University of Connecticut

Department of Economics Working Paper Series

Policy Analysis Using Multilevel Regression Models with Group Interactive Fixed Effects

by

Zhenhao Gong
Shanxi University of Finance
and Economics

Min Seong Kim
University of Connecticut

Working Paper 2024-01
January 2024

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
http://www.econ.uconn.edu/

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

Zhenhao Gong
Shanxi University of Finance and Economics

Min Seong Kim
University of Connecticut

January 2024

Abstract

The use of multilevel regression models is prevalent in policy analysis to estimate the effect of group level policies on individual outcomes. In order to allow for the time varying effect of group heterogeneity and the group specific impact of time effects, we propose a group interactive fixed effects approach that employs interaction terms of group factor loadings and common factors in this model. For this approach, we consider the least squares estimator and associated inference procedure. We examine their properties under the large n and fixed T asymptotics. The number of groups, G, is allowed to grow but at a slower rate. We also propose a test for the level of grouping to specify group factor loadings, and a GMM approach to address policy endogeneity with respect to idiosyncratic errors. Finally, we provide empirical illustrations of the proposed approach using two empirical examples.

Keywords endogeneity, GMM estimation, group heterogeneity, group level test, least squares estimation, panel, repeated cross-sections

JEL Classification Number C12, C13, C23, C54

*Email: zhenhao.gong@sxufe.edu.cn; minseong_kim@uconn.edu. We are grateful for comments from anonymous refrees, the associate editor, and the editor, as well as seminar participants at University of Montreal and UConn.

1 Introduction

In this paper, we consider policy analysis using multilevel linear regression models with panel data or repeated cross-sections. The multilevel regression model, which involves both aggregate and individual level variables, is prevalent in policy analysis when examining the effect of group level policies on individual level outcomes. See, for example, county level food stamp programs and household level employment status (Hoynes and Schanzenbach, 2012), state level cigarette tax policy and individual level childhood welfare (Simon, 2016), and county based social pension provision and household level income (Huang and Zhang, 2021). In this setup, researchers often use the additive fixed effects (or two-way fixed effects) approach based on the group fixed effects and time effects. The model is given by

$$Y_{igt} = Z_{gt}'\beta_Z^0 + W_{igt}'\beta_W^0 + \alpha_g^0 + f_t^0 + \varepsilon_{igt}, \tag{1}$$

where Y_{igt} is an outcome variable for individual i in group g at time t, W_{igt} is a $(d_w \times 1)$ vector of individual level covariates, Z_{gt} is a $(d_Z \times 1)$ vector of group level regressors including the policy variable, α_g^0 and f_t^0 are scalar group fixed effects and time effects, respectively, ε_{igt} is an idiosyncratic error, and $\beta^0 = (\beta_Z^{0\prime}, \beta_W^{0\prime})'$ is a $(d_\beta \times 1)$ vector of unknown coefficients. It is worth noting that Z_{gt} includes not only group specific variables, such as group level policies and group characteristics, but also the group averages (or other summaries) of individual level variables, such as the average income and education levels within the group.

A crucial condition for (1) to be valid is that the effect of unobserved group heterogeneity is time invariant and the impact of time effects is homogeneous across groups. The latter is referred to as the "parallel trends" assumption in the difference in differences (DID) model. However, this assumption may not be plausible in empirical studies. For instance, differential trends can appear if groups based on different regions, markets or ages exhibit heterogeneous responses to common time effects such as cyclical fluctuations (Blundell and Dias, 2009).

To address this potential problem of the additive fixed effects approach, we propose an interactive fixed effects model in the multilevel regression setting

$$Y_{igt} = Z_{gt}'\beta_Z^0 + W_{igt}'\beta_W^0 + \lambda_g^{0\prime}F_t^0 + \varepsilon_{igt}, \tag{2}$$

where a $(d_F \times 1)$ vector of unobserved common factors F_t^0 interacts with unobserved group factor loadings λ_g^0. $X_{igt} = \left(Z_{gt}', W_{igt}'\right)'$ is assumed to be exogenous with respect to ε_{igt}, but it is allowed to be arbitrarily correlated with λ_g^0 and F_t^0. As F_t^0 is multiplicative of λ_g^0, our proposed model in (2) improves upon the additive model in (1) by accounting for the time varying effect of unobserved group heterogeneity as well as the group specific impact of common factors. We refer to our model as the group interactive fixed effects model. Our model nests the additive model as a special case with $\lambda_g^0 = \left(\alpha_g^0, 1\right)'$ and $F_t^0 = (1, f_t^0)'$.

The proposed model is a variation of the standard interactive fixed effects model, in which the factor loading is individual specific. The interactive fixed effects model has become popular in the literature due to its flexibility in accommodating a rich form of unobserved heterogeneity while remaining parsimonious enough to make inference on β^0. Pesaran (2006) and Bai (2009) are two seminar papers that develop the estimation and inference procedures for this model. Pesaran (2006) introduces the common correlated effects (CCE) estimator for heterogeneous panel models, and Bai (2009) proposes the LS estimator based on the principal component (PC) method. Various methods have been studied in the literature, including the QMLE method based on the common shock model (Bai and Li, 2014), the method of LASSO (Lu and Su, 2016), and the GLS method (Bai and Liao, 2017). Moon and Weidner (2017) investigate the LS estimator for this model in the context of dynamic panel models. All those methods are based on the large n and large T asymptotics, which has been the primary focus in this literature. Alternatively, there are papers that study this model under the large n and fixed T asymptotic framework. Holtz-Eakin et al. (1988) and Ahn et al. (2001, 2013) consider the GMM method based on the quasi-differencing approach. Recently, Juodis and Sarafidis (2022) develop a novel GMM approach that approximates the unobserved common factors with observed factor proxies. In the policy analysis context, Callaway and Karami (2023) consider an interactive fixed effects method for estimating the average treatment effect with a binary treatment. They propose a two-step estimation procedure that requires time invariant instrumental variables. In our study, we extend Bai's LS estimation procedure to our group interactive fixed effects model and analyze its properties under the asymptotics where $(n, G) \to \infty$ with fixed T, where G represents the number of groups. In Bai's model, if the factor loadings were known, consistency for β would be achieved as $n \to \infty$ regardless of whether T is fixed or diverges. Therefore, the condition that $(n, T) \to \infty$ reflects the fact that the individual factor loadings are unknown and estimated. In our model, since λ_g^0 is assumed to be common within each group, we achieve consistency for a fixed T, if the number of individuals in each group grows as $n \to \infty$ such that $G/n \to 0$.

It is important to acknowledge that our proposed approach, which assumes factor loadings to be common within a group, may be vulnerable to endogeneity arising from unobserved individual heterogeneity. This is a crucial limitation of our model when compared to the standard interactive fixed effects model. Our model specification is primarily motivated by the context of policy analysis using multilevel regression models. In such cases, it is common practice to employ the additive group fixed effects regression model in (1). Our model in (2) enhances the additive model by accommodating heterogeneous responses to common factors across different groups. Moreover, our method can be applied not only to panel data but also to repeated cross-sections, whereas the standard interactive model requires panel data. Additionally, our LS method is valid under the large n and fixed T asymptotics, which differs from the typical asymptotic framework in the standard interactive fixed effects literature, which assumes both n and T to diverge. Given that the use of panel data and repeated cross-sections with a large number of individuals and a short time period is prevalent in policy analysis employing multilevel regression models, our approach serves as a valuable complement to the standard interactive model.

This paper proposes a test for the level of grouping for factor loadings. While our procedure assumes that group membership is known, in practical situations, it can be challenging to decide upon the appropriate level of grouping to specify the group factor loading. For instance, when a policy is country specific and the outcome is at the firm level, one may initially specify the factor loading at the country level. However, a finer level of grouping should be considered if there is a suspicion that, within each country, the source of endogeneity is an unobserved variable that varies across, for example, different industry sectors. We suppose that two different grouping schemes, $\mathbb{A}_0$ and $\mathbb{A}_a$, of which the latter represents a finer level of grouping that nests the former as a special case, are available. The null hypothesis of our test is that $\mathbb{A}_0$ is correctly specified, and the alternative is that $\mathbb{A}_0$ is misspecified. By exploiting the fact that, under the null, both $\mathbb{A}_0$ and $\mathbb{A}_a$ yield consistent estimators, whereas $\mathbb{A}_0$ does not under the alternative, our test compares the group interactive fixed effects estimators based on $\mathbb{A}_0$ and $\mathbb{A}_a$ and determines whether their difference is significant.

Another contribution of this paper is the extension of our approach to address the issue of policy endogeneity with respect to idiosyncratic errors. Certain sources of endogeneity, such as simultaneity and measurement error, can persist even after introducing group interaction terms. To tackle this challenge, we propose a moment condition based GMM approach which we call the "interactive fixed effects GMM" (IFE-GMM) estimator. The idea of this approach based on the standard interactive fixed effects model is discussed by Moon et al. (2017) in the context of random coefficients logit demand models, but they do not provide the asymptotic properties. Therefore, our contribution lies in providing the estimation procedure and establishing the asymptotics in the group interactive fixed effects setting. Regarding the endogeneity problem related to idiosyncratic errors, Moon et al. (2017) propose the "least-squares minimum distance (LS-MD)" method in random coefficients logit demand regression models. Subsequently, Lee et al. (2012) extend this method to the linear regression model. Lu (2023) proposes the QMLE and iterative generalized principal components (IGPC) methods to estimate spatial interactive fixed effects models in the presence of simultaneity.

We provide empirical illustrations of the proposed approach using two empirical examples. The first application builds upon Buccirossi et al. (2013), who examine the impact of country level competition policy on country-industry level total factor productivity (TFP) growth. The authors employ a multilevel additive fixed effects regression model with panel data. In our study, we reexamine their findings using the proposed group interactive fixed effects method. We also conduct the test on the level of grouping to specify the factor loading and estimate the model using the IFE-GMM estimator. The second empirical application is based on Huang and Zhang (2021), who investigate the effects of county based social pension provision on individual behaviors and social welfare in China. They employ a multilevel additive fixed effects model with repeated cross-sections, and we revisit their analysis using our group interactive fixed effects approach.

The outline of this paper is as follows. Section 2 introduces the proposed model and its LS estimator. Section 3 examines the asymptotic properties of the LS estimator and associated test statistics. In Section 4, we propose a test to determine the appropriate level of grouping for the factor loading. Section 5 studies GMM estimation to address policy endogeneity with respect to idiosyncratic errors. Empirical applications are presented in Section 6. The last section concludes. Additional theoretical results are in the appendix. Monte Carlo simulations, additional discussion, and proofs are included in the supplementary appendix.

2 Model and estimation

Let $\mathcal{A}_g = \{i : g_i = g\}$ and $n_g = \sum_{i=1}^{n} 1\{i \in \mathcal{A}_g\}$ denote the set of individuals in group g and the size of $\mathcal{A}_g$, respectively.

Assumption 1 *For all $g = 1, ..., G$, n_g is time invariant.*

Assumption 1 states that our group interactive approach requires the size of each group should remain constant over time. If this assumption does not hold in practice, we have two alternatives. First, we can reconstruct the dataset to satisfy the assumption, as demonstrated in our empirical applications in Section 6. Alternatively, we can modify our proposed estimation procedure below, based on the expectation-maximization (EM) algorithm by Stock and Watson (1998) and Bai (2009, Supplemental Material), which is developed for estimating factor models with unbalanced panel. The details of this modified procedure are provided in the supplementary appendix.

Under Assumption 1, (2) can be written as

$$Y_{ig} = Z_g\beta_Z^0 + W_{ig}\beta_W^0 + F^0\lambda_g^0 + \varepsilon_{ig}, \tag{3}$$

where $F^0 = (F_1^0, ..., F_T^0)'$, $Y_{ig} = (Y_{ig1}, ..., Y_{igT})'$, and Z_g, W_{ig} and ε_{ig} are defined in the same manner.

We propose LS estimation for our model based on Bai (2009). Let $X_{ig} = [Z_g, W_{ig}]$ and $\Lambda_G = (\lambda_1, ..., \lambda_G)'$. As λ_g is common within each group, the LS objective function can be written as

$$\mathcal{Q}(\beta, F, \Lambda_G) = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}(Y_{ig} - X_{ig}\beta - F\lambda_g)'(Y_{ig} - X_{ig}\beta - F\lambda_g), \tag{4}$$

which is minimized at the LS estimator $\left(\hat{\beta}, \hat{F}, \hat{\Lambda}_G\right)$. F and Λ_G are not separately identifiable due to their multiplicative structure, and we employ the following normalization

$$\sum_{t=1}^{T} F_tF_t' = I_{d_F} \text{ and } \frac{1}{n}\sum_{g=1}^{G} n_g\lambda_g\lambda_g' = \text{diagonal} \tag{5}$$

to uniquely determine F and Λ_G. $\hat{F}$ and $\hat{\Lambda}_G$ satisfy this normalization restriction.

Let $\bar{w}_g = n_g^{-1}\sum_{i\in\mathcal{A}_g} w_{ig}$ denote the group average of random vectors $\{w_{ig}, i \in \mathcal{A}_g\}$. Concentrating out

$$\hat{\lambda}_g(\beta, F) = (F'F)^{-1}F'\left(\frac{1}{n_g}\sum_{i\in\mathcal{A}_g}(Y_{ig} - X_{ig}\beta)\right) = F'\left(\bar{Y}_g - \bar{X}_g\beta\right) \tag{6}$$

out of (4), we have

$$\begin{aligned}\mathcal{Q}(\beta, F) &= \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(Y_{ig} - P_F\bar{Y}_g - \left(X_{ig} - P_F\bar{X}_g\right)\beta\right)'\left(Y_{ig} - P_F\bar{Y}_g - \left(X_{ig} - P_F\bar{X}_g\right)\beta\right) && (7)\\ &= \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}(Y_{ig} - X_{ig}\beta)'(Y_{ig} - X_{ig}\beta) - \frac{1}{n}\sum_{g=1}^{G} n_g\left(\bar{Y}_g - \bar{X}_g\beta\right)'FF'\left(\bar{Y}_g - \bar{X}_g\beta\right), && (8)\end{aligned}$$

where $P_F = F(F'F)^{-1}F' = FF'$ is the projection matrix. From (7), $\hat{\beta}(F)$ that minimizes $\mathcal{Q}(\beta, F)$ given F is

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

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

$$\frac{1}{n}\sum_{g=1}^{G} n_g \left(\bar{Y}_g - \bar{X}_g\beta\right)' FF' \left(\bar{Y}_g - \bar{X}_g\beta\right) = tr\left[F' \frac{\bar{\mathcal{R}}(\beta)\bar{\mathcal{R}}(\beta)'}{n} F\right], \tag{9}$$

where $\bar{\mathcal{R}}(\beta) = \left[\sqrt{n_1}\left(\bar{Y}_1 - \bar{X}_1\beta\right), ..., \sqrt{n_g}\left(\bar{Y}_g - \bar{X}_g\beta\right), ...\sqrt{n_G}\left(\bar{Y}_G - \bar{X}_G\beta\right)\right]$. It is well known that the solution to this maximization is the $(T \times d_F)$ matrix whose columns are the eigenvectors associated with the d_F largest eigenvalues of $n^{-1}\bar{\mathcal{R}}(\beta)\bar{\mathcal{R}}(\beta)'$. Therefore, we obtain $\left(\hat{\beta}, \hat{F}\right)$ based on

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

and

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

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

$$\hat{\lambda}_g = \hat{F}'\left(\bar{Y}_g - \bar{X}_g\hat{\beta}\right) \text{ and } \hat{\Lambda}_G = \left(\hat{\lambda}_1, ..., \hat{\lambda}_G\right)'. \tag{12}$$

Note that the rank condition requires $d_F \leq T - 1$. If $d_F = T$, we have $P_{\hat{F}} = I_{d_F}$ because each column of $\hat{F}$ is orthonormal. Thus, $X_i - P_{\hat{F}}\bar{X}_{g_i} = [Z_{g_i}, W_i] - \left[Z_{g_i}, \bar{W}_{g_i}\right] = \left[O, W_i - \bar{W}_{g_i}\right]$, where O denotes a $(T \times d_z)$ zero matrix. It is obvious that $\sum_{i=1}^{n}\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)'\left(X_i - P_{\hat{F}}\bar{X}_{g_i}\right)$ is not of full rank in this case. This implies, for example, that when $T = 5$, 4 is the maximum number of interaction term we can employ in the model.

Although our model nests the additive fixed effects model as discussed in Section 1, we can also explicitly include the additive fixed effects terms in the model

$$Y_{igt} = Z_{gt}'\beta_Z^0 + W_{igt}'\beta_W^0 + \lambda_g^{0\prime}F_t^0 + \alpha_g^0 + f_t^0 + \varepsilon_{igt}. \tag{13}$$

To estimate (13), we first use the within transformation to eliminate α_g^0 and f_t^0 from the model, and then apply the LS estimation method proposed above. See Bai (2009, Section 8) for details, where the procedure is discussed in the context of standard interactive fixed effects model.

3 Asymptotic theory and inference

In this section, we examine the asymptotic properties of $\hat{\beta}$ and the associated test statistics. Let $Q_n^{vw}(F) = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}(v_{ig} - P_F\bar{v}_g)'(w_{ig} - P_F\bar{w}_g)$ for random variables v and w. Define

$$\begin{aligned} B_n^{XX}(F) &= \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\mathcal{X}_{ig}^X(F)'\mathcal{X}_{ig}^X(F) \\ &= Q_n^{XX}(F) - \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} a_{g\tilde{g}}^0 \bar{X}_g' M_F \bar{X}_{\tilde{g}}, \end{aligned} \tag{14}$$

where $a_{g\tilde{g}}^0 = \lambda_g^{0\prime}\left(n^{-1}\sum_{g_1=1}^{G} n_{g_1}\lambda_{g_1}^0\lambda_{g_1}^{0\prime}\right)^{-1}\lambda_{\tilde{g}}^0$ and

$$\mathcal{X}_{ig}^X(F) = \left(X_{ig} - P_F\bar{X}_g\right) - \frac{1}{n}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}} a_{g\tilde{g}}^0 M_F \bar{X}_{\tilde{g}}. \tag{15}$$

Let $M_{\hat{F}} = I_T - P_{\hat{F}}$. To understand (14) and (15), we need to look at

$$\sqrt{n}\left(\hat{\beta} - \beta^0\right) = Q_n^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[\bar{X}_g' M_{\hat{F}} F^0 \lambda_g^0 + \left(X_{ig} - P_{\hat{F}}\bar{X}_g\right)'\varepsilon_{ig}\right], \tag{16}$$

which is directly obtained from (3) and (10). The first part of (16) comes from the estimation error in $\hat{F}$ because $M_{F^0}F^0 = 0$. For this term, we show that

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\bar{X}_g' M_{\hat{F}} F^0 \lambda_g^0 \\ &= \left\{\frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} a_{g\tilde{g}}^0 \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}}\right\}\sqrt{n}\left(\hat{\beta} - \beta^0\right) - \frac{1}{n\sqrt{n}}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} a_{g\tilde{g}}^0 \bar{X}_g' M_{\hat{F}} \bar{\varepsilon}_{\tilde{g}} + o_p(1) \end{aligned}$$

under the assumptions presented below. Combining this expression with (16), Proposition A2 states

$$\sqrt{n}\left(\hat{\beta} - \beta^0\right) = B_{nT}^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{X}_{ig}^X\left(\hat{F}\right)'\varepsilon_{ig} + o_p(1).$$

Thus, both (14) and (15) involve the effect of estimation error in $\hat{F}$.

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

Assumption 2 *(i)* $E\|X_{igt}\|^4 \le M$. *(ii) Let* $\mathcal{F} = \{F : F'F = I_{d_F}\}$. *We have* $\inf_{F\in\mathcal{F}} B_n^{XX}(F) > 0$.

Assumption 3 *(i)* $E\|F_t\|^4 \le M$. *(ii)* $E\|\lambda_g\|^4 \le M$, $n^{-1}\sum_{g=1}^{G} n_g\lambda_g\lambda_g' \to^p \sum_\Lambda > 0$ *as* $(n, G) \to \infty$.

Assumption 4 *(i) For all i, g and t, $E(\varepsilon_{igt}) = 0$ and $E\left(\varepsilon_{igt}^{8}\right) \leq M$.*
(ii) For all (t,s), $\lim_{(n,G)\to\infty} \frac{1}{G}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\sqrt{n_g n_{\tilde{g}}}\left|E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}s}\right)\right| < M$.
(iii) For all (t,s), $\lim_{(n,G)\to\infty} \frac{1}{n}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}}\left|E\left(\bar{\varepsilon}_{gt}\bar{\varepsilon}_{\tilde{g}s}\right)\right| < M$.
(iv) For all (t,s), $\lim_{(n,G)\to\infty} E\left(\frac{1}{\sqrt{G}}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}}\left[\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}s} - E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}s}\right)\right]\right)^2 < M$.

Assumption 5 *ε_{igt} is independent of $X_{j\tilde{g}s}$, $\lambda_{\tilde{g}}$ and F_s for all $i, j, g, \tilde{g}, t$ and s.*

Assumption 2 is the identification condition for the proposed LS estimator and ensures that there exist d_F distinct common factors. Assumption 3 provides the moment conditions for F_t and λ_g. Assumption 4 states the moment conditions and weak dependence conditions for ε_{igt}. These conditions are adapted from Bai (2009) and consider the group structure of our model. Weak dependence of $\sqrt{n_g}\bar{\varepsilon}_{gt}$ across groups is not restrictive, as strong dependence is absorbed in the interaction terms of the model.

Assumption 5 requires our group interactive model effectively captures the source of endogeneity to ensure that regressors are exogenous with respect to idiosyncratic errors. It is important to note that our group factor loading model may fail to satisfy this assumption if the true factor loadings are individual specific. To understand the issue, let's consider the case when the true model follows the standard interactive fixed effects structure

$$Y_{igt} = X_{igt}'\beta^0 + F_t^{0\prime}\lambda_{ig} + e_{igt}, \tag{17}$$

where λ_{ig} represents a vector of factor loadings for individual i in group g, and e_{igt} denotes the idiosyncratic error. We ignore the fact that λ_{ig} is time variant for repeated cross-sections for notational simplicity. We can define the group factor loading as the group mean of λ_{ig}, i.e., $\lambda_g^0 = E(\lambda_{ig}|i \in \mathcal{A}_g)$, and within-group individual heterogeneity γ_{ig} as the difference between the individual factor loading λ_{ig} and the group mean λ_g^0, which leads to $\lambda_{ig} = \lambda_g^0 + \gamma_{ig}$. In this setup, our regression model is written as

$$Y_{igt} = X_{igt}'\beta^0 + F_t^{0\prime}\lambda_g^0 + \varepsilon_{igt} \text{ with } \varepsilon_{igt} = F_t^{0\prime}\gamma_{ig} + e_{igt}, \tag{18}$$

which implies Assumption 5 does not hold if X_{igt} is correlated with γ_{ig}.

Assumption 6 *For all g, $n_g/n^\alpha \to c_g$ where $c_g \in (0,\infty)$ and $0 < \alpha < 1$.*

Assumption 6 allows group sizes to be different but requires them to be comparable to each other. They are also allowed to grow as n increases but at a slower rate.

Theorem 1 *Suppose that Assumptions 1-6 hold. Then, $\hat{\beta} - \beta^0 \to^p 0$ as $(n,G) \to \infty$ such that $G/n \to 0$ for fixed T.*

Theorem 1 establishes the consistency of $\hat{\beta}$. The proof is provided in the supplementary appendix. This is analogous to the consistency result in the standard interactive fixed effects model by Bai (2009, Proposition 1(i)). We can compare the rate conditions between these two estimators. In Bai's model, if the factor loadings were known, consistency would be achieved as $n \to \infty$ regardless of whether T is fixed or diverges. Hence, the rate condition $(n,T) \to \infty$ in Bai (2009) reflects the fact that the individual factor loadings are unknown and estimated. In our model, λ_g^0 is common within each group. Due to this group structure, $\hat{\beta}$ is consistent for a fixed T, if the number of individuals in each group grows as $n \to \infty$ such that $G/n \to 0$. Regarding $\hat{F}$, its average norm consistency is provided in Proposition A1 in the appendix.

If the true model follows the standard interactive fixed effects structure in (17), the proof of Theorem 1 shows the consistency of $\hat{\beta}$ requires

$$n^{-1} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{t=1}^{T} F_t^{0\prime} \gamma_{ig} X_{igt}' = n^{-1} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{t=1}^{T} F_t^{0\prime} \left(\lambda_{ig} - \lambda_g^0 \right) X_{igt}' \to^p 0,$$

which holds if the data are panel and eventually grouped at the individual level (i.e., $n_g \to 1$). However, this condition is not attainable in our large (n, G) and fixed T asymptotics as we require $G/n \to 0$ leading to $n_g \to \infty$. Thus, our approach does not achieve consistency when factor loadings are individual specific, which contrasts with Bester and Hansen (2016) who establish the asymptotic validity of the group fixed effects estimator in nonlinear panel models under the large n and large T asymptotics.

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

Assumption 7 *$(n, G) \to \infty$ such that $G/\sqrt{n} \to 0$.*

The rate condition, $G/\sqrt{n} \to 0$, in Assumption 7 requires G to grow at a slower rate than the one for consistency in Theorem 1, which is necessary to address the asymptotic bias of $\hat{\beta}$ as $(n, G) \to \infty$ with fixed T. Letting $\mathcal{X}_{ig}^X = \mathcal{X}_{ig}^X (F^0)$, we define

$$V_n = \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{\tilde{g}=1}^{G} \sum_{j \in \mathcal{A}_{\tilde{g}}} E \left[\left(\mathcal{X}_{ig}^X \right)' \varepsilon_{ig} \varepsilon_{j\tilde{g}}' \mathcal{X}_{j\tilde{g}}^X \right]. \tag{19}$$

Assumption 8 *We have $\frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^X \right)' \varepsilon_{ig} \to^d N(0, V)$ where $V = \lim_{(n,G)\to\infty} V_n$ is positive definite.*

Assumption 8 is a high level assumption of the central limit theorem for $\frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^X \right)' \varepsilon_{ig}$. This assumption allows heteroskedasticity and serial and cross sectional correlation in the idiosyncratic errors. A similar assumption is made in Bai (2009) and Lu and Su (2016). Theorem 2 states the asymptotic normality of $\hat{\beta}$.

Theorem 2 *Suppose that Assumptions 1-8 hold. We then have $\sqrt{n} \left(\hat{\beta} - \beta^0 \right) \to^d N\big(0, B_{XX}^{-1} V B_{XX}^{-1}\big)$ where $B_{XX} = \text{plim}_{(n,G)\to\infty} B_n^{XX}(F^0)$.*

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

$$\sqrt{n} \left(\hat{\beta} - \beta^0 \right) = B_{XX}^{-1} \frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^X \right)' \varepsilon_{ig} + \frac{G}{\sqrt{n}} B_{XX}^{-1} A_n + o_p \left(\sqrt{n} \left\| \hat{\beta} - \beta^0 \right\| \right) + o_p \left(\frac{G}{\sqrt{n}} \right) + O_p \left(\frac{1}{\sqrt{n}} \right), \tag{20}$$

where $\|\cdot\|$ is the Euclidean norm and

$$A_n = -\frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' M_{F^0} \Omega F^0 H \Upsilon \lambda_g^0,$$

with $\Upsilon = \left(\sum_{t=1}^{T} F_t^0 \hat{F}_t'\right)^{-1}\left(n^{-1}\sum_{g=1}^{G} n_g \lambda_g^0 \lambda_g^{0\prime}\right)^{-1}$ and $\Omega = \frac{1}{G}\sum_{g=1}^{G} n_g E\left(\bar{\varepsilon}_g \bar{\varepsilon}_g'\right)$. (20) indicates when G grows at the same rate as $\sqrt{n}$, $\frac{G}{\sqrt{n}} B_{XX}^{-1} A_n$ becomes a source of asymptotic bias unless $\sqrt{n_g}\bar{\varepsilon}_g$ is uncorrelated and homoskedastic across the groups $g = 1, ..., G$. This is related to the incidental parameters problem by Neyman and Scott (1948). Since T is assumed to be fixed in our model, the estimation error in $\hat{F}_t$ does not introduce asymptotic bias, and the only source of incidental parameters problem arises from λ_g^0. The rate condition $G/\sqrt{n} \to 0$ implies the number of λ_g^0's to be estimated becomes negligible with respect to n as $(n, G) \to \infty$, leading the bias due to the estimation error in $\hat{\lambda}_g$ to vanish asymptotically. This rate condition is relevant in empirical studies if each group includes a large number of individuals, for example, when we have a state/county level policy and household based outcome, and when we have a country level policy and firm level outcome. Under this rate condition, $\hat{\beta}$ is asymptotically normal and centered at β^0 when normalized by the sample size.

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

$$\mathcal{H}_0 : R\beta = \mathfrak{r}^0 \text{ vs. } \mathcal{H}_1 : R\beta \neq \mathfrak{r}^0, \tag{21}$$

where R is a $(d_R \times d_\beta)$ matrix and $\mathfrak{r}^0$ is a $(d_R \times 1)$ vector. To conduct the test, we first need to estimate B_n^{XX} and V_n in the variance term. We can estimate B_n^{XX} using the sample analogue

$$\hat{B}_n^{XX} = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_{ig}^X\right)'\hat{\mathcal{X}}_{ig}^X, \tag{22}$$

where

$$\hat{\mathcal{X}}_{ig}^X = \left(X_{ig} - P_{\hat{F}}\bar{X}_g\right) - \frac{1}{n}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}}\hat{a}_{g\tilde{g}} M_{\hat{F}}\bar{X}_{\tilde{g}} \text{ and } \hat{a}_{g\tilde{g}} = \hat{\lambda}_g'\left(\frac{1}{n}\sum_{g_1=1}^{G} n_{g_1}\hat{\lambda}_{g_1}\hat{\lambda}_{g_1}'\right)^{-1}\hat{\lambda}_{\tilde{g}}. \tag{23}$$

For the estimation of V_n, we consider two different cases.

Assumption 9 *(i) ε_{ig} and $\varepsilon_{j\tilde{g}}$ are independent for any i and j if $g \neq \tilde{g}$ and $n^{-1}\sum_{g=1}^{G} n_g\, E\left(\sqrt{n_g}\bar{\varepsilon}_g\right)^4 < M$, (ii) ε_{igt} are i.i.d. over i, g and t with zero mean and variance σ^2.*

The cluster covariance structure in Assumption 9(i) is commonly employed in the multilevel regression. See Moulton (1990) and Bertrand et al. (2004) for further discussion. While this assumption characterizes the cluster dependence of $\{\varepsilon_{ig}\}$ based on $\{\mathcal{A}_g, g = 1, ..., G\}$ for notational simplicity, it can be easily generalized to any level of clustering as long as the independence condition holds across clusters. A researcher selects an appropriate level of clustering by using his/her prior information about data or by conducting a test for this choice. See, for example, Ibragimov and Mueller (2014), who have developed a test about the level of clustering. The i.i.d. assumption in Assumption 9(ii) is considered to develop a group level test for factor loadings in Section 4.

Under Assumption 9(i) and (ii), V_n reduces to

$$V_n^c = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g} E\left[\left(\mathcal{X}_{ig}^X\right)'\varepsilon_{ig}\varepsilon_{jg}'\mathcal{X}_{jg}^X\right] \text{ and } V_n^s = \frac{\sigma^2}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g} E\left[\left(\mathcal{X}_{ig}^X\right)'\mathcal{X}_{ig}^X\right],$$

respectively. Thus, V_n^c represents the variance under the cluster structure based on $g = 1, ..., G$, and V_n^s represents the variance under the i.i.d. and homoskedasticity assumption. We estimate V_n^c and V_n^s with

$$\hat{V}_n^c = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_{ig}^X\right)'\hat{\varepsilon}_{ig}\hat{\varepsilon}_{jg}'\hat{\mathcal{X}}_{jg}^X \text{ and } \hat{V}_n^s = \hat{\sigma}^2\hat{B}_n^{XX},$$

where $\hat{\sigma}^2 = \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\hat{\varepsilon}_{ig}'\hat{\varepsilon}_{ig}$ with $\hat{\varepsilon}_{ig} = Y_{ig} - X_{ig}\hat{\beta} - \hat{F}\hat{\lambda}_g$.

Theorem 3 *Suppose that Assumptions 1-8 hold. Then, we have*

$$\left(\hat{B}_n^{XX}\right)^{-1}\hat{V}_n^c\left(\hat{B}_n^{XX}\right)^{-1} - B_{XX}^{-1}V_n^c B_{XX}^{-1} \to^p 0 \text{ and } \hat{\sigma}^2\left(\hat{B}_n^{XX}\right)^{-1} - \sigma^2 B_{XX}^{-1} \to^p 0$$

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

The Wald statistics are given by

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

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

The corollary below follows from Theorems 2 and 3.

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

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

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

4 Testing the level of grouping for group factor loadings

To establish the asymptotics of $\hat{\beta}$, we have assumed that group membership is known. However, this assumption may not hold in empirical applications. Researchers often do not have prior information about group membership, which makes the problem more challenging. In recent literature, there has been increasing attention given to grouped panel data models in which group membership is unknown. See, for example, Sun (2005), Hahn and Moon (2010), Bonhomme and Manresa (2015), Ando and Bai (2016), Su et al. (2016), and Lumsdaine et al. (2023). However, these approaches are not directly applicable to our model, as they require large n and large T panel data with fixed G, whereas our model accommodates large (n, G) and small T repeated cross-sections.

To address this practical issue, we develop a group level test to specify the group factor loading. This allows us to decide upon the appropriate level of grouping among different alternatives. For instance, when we estimate the effect of a country level policy on firm level outcomes, it is common to introduce interaction terms using country specific factor loadings. However, if we suspect that the sensitivity to common factors varies across different industry sectors within each country, a finer level of grouping should be considered.

Suppose that two different levels of grouping are available, $\mathbb{A}_0 = \{\mathcal{A}_1, ..., \mathcal{A}_g, ..., \mathcal{A}_G\}$ and $\mathbb{A}_a = \{\mathcal{A}_1^{(1)}, ..., \mathcal{A}_1^{(\kappa_1)}, ..., \mathcal{A}_g^{(1)} ..., \mathcal{A}_g^{(\kappa_g)}, ..., \mathcal{A}_G^{(1)} ..., \mathcal{A}_G^{(\kappa_G)}\}$, between which $\mathbb{A}_a$ represents a finer level of grouping because $\mathcal{A}_g = \cup_{\ell=1}^{\kappa_g} \mathcal{A}_g^{(\ell)}$. While we assume group membership of $\mathbb{A}_0$ is the same as the one for the group level regressors for notational simplicity, it is not necessary and can be generalized to any level of grouping. Let $\lambda_g^{(\ell)}$ denote the vector of group factor loadings based on $\mathcal{A}_g^{(\ell)}$ and $G_a = \sum_{g=1}^{G} \kappa_g$ denote the number of groups under $\mathbb{A}_a$. We assume that the rate conditions in Assumptions 6 and 7 hold for both $\mathbb{A}_0$ and $\mathbb{A}_a$. The null and alternative hypotheses are given by

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

Let $\hat{\beta}_a$ is the group interactive fixed effects estimator based on $\mathbb{A}_a$. We develop the test using the facts that (i) $\hat{\beta}$ and $\hat{\beta}_a$ are consistent under $\mathcal{H}_0$ because $\mathbb{A}_0$ is nested by $\mathbb{A}_a$ with $\lambda_g^{(\ell)} = \lambda_g$ for all $\ell = 1, ..., \kappa_g$, and (ii) only $\hat{\beta}_a$ is consistent if $\mathbb{A}_a$ is correctly specified but $\mathbb{A}_0$ is misspecified.

We define $\mathcal{X}_{a,ig}^X(F)$, $\mathcal{X}_{a,ig}^X$, $B_{a,n}^{XX}$ and $B_{a,XX}$ for the model in the same manner as $\mathcal{X}_{ig}^X(F)$, $\mathcal{X}_{ig}^X$, B_n^{XX}, and B_{XX} based on $\mathbb{A}_a$. The following result is a direct consequence of Theorem 2.

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

$$\sqrt{n}\left(\hat{\beta} - \hat{\beta}_a\right) \rightarrow^d N(0, V_{\mathcal{T}}),$$

where $V_{\mathcal{T}} = \lim_{n\to\infty} Var\left(\frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i\in\mathcal{A}_g} \left(\left(B_n^{XX}\right)^{-1} \left(\mathcal{X}_{ig}^X\right)' - \left(B_{a,n}^{XX}\right)^{-1} \left(\mathcal{X}_{a,ig}^X\right)'\right) \varepsilon_{ig}\right)$.

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

$$V_{\mathcal{T}} = \sigma^2 \left(B_{XX}^{-1} + B_{a,XX}^{-1} - B_{XX}^{-1} C_{0a,XX} B_{a,XX}^{-1} - B_{a,XX}^{-1} C_{a0,XX} B_{XX}^{-1}\right), \tag{24}$$

where $C_{0a,XX} = \lim_{(n,G)\to\infty} C_{0a,n}^{XX}$, $C_{0a,n}^{XX} = n^{-1} \sum_{g=1}^{G} \sum_{i\in\mathcal{A}_g} E\left[\left(\mathcal{X}_{ig}^X\right)' \mathcal{X}_{a,ig}^X\right]$ and $C_{a0,XX} = C_{0a,XX}'$.

Let $\hat{B}_{a,n}^{XX} = n^{-1} \sum_{g=1}^{G} \sum_{i\in\mathcal{A}_g} \left(\hat{\mathcal{X}}_{a,ig}^X\right)' \hat{\mathcal{X}}_{a,ig}^X$ and $\hat{C}_{0a,n}^{XX} = n^{-1} \sum_{g=1}^{G} \sum_{i\in\mathcal{A}_g} \left(\hat{\mathcal{X}}_{ig}^X\right)' \hat{\mathcal{X}}_{a,ig}^X$, where $\hat{\mathcal{X}}_{a,ig}^X$ is defined in the same way as $\hat{\mathcal{X}}_{ig}^X$ in (23) based on $\mathbb{A}_a$. We introduce the test statistic $\mathcal{T}$ given by

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

where

$$\hat{V}_{\mathcal{T}} = \hat{\sigma}_a^2 \left[\left(\hat{B}_n^{XX}\right)^{-1} + \left(\hat{B}_{a,n}^{XX}\right)^{-1} - \left(\hat{B}_n^{XX}\right)^{-1} \hat{C}_{0a,n}^{XX} \left(\hat{B}_{a,n}^{XX}\right)^{-1} - \left(\hat{B}_{a,n}^{XX}\right)^{-1} \hat{C}_{a0,n}^{XX} \left(\hat{B}_n^{XX}\right)^{-1} \right], \quad (25)$$

where $\hat{\sigma}_a^2$ is the estimator of σ^2 based on $\mathbb{A}_a$.

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

$$\frac{1}{n} \sum_{g=1}^{G} n_g \bar{X}_g' P_F \bar{X}_{a,g} = \frac{1}{n} \sum_{g=1}^{G} n_g \bar{X}_g' P_F \bar{X}_g. \quad (26)$$

If we apply (26) to (24), then, under the null hypothesis, the covariance of $\hat{\beta}$ and $\hat{\beta}_a$ equals the variance of $\hat{\beta}$. That is,

$$\lim_{n\to\infty} \frac{\sigma^2}{n} \sum_{g=1}^{G} \sum_{i\in\mathcal{A}_g} E\left[\left(\mathcal{X}_{ig}^X\right)' \mathcal{X}_{a,ig}^X \right] = \sigma^2 B_{XX}$$

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

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

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 $\hat{F}_a$ denote the estimator of $F^0 H$ based on $\mathbb{A}_a$. The crucial condition we need to rely on for $\tilde{\mathcal{T}}$ is

$$\frac{1}{n} \sum_{g=1}^{G} \sum_{i\in\mathcal{A}_g} \mathcal{X}_{ig}^X \left(\hat{F}\right)' \mathcal{X}_{a,ig}^X \left(\hat{F}_a\right) \approx \frac{1}{n} \sum_{g=1}^{G} \sum_{i\in\mathcal{A}_g} \mathcal{X}_{ig}^X \left(\hat{F}\right)' \mathcal{X}_{ig}^X \left(\hat{F}\right),$$

which requires $\hat{F} \approx \hat{F}_a$. However, even when the null is true, this approximation can be poor due to the estimation errors in $\hat{F}$ and $\hat{F}_a$. If this approximation does not work well, $\tilde{\mathcal{T}}$ suffers from poor finite sample properties. In contrast, $\mathcal{T}$ does not impose such a restriction in $\hat{V}_{\mathcal{T}}$ to accommodate estimation uncertainty in $\hat{F}$ and $\hat{F}_a$. Another advantage of using $\hat{V}_{\mathcal{T}}$ is that it is positive semi-definite by construction, which is an important property for the practical use of variance estimators. $\tilde{V}_{\mathcal{T}}$ may not yield a positive semi-definite estimate.

The asymptotics of $\mathcal{T}$ under the null hypothesis is characterized as follows.

Theorem 4 *Suppose that Assumptions 1-8 and 9(ii) hold. If $\mathcal{H}_0$ is true, then $\mathcal{T} \to^d \chi^2(d_\beta)$.*

We note that the rejection of $\mathcal{H}_0$ does not mean that $\hat{\beta}_a$ is consistent. If the appropriate level of grouping is in doubt and various levels of grouping are available in practical situations, we suggest researchers use the finest level of grouping for $\mathbb{A}_a$ and explore other grouping schemes for $\mathbb{A}_0$ to test their validity.

5 Policy endogeneity with respect to idiosyncratic errors: IFE-GMM approach

The validity of the LS estimation discussed thus far relies on the assumption that the regressors are exogenous with respect to idiosyncratic errors. However, in empirical applications, this assumption may not hold. For instance, simultaneity often appears between the policy and outcome variables, in which case endogeneity persists even when the group interactive fixed effects are introduced. To address this issue, we consider a moment condition based GMM estimator, which we refer to as the "interactive fixed effects GMM (IFE-GMM)" estimator.

Let $Z_{gt} = \left(Z_{gt}^{(1)}, ..., Z_{gt}^{(d_z)}\right)'$. $Z_{gt}^{(1)}$ represents a scalar policy variable that is potentially endogenous with respect to ε_{igt}. Suppose that we have a vector of instruments for $Z_{gt}^{(1)}$, denoted as $\tilde{\Psi}_{gt}$, and define a $(d_\Psi \times 1)$ vector $\Psi_{igt} = \left(\tilde{\Psi}_{gt}', Z_{gt}^{(2)}, ..., Z_{gt}^{(d_z)}, W_{igt}'\right)'$. We assume that Ψ_{igt} satisfies the following conditions.

Assumption 10 *(i) ε_{igt} is independent of $\Psi_{j\tilde{g}s}$ for all $i, j, g, \tilde{g}, t$ and s. (ii) rank$\left(Q_n^{X\Psi}(F)\right) = d_\beta$ and rank$\left(B_n^{X\Psi}(F)\right) = d_\beta$ for any $F \in \mathcal{F}$.*

Assumption 10 states the exogeneity condition of Ψ_{igt} with respect to idiosyncratic errors and rank conditions. The advantage of our IFE-GMM approach is that it has a larger set of potential instruments compared to the approach without interaction terms. In the latter case, the independence condition is required to hold not only for ε_{igt} but also for $\lambda_g^{0\prime} F_t^0$. This can be restrictive in policy analysis, as instruments for a policy variable often exhibit a correlation with group characteristics. See, for example, Besley and Case (2000). They use the fraction of female legislators in state lower and upper houses as an instrument for the manual rate to study the impact of state workers' compensation benefits on the employment and earnings of construction workers. It is natural to expect such a political variable to be correlated with unobserved (possibly time varying) state characteristics.

The IFE-GMM estimator is given by

$$\left(\hat{F}(\beta), \hat{\Lambda}_G(\beta)\right) = \underset{(F,\Lambda_G)}{\text{argmin}} \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} (Y_{ig} - X_{ig}\beta - F\lambda_g)' (Y_{ig} - X_{ig}\beta - F\lambda_g), \tag{27}$$

and

$$\hat{\beta}_{gmm}(F, \Lambda_G) = \underset{\beta}{\text{argmin}} \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left[\Psi_{ig}'(Y_{ig} - X_{ig}\beta - F\lambda_g)\right]' \Omega_n^{-1} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left[\Psi_{ig}'(Y_{ig} - X_{ig}\beta - F\lambda_g)\right], \tag{28}$$

where Ω_n is a positive definite $(d_\Psi \times d_\Psi)$ weight matrix. Note that while $\hat{\beta}_{gmm}$ is obtained via the GMM criterion based on the moment conditions in Assumption 10, F^0 and Λ_G^0 are estimated via the LS criterion and principal component method. We have $\hat{\lambda}_g(\beta, F) = F'\left(\bar{Y}_g - \bar{X}_g\beta\right)$ from (27). Plugging this in (28), we have the GMM estimator

$$\hat{\beta}_{gmm} = \left[Q_n^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_n^{-1}Q_n^{\Psi X}\left(\hat{F}_{gmm}\right)\right]^{-1} Q_n^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_n^{-1}Q_n^{\Psi Y}\left(\hat{F}_{gmm}\right), \tag{29}$$

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

$$\frac{1}{n}\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)'\hat{F}_{gmm}=\hat{F}_{gmm}\hat{\Gamma}_{gmm}. \tag{30}$$

$\hat{\Gamma}_{gmm}$ is a diagonal matrix of the d_F largest eigenvalues of $n^{-1}\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)\bar{\mathcal{R}}\left(\hat{\beta}_{gmm}\right)'$. As in LS estimation, we can obtain $\hat{\beta}_{gmm}$ by iterating (29) and (30) to convergence. In case $d_\Psi = d_\beta$, we have

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

It is easy to see that this approach includes the LS estimator in Section 2 as a special case with $\Psi_{ig}=X_{ig}$. That is, $\hat{\beta}_{gmm}=\hat{\beta}$ if $\Psi_{ig}=X_{ig}$.

As noted by Moon et al. (2017), a drawback of the GMM approach is that it can lead to a local minimum, which is in contrast to their LS-MD estimator. To minimize the risk of falsely choosing a local minimum, we should conduct the iterations of (29) and (30) using multiple initial values. The potential local minima problem is not a unique issue for this method. The LS estimator has the same problem.

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

Assumption 11 *(i) $E\left\|\Psi_{igt}\right\|^4\leq M$. (ii) $\Omega_n\rightarrow^p\Omega$ and Ω is positive definite.*

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

Theorem 5 *Under Assumptions 1-4, 6, 10 and 11, we have $\hat{\beta}_{gmm}-\beta^0\rightarrow^p 0$ as $(n,G)\rightarrow\infty$ such that $G/n\rightarrow\infty$ for fixed T.*

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

Assumption 12 *Let $\mathcal{X}_{ig}^{\Psi}=\mathcal{X}_{ig}^{\Psi}\left(F^0\right)$. We have $\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\mathcal{X}_{ig}^{\Psi}\right)'\varepsilon_{ig}\rightarrow^d N\left(0,V_{gmm}\right),$ where $V_{gmm}=\lim_{(n,G)\rightarrow\infty}\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{\tilde{g}=1}^{G}\sum_{j\in\mathcal{A}_{\tilde{g}}}E\left[\left(\mathcal{X}_{ig}^{\Psi}\right)'\varepsilon_{ig}\varepsilon_{j\tilde{g}}'\mathcal{X}_{j\tilde{g}}^{\Psi}\right]$ is positive definite.*

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

Theorem 6 *Under Assumptions 1-4, 6, 7 and 10-12, we have*

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

where $Q_{X\Psi}=\text{plim}_{n\rightarrow\infty}Q_n^{X\Psi}\left(F^0\right),\ Q_{\Psi X}=Q_{X\Psi}',\ B_{\Psi X}=\text{plim}_{n\rightarrow\infty}B_n^{\Psi X}\left(F^0\right)$ and $B_{X\Psi}=B_{\Psi X}'$.

The proof is in the supplementary appendix. Inference on β^0 can be conducted based on Theorem 6. This procedure is analogous to the one for the LS method in Section 3, so it is omitted here to save space.

Regarding the choice of Ω_n, it is well known that in a standard GMM framework, the asymptotic variance of the sample moments is optimal and minimizes the asymptotic variance of the GMM estimator (Hansen, 1982). However, this optimality scheme does not apply to our method. In our case, the estimation error in $\hat{F}_{gmm}$ affects the asymptotic variance of $\hat{\beta}_{gmm}$ through $B_{\Psi X}$, so the choice of $\Omega_n=V_{gmm}$ does not yield the usual variance form of the efficient GMM estimator. It will be interesting to study the optimal Ω_n in IFE-GMM estimation and we leave this to future research. In our empirical illustrations in Section 6, we set $\Omega_n=Q_n^{\Psi\Psi}\left(\hat{F}_{gmm}\right)$, motivated by 2SLS estimation.

6 Empirical illustrations

In this section, we apply our group interactive fixed effects approach to two empirical examples. The first example builds on Buccirossi et al. (2013, BCDSV hereafter), examining the impact of country level competition policy on country-industry level total factor productivity (TFP) growth. The second example is based on Huang and Zhang (2021, HZ hereafter), investigating the effects of county based social pension provision on individual behaviors and social welfare in China. We revisit the analyses of these studies, which employ multilevel additive fixed effects models with panel and repeated cross-sections, using our group interactive fixed effects approach.

The choice of the number of interactive fixed effects terms is clearly important. To address this, we modify the CP and IC criteria provided by Bai and Ng (2002) to reflect the fact that we consider the large n and fixed T asymptotics and we have the group structure

$$IC\left(r\right)=\log\hat{\mathcal{Q}}(r)+r\frac{\log\left(G\right)}{n},\ CP\left(r\right)=\hat{\mathcal{Q}}(r)+r\hat{\mathcal{Q}}\left(\bar{r}\right)\frac{\log\left(G\right)}{n},$$

where $\hat{\mathcal{Q}}(r)=\mathcal{Q}\left(\hat{\beta},\hat{F},\hat{\Lambda}_G\right)$ based on r factors and $\bar{r}$ represents the largest number of interactive fixed effects terms that we assume. If T is not much smaller than G, we can also include T in the criteria as follows

$$\widetilde{IC}\left(r\right)=\log\left(\hat{\sigma}^2\left(r\right)\right)+r\frac{\log\left(GT\right)}{nT},\ \widetilde{CP}\left(r\right)=\hat{\sigma}^2\left(r\right)+r\hat{\sigma}^2\left(\bar{r}\right)\frac{\log\left(GT\right)}{nT}.$$

To avoid the local minima problem, we choose 50 different initial values of β around the additive fixed effects estimator.

6.1 Competition policy and productivity growth

There is broad consensus in economics that competition tends to enhance economic efficiency. However, there is no such agreement on the effectiveness of competition policy. For example, Baker (2002) argues that the benefit of antitrust enforcement outweighs the cost, while Crandall and Winston (2003) claim that antitrust law has been ineffective in the US. In this regard, BCDSV examine the importance of competition policy in improving productivity growth. Using the panel and multilevel data, they provide empirical evidence that the effect of competition policy on total factor productivity (TFP) growth is positive and significant. As a measure of competition policy, they construct the Competition Policy Indicator (CPI) that summarizes the key features of a country's competition policy.

We revisit their analysis using our group interactive fixed effects approach. Our regression model is given by

$$\Delta TFP_{igt}=\beta_0^0 CPI_{gt-1}+\beta_1^0\Delta TFP_{L(i)t}+\beta_2^0\frac{TFP_{L(i)t}}{TFP_{igt}}+\beta_3^0 W_{igt-1}+\beta_4^0 Z_{gt-1}+\lambda_g^{0\prime}F_t^0+\varepsilon_{igt},\quad(31)$$

where ΔTFP_{igt} is the TFP growth of industry i in country g at time t, CPI_{gt-1} is the CPI of country g at time $t-1$. Thus, β_0^0 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 $TFP_{L(i)t}/TFP_{igt}$

denote the technology transfer from a technological frontier country and the productivity gap to the technological frontier, respectively. W_{igt-1} is a vector of country-industry specific covariates including trade openness and country-industry specific trends, and Z_{gt-1} denotes a country level product market regulation variable (prm). We estimate (31) using the proposed LS method and IFE-GMM method.

Our estimation is based on the 1995-2005 balanced panel data. The dataset consists of 22 industries in 7 countries (Czech Republic, Germany, Italy, Japan, Sweden, UK, and US). The data is available at https://dataverse.harvard.edu/dataverse/restat. BCDSV include 5 more countries (Canada, France, Hungary, Netherlands and Spain), but we exclude them to obtain the balanced panel. Note that the omission of these 5 countries yields a mild change in additive fixed effects estimates. We follow ISIC Rev.3 for industry classification.

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

<Test on the level of grouping>

$\mathcal{H}_0$: Country level of grouping is correctly specified.			
$\mathcal{H}_a$: Country level of grouping is misspecified.			
Number of interactive terms (d_F)	1	2	3
$\mathcal{T}$	10.550	5.861	7.332
critical value	$\chi^2_{0.95}(6) = 12.592$		

As presented in the table, the test does not reject the null hypothesis with various choice of d_F at 5% level. According to this result, we set the group factor loadings at country level. Thus, in this application, we have $n = 154$, $T = 10$, and $G = 7$. Our procedure to choose the number of interaction terms yields $r = 2$ with $IC(r)$ and $CP(r)$ and $r = 1$ with $\widetilde{IC}(r)$ and $\widetilde{CP}(r)$.

In addition to the proposed approach based on (31), we consider two additive fixed effects models for the purpose of comparison. One is based on the individual fixed effects, which is employed by BCDSV, and the other is based on the group fixed effects.

Table 1 reports the coefficients and their t-statistics based on the country based cluster standard errors. We observe that the magnitudes of coefficients for CPI obtained from our approach are substantially smaller than the ones based on the additive fixed effects models. The former are between 0.029 and 0.037 with $d_F = 1\sim 3$, while the latter are 0.088 with the group effects and 0.070 with the individual effects.

Table 2 reports the coefficients and their t-statistics using our IFE-GMM method and the 2SLS method. We use the political variables developed by Cusack and Fuchs (2002) as instruments. They include Market regulation (per403), Economic planning (per404), Welfare state limitations planning (per505) and European Community (per108). BCDSV also use them as instruments. The qualitative results are similar to the LS estimation case. Compared to the additive fixed effects estimates, the magnitudes of the coefficients for CPI are substantially reduced when IFE-GMM method are employed with various choices of d_F.

6.2 Social pensions policy on household health status

As public pension programs have become crucial components of social security systems in many countries with the increasing aging population, it has become a critical research question to understand the impact of social pensions on the welfare of elderly individuals. In this context, HZ examine the effect of the New Rural Pension Scheme (NRPS) on individual behaviors and social welfare in China. The NRPS is a county-based rural pension program that started in 2009 and expanded to cover all counties in China by the end of 2012. It is an extensive social pension program, which covers the largest population in human history. In 2011, it offered pension benefits to 89 million rural pensioners, and 326 million rural residents participated in this program. The program specifically targets age-eligible rural seniors, aged 60 years and older, who are eligible to receive a non-trivial monthly pension. The paper finds that the NRPS leads to higher household income and food expenditure, less farm work, better health, and lower mortality among age-eligible seniors.

HZ consider the following additive fixed effects regression model

$$Y_{igt} = \beta_0^0 NRPS_{gt} + W_{igt}'\beta_W^0 + \alpha_g^0 + f_t^0 + \varepsilon_{igt}, \tag{32}$$

where Y_{igt} denotes an outcome variable (e.g., household income, expenditure, labor supply, and health outcomes) of individual i in county g in year t, $NRPS_{gt}$ is an indicator variable representing whether county g implemented the NRPS in year t, W_{igt} is a vector of individual level demographic controls, including gender, age, age squared, and dummies for the education level, and α_g^0 and f_t^0 are the county fixed effects and year effects, respectively. Thus, the parameter of interest β_0^0 captures the effect of county level social pensions policy on county-individual specific outcomes.

HZ estimate the model using yearly repeated cross-sections from 2010 to 2013. The data for their analysis is constructed based on China Family Panel Studies (CFPS) and the China Health and Retirement Longitudinal Studies (CHARLS). These surveys are the Chinese equivalent of the Panel Study of Income Dynamics and the Health and Retirement Survey in the US, respectively. Since the CFPS and CHARLS surveys are biennial, they employ the 2010 and 2012 waves of the CFPS and the 2011 and 2013 waves of the CHARLS for their study. The CFPS covers 162 counties and the CHARLS covers 150 counties.

We revisit their study using the proposed group interactive fixed effects approach, focusing on estimating the effect of the NRPS on the health status of rural seniors. Our regression model is given by

$$Y_{igt} = \beta_0^0 NRPS_{gt} + W_{igt}'\beta_W^0 + \lambda_g^{0\prime} F_t^0 + \varepsilon_{igt}, \tag{33}$$

where Y_{igt} represents the “unhealthiness score" developed by HZ to measure the poor health, disability and malnutrition of individual i in county g in year t.

Since our estimation method requires the group size to remain time invariant, we initially consider 10 counties in HZ's data that appear in both the CFPS and CHARLS surveys. We then exclude one county from this set which contains a too small number of observations. Subsequently, we randomly select the same number of individuals for each year in each county. After this adjustment, our dataset consists of 9 counties with 292 individuals each year. Since $T = 4$, the maximum number of common factors that we can employ in this model is 3 and our selection of the number of interaction terms yields $r = 2$ with both $IC(r)$ and $CP(r)$.

Table 3 presents the estimation results. In addition to the group interactive fixed effects approach, we also employ the additive fixed effects model (32) for comparative purposes. The first two columns of the table report the coefficients and t-statistics for (33) based on our LS estimator. For the additive models, we provide two estimates. The third column presents the estimation results using the size adjusted sample that we use for our method, while the fourth column corresponds to the estimates without such adjustment on the sample size.

From the table, we first observe that the NRPS coefficient on the unhealthiness score is negative and significant when using the additive fixed effects model. Specifically, when employing the adjusted sample, the estimate is -0.138 with a t-statistic of -2.238. Without the adjustment, the estimate is -0.179 with a t-statistic of -2.550. These results are similar to the ones presented in HZ.

When using our group interactive fixed effects method, the magnitude of the NRPS coefficient becomes substantially smaller and insignificant. The table indicates that the estimate is 0.036 with a t-statistic of 0.556 when $d_F = 2$, and it is 0.064 with a t-statistic of 0.741 when $d_F = 3$. These results suggest that the significance of the NRPS coefficient in the additive fixed effects model is possibly due to endogeneity associated with unobserved time varying group heterogeneity, and our group interactive fixed effects model may effectively mitigate this endogeneity issue.

7 Conclusion

The multilevel regression model is widely used for studying the effect of group level policies on individual level outcomes. In this setting, researchers often employ the additive fixed effects regression to account for the 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 hold in many applications. To address this important issue, we propose the group interactive fixed effects model in the multilevel regression setting. This model accounts for group specific impact of time effects as well as time varying effect of group heterogeneity. We also develop a test to determine the level of grouping for factor loadings. To address the policy endogeneity with respect to idiosyncratic errors in our model, we propose the GMM method.

For future research, it would be interesting to explore theoretical issues in our model under the large n and large T panel framework, including (i) examining the trade-off between the misspecification bias due to the group factor loading structure and incidental parameters bias, (ii) extending our model to dynamic regression models, and (iii) investigating the problem of unknown group membership problem. Another interesting extension would be to allow the heterogeneity of regression coefficients across groups when G is large, particularly to examine the heterogeneous effect of a policy variable on different groups.

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] Ahn, S.C., Y. Lee, and P. Schmidt. (2013). Panel data models with multiple time-varying individual effects. *Journal of econometrics*, 174(1), pp.1-14.

[3] 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.

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

[5] Bai, J., and K. Li (2014). Theory and methods of panel data models with interactive effects. *Annals of Statistics*, 42(1), 142-170.

[6] Bai, J., and Y. Liao (2017). Inferences in Panel Data with Interactive Effects Using Large Covariance Matrices, 200(1), 59-78.

[7] Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191-221.

[8] Baker, J.B. (2002). The case for antitrust enforcement. *Journal of Economic perspectives*, 17(4), 27-50.

[9] 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.

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

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

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

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

[14] 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.

[15] Callaway, B., and S. Karami (2023). Treatment effects in interactive fixed effects models with a small number of time periods. *Journal of Econometrics*, 233(1), 184-208.

[16] Crandall, R.W., and Winston, C. (2002). Does antitrust policy improve consumer welfare? Assessing the evidence. *Journal of Economic Perspective*s, 17(4), 3-26.

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

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

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

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

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

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

[23] Huang, W., and C. Zhang (2021). The Power of Social Pensions: Evidence from China's New Rural Pension Scheme, *American Economic Journal: Applied Economics*, 13(2) 179-205.

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

[25] Juodis, A., and V. Sarafidis. (2022) A linear estimator for factor-augmented fixed-T panels with endogenous regressors. *Journal of Business & Economic Statistics*, 40(1), 1-15.

[26] 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.

[27] Lu, L. (2023). Simultaneous spatial panel data models with common shocks. *Journal of Business & Economic Statistics*, 41(2), 608-623.

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

[29] Lumsdaine, R.L., R. Okui, and W. Wang (2023). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, 233(1), 45-65.

[30] Moon, H., M. Shum, and M. Weidner (2017). Interactive Fixed effects in the BLP Random Coefficients Demand Model, *Journal of Econometrics*, 613-644.

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

[32] 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.

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

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

[35] Simon, D. (2016). Does Early Life Exposure to Cigarette Smoke Permanently Harm Childhood Welfare? Evidence from Cigarette Tax Hikes, *American Economic Journal: Applied Economics* 8(4), 128-159.

[36] Stock, J. H., and M. W. Watson (1998). Diffusion Indexes, Working Paper 6702, NBER.

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

[38] Su, L., Z. Shi, and P. C. B. Phillips (2016). Identifying Latent Structures in Panel Data, *Econometrica*, 84(6), 2215-2264.

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

8 Appendix

Proposition A1 *Let* $H = \left(n^{-1}\sum_{g=1}^{G} n_g \lambda_g^0 \lambda_g^{0\prime}\right)\left(\sum_{t=1}^{T} F_t^0 \hat{F}_t^{\prime}\right)\hat{\Gamma}^{-1}$. *Suppose that Assumptions 1-6 hold. Then, we have*

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

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

Proposition A2 *Let*

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

with $\Upsilon = \left(\sum_{t=1}^{T} F_t^0 \hat{F}_t^{\prime}\right)^{-1}\left(n^{-1}\sum_{g=1}^{G} n_g \lambda_g^0 \lambda_g^{0\prime}\right)^{-1}$. *Under Assumptions 1-7,*

$$\begin{aligned}\sqrt{n}\left(\hat{\beta} - \beta^0\right) = {} & B_n^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\left(X_{ig} - P_{\hat{F}}\bar{X}_g\right) - \frac{1}{n}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}} a_{g\tilde{g}}^0 M_{\hat{F}} \bar{X}_{\tilde{g}}\right)^{\prime}\varepsilon_{ig} \\ & + \frac{G}{\sqrt{n}} B_n^{XX}\left(\hat{F}\right)^{-1} A_n + o_p\left(\sqrt{n}\left\|\hat{\beta} - \beta\right\|\right) + o_p\left(\frac{G}{\sqrt{n}}\right) + O_p\left(\frac{1}{n}\right).\end{aligned}$$

Proposition A3 *Let* $H_{gmm} = \left(n^{-1}\sum_{g=1}^{G} n_g \lambda_g^0 \lambda_g^{0\prime}\right)\left(\sum_{t=1}^{T} F^0 \hat{F}_{gmm}^{\prime}\right)\hat{\Gamma}_{gmm}^{-1}$. *Suppose that Assumptions 1-4, 6, 10 and 11 hold. Then, we have*

$$\left\|\hat{F}_{gmm} - F^0 H_{gmm}\right\| = O_p\left(\left\|\hat{\beta}_{gmm} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right),$$

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

The proofs of Theorems 1-6 and Propositions A1-A3 are provided in the supplementary appendix

Table 1: The effect of competition policy on TFP growth: LS estimation

Dependent Variable	ΔTFP_{igt}				
		Group IFE		Group AFE	Individual
d_F	1	2	3		AFE
CPI_{gt-1}	0.037	0.029	0.031	0.088	0.070
	(4.660)	(3.725)	(3.468)	(4.507)	(5.110)
$\Delta TFP_{L(i),t}$	0.107	0.111	0.121	0.114	0.079
	(4.862)	(4.996)	(5.487)	(5.675)	(4.053)
$\left(TFP_{L(i)g,t}/TFP_{igt}\right)$	0.000	0.000	0.001	0.000	0.009
	(0.842)	(0.724)	(1.176)	(0.992)	(4.531)
Industry trend_{igt}	0.001	0.006	0.007	0.006	0.038
	(1.784)	(5.827)	(5.315)	(6.943)	(5.487)
Import penetration_{igt}	0.003	0.004	0.004	0.004	0.009
	(3.305)	(3.855)	(3.668)	(3.439)	(3.718)
pmr_{gt}	-0.009	0.008	-0.000	-0.025	-0.039
	(-4.290)	(-3.363)	(-3.445)	(-4.002)	(-8.302)

Note: The numbers in parentheses represent t statistics based on country level cluster standard errors.

Table 2: The effect of competition policy on TFP growth: IV estimation

Dependent Variable	ΔTFP_{igt}				
		IFE-GMM		Group AFE 2SLS	Individual AFE 2SLS
d_F	1	2	3		
CPI_{gt-1}	0.022	-0.013	0.009	0.213	0.178
	(0.884)	(-0.209)	(0.268)	(2.739)	(1.954)
$\Delta TFP_{L(i)gt}$	0.110	0.112	0.120	0.112	0.076
	(4.920)	(5.320)	(5.483)	(5.744)	(4.202)
$\left(TFP_{L(i)gt}/TFP_{igt}\right)$	0.000	0.000	0.000	0.000	0.008
	(0.502)	(0.644)	(0.909)	(0.976)	(3.808)
Industry trend_{igt}	-0.000	0.009	0.009	0.005	0.029
	(-0.077)	(2.092)	(2.272)	(4.612)	(1.381)
Import penetration_{igt}	0.004	0.004	0.004	0.004	0.006
	(2.360)	(2.885)	(2.972)	(3.544)	(2.356)
pmr_{gt}	-0.005	0.004	-0.003	-0.060	-0.066
	(-0.681)	(0.205)	(-0.311)	(-2.672)	(-2.861)

Note: The numbers in parentheses represent t statistics based on country level cluster standard errors.

Table 3: The effect of NRPS on household income: LS estimation

Dependent Variable	Unhealthiness Score			
	Group IFE		Group AFE	
d_F	2	3	Adjusted	Unadjusted
NRPS	0.036	0.064	-0.138	-0.179
	(0.556)	(0.741)	(-2.238)	(-2.550)
Gender	-0.087	-0.082	-0.082	-0.154
(Male=1, Female=0)	(-2.887)	(-2.954)	(-2.471)	(-2.660)
Age	-0.027	-0.025	-0.088	-0.007
	(-5.106)	(-5.401)	(-1.120)	(-0.100)
Age^2	0.040	0.039	0.082	0.0263
	(6.038)	(6.442)	(1.513)	(0.500)
Education	Yes	Yes	Yes	Yes

Note: The numbers in parentheses represent t statistics based on county level cluster standard errors.

Policy Analysis Using Multilevel Regression Models with Group Interactive Fixed Effects

Supplementary Appendix

Zhenhao Gong
Shanxi University of Finance and Economics

Min Seong Kim
University of Connecticut

S.1 Monte Carlo simulation

This section reports simulation results to document the properties of the proposed procedures. The data is generated from the multilevel regression model:

$$Y_{igt} = \beta_Z^0 Z_{gt} + \beta_W^0 W_{igt} + u_{igt}, \tag{S.1}$$

where Z_{gt} is a scalar group level regressor and W_{igt} is a scalar individual regressor. We set $\beta_Z^0 = \beta_W^0 = 0$.

We first examine the performance of our method using the interactive fixed effects structure. We generate u_{igt} based on the following DGP:

$$\begin{aligned} u_{igt} &= \lambda_{ig}^{0\prime} F_t^0 + \varepsilon_{igt}, \\ \lambda_{ig}^0 &= \lambda_g^0 + \delta \cdot \gamma_{ig} \text{ with } \lambda_g^0 = \left(\lambda_{1,g}^0, \lambda_{2,g}^0\right)', \ \gamma_{ig} = \left(\gamma_{1,ig}, \gamma_{2,ig}\right)', \\ \lambda_{c,g}^0 &\sim^{iid} U(0,1) \text{ and } \gamma_{c,ig} \sim^{iid} U(-0.5, 0.5) \text{ for } c = 1, 2, \\ F_t^0 &\sim^{iid} N(0, I_2), \ \varepsilon_{igt} \sim^{iid} N(0, 4). \end{aligned} \tag{S.2}$$

λ_{ig}^0 is the factor loading for individual i in group g, which consists of two components, i.e., λ_g^0 : group factor loading, and γ_{ig} : within-group individual heterogeneity. Since γ_{ig} has zero means, the group factor loading is the group mean of individual factor loadings in this setup. δ determines the impact of individual heterogeneity on the factor loading. If $\delta = 0$, we have $\lambda_{ig}^0 = \lambda_g^0$ and (S.2) reduces to our group interactive fixed effects structure.

Z_{gt} and W_{igt} are generated from the following processes:

$$Z_{gt} = 0.5\lambda_g^{0\prime} F_t^0 + 0.5 F_t^{0\prime} 1_2 + 0.5 \lambda_g^{0\prime} 1_2 + v_{gt}^Z, \tag{S.3}$$

$$W_{igt} = 0.5\lambda_{ig}^{0\prime} F_t^0 + 0.5 F_t^{0\prime} 1_2 + 0.5 \lambda_{ig}^{0\prime} 1_2 + 0.5 Z_{gt} + v_{igt}^W, \tag{S.4}$$

where $1_2 = (1,1)'$ and $\left(v_{gt}^Z, v_{igt}^W\right)' \sim^{iid} N(0, 4I_2)$. (S.3) and (S.4) show that the group level regressor Z_{gt} is corrleated with λ_g^0 whereas individual level regressor W_{igt} is correlated with λ_{ig}^0. Additionally, Z_{gt} is correlated with W_{igt}, allowing the endogeneity of W_{igt} to affect our estimation of β_Z^0. We make inference on β_Z^0 and β_W^0 based on the procedure in Section 2. The number of replications is 5000.

Table S.1 provides the bias, standard deviation (SD), and empirical rejection probability (ERP) of our group interactive fixed effects approach with $\delta = 0$. We use individual based cluster variance estimation to obtain the ERPs. $\tilde{d}_F$ denotes the number of interaction terms used to estimate the model. The table reveals a few findings. Firstly, when $\tilde{d}_F$ is the same or larger than the number of interaction terms in the DGP ($d_F = 2$), our method performs well and produces valid estimation and inference results. However, when our regression model includes only one interactive term ($\tilde{d}_F = 1$), our approach fails to yield valid results. Moon and Weidner (2015) show that the limiting distribution of the standard interactive fixed effects estimator is independent of the number of interaction terms in the regression model as long as this number is not smaller than the true number of interaction terms. Our simulation results suggest that our estimator may have the same property. The performance of our method becomes more accurate as n and T increase.

Table S.2 examines the performance of our method when group sizes are heterogenous. The table shows that though the presence of heterogeneity in group sizes has a slight negative impact on the performance of our method in finite samples, it exhibits solid performance as the sample size increases. For example, when there are 7 groups with a group size of $n_g = 100$ and 10 groups with a group size of $n_g = 30$, the ERP for β_0^Z is 0.070 for a sample size of $(n,T) = (1000, 5)$ and $\tilde{d}_F = 2$. However, as the sample size increases to $(n,T) = (1000, 10)$, the ERP becomes very close to the nominal size.

Table S.3 compares the proposed method with the standard interactive fixed effects approach using different values of δ. When we set $\delta = 0$, our model is correctly specified, and the group interactive fixed effects approach performs very well. However, when $\delta = 0.5$ and 1, in which case our model is misspecified, our estimation and inference for β_W^0 and β_Z^0 become invalid. This is because the group factor loading structure fails to accommodate individual heterogeneity in λ_{ig}^0, leading to an endogeneity bias of group interactive fixed effects estimator. The performance becomes worse as the degree of individual heterogeneity increases. For example, when $\delta = 0.5$ and $(n,T) = (1000, 10)$, the bias, SD and ERP of $\hat{\beta}_Z$ are -0.002, 0.013 and 0.067, respectively, and they become -0.010, 0.014 and 0.133 when $\delta = 1.0$. Regarding the standard interactive fixed effects approach, we find that it does not yield accurate inference results when T is small. For example, when $(n,T) = (1000, 5)$, the ERP for β_Z^0 is 0.111 with $\delta = 0$ and 0.085 with $\delta = 1$. They improve to 0.075 and 0.059, respectively, when $(n,T) = (1000, 20)$. This result is well expected because the consistency of the standard interactive fixed effects estimator requires both n and T to diverge.

In Table S.4, we compare our method with the additive fixed effects approach. For this comparison, we generate u_{igt} based on the interactive model in (S.2) and the additive model given by

$$u_{igt} = \alpha_g^0 + f_t^0 + \varepsilon_{igt}, \tag{S.5}$$
$$\alpha_g^0 \sim^{iid} U(0,1), \; f_t^0 \sim^{iid} N(0,1), \; \varepsilon_{igt} \sim^{iid} N(0,4).$$

The table demonstrates that our approach performs almost as well as the additive fixed effects approach when u_{igt} is generated from the additive model. For instance, when $(n, T) = (2000, 10)$, the bias, SD and ERP of $\hat{\beta}_Z$ are 0.000, 0.009 and 0.055, respectively. These values are very close to 0.000, 0.009 and 0.051, obtained by the additive fixed effects approach. However, when u_{igt} is generated from the interactive model in (S.2), only our approach performs properly. For example, when $(n, T) = (2000, 10)$, the ERP based on $\hat{\beta}_Z$ is 0.052, while the one based on the additive fixed effects approach is 0.421. Therefore, there is a remarkable improvement of our method compared to the conventional additive fixed effects approach when u_{igt} exhibits the interactive structure, while its loss of accuracy when u_{igt} is generated from the additive model is negligible.

In Table S.5, we present the performance of our group level test. We consider two candidate group structures, $\mathbb{A}_0 = \{\mathcal{A}_1, ..., \mathcal{A}_g, ..., \mathcal{A}_G\}$ and $\mathbb{A}_a = \{\mathcal{A}_1^{(1)}, \mathcal{A}_1^{(2)}, ..., \mathcal{A}_g^{(1)}, \mathcal{A}_g^{(2)}, ..., \mathcal{A}_G^{(1)}, \mathcal{A}_G^{(2)}\}$, where $\mathcal{A}_g^{(1)}$ and $\mathcal{A}_g^{(2)}$ in $\mathbb{A}_a$ have the same group size for $g = 1, ..., G$. First, we examine the size property of the test. In this simulation, we generate data based on (S.1)-(S.4) using the null group structure $\mathbb{A}_0$, where the group structure for the factor loading is identical to the one for the group level regressor. The table presents ERPs at a 5% nominal level and we observe that the size of our test is well controlled.

Next, we investigate the power of the test. To examine the power properties, we generate the factor loading based on the finer level of grouping $\mathbb{A}_a$, while the group level regressor is still generated based on $\mathbb{A}_0$. We simulate the regressors from the following processes

$$\begin{aligned}
Z_{gt} &= 0.5\bar{\lambda}_g' F_t^0 + 0.5F_t^{0\prime} 1_2 + 0.5\bar{\lambda}_g' 1_2 + v_{gt}^Z, \\
W_{igt}^{(\ell)} &= 0.5\lambda_g^{(\ell)\prime} F_t^0 + 0.5F_t^{0\prime} 1_2 + 0.5\lambda_g^{(\ell)\prime} 1_2 + 0.5Z_{gt} + v_{igt}^W, \ \ell = 1, 2,
\end{aligned}$$

where $v_{gt}^Z, v_{igt}^W \sim^{iid} N(0, 4)$, $\bar{\lambda}_g = \frac{1}{2}\sum_{\ell=1}^2 \lambda_g^{(\ell)}$, and $\lambda_g^{(\ell)} = \left(\lambda_{1,g}^{(\ell)}, \lambda_{2,g}^{(\ell)}\right)'$ is generated from

$$\lambda_{c,g}^{(1)} \sim^{iid} U(0, 1) \text{ and } \lambda_{c,g}^{(2)} = (1 + \tau)\lambda_{c,g}^{(1)}, \ c = 1, 2. \tag{S.6}$$

(S.6) implies the group factor loading becomes more heterogenous between $\mathcal{A}_g^{(1)}$ and $\mathcal{A}_g^{(2)}$ as τ increases. Table S.5 demonstrates that our test exhibits good power properties with respect to (S.6). The power of the test monotonically increases as the degree of heterogeneity between $\mathcal{A}_g^{(1)}$ and $\mathcal{A}_g^{(2)}$ grows. Since our simulation shows that the size of the test is well controlled when $G = 20$, $G_a = 40$ and $\tilde{d}_F = 2$, we calculate the power without adjusting the size.

In Table S.6, we examine the properties of the LS estimator and the IFE-GMM estimator when the group level regressor is endogenous with respect to idiosyncratic errors. We generate data based on (S.1), (S.2) and (S.4), and consider the following specifications

$$\Psi_{gt}^Z = 0.5\lambda_g^{0\prime} F_t^0 + 0.5F_t^{0\prime} 1_2 + 0.5\lambda_g^{0\prime} 1_2 + v_{gt}^\Psi \text{ with } v_{gt}^\Psi \sim^{iid} N(0, 4),$$

and

$$Z_{gt} = \Psi_{gt}^Z + \bar{\varepsilon}_{gt} + v_{gt}^Z, \tag{S.7}$$

$$W_{igt} = 0.5\lambda_{ig}^{0\prime} F_t^0 + 0.5F_t^{0\prime} 1_2 + 0.5\lambda_{ig}^{0\prime} 1_2 + 0.5\Psi_{gt}^Z + v_{igt}^W, \tag{S.8}$$

where $\left(v_{gt}^Z, v_{igt}^W\right)' \sim^{iid} N(0, 4I_2)$ and $\bar{\varepsilon}_{gt} = n_g^{-1}\sum_{i\in\mathcal{A}_g} \varepsilon_{igt}$ is the group average of idiosyncratic errors. The table shows that the LS estimator of β_Z^0 suffers from substantial bias and size distortion, while the IFE-GMM successfully addresses this issue. For example, when $(n, T) = (2000, 10)$ with $G = 50$, the bias and ERP of the LS estimator are 0.014 and 0.632, respectively. However, when we use the proposed IFE-GMM method, they become 0.000 and 0.057, respectively.

Table S.1: Bias, Standard Deviation and ERP of group interactive fixed effects estimators ($n_g = n/G$, $d_F = 2$, $\delta = 0$)

			$\beta_W^0 = 0$			$\beta_Z^0 = 0$		
	n	T	Bias	SD	ERP	Bias	SD	ERP
	1000	5	0.009	0.017	0.127	0.008	0.026	0.169
$G = 20$	1000	10	0.009	0.012	0.187	0.006	0.017	0.183
$\tilde{d}_F = 1$	2000	5	0.009	0.014	0.174	0.007	0.022	0.247
	2000	10	0.009	0.010	0.270	0.006	0.014	0.260
	1000	5	0.000	0.014	0.054	0.000	0.022	0.060
$G = 20$	1000	10	0.000	0.010	0.055	0.000	0.013	0.054
$\tilde{d}_F = 2$	2000	5	0.000	0.010	0.053	0.000	0.015	0.062
	2000	10	-0.000	0.007	0.044	-0.000	0.009	0.053
	1000	5	0.000	0.014	0.052	0.002	0.030	0.084
$G = 20$	1000	10	-0.000	0.010	0.056	0.000	0.015	0.074
$\tilde{d}_F = 3$	2000	5	-0.000	0.010	0.051	0.001	0.020	0.077
	2000	10	0.000	0.007	0.042	-0.000	0.011	0.064
	1000	5	0.009	0.017	0.129	0.007	0.021	0.118
$G = 50$	1000	10	0.009	0.012	0.188	0.006	0.014	0.127
$\tilde{d}_F = 1$	2000	5	0.009	0.014	0.193	0.007	0.018	0.180
	2000	10	0.009	0.010	0.285	0.006	0.011	0.195
	1000	5	0.000	0.015	0.057	0.000	0.021	0.062
$G = 50$	1000	10	0.000	0.010	0.053	0.001	0.013	0.053
$\tilde{d}_F = 2$	2000	5	0.000	0.010	0.056	0.000	0.015	0.060
	2000	10	-0.000	0.007	0.048	0.000	0.009	0.051
	1000	5	0.000	0.015	0.053	0.001	0.027	0.074
$G = 50$	1000	10	0.000	0.010	0.057	0.001	0.014	0.065
$\tilde{d}_F = 3$	2000	5	0.000	0.010	0.055	0.000	0.019	0.075
	2000	10	0.000	0.007	0.052	-0.000	0.010	0.064

The individual based cluster variance estimation is used to obtain ERPs. The nominal size is 0.05.

Table S.2: Bias, Standard Deviation and ERP of group interactive fixed effects estimators with heterogenous group sizes ($d_F = 2$, $\delta = 0$)

	n	T	$\beta_W^0 = 0$ Bias	SD	ERP	$\beta_Z^0 = 0$ Bias	SD	ERP
$G = 17$	7 groups with $n_g = 100$ and 10 groups with $n_g = 30$							
$\tilde{d}_F = 2$	1000	5	0.000	0.015	0.055	0.000	0.023	0.070
	1000	10	0.000	0.010	0.054	0.000	0.013	0.055
$\tilde{d}_F = 3$	1000	5	0.000	0.015	0.054	0.002	0.032	0.088
	1000	10	0.000	0.010	0.055	-0.000	0.016	0.070
$G = 17$	7 groups with $n_g = 200$ and 10 groups with $n_g = 60$							
$\tilde{d}_F = 2$	2000	5	0.000	0.010	0.050	0.000	0.016	0.063
	2000	10	-0.000	0.007	0.048	0.000	0.009	0.053
$\tilde{d}_F = 3$	2000	5	0.000	0.010	0.050	0.001	0.022	0.087
	2000	10	-0.000	0.007	0.048	0.001	0.011	0.072
$G = 35$	10 groups with $n_g = 50$ and 25 groups with $n_g = 20$							
$\tilde{d}_F = 2$	1000	5	-0.000	0.014	0.055	0.001	0.022	0.066
	1000	10	0.000	0.010	0.051	-0.000	0.013	0.054
$\tilde{d}_F = 3$	1000	5	-0.001	0.014	0.057	0.002	0.028	0.082
	1000	10	0.000	0.010	0.054	0.000	0.015	0.072
$G = 35$	10 groups with $n_g = 100$ and 25 groups with $n_g = 40$							
$\tilde{d}_F = 2$	2000	5	0.000	0.010	0.056	-0.000	0.015	0.060
	2000	10	0.000	0.007	0.055	0.000	0.009	0.050
$\tilde{d}_F = 3$	2000	5	0.000	0.010	0.057	-0.000	0.020	0.078
	2000	10	0.000	0.007	0.056	0.000	0.010	0.065

The individual based cluster variance estimation is used to obtain ERPs. The nominal size is 0.05.

Table S.3: Comparison between the group interactive fixed effects estimator and standard interactive fixed effects estimator ($G = 20$, $n_g = n/G$, $d_F = \tilde{d}_F = 2$)

		$\beta_W^0 = 0$			$\beta_Z^0 = 0$		
n	T	Bias	SD	ERP	Bias	SD	ERP
				$\delta = 0$			
				Group Interactive FE			
1000	5	0.000	0.014	0.054	0.000	0.022	0.060
1000	10	0.000	0.010	0.054	0.000	0.013	0.054
1000	20	0.000	0.007	0.045	0.000	0.009	0.052
				Standard Interactive FE			
1000	5	0.006	0.019	0.086	0.003	0.025	0.111
1000	10	0.004	0.012	0.078	0.002	0.015	0.092
1000	20	0.002	0.007	0.062	0.001	0.009	0.075
				$\delta = 0.5$			
				Group Interactive FE			
1000	5	0.006	0.014	0.073	-0.002	0.022	0.063
1000	10	0.005	0.010	0.090	-0.002	0.013	0.067
1000	20	0.005	0.007	0.119	-0.003	0.009	0.061
				Standard Interactive FE			
1000	5	0.006	0.020	0.093	0.002	0.024	0.099
1000	10	0.004	0.012	0.080	0.001	0.015	0.089
1000	20	0.002	0.007	0.061	0.001	0.009	0.065
				$\delta = 1$			
				Group Interactive FE			
1000	5	0.021	0.018	0.322	-0.010	0.022	0.094
1000	10	0.020	0.013	0.505	-0.010	0.014	0.133
1000	20	0.020	0.009	0.748	-0.010	0.009	0.223
				Standard Interactive FE			
1000	5	0.008	0.020	0.105	0.001	0.023	0.085
1000	10	0.005	0.012	0.093	0.001	0.014	0.075
1000	20	0.003	0.007	0.066	0.001	0.009	0.059

The individual based cluster variance estimation is used to obtain ERPs. The nominal size is 0.05.

Table S.4: Comparison between the group interactive fixed effects estimator and additive fixed effects estimator ($G = 20$, $n_g = n/G$, $d_F = \tilde{d}_F = 2$, $\delta = 0$)

			$\beta_W^0 = 0$			$\beta_Z^0 = 0$	
n	T	Bias	SD	ERP	Bias	SD	ERP
			Additive Structure in (S.5)				
			Group Interactive FE				
1000	5	0.000	0.014	0.054	-0.000	0.021	0.058
1000	10	-0.000	0.010	0.055	0.000	0.013	0.056
2000	5	-0.000	0.010	0.054	0.000	0.015	0.063
2000	10	-0.000	0.007	0.047	0.000	0.009	0.055
			Additive FE				
1000	5	0.000	0.014	0.051	-0.000	0.018	0.044
1000	10	-0.000	0.010	0.056	0.000	0.012	0.057
2000	5	-0.000	0.010	0.052	0.000	0.013	0.059
2000	10	-0.000	0.007	0.048	0.000	0.009	0.051
			Interactive Structure in (S.2)				
			Group Interactive FE				
1000	5	0.000	0.014	0.054	0.000	0.022	0.060
1000	10	0.000	0.010	0.055	0.000	0.013	0.054
2000	5	0.000	0.010	0.051	0.000	0.015	0.060
2000	10	-0.000	0.007	0.044	0.000	0.009	0.052
			Additive FE				
1000	5	0.016	0.016	0.215	0.012	0.030	0.249
1000	10	0.017	0.012	0.401	0.012	0.020	0.296
2000	5	0.015	0.013	0.338	0.013	0.027	0.369
2000	10	0.017	0.010	0.622	0.012	0.018	0.421

The individual based cluster variance estimation is used to obtain ERPs. The nominal size is 0.05.

Table S.5: Group level test for group factor loadings ($n_g = n/G$, $d_F = 2$, $\delta = 0$)

n	T	Empirical Size	
		$G = 20,\ G_a = 40,\ d_{\tilde{F}} = 2$	$G = 20, G_a = 40, d_{\tilde{F}} = 3$
1000	5	0.064	0.067
1000	10	0.058	0.112
2000	5	0.058	0.057
2000	10	0.052	0.106
		$G = 20,\ G_a = 100,\ d_{\tilde{F}} = 2$	$G = 20,\ G_a = 100,\ d_{\tilde{F}} = 3$
1000	5	0.096	0.083
1000	10	0.067	0.125
2000	5	0.072	0.074
2000	10	0.059	0.118
n	T	Power	
		$G = 20,\ G_a = 40,\ d_{\tilde{F}} = 2$	
		$\tau = 0.05$	$\tau = 0.1$
1000	5	0.594	0.650
1000	10	0.766	0.850
2000	5	0.599	0.709
2000	10	0.783	0.931
		$\tau = 0.3$	$\tau = 0.5$
1000	5	0.932	0.984
1000	10	0.997	1.000
2000	5	0.973	0.993
2000	10	0.999	1.000

The nominal size is 0.05.

Table S.6: Comparison between the LS estimator and IFE-GMM estimator in the presence of endogeneity with respect to idiosyncratic errors ($n_g = n/G$, $d_F = \tilde{d}_F = 2$)

		$\beta_W^0 = 0$			$\beta_Z^0 = 0$		
n	T	Bias	SD	ERP	Bias	SD	ERP
				$G = 20$			
				Group Interactive FE			
1000	5	-0.002	0.014	0.055	0.010	0.015	0.130
1000	10	-0.003	0.010	0.063	0.011	0.009	0.240
2000	5	-0.001	0.010	0.052	0.005	0.010	0.088
2000	10	-0.002	0.007	0.060	0.006	0.006	0.150
				IFE-GMM			
1000	5	0.000	0.014	0.049	-0.000	0.023	0.067
1000	10	-0.000	0.010	0.050	-0.000	0.013	0.054
2000	5	-0.000	0.010	0.046	-0.000	0.016	0.058
2000	10	-0.000	0.007	0.051	-0.000	0.009	0.047
				$G = 50$			
				Group Interactive FE			
1000	5	-0.006	0.014	0.077	0.026	0.014	0.506
1000	10	-0.008	0.010	0.141	0.026	0.008	0.885
2000	5	-0.003	0.010	0.064	0.013	0.010	0.278
2000	10	-0.004	0.007	0.103	0.014	0.006	0.632
				IFE-GMM			
1000	5	0.001	0.014	0.056	0.001	0.021	0.062
1000	10	0.000	0.010	0.055	-0.000	0.013	0.051
2000	5	0.000	0.010	0.047	-0.000	0.015	0.059
2000	10	-0.000	0.007	0.057	0.000	0.009	0.057

The individual based cluster variance estimation is used to obtain ERPs. The nominal size is 0.05.

S.2 When group sizes are time varying

The main results of this paper rely on the assumption that the group sizes remain constant over time (Assumption 1). This assumption may not hold in practice. For example, when we use unbalanced panel or repeated cross-sections, it is likely that group sizes change over time. In such cases, we can reconstruct the dataset to satisfy the assumption, as demonstrated in our empirical applications in Section 6. Alternatively, we can consider a modification of our estimation procedure based on the expectation-maximization (EM) algorithm by Bai (2009, Supplemental Material), which is proposed to estimate the standard interactive fixed effects models with unbalanced panel. Bai's approach is an extension of the EM algorithm by Stock and Watson (1998), which is developed for the pure factor model.

Let n_{gt} denote the size of group g at time t. We also define $n_g = \max\{n_{g1}, ..., n_{gT}\}$, $n = \sum_{g=1}^{n} n_g$. Additionally, we define $n_g^S = \sum_{l=1}^{g} n_{l-1} + 1$ and $n_g^E = \sum_{l=1}^{g} n_l$ with $n_0 = 0$. For each time period, we can index each individual so that all individuals who belong to group g are indexed as $i \in \left[n_g^S, n_g^E\right]$. Let $I_{igt} = 1$ if individaul i exists in group g at time t and $I_{igt} = 0$ otherwise for $i = 1, ..., n$.

The following steps describe our estimation procedure using the EM algorithm.

1. Given F and Λ_G, we estimate β by

$$\hat{\beta} = \left(\sum_{t=1}^{T} \sum_{g=1}^{G} \sum_{i=n_g^S}^{n_g^E} 1\{i \in \mathcal{A}_g\} X_{igt} X_{igt}' \right)^{-1} \sum_{t=1}^{T} \sum_{g=1}^{G} \sum_{i=n_g^S}^{n_g^E} 1\{i \in \mathcal{A}_g\} X_{igt} \left(Y_{igt} - \lambda_g' F_t\right) \quad \text{(S.9)}$$

2. Given β, we let $R_{igt} = Y_{igt} - X_{igt}'\beta$. Then, $R_{igt} = \lambda_g' F_t + \varepsilon_{igt}$ is a pure factor model with group factor loadings, which has missing values. We impute them in each round of iteration based on the EM algorithm by Stock and Watson (1998).

   (a) In round a, we define

   $$R_{igt}^{(a)} = R_{igt} 1\{I_{igt} = 1\} + \hat{\lambda}_g^{(a-1)\prime} \hat{F}_t^{(a-1)} 1\{I_{igt} = 0\},$$

   where $\hat{\lambda}_g^{(0)} = \hat{F}_t^{(0)} = 0$ and $1\{\cdot\}$ is an indicator function.

   (b) $\hat{F}^{(a)} = \left(\hat{F}_1^{(a)}, ..., \hat{F}_T^{(a)}\right)'$ is the $(T \times d_F)$ matrix whose columns are the eigenvectors associated with the d_F largest eigenvalues of $n^{-1}\bar{\mathcal{R}}^{(a)}\bar{\mathcal{R}}^{(a)\prime}$ subject to $\sum_{t=1}^{T} F_t F_t' = I_{d_F}$, where

   $$\bar{\mathcal{R}}^{(a)} = \left[\sqrt{n_1}\bar{R}_1^{(a)}, ..., \sqrt{n_g}\bar{R}_g^{(a)}, ..., \sqrt{n_G}\bar{R}_G^{(a)}\right]$$

   with $\bar{R}_g^{(a)} = \left(n_g^{-1} \sum_{i=n_g^S}^{n_g^E} R_{ig1}^{(a)}, ..., n_g^{-1} \sum_{i=n_g^S}^{n_g^E} R_{igT}^{(a)}\right)'$.

   (c) $\hat{\Lambda}_G^{(a)} = \left(\hat{\lambda}_1^{(a)}, .., \hat{\lambda}_g^{(a)}, .., \hat{\lambda}_G^{(a)}\right)'$ is obtained from

   $$\hat{\lambda}_g^{(a)} = \hat{F}^{(a)\prime} \bar{R}_g^{(a)}.$$

(d) We iterate (a) - (c) to convergence. Let $\hat{F}$ and $\hat{\Lambda}_G$ be the the estimates in the final round.

3. We plug $\hat{F}$ and $\hat{\Lambda}_G$ from step 2 in (S.9) and repeat steps 1 and 2 to update $\hat{\beta}$, $\hat{F}$ and $\hat{\Lambda}_G$. We iterate the procedure to convergence.

S.3 Proofs

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 $\|A\| = \sqrt{tr\,(A'A)}$.

Proof of Theorem 1. Suppose that $\beta^0 = 0$ without loss of generality. $\left(\hat{\beta}, \hat{F}\right)$ minimizes

$$\tilde{\mathcal{Q}}(\beta, F) = \mathcal{Q}\,(\beta, F) - \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\varepsilon_{ig} - P_{F^0}\bar{\varepsilon}_g\right)'\left(\varepsilon_{ig} - P_{F^0}\bar{\varepsilon}_g\right) \tag{S.10}$$

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

$$\begin{aligned}\mathcal{Q}^*(\beta, F) = {} & \frac{1}{n}\sum_{g=1}^{G} n_g \lambda_g^{0\prime} F^{0\prime} M_F F^0 \lambda_g^0 - \frac{2}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\lambda_g^{0\prime}F^{0\prime}M_F\bar{X}_g\beta \\ & + \beta'\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_{ig} - P_F\bar{X}_g\right)'\left(X_{ig} - P_F\bar{X}_g\right)\beta.\end{aligned} \tag{S.11}$$

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

$$\mathcal{Q}^*(\beta^0, F^0H) = 0. \tag{S.12}$$

The first step is to show that

$$\begin{aligned}& \tilde{\mathcal{Q}}(\beta, F) - \mathcal{Q}^*(\beta, F) \\ & = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[2\lambda_g^{0\prime}F^{0\prime}M_F\bar{\varepsilon}_g - 2\left(\varepsilon_{ig}'X_{ig} - \bar{\varepsilon}_g'P_F\bar{X}_g\right)\beta - \bar{\varepsilon}_g'\left(P_F - P_{F^0}\right)\bar{\varepsilon}_g\right] \\ & = o_p(1)\end{aligned} \tag{S.13}$$

for all bounded β and $F \in \mathcal{F} = \{F : F'F = I_{d_F}\}$ as $(n, G) \to \infty$ with T fixed such that $G/n \to 0$. For the first term, we have

$$\begin{aligned}& \left|\frac{1}{n}\sum_{g=1}^{G} n_g\lambda_g^{0\prime}F^{0\prime}M_F\bar{\varepsilon}_g\right| \\ & \leq \left|\frac{1}{n}\sum_{g=1}^{G} n_g\lambda_g^{0\prime}F^{0\prime}\bar{\varepsilon}_g\right| + \left|\frac{1}{n}\sum_{g=1}^{G} n_g\left(\lambda_g^{0\prime}F^{0\prime}F\right)\left(F'\bar{\varepsilon}_g\right)\right| \\ & = a1 + a2.\end{aligned}$$

It is easy to show that $a1 = O_p\left(n^{-\alpha/2}\right)$ under Assumptions 1-5. For $a2$,

$$a2 \leq \left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\lambda_g^{0\prime} F^{0\prime} F\right\|^2\right)^{1/2} \left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|F'\bar{\varepsilon}_g\right\|^2\right)^{1/2}$$
$$= O_p(1)\left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|F'\bar{\varepsilon}_g\right\|^2\right)^{1/2},$$

and

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

Thus, $a1$ and $a2$ are $o_p(1)$ under our rate condition. Using the same procedure, we can show that the second and third terms in (S.13) are also $o_p(1)$. Therefore, (S.13) holds.

The second step is to show

$$\mathcal{Q}^*(\beta, F) > 0 \tag{S.14}$$

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

Proof of Proposition A1. Define $\Lambda^0 = (\underbrace{\lambda_1^0, ..., \lambda_1^0}_{n_1}, ..., \underbrace{\lambda_G^0, ..., \lambda_G^0}_{n_G})'$. From (11), we have

$$\begin{aligned}
&\hat{F}\hat{\Gamma} - F^0\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)\left(F^{0\prime}\hat{F}\right) \\
&= \frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g\left(\beta^0 - \hat{\beta}\right)\left(\beta^0 - \hat{\beta}\right)' \bar{X}_g'\hat{F} \\
&+ \frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g\left(\beta^0 - \hat{\beta}\right)\lambda_g^{0\prime} F^{0\prime}\hat{F} + \frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g\left(\beta^0 - \hat{\beta}\right)\bar{\varepsilon}_g'\hat{F} \\
&+ \frac{1}{n}\sum_{g=1}^{G} n_g F^0\lambda_g^0\left(\beta^0 - \hat{\beta}\right)' \bar{X}_g'\hat{F} + \frac{1}{n}\sum_{g=1}^{G} n_g \bar{\varepsilon}_g\left(\beta^0 - \hat{\beta}\right)' \bar{X}_g'\hat{F} \\
&+ \frac{1}{n}\sum_{g=1}^{G} n_g F^0\lambda_g^0\bar{\varepsilon}_g'\hat{F} + \frac{1}{n}\sum_{g=1}^{G} n_g \bar{\varepsilon}_g\lambda_g^{0\prime} F^{0\prime}\hat{F} + \frac{1}{n}\sum_{g=1}^{G} n_g \bar{\varepsilon}_g\bar{\varepsilon}_g'\hat{F} \\
&= I1 + I2 + \cdots + I8.
\end{aligned} \tag{S.15}$$

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

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

For $I1$,

$$\begin{aligned}
\|I1\| &\leq \frac{1}{n}\sum_{g=1}^{G} n_g \left\|\bar{X}_g\right\|^2 \left\|\beta^0 - \hat{\beta}\right\|^2 \underbrace{\left\|\hat{F}\right\|}_{=1} \\
&= O_p\left(\left\|\hat{\beta} - \beta^0\right\|^2\right).
\end{aligned}$$

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

For $I6$, we have

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

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

For $I8$, we have

$$\begin{aligned}
\|I8\| &= \left\|\frac{1}{n}\sum_{g=1}^{G} n_g \bar{\varepsilon}_g \bar{\varepsilon}_g' \hat{F}\right\| \\
&\leq O_p\left(\frac{G}{n}\right)\frac{1}{G}\sum_{g=1}^{G}\left\|\sqrt{n_g}\bar{\varepsilon}_g\right\|^2\left\|\hat{F}\right\| \\
&= O_p\left(\frac{G}{n}\right).
\end{aligned}$$

Combining $I1$-$I8$, we obtain

$$\begin{aligned}
&\hat{F}\hat{\Gamma}\left(F^{0\prime}\hat{F}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{n}\right)^{-1} - F^0 \\
&= O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right). \qquad \text{(S.16)}
\end{aligned}$$

Premultiplying $\hat{F}'$ in (S.16), we obtain

$$\hat{\Gamma} = \left(\hat{F}'F^0\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)\left(F^{0\prime}\hat{F}\right) + o_p(1). \tag{S.17}$$

As in Bai (2009, Proposition I(ii)), $F^{0\prime}\hat{F}$ is invertible, so $\hat{\Gamma}$ is invertible. Thus, from (S.16) we have

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

■

Lemma S1 *Under Assumptions 1-5, for each g*

$$\sqrt{n_g}\bar{\varepsilon}_g'\left(\hat{F} - F^0 H\right) = o_p\left(\left\|\beta^0 - \hat{\beta}\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right)$$

as $(n, G) \to \infty$ *such that* $G/n \to 0$.

Lemma S2 *Under Assumptions 1-5,*

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

Proof of Lemma S2. Note that

$$\begin{aligned}
HH' - \left(F^{0\prime}F^0\right)^{-1} &= H\left[H'F^{0\prime}F^0H - I\right]\left(H'F^{0\prime}F^0H\right)^{-1}H' \\
&= H\left[H'F^{0\prime}\left(F^0H - \hat{F}\right) + \left(F^0H - \hat{F}\right)'\left(\hat{F} - F^0H\right)\right. \\
&\quad \left. + \left(F^0H - \hat{F}\right)'F^0H\right]\left(H'F^{0\prime}F^0H\right)^{-1}H'.
\end{aligned} \tag{S.18}$$

Thus, it follows from Proposition A1 that

$$HH' - \left(F^{0\prime}F^0\right)^{-1} = O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right). \tag{S.19}$$

■

Lemma S3 *Under Assumptions 1-5,*

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

Proof of Lemma S3. We note that

$$\begin{aligned}
P_{\hat{F}} - F_{F^0} &= \hat{F}\hat{F}' - F^0\left(F^{0\prime}F^0\right)^{-1}F^{0\prime} \\
&= \hat{F}\hat{F}' - \left(F^0H\right)\left(F^0H\right)' + F^0HH'F^{0\prime} - F^0\left(F^{0\prime}F^0\right)^{-1}F^{0\prime} \\
&= O_p\left(\left\|\hat{\beta}-\beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right)
\end{aligned}$$

due to Proposition A1 and Lemma S2. ■

Proof of Proposition A2. Note that

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}-\beta^0\right) = &\left[\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_{ig}-P_{\hat{F}}\bar{X}_g\right)'\left(X_{ig}-P_{\hat{F}}\bar{X}_g\right)\right]^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\Bigg(X_{ig}'M_{\hat{F}}F^0\lambda_g^0 \\
&+\left(X_{ig}-P_{\hat{F}}\bar{X}_g\right)'\varepsilon_{ig}\Bigg).
\end{aligned} \tag{S.20}$$

Let $\Upsilon = \left(F^{0\prime}\hat{F}\right)^{-1}\left(\Lambda^{0\prime}\Lambda^0/n\right)^{-1}$. For the first term of (S.20), using $M_{\hat{F}}\hat{F}=0$, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}X_{ig}'M_{\hat{F}}F^0\lambda_g^0 &= \frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}X_{ig}'M_{\hat{F}}\left[F^0-\hat{F}\hat{\Gamma}\left(F^{0\prime}\hat{F}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)^{-1}\right]\lambda_g^0 \\
&= \frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}X_{ig}'M_{\hat{F}}\left[F^0\left(\frac{\Lambda^{0\prime}\Lambda^0}{n}\right)\left(F^{0\prime}\hat{F}\right)-\hat{F}\hat{\Gamma}\right]\Upsilon\lambda_g^0 \\
&= -\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}X_{ig}'M_{\hat{F}}\left(I1+\cdots+I8\right)\Upsilon\lambda_g^0 \\
&:= J1+\cdots+J8.
\end{aligned}$$

For $J1$,

$$\begin{aligned}
\|J1\| &= \left\|\frac{1}{\sqrt{n}}\sum_{g=1}^{G}n_g\bar{X}_g'M_{\hat{F}}\left(I1\right)\Upsilon\lambda_g^0\right\| \\
&\leq \frac{1}{n\sqrt{n}}\left(\frac{1}{n}\sum_{g=1}^{G}n_g\left\|\bar{X}_g'\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{g=1}^{G}n_g\left\|\lambda_g^0\right\|^2\right)^{1/2}\|\Upsilon\|\,o_p\left(\left\|\hat{\beta}-\beta^0\right\|\right) \\
&= o_p\left(\left\|\sqrt{n}\left(\hat{\beta}-\beta^0\right)\right\|\right),
\end{aligned} \tag{S.21}$$

in which we use

$$\begin{aligned}
\left\|\bar{X}_g'M_{\hat{F}}\right\|^2 &= tr\left(\bar{X}_g'\bar{X}_g\right) - tr\left(\bar{X}_g'\hat{F}\hat{F}'\bar{X}_g\right) \\
&= \left\|\bar{X}_g\right\|^2 - \left\|\hat{F}'\bar{X}_g\right\|^2 \leq \left\|\bar{X}_g\right\|^2.
\end{aligned} \tag{S.22}$$

For $J2$, we have

$$\begin{aligned} J2 &= \frac{1}{n^2} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \lambda_{\tilde{g}}^{0\prime} \left(\frac{\Lambda^{0\prime}\Lambda^0}{n} \right)^{-1} \lambda_g^0 \sqrt{n} \left(\hat{\beta} - \beta^0 \right) \\ &= \frac{1}{n^2} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} a_{g\tilde{g}}^0 n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \sqrt{n} \left(\hat{\beta} - \beta^0 \right) \\ &= O_p \left(\sqrt{n} \left(\hat{\beta} - \beta^0 \right) \right), \end{aligned} \tag{S.23}$$

because

$$\begin{aligned} &\left\| \frac{1}{n^2} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \left\{ \lambda_g^{0\prime} \left(\frac{\Lambda^{0\prime}\Lambda^0}{n} \right)^{-1} \lambda_{g_i}^0 \right\} \right\| \\ &\leq \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\| X_{ig}' \right\| \left\| \lambda_g^0 \right\| \right)^2 \left\| \left(\frac{\Lambda^{0\prime}\Lambda^0}{n} \right)^{-1} \right\| = O_p(1). \end{aligned}$$

For $J3$,

$$\begin{aligned} \|J3\| &= \left\| \frac{1}{n\sqrt{n}} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \left(\beta^0 - \hat{\beta} \right) \bar{\varepsilon}_{\tilde{g}}' \hat{F} \Upsilon \lambda_g^0 \right\| \\ &\leq \frac{\sqrt{G}}{\sqrt{n}} \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\| \bar{X}_g' \right\|^2 \right) \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\| \lambda_g^0 \right\|^2 \right)^{1/2} \left(\frac{1}{G} \sum_{\tilde{g}=1}^{G} \left\| \sqrt{n} \bar{\varepsilon}_{\tilde{g}}' \right\|^2 \right)^{1/2} \\ &\times \|\Upsilon\| \left\| \sqrt{n} \left(\beta^0 - \hat{\beta} \right) \right\| \\ &= o_p \left(\sqrt{n} \left(\beta^0 - \hat{\beta} \right) \right). \end{aligned}$$

For $J4$,

$$\begin{aligned} \|J4\| &= \left\| \frac{1}{n\sqrt{n}} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \left(F^0 - \hat{F} H^{-1} \right) \lambda_{\tilde{g}}^0 \left(\beta^0 - \hat{\beta} \right)' \bar{X}_{\tilde{g}}' \hat{F} \Upsilon \lambda_g^0 \right\| \\ &\leq \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\| \bar{X}_g' \right\|^2 \right) \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\| \lambda_g^0 \right\|^2 \right)^{1/2} \left(\frac{1}{n} \sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \left\| \lambda_{\tilde{g}}^0 \right\|^2 \right)^{1/2} \\ &\times \left\| \hat{F} \right\| \|\Upsilon\| \left\| F^0 - \hat{F} H^{-1} \right\| \left\| \sqrt{n} \left(\beta^0 - \hat{\beta} \right) \right\| \\ &= o_p \left(\left\| \sqrt{n} \left(\beta^0 - \hat{\beta} \right) \right\| \right). \end{aligned}$$

For $J5$,

$$\begin{aligned}\|J5\| &= \left\| \frac{1}{n\sqrt{n}} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} n_g \bar{X}_g' M_{\hat{F}} n_{\tilde{g}} \bar{\varepsilon}_{\tilde{g}} \left(\beta^0 - \hat{\beta}\right)' \bar{X}_{\tilde{g}}' \hat{F} \Upsilon \lambda_g^0 \right\| \\ &\leq O\left(\frac{G}{n}\right) \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\|\bar{X}_g'\right\|^2\right)^{1/2} \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\|\lambda_g^0\right\|^2\right)^{1/2} \left(\frac{1}{G} \sum_{\tilde{g}=1}^{G} \left\|\sqrt{n_{\tilde{g}}} \bar{\varepsilon}_{\tilde{g}}\right\|^2\right)^{1/2} \\ &\times \left(\frac{1}{n} \sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \left\|\bar{X}_{\tilde{g}}'\right\|\right)^{1/2} \left\|\sqrt{n}\left(\beta^0 - \hat{\beta}\right)'\right\| \left\|\hat{F}\right\| \|\Upsilon\| \\ &= o_p\left(\left\|\sqrt{n}\left(\beta^0 - \hat{\beta}\right)\right\|\right).\end{aligned}$$

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

$$\begin{aligned}\|J6\| &= \left\| \frac{1}{n\sqrt{n}} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \left(F^0 - \hat{F} H^{-1}\right) \lambda_{\tilde{g}}^0 \bar{\varepsilon}_{\tilde{g}}' \hat{F} \Upsilon \lambda_g^0 \right\| \\ &\leq \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\|\bar{X}_g'\right\|^2\right)^{1/2} \left(\frac{1}{n} \sum_{g=1}^{G} \left\|\lambda_g^0\right\|^2\right)^{1/2} \left\|\hat{F}\right\| \|\Upsilon\| \qquad \text{(S.24)} \\ &\times \left\| \frac{1}{\sqrt{n}} \sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \lambda_{\tilde{g}}^0 \bar{\varepsilon}_{\tilde{g}}' \right\| \left\| F^0 - \hat{F} H^{-1} \right\| \qquad \text{(S.25)} \\ &= O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right).\end{aligned}$$

where

$$P\left(\left\| \frac{1}{\sqrt{n}} \sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \lambda_{\tilde{g}}^0 \bar{\varepsilon}_{\tilde{g}}' \right\| > \Delta\right) \leq O(1) \frac{1}{n} \sum_{\tilde{g}_1=1}^{G} \sum_{\tilde{g}_2=1}^{G} n_{\tilde{g}_1} n_{\tilde{g}_2} E\left(\bar{\varepsilon}_{\tilde{g}_1}' \bar{\varepsilon}_{\tilde{g}_2}\right) < M.$$

For $J7$, since $a_{\tilde{g}g}^0$ is a scalar,

$$J7 = -\frac{1}{n\sqrt{n}} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} a_{\tilde{g}g}^0 \bar{X}_g' M_{\hat{F}} \bar{\varepsilon}_{\tilde{g}}.$$

Let

$$A_n = -\frac{1}{n} \sum_{g=1}^{G} n_g \bar{X}_g' M_{F^0} \Omega F^0 H \Upsilon \lambda_g^0$$

where $\Omega = \frac{1}{G} \sum_{\tilde{g}=1}^{G} E\left(n_{\tilde{g}} \bar{\varepsilon}_{\tilde{g}} \bar{\varepsilon}_{\tilde{g}}'\right)$.

Then, $J8$ can be rewritten as

$$\begin{aligned}
J8 &= -\frac{G}{\sqrt{n}}\frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' M_{\hat{F}} \left(\frac{1}{G}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}'\right) \hat{F}\Upsilon\lambda_g^0 \\
&= -\frac{G}{\sqrt{n}}\frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' M_{\hat{F}} \Omega \hat{F}\Upsilon\lambda_g^0 \\
&- \sqrt{\frac{G}{n}}\frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' M_{\hat{F}} \left(\frac{1}{\sqrt{G}}\sum_{\tilde{g}=1}^{G} \left(n_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}' - \Omega\right)\right) \hat{F}\Upsilon\lambda_g^0 \\
&= J81 + J82.
\end{aligned}$$

For $J81$,

$$\begin{aligned}
J81 &= -\frac{G}{\sqrt{n}}\frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' M_{F^0} \Omega F^0 H \Upsilon\lambda_g^0 \\
&- \frac{G}{\sqrt{n}}\frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' \left(M_{\hat{F}} - M_{F^0}\right) \Omega F^0 H \Upsilon\lambda_g^0 \\
&- \frac{G}{\sqrt{n}}\frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' M_{\hat{F}} \Omega \left(\hat{F} - F^0 H\right) \Upsilon\lambda_g^0 \\
&= \frac{G}{\sqrt{n}} A_n - \frac{G}{\sqrt{n}} b_{11} - \frac{G}{\sqrt{n}} b_{12}.
\end{aligned}$$

For b_{11},

$$\begin{aligned}
b_{11} &= \frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' \left(P_{F^0} - P_{\hat{F}}\right) \Omega F^0 H \Upsilon\lambda_g^0 \\
&= O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right) \qquad \text{(S.26)}
\end{aligned}$$

under our rate condition because of Lemma S3. It is also easy to show that $b_{12} = o_P(1)$. Thus,

$$J81 = \frac{G}{\sqrt{n}} A_n + o_P\left(\frac{G}{\sqrt{n}}\right).$$

For $J82$, we note that

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{g=1}^{G} n_g \bar{X}_g' M_{\hat{F}} \left(\frac{1}{\sqrt{G}}\sum_{\tilde{g}=1}^{G} \left(n_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}' - \Omega\right)\right) \hat{F}\Upsilon\lambda_g^0\right\| \\
&\leq \left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\bar{X}_g'\right\|^2\right)^{1/2} \left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\lambda_g^0\right\|^2\right)^{1/2} \left\|\frac{1}{\sqrt{G}}\sum_{\tilde{g}=1}^{G} \left(n_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}\bar{\varepsilon}_{\tilde{g}}' - \Omega\right)\right\| \left\|\hat{F}\right\| \left\|\Upsilon\right\| \\
&= O_p(1),
\end{aligned}$$

because for all $t, s = 1, ..., T$

$$
\begin{aligned}
&P\left(\left|\frac{1}{\sqrt{G}}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}}\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}s} - E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}s}\right)\right)\right| > \Delta\right) \\
&\leq \frac{1}{\Delta^2}E\left(\frac{1}{\sqrt{G}}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}}\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}s} - E\left(\bar{\varepsilon}_{\tilde{g}t}\bar{\varepsilon}_{\tilde{g}s}\right)\right)\right)^2 \\
&= O(1)
\end{aligned}
$$

as $(n, G) \to \infty$ under Assumption 4(iv). Therefore, $J82 = O_P\left(\sqrt{\frac{G}{n}}\right)$ and we have

$$
J8 = \frac{G}{\sqrt{n}}A_n + o_p\left(\frac{G}{\sqrt{n}}\right).
$$

From $J1$-$J8$, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g} X_{ig}' M_{\hat{F}} F^0 \lambda_g^0 \\
&= \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} a_{g\tilde{g}}^0 n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \sqrt{n}\left(\hat{\beta} - \beta^0\right) - \frac{1}{n\sqrt{n}}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} a_{\tilde{g}g}^0 \bar{X}_g' M_{\hat{F}} \bar{\varepsilon}_{\tilde{g}} \\
&+ \frac{G}{\sqrt{n}}A_n + o_p\left(\sqrt{n}\left(\beta^0 - \hat{\beta}\right)\right) + o_p\left(\frac{G}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right). \qquad \text{(S.27)}
\end{aligned}
$$

Combining this result with (S.20), we have

$$
\begin{aligned}
&\left(\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_{ig} - P_{\hat{F}}\bar{X}_g\right)'\left(X_{ig} - P_{\hat{F}}\bar{X}_g\right) - \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} a_{g\tilde{g}}^0 n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}}\right)\sqrt{n}\left(\hat{\beta} - \beta^0\right) \\
&= \frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\left(X_{ig} - P_{\hat{F}}\bar{X}_g\right)' - \frac{1}{n}\sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} a_{\tilde{g}g}^0 \bar{X}_{\tilde{g}}' M_{\hat{F}}\right)'\varepsilon_{ig} + \frac{G}{\sqrt{n}}A_n \\
&+ o_p\left(\sqrt{n}\left(\beta^0 - \hat{\beta}\right)\right) + o_p\left(\frac{G}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

By premultiplying $B_n^{XX}\left(\hat{F}\right)^{-1}$, we have

$$
\begin{aligned}
\sqrt{n}\left(\hat{\beta} - \beta^0\right) &= B_n^{XX}\left(\hat{F}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\left(X_{ig} - P_{\hat{F}}\bar{X}_g\right)' - \frac{1}{n}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}} a_{\tilde{g}g}^0 \bar{X}_{\tilde{g}}' M_{\hat{F}}\right)'\varepsilon_{ig} \\
&+ \frac{G}{\sqrt{n}}B_n^{XX}\left(\hat{F}\right)^{-1}A_n + o_p\left(\sqrt{n}\left(\beta^0 - \hat{\beta}\right)\right) + o_p\left(\frac{G}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

■

Proof of Proposition A3. The proof is omitted because it is the same as the proof of Proposition A1. ■

Proof of Theorem 2. Let $\mathcal{X}_{ig}^{X} = \mathcal{X}_{ig}^{X}(F^0)$ and $B_n^{XX} = B_n^{XX}(F^0)$ for notational simplicity. From Proposition A2, we have

$$\sqrt{n}\left(\hat{\beta} - \beta^0\right) = B_n^{XX}\left(\hat{F}\right)^{-1} \frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \mathcal{X}_{ig}^{X}\left(\hat{F}\right)' \varepsilon_{ig} + o_p(1)$$

as $(n, G) \to \infty$ such that $G/\sqrt{n} \to 0$. Thus, we need to show

$$\text{(i) } \frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \mathcal{X}_{ig}^{X}\left(\hat{F}\right)' \varepsilon_{ig} - \frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^{X}\right)' \varepsilon_{ig} = o_p(1),$$
$$\text{(ii) } B_n^{XX}\left(\hat{F}\right) - B_n^{XX} = o_p(1)$$

to complete the proof.

Part (i) We have

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \mathcal{X}_{ig}^{X}\left(\hat{F}\right)' \varepsilon_{ig} - \frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^{X}\right)' \varepsilon_{ig} \\
&= \frac{1}{\sqrt{n}} \sum_{g=1}^{G} n_g \bar{X}_g' \left(P_{\hat{F}} - P_{F^0}\right) \bar{\varepsilon}_g - \frac{1}{n\sqrt{n}} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} a_{\tilde{g}g}^0 \bar{X}_{\tilde{g}}' \left(M_{F^0} - M_{\hat{F}}\right)' \bar{\varepsilon}_g \\
&= \mathcal{H}1 + \mathcal{H}2.
\end{aligned}$$

For $\mathcal{H}1$,

$$\begin{aligned}
\|\mathcal{H}1\| &\leq O_p\left(\sqrt{G}\right)\left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\bar{X}_g'\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\sqrt{n_g}\bar{\varepsilon}_g\right\|^2\right)^{1/2}\left\|\hat{F}\hat{F}' - P_{F^0}\right\| \\
&= O_p\left(\sqrt{G}\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{\sqrt{G}}{\sqrt{n}}\right) + O_p\left(\frac{G\sqrt{G}}{n}\right) \\
&= o_p\left(\sqrt{n}\left\|\hat{\beta} - \beta^0\right\|\right) + o_p\left(\frac{G}{\sqrt{n}}\right) \qquad \text{(S.28)}
\end{aligned}$$

if $(n, G) \to \infty$ such that $G/\sqrt{n} \to 0$.

For $\mathcal{H}2$, we define $K_g = n^{-1}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}} a_{\tilde{g}g}^0 \bar{X}_{\tilde{g}}'$. Replacing $\bar{X}_g$ with K_g in (S.28), we can use the same procedure for $\mathcal{H}2$, and we have

$$\|\mathcal{H}2\| = o_p\left(\sqrt{n}\left\|\hat{\beta} - \beta^0\right\|\right) + o_p\left(\frac{G}{\sqrt{n}}\right). \qquad \text{(S.29)}$$

From (S.28) and (S.29), we have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \mathcal{X}_{ig}^{X} \left(\hat{F}\right)' \varepsilon_{ig} \\
&= \frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^{X}\right)' \varepsilon_{ig} + o_p\left(\sqrt{n}\left\|\hat{\beta} - \beta^0\right\|\right) + o_p\left(\frac{G}{\sqrt{n}}\right).
\end{aligned}
$$

as $(n, G) \to \infty$ such that $G/\sqrt{n} \to 0$.

Part (ii) We have

$$
\begin{aligned}
&B_n^{XX}\left(\hat{F}\right) - B_n^{XX} \\
&= \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} n_g \bar{X}_g' \left(P_{F^0} - P_{\hat{F}}\right) \bar{X}_g \\
&- \frac{1}{n^2} \sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} a_{g\tilde{g}}^0 n_g n_{\tilde{g}} \bar{X}_g' \left(M_{\hat{F}} - M_{F^0}\right) \bar{X}_{\tilde{g}} \\
&= \mathcal{G}1 + \mathcal{G}2,
\end{aligned}
$$

Using similar procedures in (S.26), we have

$$
\begin{aligned}
\mathcal{G}1 &= O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right), \\
\mathcal{G}2 &= O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right).
\end{aligned}
$$

Combining (i) and (ii), we have

$$
\begin{aligned}
\sqrt{n}\left(\hat{\beta} - \beta\right) &= \left(B_n^{XX} + o_p(1)\right)^{-1} \left[\frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^{X}\right)' \varepsilon_{ig} + \frac{G}{\sqrt{n}} A_n \right. \\
&\left. + o_p\left(\sqrt{n}\left(\beta^0 - \hat{\beta}\right)\right) + o_p\left(\frac{G}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right] \\
&= \left(B_n^{XX}\right)^{-1} \frac{1}{\sqrt{n}} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^{X}\right)' \varepsilon_{ig} + o_p(1)
\end{aligned}
$$

under Assumption 7. ∎

Lemma S4 *Under Assumptions 1-6,*

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

Proof of Lemma S4. Note that

$$
\begin{aligned}
&\frac{1}{\sqrt{G}}\left(\hat{\Lambda}_G' - H^{-1}\Lambda_G^{0\prime}\right)\\
&= \frac{1}{\sqrt{G}}\hat{F}'\left[\left(\bar{Y}_1 - \bar{X}_1\hat{\beta}\right), \cdots, \left(\bar{Y}_G - \bar{X}_G\hat{\beta}\right)\right] - \frac{1}{\sqrt{G}}H^{-1}\Lambda_G^{0\prime}\\
&= \frac{1}{\sqrt{G}}\hat{F}'\left[\left(\bar{X}_1\left(\beta^0 - \hat{\beta}\right) + F^0\lambda_1^0 + \bar{\varepsilon}_1\right), \cdots, \left(\bar{X}_G\left(\beta^0 - \hat{\beta}\right) + F^0\lambda_G^0 + \bar{\varepsilon}_G\right)\right] - \frac{1}{\sqrt{G}}H^{-1}\Lambda_G^{0\prime}\\
&= \frac{\hat{F}'}{\sqrt{G}}\left(F^0 - \hat{F}H^{-1}\right)\Lambda_G^{0\prime} + \frac{\hat{F}'}{\sqrt{G}}\left[\bar{\varepsilon}_1, \cdots, \bar{\varepsilon}_G\right] + \frac{\hat{F}'}{\sqrt{G}}\left[\bar{X}_1, \cdots, \bar{X}_G\right]\left(\beta^0 - \hat{\beta}\right)\\
&= S1 + S2 + S3,
\end{aligned}
$$

where

$$
\begin{aligned}
\|S1\| &\le \left\|\hat{F}'\right\|\left\|F^0 - \hat{F}H^{-1}\right\|\frac{\|\Lambda_G^0\|}{\sqrt{G}}\\
&= O_p\left(\left\|\hat{\beta} - \beta^0\right\|\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{G}{n}\right),\\
\|S2\| &\le \frac{\left\|\hat{F}'\right\|}{\sqrt{G}}\left\|\bar{\varepsilon}_1, \cdots, \bar{\varepsilon}_G\right\|\\
&= O\left(\sqrt{\frac{G}{n}}\right)\left\|\hat{F}'\right\|\sqrt{\frac{1}{G}\sum_{g=1}^{G} n_g\bar{\varepsilon}_g'\bar{\varepsilon}_g}\\
&= O_p\left(\sqrt{\frac{G}{n}}\right),\\
\|S3\| &\le \frac{\left\|\hat{F}'\right\|}{\sqrt{G}}\left\|\left[\bar{X}_1, \cdots, \bar{X}_G\right]\right\|\left\|\beta^0 - \hat{\beta}\right\|\\
&= O_p\left(\left\|\beta^0 - \hat{\beta}\right\|\right)
\end{aligned}
$$

Therefore,

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

■

Proof of Theorem 3. Part (i) Due to Proposition A2 and Theorem 2, we need to prove

$$
\begin{aligned}
&\text{(a) } \hat{B}_n^{XX} - B_n^{XX}\left(\hat{F}\right) = o_p(1),\\
&\text{(b) } \hat{V}_n^c - V_n^c = o_p(1).
\end{aligned}
$$

For (a), it is straightforward to show that

$$
\begin{aligned}
&\hat{B}_n^{XX} - B_n^{XX}\left(\hat{F}\right) \\
&= \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\left(a_{g\tilde{g}}^0 - \hat{a}_{g\tilde{g}}\right) n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \\
&= \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\left[\lambda_g^{0\prime}\left(H^{-1}\right)'\left(\frac{H^{-1}\Lambda^{0\prime}\Lambda^0\left(H^{-1}\right)'}{n}\right)^{-1}\left(H^{-1}\lambda_{\tilde{g}}^0 - \hat{\lambda}_{\tilde{g}}\right)\right] n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \\
&+ \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\left[\lambda_g^{0\prime}\left(H^{-1}\right)'\left(\left(\frac{H^{-1}\Lambda^{0\prime}\Lambda^0\left(H^{-1}\right)'}{n}\right)^{-1} - \left(\frac{\hat{\Lambda}'\hat{\Lambda}}{n}\right)^{-1}\right)\hat{\lambda}_{\tilde{g}}\right] n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \\
&+ \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\left[\left(\lambda_g^{0\prime}\left(H^{-1}\right)' - \hat{\lambda}_g'\right)\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{n}\right)^{-1}\hat{\lambda}_{\tilde{g}}\right] n_g n_{\tilde{g}} \bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}} \\
&= o_p(1),
\end{aligned}
$$

using $n^{-2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} n_g n_{\tilde{g}} \left\|\bar{X}_g' M_{\hat{F}} \bar{X}_{\tilde{g}}\right\|^2 = O_p(1)$ and Lemma S4.

For (b), let

$$
\tilde{V}_n^c = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\mathcal{X}_{ig}^X\right)'\varepsilon_{ig}\varepsilon_{jg}'\mathcal{X}_{jg}^X \text{ and } \ddot{V}_n^c = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_{ig}^X\right)'\varepsilon_{ig}\varepsilon_{jg}'\hat{\mathcal{X}}_{jg}^X.
$$

We have

$$
\hat{V}_n^c - V_n^c = \left(\hat{V}_n^c - \ddot{V}_n^c\right) + \left(\ddot{V}_n^c - \tilde{V}_n^c\right) + \left(\tilde{V}_n^c - V_n^c\right) \tag{S.30}
$$

and need to show that each term in (S.30.) is $o_p(1)$.

We note that

$$
\begin{aligned}
\hat{V}_n^c - \ddot{V}_n^c = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g}\left(\hat{\mathcal{X}}_{ig}^X\right)'&\left[X_{ig}\left(\beta^0 - \hat{\beta}\right)\left(\beta^0 - \hat{\beta}\right)' X_{jg}'\right. \\
&+ \left(F^0\lambda_g^0 - \hat{F}\hat{\lambda}_g\right)\left(\beta^0 - \hat{\beta}\right)' X_{jg}' + \varepsilon_{ig}\left(\beta^0 - \hat{\beta}\right)' X_{jg}' \\
&+ X_{ig}\left(\beta^0 - \hat{\beta}\right)\left(F^0\lambda_g^0 - \hat{F}\hat{\lambda}_g\right)' + \left(F^0\lambda_g^0 - \hat{F}\hat{\lambda}_g\right)\left(F^0\lambda_g^0 - \hat{F}\hat{\lambda}_g\right)' \\
&\left. + \varepsilon_{ig}\left(F^0\lambda_g^0 - \hat{F}\hat{\lambda}_g\right)' + X_{ig}\left(\beta^0 - \hat{\beta}\right)\varepsilon_{jg} + \left(F^0\lambda_g^0 - \hat{F}\hat{\lambda}_g\right)\varepsilon_{jg}\right]\hat{\mathcal{X}}_{jg}^X,
\end{aligned} \tag{S.31}
$$

and

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{j \in \mathcal{A}_g} \left(\hat{\mathcal{X}}_{ig}^{X} \right)' X_{ig} \left(\beta^0 - \hat{\beta} \right) \left(\beta^0 - \hat{\beta} \right)' X_{jg}' \hat{\mathcal{X}}_{jg}^{X} \right\| \\
&\leq \frac{1}{G} \sum_{g=1}^{G} \left\| \frac{1}{n_g} \sum_{i \in \mathcal{A}_g} \left(\hat{\mathcal{X}}_{ig}^{X} \right)' X_{ig} \right\|^2 O_P \left(\left\| \sqrt{\frac{n}{G}} \left(\beta^0 - \hat{\beta} \right) \right\|^2 \right) \\
&= o_p (1) .
\end{aligned}$$

Using similar arguments, we can show that the other terms in (S.31) are $o_p(1)$.

For the second term in (S.30),

$$\begin{aligned}
\ddot{V}_n^c - \tilde{V}_n^c &= \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{j \in \mathcal{A}_g} \left(\hat{\mathcal{X}}_{ig}^{X} - \mathcal{X}_{ig}^{X} \right)' \varepsilon_{ig} \varepsilon_{jg}' \left(\hat{\mathcal{X}}_{jg}^{X} - \mathcal{X}_{jg}^{X} \right) \\
&+ \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{j \in \mathcal{A}_g} \left(\hat{\mathcal{X}}_{ig}^{X} - \mathcal{X}_{ig}^{X} \right)' \varepsilon_{ig} \varepsilon_{jg}' \mathcal{X}_{jg}^{X} \\
&+ \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{j \in \mathcal{A}_g} \left(\mathcal{X}_{ig}^{X} \right)' \varepsilon_{ig} \varepsilon_{jg}' \left(\hat{\mathcal{X}}_{jg}^{X} - \mathcal{X}_{jg}^{X} \right) \\
&= R1 + R2 + R3 \qquad \text{(S.32)}
\end{aligned}$$

First, we have

$$\begin{aligned}
R1 &= \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{j \in \mathcal{A}_g} \bar{X}_g' \left(P_{F^0} - P_{\hat{F}} \right) \varepsilon_{ig} \varepsilon_{jg}' \left(P_{F^0} - P_{\hat{F}} \right) \bar{X}_g \\
&+ \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{j \in \mathcal{A}_g} \frac{1}{n} \sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \bar{X}_{\tilde{g}}' \left(a_{g\tilde{g}}^0 M_{F^0} - \hat{a}_{g\tilde{g}} M_{\hat{F}} \right) \varepsilon_{ig} \varepsilon_{jg}' \left(P_{F^0} - P_{\hat{F}} \right) \bar{X}_g \\
&+ \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{j \in \mathcal{A}_g} \bar{X}_g' \left(P_{F^0} - P_{\hat{F}} \right) \varepsilon_{ig} \varepsilon_{jg}' \frac{1}{n} \sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \left(a_{g\tilde{g}}^0 M_{F^0} - \hat{a}_{g\tilde{g}} M_{\hat{F}} \right) \bar{X}_{\tilde{g}} \\
&+ \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{A}_g} \sum_{j \in \mathcal{A}_g} \left[\frac{1}{n} \sum_{\tilde{g}_1=1}^{G} n_{\tilde{g}_1} \bar{X}_{\tilde{g}_1}' \left(a_{g\tilde{g}_1}^0 M_{F^0} - \hat{a}_{g\tilde{g}_1} M_{\hat{F}} \right) \right] \varepsilon_{ig} \varepsilon_{jg}' \left[\frac{1}{n} \sum_{\tilde{g}_2=1}^{G} n_{\tilde{g}_2} \left(a_{g\tilde{g}_2}^0 M_{F^0} - \hat{a}_{g\tilde{g}_2} M_{\hat{F}} \right) \bar{X}_{\tilde{g}_2} \right] \\
&= R11 + R12 + R13 + R14.
\end{aligned}$$

For $R11$,

$$\begin{aligned}
\| R11 \| &\leq \left(\frac{1}{n} \sum_{g=1}^{G} n_g \left\| \bar{X}_g' \right\|^2 \left\| \sqrt{n_g} \bar{\varepsilon}_g' \right\|^2 \right) \left\| P_{F^0} - P_{\hat{F}} \right\|^2 \\
&= o_p (1) .
\end{aligned}$$

For $R12,$

$$R12 = \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G} n_g^2 n_{\tilde{g}} \bar{X}_{\tilde{g}}' \left(a_{g\tilde{g}}^0 M_{F^0} - \hat{a}_{g\tilde{g}} M_{\hat{F}}\right) \bar{\varepsilon}_g \bar{\varepsilon}_g' \left(P_{F^0} - P_{\hat{F}}\right) \bar{X}_g$$
$$= \frac{1}{n}\sum_{g=1}^{G} n_g \left(\frac{1}{n}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}} a_{g\tilde{g}}^0 \bar{X}_{\tilde{g}}'\right) \left(P_{\hat{F}} - P_{F^0}\right) \left(\sqrt{n_g}\bar{\varepsilon}_g\right) \left(\sqrt{n_g}\bar{\varepsilon}_g'\right) \left(P_{F^0} - P_{\hat{F}}\right) \bar{X}_g$$
$$+ \frac{1}{n}\sum_{g=1}^{G} n_g \left(\frac{1}{n}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \left(a_{g\tilde{g}}^0 - \hat{a}_{g\tilde{g}}\right) \bar{X}_{\tilde{g}}'\right) M_{\hat{F}} \left(\sqrt{n_g}\bar{\varepsilon}_g\right) \left(\sqrt{n_g}\bar{\varepsilon}_g'\right) \left(P_{F^0} - P_{\hat{F}}\right) \bar{X}_g$$
$$= \mathcal{L}1 + \mathcal{L}2$$

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

$$\|\mathcal{L}2\| \leq O(1)\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\frac{1}{n}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \left(a_{g\tilde{g}}^0 - \hat{a}_{g\tilde{g}}\right) \bar{X}_{\tilde{g}}'\right\| \left\|\sqrt{n_g}\bar{\varepsilon}_g\right\|^2 \left\|\bar{X}_g\right\| \left\|P_{F^0} - P_{\hat{F}}\right\|$$
$$\leq O(1)\left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\frac{1}{n}\sum_{\tilde{g}=1}^{G} n_{\tilde{g}} \left(a_{g\tilde{g}}^0 - \hat{a}_{g\tilde{g}}\right) \bar{X}_{\tilde{g}}'\right\|^2\right)^{1/2}$$
$$\times \left(\frac{1}{n}\sum_{g=1}^{G} n_g \left\|\sqrt{n_g}\bar{\varepsilon}_g\right\|^4 \left\|\bar{X}_g\right\|^2\right)^{1/2} \left\|P_{F^0} - P_{\hat{F}}\right\|$$
$$= o_p(1).$$

Thus, $R12 = o_p(1)$. Using similar procedures, we can prove that $R13 = o_p(1)$ and $R14 = o_p(1)$. Thus, $R1 = o_p(1)$. The proofs for $R2 = o_p(1)$ and $R3 = o_p(1)$ are omitted because it is the same as the one for $R1 = o_p(1)$.

For the last term in (S.30), since the convergence is elementwise, let's assume that $\tilde{V}_n^c - V_n^c$ is a scalar. Then, under independence among groups in Assumption 9(i), we have

$$P\left(\left|\tilde{V}_n^c - V_n^c\right| > \Delta\right)$$
$$\leq \frac{1}{\Delta^2}\frac{1}{n^2}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\sum_{j\in\mathcal{A}_g} E\left(\left(\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\mathcal{X}_j^X - E\left[\left(\mathcal{X}_i^X\right)'\varepsilon_i\varepsilon_j'\mathcal{X}_j^X\right]\right)^2$$
$$= o(1).$$

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

$$\hat{\sigma}^2 - \sigma^2 \to^p 0. \tag{S.33}$$

We have

$$\hat{\sigma}^2 = \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\Big[\left(\beta^0-\hat{\beta}\right)' X_{ig}'X_{ig}\left(\beta^0-\hat{\beta}\right) + 2\left(\beta^0-\hat{\beta}\right)' X_{ig}'\left(F^0\lambda_g^0-\hat{F}\hat{\lambda}_g\right) + 2\left(\beta^0-\hat{\beta}\right)' X_{ig}'\varepsilon_{ig}$$
$$+\left(F^0\lambda_g^0-\hat{F}\hat{\lambda}_g\right)'\left(F^0\lambda_g^0-\hat{F}\hat{\lambda}_g\right) + 2\left(F^0\lambda_g^0-\hat{F}\hat{\lambda}_g\right)'\varepsilon_{ig} + \varepsilon_{ig}'\varepsilon_{ig}\Big]$$
$$= T1 + ... + T6.$$

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

For $T6$, we have

$$T6 = \frac{1}{nT}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\varepsilon_{ig}'\varepsilon_{ig} \to^p = \frac{1}{T}E\varepsilon_{ig}'\varepsilon_{ig} = \frac{1}{T}\sum_{t=1}^{T}E\varepsilon_{igt}^2 = \sigma^2,$$

which completes the proof of (S.33). ■

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

Let

$$\tilde{C}_{0a,n}^{XX} = \frac{1}{n}\sum_{g=1}^{G}\sum_{\ell=1}^{\kappa_g}\sum_{i\in\mathcal{A}_g^{(\ell)}}\mathcal{X}_{i\ell g}^{X}\mathcal{X}_{a,i\ell g}^{X}.$$

We have

$$\hat{C}_{0a,n}^{XX} - C_{0a,n}^{XX} = \left(\hat{C}_{0a,n}^{XX} - \tilde{C}_{0a,n}^{XX}\right) + \left(\tilde{C}_{0a,n}^{XX} - C_{0a,n}^{XX}\right). \tag{S.34}$$

The second term is $o_p(1)$ by the law of large numbers. For the first term, under the null, we have

$$
\begin{aligned}
&\hat{C}^{XX}_{0a,n} - \tilde{C}^{XX}_{0a,n} \\
&= \frac{1}{n}\sum_{g=1}^{G}\sum_{\ell=1}^{\kappa_g}\sum_{i\in\mathcal{A}_g^{(\ell)}}\left[\left(X_{i\ell g} - P_{\hat{F}}\bar{X}_g\right)'\left(X_{i\ell g} - P_{\hat{F}_a}\bar{X}_{\ell g}\right) - \left(X_{i\ell g} - P_{F^0}\bar{X}_g\right)'\left(X_{i\ell g} - P_{F^0}\bar{X}_{\ell g}\right)\right] \\
&+ \frac{1}{n}\sum_{g=1}^{G}\sum_{\ell=1}^{\kappa_g} n_{\ell g}\left[\left(\frac{1}{n}\sum_{\tilde{g}_1=1}^{G} n_{\tilde{g}_1}\hat{a}_{g\tilde{g}_1}M_{\hat{F}}\bar{X}_{\tilde{g}_1}\right)'\left(\frac{1}{n}\sum_{\tilde{g}_2=1}^{G}\sum_{\tilde{\ell}_2=1}^{\kappa_{\tilde{g}_2}} n_{\tilde{\ell}_2\tilde{g}_2}\hat{a}_{\ell g,\tilde{\ell}_2\tilde{g}_2}M_{\hat{F}_a}\bar{X}_{\tilde{\ell}_2\tilde{g}_2}\right)\right. \\
&\left. - \left(\frac{1}{n}\sum_{\tilde{g}_1=1}^{G} n_{\tilde{g}_2}a^0_{g\tilde{g}_2}M_{F^0}\bar{X}_{\tilde{g}_2}\right)'\left(\frac{1}{n}\sum_{\tilde{g}_2=1}^{G} n_{\tilde{g}_2}a^0_{g\tilde{g}_2}M_{F^0}\bar{X}_{\tilde{g}_2}\right)\right] \\
&- \frac{1}{n}\sum_{g=1}^{G}\sum_{\ell=1}^{\kappa_g}\sum_{i\in\mathcal{A}_g^{(\ell)}}\left[\left(X_{i\ell g} - P_{\hat{F}}\bar{X}_g\right)'\left(\frac{1}{n}\sum_{\tilde{g}_2=1}^{G}\sum_{\tilde{\ell}_2=1}^{\kappa_{\tilde{g}_2}} n_{\tilde{\ell}_2\tilde{g}_2}\hat{a}_{\ell g,\tilde{\ell}_2\tilde{g}_2}M_{\hat{F}_a}\bar{X}_{\tilde{\ell}_2\tilde{g}_2}\right)\right. \\
&\left. - \left(X_{i\ell g} - P_{F^0}\bar{X}_g\right)'\left(\frac{1}{n}\sum_{\tilde{g}_2=1}^{G} n_{\tilde{g}_2}a^0_{g\tilde{g}_2}M_{F^0}\bar{X}_{\tilde{g}_2}\right)\right] \\
&- \frac{1}{n}\sum_{g=1}^{G}\sum_{\ell=1}^{\kappa_g}\sum_{i\in\mathcal{A}_g^{(\ell)}}\left[\left(\frac{1}{n}\sum_{\tilde{g}_1=1}^{G} n_{\tilde{g}_1}\hat{a}_{g\tilde{g}_1}M_{\hat{F}}\bar{X}_{\tilde{g}_1}\right)'\left(X_{i\ell g} - P_{\hat{F}_a}\bar{X}_{\ell g}\right)\right. \\
&\left. - \left(\frac{1}{n}\sum_{\tilde{g}_1=1}^{G} n_{\tilde{g}_2}a^0_{g\tilde{g}_2}M_{F^0}\bar{X}_{\tilde{g}_2}\right)'\left(X_{i\ell g} - P_{F^0}\bar{X}_{\ell g}\right)\right] \\
&= U1 + U2 + U3 + U4.
\end{aligned}
$$

For $U1$, we can show that

$$
\begin{aligned}
U1 &= \frac{1}{n}\sum_{g=1}^{G}\sum_{\ell=1}^{\kappa_g} n_g^{(\ell)}\bar{X}'_{\ell g}\left(P_{F^0} - P_{\hat{F}_a}\right)\bar{X}_{\ell g} \\
&+ \frac{1}{n}\sum_{g=1}^{G} n_g\bar{X}'_g\left(P_{F^0} - P_{\hat{F}}\right)\bar{X}_g \\
&- \frac{1}{n}\sum_{g=1}^{G} n_g\bar{X}'_g\left(P_{F^0} - P_{\hat{F}}P_{\hat{F}_a}\right)\bar{X}_g \\
&= o_p(1)
\end{aligned}
$$

under the null as $(G_a, n) \to \infty$ such that $G_a/\sqrt{n} \to 0$ based on Lemma S2. Using similar arguments, it is easy to prove $U2$, $U3$ and $U4$ are also $o_p(1)$. ■

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

$$\begin{aligned}\tilde{\mathcal{Q}}_{gmm}(\beta, F) &= \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[M_F F^0\lambda_g^0 + (\varepsilon_{ig} - P_F\bar{\varepsilon}_g) - \left(X_{ig} - P_F\bar{X}_g\right)\beta\right]'\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)\Omega_n^{-1} \\ &\quad\times\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)'\left[M_F F^0\lambda_g^0 + (\varepsilon_{ig} - P_F\bar{\varepsilon}_g) - \left(X_{ig} - P_F\bar{X}_g\right)\beta\right] \\ &\quad-\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\varepsilon_{ig} - P_{F^0}\bar{\varepsilon}_g\right)'\left(\Psi_{ig} - P_{F^0}\bar{\Psi}_g\right)\Omega_n^{-1}\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig} - P_{F^0}\bar{\Psi}_g\right)'\left(\varepsilon_{ig} - P_{F^0}\bar{\varepsilon}_g\right),\end{aligned}$$

$$\begin{aligned}\mathcal{Q}^*_{gmm}(\beta, F) &= \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\lambda_g^{0\prime}F^{0\prime}M_F'\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)\Omega_n^{-1}\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)'M_F F^0\lambda_g^0 \\ &\quad-\frac{2}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\lambda_g^{0\prime}F^{0\prime}M_F'\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)\Omega_n^{-1}\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)'\left(X_{ig} - P_F\bar{X}_g\right)\beta \\ &\quad+\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\beta'\left(X_{ig} - P_F\bar{X}_g\right)'\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)\Omega_n^{-1}\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)'\left(X_{ig} - P_F\bar{X}_g\right)\beta,\end{aligned}$$

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

and

$$\begin{aligned}\mathcal{Q}^*_{LS}(F) &= \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\lambda_g'F^{0\prime}M_F F^0\lambda_g - \frac{2}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\lambda_g'F^{0\prime}M_F\left(X_{ig} - P_F\bar{X}_g\right)\mathbb{C}_{nT}(F) \\ &\quad+\mathbb{C}_{nT}(F)'\left(\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(X_{ig} - P_F\bar{X}_g\right)'\left(X_{ig} - P_F\bar{X}_g\right)\right)\mathbb{C}_{nT}(F),\end{aligned}$$

where

$$\begin{aligned}\mathbb{C}_n(F) &= \mathbb{D}_n(F)\left[\frac{1}{n}\sum_{g=1}^{G}\sum_{j\in\mathcal{A}_g}\left(\Psi_{jg} - P_F\bar{\Psi}_g\right)M_F F^0\lambda_g\right], \\ \mathbb{D}_n(F) &= \left[Q_n^{X\Psi}(F)\Omega_n^{-1}Q_n^{\Psi X}(F)\right]^{-1}Q_n^{X\Psi}(F)\Omega_n^{-1}.\end{aligned}$$

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

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

$$\mathcal{Q}^*_{gmm}(\beta, F) = \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left[\lambda_g^{0\prime}F^{0\prime}M_F' - \beta'\left(X_{ig} - P_F\bar{X}_g\right)'\right]\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)\Omega_n^{-1}$$
$$\times\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)'\left[M_F F^0\lambda_g^0 - \left(X_{ig} - P_F\bar{X}_g\right)\beta\right]$$
$$> 0$$

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

$$\begin{aligned}
&\tilde{\mathcal{Q}}_{gmm}(\beta, F) - \mathcal{Q}^*_{gmm}(\beta, F)\\
&= \frac{2}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\lambda_g^{0\prime}F^{0\prime}M_F\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)\Omega_n^{-1}\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)'\left(\varepsilon_{ig} - P_F\bar{\varepsilon}_g\right)\\
&+ \frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\varepsilon_{ig} - P_F\bar{\varepsilon}_g\right)'\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)\Omega_n^{-1}\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)'\left(\varepsilon_{ig} - P_F\bar{\varepsilon}_g\right)\\
&- \frac{2}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\varepsilon_{ig} - P_F\bar{\varepsilon}_g\right)'\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)\Omega_n^{-1}\frac{1}{n}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig} - P_F\bar{\Psi}_g\right)'\left(X_{ig} - P_F\bar{X}_g\right)\beta\\
&= o_p(1).
\end{aligned}$$

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

Proof of Theorem 6. Note that

$$\begin{aligned}
&\left[Q_n^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_n^{-1}Q_n^{\Psi X}\left(\hat{F}_{gmm}\right)\right]\sqrt{n}\left(\hat{\beta}_{gmm} - \beta^0\right)\\
&= Q_n^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_n^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\Psi_{ig}'M_{\hat{F}_{gmm}}F^0\lambda_g - \left(\Psi_{ig} - P_{F^0}\bar{\Psi}_g\right)'\varepsilon_{ig}\right).
\end{aligned} \tag{S.35}$$

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

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\Psi_{ig}'M_{\hat{F}_{gmm}}F^0\lambda_g\\
&= \frac{1}{n^2}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}n_g n_{\tilde{g}}a^0_{g\tilde{g}}\bar{\Psi}_g'M_{\hat{F}_{gmm}}\bar{X}_{\tilde{g}}\sqrt{nT}\left(\hat{\beta}_{gmm} - \beta\right) - \frac{1}{n\sqrt{n}}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}n_g n_{\tilde{g}}a^0_{g\tilde{g}}\bar{\Psi}_g'M_{\hat{F}_{gmm}}\bar{\varepsilon}_{\tilde{g}}\\
&+ o_p\left(\sqrt{n}\left\|\hat{\beta}_{gmm} - \beta\right\|\right) + O_p\left(\frac{G}{\sqrt{n}}\right).
\end{aligned}$$

Applying this result to (S.35), under the rate condition in Assumption 5, we have

$$
\begin{aligned}
& Q_n^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_n^{-1}B_n^{\Psi X}\left(\hat{F}_{gmm}\right)\sqrt{n}\left(\hat{\beta}_{gmm}-\beta\right) \\
&= Q_n^{X\Psi}\left(\hat{F}_{gmm}\right)\Omega_n^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\mathcal{X}_{ig}^{\Psi}\left(\hat{F}_{gmm}\right)'\varepsilon_{ig}+o_p(1).
\end{aligned}
$$

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

$$
\sqrt{n}\left(\hat{\beta}_{gmm}-\beta\right)=\left(Q_{X\Psi}\Omega_n^{-1}B_{\Psi X}\right)^{-1}Q_{X\Psi}\Omega_n^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i\in\mathcal{A}_g}\left(\mathcal{X}_{ig}^{\Psi}\right)'\varepsilon_{ig}+o_p(1).
$$

Applying Assumptions 11(ii) and 12 to this result completes the proof. ■

References

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

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

[3] Stock, J. H., and M. W. Watson (1998). Diffusion Indexes, Working Paper 6702, NBER.