{aligned}\hat{B}_{nT}^{XX}-B_{nT}^{XX}\left(\hat{F}\right)&=\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_{ij}-\hat{a}_{ij}\right)X_i'M_{\hat{F}}X_j\\&\leq\frac{1}{\sqrt{T}}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|a_{ij}-\hat{a}_{ij}\right\|^2\right)^{1/2}\left(\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|X_i'M_{\hat{F}}X_j\right\|^2\right)^{1/2}\\&=o_p(1),\end{aligned}$$

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

$$\begin{aligned}\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|a_{ij}-\hat{a}_{ij}\right\|^2&\leq\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\lambda_{g_i}'\left(\Lambda'\Lambda/n\right)^{-1}H\left(H^{-1}\lambda_{g_j}-\hat{\lambda}_{g_j}\right)\right\|^2\\&+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\lambda_{g_i}'H'^{-1}\left[\left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1}-H\left(\Lambda'\Lambda/n\right)^{-1}H\right]\hat{\lambda}_{g_j}\right\|\\&+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\left(\hat{\lambda}_{g_i}-\lambda_{g_i}'H^{-1}\right)\left(\hat{\Lambda}'\hat{\Lambda}/n\right)^{-1}\hat{\lambda}_{g_j}\right\|\\&=O_p(1)\end{aligned}$$

Part (ii) To be added. ■

Proof of Theorem 5. For simplicity, we suppose $\beta^0 = 0$. Define

$$\begin{aligned}\tilde{\mathcal{Q}}_{cf}\left(\beta,\phi,F^0H\right) &= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[Y_i - P_{F^0H}\bar{Y}_g - \left(X_i - P_{F^0H}\bar{X}_g\right)\beta - \left(\hat{\eta}_i - P_{F^0H}\bar{\bar{\hat{\eta}}}_g\right)\phi\right]' \\ &\times\left[Y_i - P_{F^0H}\bar{Y}_g - \left(X_i - P_{F^0H}\bar{X}_g\right)\beta - \left(\hat{\eta}_i - P_{F^0H}\bar{\bar{\hat{\eta}}}_g\right)\phi\right] \\ &- \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(e_i - P_{F^0H}\bar{e}_{g_i}\right)'\left(e_i - P_{F^0H}\bar{e}_{g_i}\right), \\ \tilde{\mathcal{Q}}_{cf}^{o}\left(\beta,\phi,F^0H\right) &= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[Y_i - P_{F^0H}\bar{Y}_g - \left(X_i - P_{F^0H}\bar{X}_g\right)\beta - \left(\eta_i - P_{F^0H}\bar{\bar{\eta}}_g\right)\phi\right]' \\ &\times\left[Y_i - P_{F^0H}\bar{Y}_g - \left(X_i - P_{F^0H}\bar{X}_g\right)\beta - \left(\eta_i - P_{F^0H}\bar{\bar{\eta}}_g\right)\phi\right] \\ &- \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(e_i - P_{F^0H}\bar{e}_{g_i}\right)'\left(e_i - P_{F^0H}\bar{e}_{g_i}\right),\end{aligned}$$

and

$$\begin{aligned}\mathcal{Q}_{cf}^{*}\left(\beta,\phi,F^0H\right) &= \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\phi^0-\phi\right)'\left(\eta_i - P_{F^0H}\bar{\eta}_{g_i}\right)'\left(\eta_i - P_{F^0H}\bar{\eta}_{g_i}\right)\left(\phi^0-\phi\right) \\ &+ \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\beta'\left(X_i - P_{F^0H}\bar{X}_g\right)'\left(X_i - P_{F^0H}\bar{X}_g\right)\beta \\ &- \frac{2}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\phi^0-\phi\right)'\left(\eta_i - P_{F^0H}\bar{\eta}_{g_i}\right)'\left(X_i - P_{F^0H}\bar{X}_g\right)\beta.\end{aligned}$$

$\mathcal{Q}_{cf}^{o}\left(\beta,\phi,FH^0\right)$ is an infeasible version of objective function that is based on true η_i and FH^0. We also define

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

and

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

where

$$\hat{\beta}_{cf}^{o}\left(F\right) = \left[Q_{nT}^{XX}\left(F\right) - Q_{nT}^{X\eta}\left(F\right)\left(Q_{nT}^{\eta\eta}\left(F\right)\right)^{-1}Q_{nT}^{\eta X}\left(F\right)\right]^{-1}\left[Q_{nT}^{XY}\left(F\right) - Q_{nT}^{X\eta}\left(F\right)\left(Q_{nT}^{\eta\eta}\left(F\right)\right)^{-1}Q_{nT}^{\eta Y}\left(F\right)\right]$$

is the infeasible version of control function estimator given F based on the true η. We can obtain this formula using the Frisch-Waugh theorem.

The proof consists of three steps. In the first step, we show that $\tilde{\mathcal{Q}}_{LS}(F) - \mathcal{Q}^*_{LS}(F) = o_p(1)$ and $\mathcal{Q}^*_{LS}(F)$ is uniquely minimized at $F = F^0H$. The second step is to prove that $\tilde{\mathcal{Q}}^o_{cf}(\beta,\phi,F^0H) - \mathcal{Q}^*_{cf}(\beta,\phi,F^0H) = o_p(1)$ and $\mathcal{Q}^*_{cf}(\beta,\phi,F^0H)$ is uniquely minimized at $(\beta,\phi) = (\beta^0,\phi^0)$. The last step is to show that $\tilde{\mathcal{Q}}_{cf}(\beta,\phi,F^0H) - \tilde{\mathcal{Q}}^o_{cf}(\beta,\phi,F^0H) = o_p(1)$. We omit the details because they are analogous to the proof of Theorem 1 Part (i). The proof of Part (ii) is the same as the proof in Theorem 1 Part (ii). ■

Proof of Theorem 6. Part (i) The proof is similar to the proof of consistency of the LS estimator. Without loss of generality, we set $\beta^0 = 0$. Let

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

and

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

where

$$\mathbb{C}^{gmm}_{nT}(F) = \left[Q^{X\Psi}_{nT}(F)\Omega^{-1}_{nT}Q^{\Psi X}_{nT}(F)\right]^{-1}Q^{X\Psi}_{nT}(F)\Omega^{-1}_{nT}\left[\frac{1}{nT}\sum_{j=1}^{n}\left(\Psi_j - P_F\bar{\Psi}_{g_j}\right)M_F F^0\lambda_{g_j}\right].$$

The proof consists of two steps. In the first step, we show that $\mathcal{Q}^*_{iv}(\beta^0,F^0H) = 0$, $\tilde{\mathcal{Q}}_{iv}(\beta,F) - \mathcal{Q}^*_{iv}(\beta,F) = o_p(1)$ for all bounded β and $F \in \mathcal{F}$ and $\mathcal{Q}^*_{iv}(\beta,F) > 0$ if $(\beta,F) \neq (\beta^0,F^0H)$ .which is minimized at (β^0,F^0H). In the second step, we show that $\mathcal{Q}^*_{LS}(F^0H) = 0$, $\tilde{\mathcal{Q}}_{LS}(F) - \mathcal{Q}^*_{LS}(F) = o_p(1)$ and $\mathcal{Q}^*_{LS}(F) > 0$ if $F \neq F^0H$. The first step is omitted because it is analogous to the proof of Theorem 1. Part (i).

For the second step, we can show that $\mathcal{Q}^*_{LS}(F) > 0$ if $F \neq F^0 H$ using (A.3) by replacing β with $\mathbb{C}_{nT}(F)$. It is easy to show that $\mathcal{Q}^*_{LS}(F^0 H) = 0$, $\tilde{\mathcal{Q}}_{LS}(F) - \mathcal{Q}^*_{LS}(F) = o_p(1)$.

Part (ii) The proof is omitted because it is the same as the proof of Theorem 1 Part (ii). ■

Proof of Theorem 7. Note that

$$
\begin{aligned}
&\left[Q^{X\Psi}_{nT}\left(\hat{F}_{gmm}\right)\Omega^{-1}_{nT}Q^{\Psi X}_{nT}\left(\hat{F}_{gmm}\right)\right]\left(\hat{\beta}_{gmm} - \beta^0\right) \\
&= Q^{X\Psi}_{nT}\left(\hat{F}_{gmm}\right)\Omega^{-1}_{nT}\frac{1}{nT}\sum_{i=1}^{n}\left(\Psi_i' M_{\hat{F}_{gmm}} F^0 \lambda_{g_i} - \left(\Psi_i - P_{F^0}\bar{\Psi}_{g_i}\right)' \varepsilon_i\right). \qquad \text{(A.23)}
\end{aligned}
$$

Using a similar procedure in the proof of Proposition A1, we can show that

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

Applying this result to (A.23), under the rate condition in Assumption 6, we have

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

Then, using similar procedure in the proof of Theorem 2, we can show that

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

Applying Assumptions 14(ii) and 15 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] Hausman, J. (1978). Specification Tests in Econometrics, *Econometrica*, 46(6), 1251-1271.

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

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

[22] Kiefer, N. M. (1980). A Time Series-Cross Section Model with Fixed Effects with an Intertemporal Factor Structure, Unpublished Manuscript, Cornell University.

[23] 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

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

[25] Moon, H., M. Shum, and M. Weidner (2014). Interactive Fixed effects in the BLP Random Coefficients Demand Model, *CeMMAP working paper series.*

[26] Moon, H., and M. Weidner (2014). Dynamic Linear Panel Regression Models with Interactive Fixed Effects, *Econometric Theory*, forthcoming.

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

[28] 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.

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

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

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

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

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

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