Estimating Panel Data Models With Common Factors: A Mundlak Projection Approach*

Qu Feng † Wei Wang‡ Wenjie Wang§ Mengying Yuan¶

August 6, 2024

Abstract

Panel data models with factor structures or interactive fixed effects have been researched for decades. In this paper, we proposed the two-way Mundlak-type projection estimator for such models, allowing for linear or interactive relationships between the factor structures and the regressors. We further consider the endogenous case, by combining the Mundlak-type estimation with the control function approach to simultaneously control all the correlations between the regressors and the composite errors. All of our estimation procedures do not require iteration. For inference, we apply the dependent wild bootstrap to obtain consistent covariance estimators and bootstrapped test statistics. In addition, we extend the robust inference method under the fixed-b asymptotics to the interactive panel data models. The asymptotic distributions of the proposed estimators are derived as long as one dimension of panels tends to infinity, and Monte Carlo simulations are conducted to verify the theoretical results in finite samples. Finally, we apply our proposed methods to the estimation of the aggregate production function.

Keywords: Panel Data, Interactive Fixed Effects, Mundlak projection, Dependent Wild bootstrap, Endogeneity.

JEL Classification: C23, C33.

*An earlier version of this paper was presented at the Shanghai Jiaotong University(2023), CUHK-Shenzhen(2023), CUHK(2023), the 2023 AMSE, East and Southeast Asia in Singapore. Constructive comments from the audiences at these conferences are highly acknowledged.

†Economics Division, School of Social Sciences, Nanyang Technological University. SHHK-04-84, 48 Nanyang Avenue, Singapore 639818. Email Address: qfeng@ntu.edu.sg.

‡School of Statistics and Mathematics, Shandong University of Finance and Economics, No.7366, East Erhuan Road, 250014, Jinan, Shandong, China. Email Address: wangwei_0115@outlook.com.

§Division of Economics, School of Social Sciences, Nanyang Technological University. HSS-04-65, 14 Nanyang Drive, Singapore 637332. E-mail address: wang.wj@ntu.edu.sg.

¶School of Economics and Finance, South China University of Technology, mengying001@e.ntu.edu.sg.

1 Introduction

Panel data models have become popular in economics and other fields for long periods. An important superiority of panel data models is that the parameters of interest can be consistently estimated while controlling unobservable cross-sectional or time heterogeneity. For decades, many empirical researchers have used factor structures to capture the unobservable heterogeneity in panel data, which are called panel models with the interactive effects or the factor augmented panels in the literature. In particular, the unobserved time-varying factors in such models have heterogeneous impact across individuals, while causing cross-sectional dependence in the composite errors.

Since both individual specific effects and time-varying effects in the factor structures are unobservable, they can be treated as either fixed parameters or random variables. Hsiao (2018) summarized four formulations of the models and distinguishes differences in various assumptions:

(i) When both individual effects and time effects are treated as random (Sarafidis and Robertson, 2009), it is reasonable to treat them as a component of errors. The correlations between regressors and factor structures lead to inconsistent estimator by ordinary least squared while the instrumental variable estimation works.

(ii) If both are treated as fixed parameters, the common factors can be treated as additional regressors (Pesaran, 2006; Bai, 2009; Moon and Weidner, 2015). Thus, there is no need to discuss the specific correlation between regressors and factor structure.

(iii) In the case of random individual effects with fixed time effects (Sarafidis and Wansbeek, 2012; Bai 2013; Robertson and Sarafidis, 2015), which corresponds to the case with a large number of individuals and a fixed time span, the correlations between the regressors and common factors can be controlled by treating the common factors as fixed constants, and the quasi maximum likelihood estimations work if the random individual effects is not correlated with the regressors.

(iv) In the case of random time effects with fixed individual effects (Hsiao, 2018), it is reasonable to treat the factor loadings as fixed and assume that the regressors is uncorrelated with the common factors, controlling the correlations between the regressor and factor structures.

The above arguments show that the key point of estimation lies in how to deal with the factor structures when correlation between regressors and factor structure exists. Estimation methods can be classified accordingly. The most direct method is to estimate the interactive effects under case (ii), where the factor structure is regarded as fixed parameters. Bai (2009) proposed the iterative fixed effects (IFE) estimator under large panels and show the theoretical guarantee. Jiang et al. (2021) show convergence issues of the recursive estimation procedure of the IFE. In addition, Moon and Weidner (2015) show that the estimator of Bai (2009) can be interpreted as a quasi maximum likelihood estimator (QMLE), the consistency of which is maintained even when the number of factors is not specified correctly, as long as it is larger than or equal to the true number of factors. Moon and Weidner (2017) proposed a bias-corrected QMLE estimator for dynamic panel data models with homogeneous slopes, while Moon and Weidner (2019) used a nuclear norm regularization to obtain computational advantage. Furthermore, Bai and Li (2014) proposed maximum likelihood estimation for the models.

Another branch of literature proposed estimation methods that eliminate the interactive effects directly. Holtz-Eakin et al. (1988) suggested eliminating the unobserved factor component using the quasi-differencing transformation. Ahn, Lee and Schmidt (2013) proposed the generalized method of moments (GMM) method. The GMM method is based on a nonlinear transformation known as quasi-differencing that eliminates the interactive effects. Estimating the common factors and then removing the interactive effects is a compromise. Pesaran (2006) proposed the common correlated estimator (CCE), which only allows the correlation between the common factors f_t and regressors. The CCE estimator uses the cross-sectional averages of both regressors and dependent variables as proxies for the unobservable common factors. In addition, Hsiao, Shi and Zhou (2021) proposed to find the null space of the factors or the loadings and constructed the transformed estimators (TE) to get rid of the interactive effects. Morkutė,Sarafidis, Yamagata and Cui (2021) projected out the common factors from the exogenous covariates of the model under case (i). Juodis and Sarafidis (2022) considered the case where regressors are allowed to be correlated with the factors and its loadings under the fixed T setup and proposed two methods to construct the factor proxies by observed variables.

Last strategy aims to control the correlation between the regressors and interactive effects. Then, IV or GMM-type estimation methods are applied. Ahn, Lee and Schmidt (2001) proposed a GMM estimator to remove the correlation under case (iii). Sarafidis and Robertson (2009) and Robertson and Sarafidis (2015) proposed IV or GMM-based method by regarding the loadings and factors as random variables and assuming there exist instruments that are both correlated with the regressors and uncorrelated with the composite error terms. Bai (2024) proposed a MLE of with the

Mundlak or Chamberlain-type projection in order to control the correlation between the regressors and loadings. Similarly, Hayakawa (2013) provided a GMM estimator based on the Mundlak-Chamberlain type projection. Juodis and Sarafidis (2018) gave a summary of the existing literature on the dynamic panel data estimators with multi-factor errors and proposed a more general projection specification form.

The above approaches can be adapted to various specifications. Westerlund (2019), Westerlund and Urbain (2015) compare the properties of CCE estimator, and IFE estimator. Under certain assumptions, the relative properties of these two approaches are different and no general conclusion is drawn regarding which one is dominant. Similar study regarding multi-factor error structure includes Phillips and Sul (2003), who proposed a seemingly unrelated median unbiased estimator to estimate autoregressive model with cross-sectional dependence and heterogeneous coefficients. Kneip, Sickles, and Song (2012) estimated the unobservable common effects by smooth spline based on the assumption that the unobservable common effects are a smooth function of time.

In this paper, we aim to extend the one-way Mundlak projection approach for the panel data models with the interactive effects as noted or used in Bai (2009) and Bai (2024), which only allowed for the linear correlation between the loadings and regressors. In addition, Keilbara et al. (2023) projected the loadings onto the regressors by a non-parametric form. Different from these papers, we regard both individuals effects and time effects as random variables and allow the linear or interactive correlations between the regressors and the factor structures by proposing a two-way Mundlak projection method for the panel model with interactive effects. Our paper is also different from the two-way Mundlak projection estimator in Wooldridge (2021), which focused on the two-way fixed effects panel data model and estimated the interaction of time fixed effects and individual fixed effect by using directly the interactions of the cross-sectional and time averages of the regressors. While both factor loadings and common factors are correlated with the regressors in a non-parametric way, they can be transformed into a interactive form (Freeman and Weidner, 2023). Thus, our two-way Mundlak projection method is more general. Secondly, our method does not require the iteration step to estimate the factors or loadings, which may cause the asymptotic bias or require additional rank conditions. In the case of linear correlation between the regressors and factor structures, our approach can be regarded as an extension of the CCE approach. For case of interactive correlation, it can be regarded as a combination of the CCE and IFE approaches. Thirdly, our estimator, named the Mundlak least squared estimator (MLS), is consistent and asymptotic normal under N and/or T tending to infinity.

Last, we consider statistical inference for the proposed estimator under our framework, in which a dependent wild bootstrap procedure (Gao, Peng and Yan, 2023) and a fixed-b type robust inference procedure (Vogelsang, 2012) are proposed.

Furthermore, in the paper we consider another source of endogeneity, originating from the correlation between the regressors and the errors, which has received much attention recently. For example, Robertson and Sarafidis (2015) proposed a new instrumental variables approach for consistent and asymptotically efficient estimation of panel data models with weakly exogenous or endogenous regressors under a multi-factor error structure. Hong, Su and Jiang (2023) proposed a profile GMM estimation method for panel data models with interactive fixed effects. Juodis and Sarafidis (2022) put forward a novel method-of-moments approach for estimation of factor-augmented panel data models with endogenous regressors and fixed T. Morkutė et al. (2021) projected out the common factors from the exogenous covariates of the model, and constructed instruments based on defactored covariates. Hsiao, Zhou, Kong (2023) extended the transformed estimation approach to the endogenous case.

The rest of the paper is organized as follows. Section 2 introduces the models, the Mundlak-type projection estimators in the case with linear correlation, and the asymptotic distribution of the estimators. In Section 3, we study the Munlak-type projection in the general case with a interactive correlation between the regressors and the factor structures. In Section 4, we further extend the Mundlak-type projection estimator to the endogenous case. For inference, we propose the dependent wild bootstrap procedure and the robust inference procedure in Sections 5 and 6, respectively. Section 7 gives the design of Monte Carlo simulations, while the empirical application is shown in Section 8. We conclude this study in Section 9. The simulation results are shown in Appendix A and mathematical proofs are provided in the Appendix B. Additional simulation results are shown in Appendix C (not for publications).

2 Model and the Mundlak Estimators

2.1 Background and Motivation

The dependent variable y_{it}, observed on i-th individual at time t, for $i = \{1, 2, ..., N\}$ and $t = \{1, 2, ..., T\}$, is generated by

$$y_{it} = \beta_0 + x_{it}'\beta + e_{it}, \tag{1}$$

where β_0 is an intercept, $x_{it} = (x_{it,1}, \cdots, x_{it,p})'$ is a $p \times 1$ vector of explanatory variables, with a $p \times 1$ vector of slopes β. The errors e_{it} are cross-sectional correlated with a multi-factor structure, i.e.,

$$e_{it} = \lambda_i' f_t + \varepsilon_{it}, \tag{2}$$

where f_t is a $r \times 1$ vector of unobservable common factors with $r \times 1$ vector of factor loadings λ_i, and ε_{it} is the idiosyncratic error. We assume that the dimension r is fixed.

Although the common factors f_t and factor loadings λ_i are strictly exogenous with respect to the idiosyncratic errors ε_{it} under standard assumptions, the regressors x_{it} could be correlated with e_{it}, due to the correlation between x_{it} and f_t, λ_i, ε_{it}, resulting in the inconsistency of the OLS estimator $\hat{\beta}_{OLS}$:

$$\hat{\beta}_{OLS} - \beta = (\textstyle\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} x_{it}')^{-1} (\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \lambda_i' f_t + \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \varepsilon_{it}).$$

Specifically, there may exist two sources of endogeneity that result in the correlations between the regressors x_{it} and the factor structure $\lambda_i' f_t$ in the errors. In this paper, we propose Mundlak projection approaches and control function approaches to deal with those correlations.

For time t, let $\overline{x}_{\cdot t} = (1, \frac{1}{N} \sum_{i=1}^{N} x_{it,1}, \cdots, \frac{1}{N} \sum_{i=1}^{N} x_{it,p})'$ denote the cross-sectional average of the regressors. First, suppose that the factor loadings λ_i are uncorrelated with x_{it}. Then, to control for the correlation between x_{it} and f_t, we may simply project the common factor f_t onto the space of the cross-sectional average $\overline{x}_{\cdot t}$, following Mundlak's approach (Mundlak, 1978), i.e., we let

$$\underset{r \times 1}{f_t} = \underset{r \times (p+1)}{B} \underset{(p+1) \times 1}{\overline{x}_{\cdot t}} + \xi_t, \tag{3}$$

where B is the coefficient matrix and ξ_t is the random projection error. In order to display the spirit of our Mundlak projection approach, equation (3) restricts the linear correlation between f_t and $\overline{x}_{\cdot t}$.

Remark 1: We note that in Pesaran (2006)'s model with common correlated effects $x_{it} = \Gamma_i' f_t + v_{it}$, where Γ_i is not correlated with λ_i, the common factor f_t is assumed to be fully determined by the cross-sectional average of regressor $\overline{x}_{\cdot t}$ (i.e., the variance of v_{it} is assumed to converge to zero), as $N \to \infty$. By contrast, in equation (3), the common factor f_t can still be affected by the idiosyncratic error ξ_t even as $N \to \infty$ (i.e., we allow for the case where the variance of ξ_t is fixed). The difference between $f_t = B\overline{x}_{\cdot t} + \xi_t$, common factor f_t has additional information ξ_t. In addition, $x_{it} = \Gamma_i' f_t + v_{it}$,

x_{it} **includes all the information of** f_t. **above equation , we allow for the factor number** r **is 'larger than the number of regressor** p.

Plugging the equation (3) into factor structure (2), the model (1) becomes

$$y_{it} \quad = \quad \beta_0 + x_{it}'\beta + \lambda_i' B\overline{x}_{\cdot t} + u_{it}, \tag{4}$$

where $u_{it} = \lambda_i'\xi_t + \varepsilon_{it}$. Then, under the previous assumption that λ_i is uncorrelated with x_{it}, $\lambda_i'\xi_t$ will be uncorrelated with $\overline{x}_{\cdot t}$ and x_{it}. Therefore, the pooled OLS estimator of β in (4) is consistent under standard regularity condition, by using the argument of partitioned regression.

More specifically, for $j = \{1, \cdots, p\}$, let $\bar{x}_{\cdot t,j} = \frac{1}{N}\sum_{i=1}^{N} x_{it,j}$, and $\bar{X}_{\cdot,j} = (\bar{x}_{\cdot 1,j}, \cdots, \bar{x}_{\cdot T,j})'$. Stacking time t, the regression (4) can be written as the following vector form:

$$Y_i = \iota_T\beta_0 + X_i\beta + \bar{X}B'\lambda_i + u_i, \tag{5}$$

where $\underset{T\times 1}{Y_i} = (y_{i1}, y_{i2}..., y_{iT})'$, $\underset{T\times p}{X_i} = (x_{i1}, x_{i2}..., x_{iT})'$, $\underset{T\times(p+1)}{\bar{X}} = (l_T, \bar{X}_{\cdot,1}, \cdots, \bar{X}_{\cdot,p})$ with $\underset{T\times 1}{l_T} = (1, 1, \cdots, 1)'$, and $\underset{T\times 1}{u_i} = (u_{i1}, u_{i2}, ..., u_{iT})'$. We pre-multiple the equation (5) by $M_{\bar{X}} = I_T - \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'$, and the model becomes

$$M_{\bar{X}}Y_i = M_{\bar{X}}X_i\beta + M_{\bar{X}}u_i.$$

Then, the standard (one-way) Mundlak-type least squared estimator is defined as

$$\widehat{\beta}_{M1} = (\textstyle\sum_{i=1}^{N} X_i' M_{\bar{X}} X_i)^{-1}(\sum_{i=1}^{N} X_i' M_{\bar{X}} Y_i). \tag{6}$$

However, if the factor loadings λ_i is in fact correlated with x_{it}, the above one-way Mundlak estimator $\widehat{\beta}_{M1}$ is no longer consistent. In the following section, we propose a two-way Mundlak projection estimation procedure to deal with such a more general setting.

2.2 The Two-way Mundlak Projection in Linear Form (CLF)

To control for the correlation between x_{it} and λ_i, we further project λ_i onto the space of the time average of x_{it}. Specifically, for each individual i, denote the time average of regressors as $\overline{x}_{i\cdot} = (1, \frac{1}{T}\sum_{t=1}^{T} x_{it,1}, \cdots, \frac{1}{T}\sum_{t=1}^{T} x_{it,p})'$. Then, we have

$$\underset{r\times 1}{\lambda_i} = \underset{r\times(p+1)}{A}\ \underset{(p+1)\times 1}{\overline{x}_{i\cdot}} + \eta_i, \tag{7}$$

where, A is a constant coefficient matrix and η_i is the projection error.

Thus, our panel model now consists of equations (1), (2), (3), and (7). Plugging (3) and (7) into the factor structure (2), we can re-write (1) as

$$\begin{aligned} y_{it} &= \beta_0 + x'_{it}\beta + \overline{x}'_{\cdot t}B'\eta_i + \overline{x}'_{i\cdot}A'\xi_t + \overline{x}'_{i\cdot}A'B\overline{x}_{\cdot t} + \eta'_i\xi_t + \varepsilon_{it} \\ &\equiv \beta_0 + x'_{it}\beta + \overline{x}'_{\cdot t}\rho_i + \overline{x}'_{i\cdot}\delta_t + \overline{x}'_{i\cdot}A'B\overline{x}_{\cdot t} + u_{it}, \end{aligned} \tag{8}$$

where $\underset{(p+1)\times 1}{\rho_i} = B'\eta_i$, $\underset{(p+1)\times 1}{\delta_t} = A'\xi_t$, and $u_{it} = \eta'_i\xi_t + \varepsilon_{it}$.

Similar to the argument of partitioned regression in Section 2.1, we implement the following steps to remove the nuisance parameters and obtain a direct estimation procedure of the parameter of interest β. First, by stacking time t, we can re-write the regression (8) as

$$Y_i = \beta_0 \cdot l_T + X_i\beta + \bar{X}\rho_i + \delta\overline{x}_{i\cdot} + \bar{X}B'A\overline{x}_{i\cdot} + u_i, \tag{9}$$

where $\underset{T\times(p+1)}{\delta} = (\delta_1, \delta_2, ...\delta_T)'$.

Second, we pre-multiply the equation (9) by $M_{\bar{X}} = I_T - \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'$,

$$M_{\bar{X}}Y_i = M_{\bar{X}}X_i\beta + M_{\bar{X}}\delta\overline{x}_{i\cdot} + M_{\bar{X}}u_i.$$

Third, we stack individual i, and rewrite the above model as

$$M_{\bar{X}}Y = \beta_1 \cdot M_{\bar{X}}X^1 + \beta_2 \cdot M_{\bar{X}}X^2 + ... + \beta_p \cdot M_{\bar{X}}X^p + M_{\bar{X}}\delta\underline{X}' + M_{\bar{X}}U,$$

where $\underset{T\times N}{Y} = (Y_1, Y_2, ..., Y_N)$, $\underset{T\times N}{X^j}$ is a $T \times N$ matrix being the j^{th} regressor matrix associated with the parameter β_j, $\underset{N\times(p+1)}{\underline{X}} = (\overline{x}_{1\cdot}, \overline{x}_{2\cdot}, ..., \overline{x}_{N\cdot})'$, $\underset{T\times N}{U} = (u_1, u_2, ..., u_N) = \xi\eta' + \varepsilon$.

Then, we post-multiply the equation by $M_{\underline{X}} = I_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$ and obtain

$$M_{\bar{X}}YM_{\underline{X}} = \beta_1 \cdot M_{\bar{X}}X^1M_{\underline{X}} + ... + \beta_p \cdot M_{\bar{X}}X^pM_{\underline{X}} + M_{\bar{X}}UM_{\underline{X}}. \tag{10}$$

Finally, we collect all the transformed regressors for individual i at period t, $\underset{NT\times p}{\tilde{X}} = [vec(M_{\bar{X}}X^1M_{\underline{X}}), \cdots, vec(M_{\bar{X}}X^pM_{\underline{X}})]$. Similarly, let $\underset{NT\times 1}{\tilde{Y}} = vec(M_{\bar{X}}YM_{\underline{X}})$, and $\underset{NT\times 1}{\tilde{U}} = vec(M_{\bar{X}}UM_{\underline{X}})$. Thus, equation (10) can be further parameterized by

$$\tilde{Y} = \tilde{X}\beta + \tilde{U}, \tag{11}$$

and then our two-way Mundlak least squared estimator is defined as

$$\widehat{\beta}_{M2} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}. \tag{12}$$

Remark 2: *Our approach is robust to the case with additional individual and time fixed effects. On the other hand, the interactive fixed effect (IFE) estimator proposed by Bai (2009) and other related approaches need additional treatment in this case. For example, consider the interactive fixed effect model with two-way fixed effects:*

$$y_{it} = \beta_0 + x_{it}'\beta + \lambda_i' f_t + \alpha_i + \phi_t + \varepsilon_{it},$$

where α_i is the individual fixed effect and ϕ_t is the time fixed effect. To deal with α_i, we can write down an additional projection similar to (7), i.e.,

$$\alpha_i = \underset{r\times(p+1)}{\tilde{A}} \underset{(p+1)\times 1}{\overline{x}_{i\cdot}} + \tilde{\eta}_i. \tag{13}$$

Then, by adding (13) to the right hand side of (8), we obtain

$$y_{it} = \beta_0 + x_{it}'\beta + \overline{x}_{\cdot t}'\rho_i + \overline{x}_{i\cdot}'(\delta_t + \tilde{A}) + \overline{x}_{i\cdot}'A'B\overline{x}_{\cdot t} + (\tilde{\eta}_i + u_{it}). \tag{14}$$

Thus, the two-way Mundlak estimator remains consistent in this case. The projection for ϕ_t is similar.

2.3 Asymptotic Normality

In this section, we show the asymptotic properties of the two-way Mundlak type estimator $\widehat{\beta}_{M2}$. First, we give the following regularity conditions.

Assumption 1 *A and B are finite* ***and full row rank matrices.***

Assumption 2 *(i) $E||\tilde{x}_{it}||^4 < \infty$; (ii) the matrix $(NT)^{-1}\tilde{X}'\tilde{X}$ converges to non-singular matrix, as $(T, N) \to \infty$.*

Assumption 3 *(i) u_{it} is independent of $\tilde{x}_{it}$; (ii) Given $\tilde{x}_{it}$, let $U_t = (u_{1t,\cdots,}u_{Nt})'$, and $\bar{U}_t = \frac{1}{\sqrt{N}}\sum_{i=1}^N u_{it}$, in which U_t follows a process $U_t = g(\theta_t, \theta_{t-1}, \cdots)$ with $\theta_t = (\theta_{1t}, \cdots, \theta_{Nt})'$ being a sequence of independent and identically distributed random vectors, $E(\theta_t|\tilde{x}_{it}) = 0_N$, and $g(\cdot)$ is a measurable function. In addition, let $\bar{U}_t^\star = \frac{1}{\sqrt{N}}U_t^{*\prime}1_N$, where $U_t^\star = g(\theta_t, \cdots, \theta_1, \theta_0', \theta_{-1}', \cdots)$ is the coupled version of U_t, and θ_t' is an independent copy of θ_t. Suppose that $\sum_{t=0}^\infty t^2\delta_{t,\kappa}^U < \infty$, for $\kappa \geq 4$, where $\delta_{t,\kappa}^U = ||\bar{U}_t - \bar{U}_t^\star||_\kappa$.*

Remark 3: *(1) The full row rank matrix of A in Assumption 1 implies the factor loadings λ_i have not overlap, such that the rank of matrix $\frac{1}{N}\sum_{i=1}^N \lambda_i'\lambda_i$ equals r. Similar, the full row rank matrix of B implies the $\frac{1}{N}\sum_{i=1}^N f_t'f_t$ is full rank.* ***If the matrices of***

A ***and*** B ***are not full column rank, it implies that the number of true factors is smaller than*** r ***and our estimation also work. We allow for the factor number is larger than the number of regressors;*** *(2) Assumption 2 directly set the identification condition on the transformed regressors* $\tilde{x}_{it}$*, which is common in the literature. (3) Since* $u_{it} = \eta_i'\xi_t + \varepsilon_{it}$*, to show the estimated estimator is consistent, we only need* $E(u_{it}|\tilde{x}_{it}) = 0$*, allowing for* ξ_t*,*η_i *and* ε_{it} *are mutual dependent. In Pesaran (2006), the regressors are driven by common factors* $x_{it} = \Gamma_i' f_t + v_{it}$*, in which* v_{it} *is not correlated with* ε_{it}*. Thus, we relax it; (4) To obtain the asymptotic distribution of the estimator, Assumption 3(i) sets strictly assumption that* u_{it} is independent of $\tilde{x}_{it}$*, instead of* $E(u_{it}|\tilde{x}_{it}) = 0$*, which is necessary as Bai (2009) and Gao et al.(2023); (6) Since* $u_{it} = \eta_i'\xi_t + \varepsilon_{it}$*, for all* (i, j) *and* (t, s)*, there exists heteroskedasticity across each* i *and* t*, serial correlation for each* i *and sectional dependence among individuals for each* t*. Assumption 3(ii) assume the generating process of composite error* u_{it}*, which is borrowed from Assumption 1 of Gao, Peng and Yan (2023). It's more general and allows for both cross-sectional dependence and serial correlation and conditional heteroscedasticity in the composite error* u_{it}*. As noted in Example 1.2 of Gao, Peng and Yan (2023), Assumption 3 implies Assumption C of the errors in Bai (2009), allowing for dependence in the both dimensions. Thus, according to Assumptions 2 and 3, and Theorem 2.1 in Gao, Peng and Yan (2023), let* $\underset{NT\times p}{\mathbb{X}} = (M_{\underline{X}} \otimes M_{\bar{X}})\tilde{X}$*, and the* $(i-1)T + t$ *row of* $\mathbb{X}$ *is denoted by* $\underset{p\times 1}{\mathbb{X}_{ti}}$*,*

$$(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{X}_{ti}u_{it} \xrightarrow{d} N(0, \Phi_{NT}),$$

where Φ_{NT} *is* $p \times p$ *dimensional non-singular positive matrix.*

Under the above assumptions, we obtain the following Theorem 1.

Theorem 1 *Under Assumptions 1-3, as* $(T, N) \to \infty$*, then*

$$\sqrt{NT}(\widehat{\beta}_{M2} - \beta) \xrightarrow{d} N(0, V_{\beta 2}).$$

where $V_{\beta 2} = \Psi_{NT}^{-1}\Phi_{NT}\Psi_{NT}^{-1}$ *with* $\Psi_{NT} = \underset{(N,T)\to\infty}{plim} \frac{1}{NT}\tilde{X}'\tilde{X}$ *and*

$$\Phi_{NT} = \underset{(T,N)\to\infty}{plim} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} E(u_{it}u_{js}')\mathbb{X}_{ti}\mathbb{X}_{sj}',$$

where $u_{it} = \eta_i'\xi_t + \varepsilon_{it}$*.*

Theorem 1 states the asymptotic distribution of $\widehat{\beta}_{M2}$ as $(T,N)\to\infty$. **Based on the discussions in Gao et al. (2023), the panel HAC estimation is consistent estimator of $V_{\beta 2}$. Specifically, let $\underset{p\times 1}{\hat{\nu}_{it}}=\tilde{x}_{it}\hat{\hat{u}}_{it}$ with $\hat{\hat{u}}_{it}=\tilde{y}_{it}-\tilde{x}_{it}'\hat{\beta}_{M2}$. Define $\hat{\nu}_t=\sum_{i=1}^{N}\tilde{x}_{it}\hat{\hat{u}}_{it}$ and compute a HAC estimator as**

$$\hat{\hat{\Omega}}=\hat{\hat{\Gamma}}_0+\textstyle\sum_{s=1}^{T-1}k(s/M)(\hat{\hat{\Gamma}}_s+\hat{\hat{\Gamma}}_s'),$$

where $\hat{\hat{\Gamma}}_s=\frac{1}{T}\sum_{t=s+1}^{T}\hat{\nu}_t\hat{\nu}_{t-s}'$. Defining$\hat{\hat{\Omega}}$ equals

$$\hat{\hat{\Omega}}=\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}K_{ts}\hat{\nu}_t\hat{\nu}_s',$$

where $K_{ts}=k(|t-s|/M)$ is the Bartlett kernel with bandwidth $M\approx T^{1/3}$. Thus, the estimation of variance-covariance matrix of $\hat{\beta}_{M2}$ has the following sandwich form,

$$\hat{V}_{\beta 2}=T(\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}')^{-1}\hat{\hat{\Omega}}(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}')^{-1}.$$

According to Theorem 2.3 of Gao et al. (2023), $E(\hat{V}_{\beta 2})=V_{\beta 2}+o_p(1)$, as $(N,T)\to\infty$. Under our two-way Mundlak projection transformation, the panel data model with the interactive fixed effects becomes the pooled panel data model. Thus, it's convenient to apply the wild bootstrap and robust inference in the spirit of Gao et al. (2023) in Section 4 and Vogelsang (2012) in Section 5 below.

We can also derive the asymptotic distribution of $\widehat{\beta}_{M2}$ under fixed T and $N\to\infty$, or fixed N and $T\to\infty$. For example, if T is finite and under Assumptions, thus the Corollary 1 can be derived. The case of fixed N and $T\to\infty$ can be derived similarly.

Corollary 1 ***Under the Assumptions*** $N^{-1/2}\sum_{i=1}^{N}\mathbb{X}_{ti}u_{it}\xrightarrow{d}N(0,\Phi_N)$ ***with*** $\Phi_N=\underset{N\to\infty}{plim}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}E(u_{it}u_{js}')\mathbb{X}_{ti}\mathbb{X}_{sj}'$, ***and*** $\frac{1}{N}\tilde{X}'\tilde{X}$ ***converges to nonsingular matrix as*** $N\to\infty$, ***then***

$$\sqrt{N}(\widehat{\beta}_{M2}-\beta)\xrightarrow{d}N(0,V_{\beta 2}),$$

with $\Psi_N=\underset{N\to\infty}{plim}\frac{1}{N}\tilde{X}'\tilde{X}$ ***and*** $V_{\beta 2}=\Psi_N^{-1}\Phi_N\Psi_N^{-1}$.

3 The Two-way Mundlak Projection in the Interactive Form (CIF)

In general, the regressors x_{it} can be correlated with the factor structure $\lambda_i' f_t$ in an interactive form. In such cases, the validity of the approach described in Section 2 would require more stringent assumptions. Specifically, under the model

$$y_{it} = \beta_0 + x_{it}'\beta + \overline{x}_{\cdot t}'\rho_i + \overline{x}_{i\cdot}'\delta_t + \overline{x}_{i\cdot}'A'B\overline{x}_{\cdot t} + \eta_i'\xi_t + \varepsilon_{it},$$

if the factor structure $\lambda_i' f_t$ is correlated with x_{it}, then $\eta_i'\xi_t$ in the composite errors u_{it} also be correlated with x_{it}.

In this section, we first show that under certain specific situations, the two-way Mundlak projection method proposed in Section 2 still works even with the interactive form. Second, we propose a hybrid approach by combining the two-way Mundlak projection method with instrument variables. The validity of the new approaches under the general setup with the interactive form is show.

Remark 4: *In addition, we note that the averaging method in p.647 of Hsiao (2018) and other one-way or two-way transformed estimation are also not applicable in the general case with the interactive form. More specifically, given that $E(\eta_i) = 0$ or $E(\xi_t) = 0$, averaging both sides of equation (8) along the time and/or individual dimension gives*

$$\begin{aligned}
\overline{y}_{i\cdot} &= (\beta_0 + c'd) + \overline{x}_{i\cdot}'\beta + \overline{x}_{\cdot\cdot}'\rho_i + \overline{x}_{i\cdot}'\bar{\delta}_{\cdot} + \overline{x}_{i\cdot}'A'B\overline{x}_{\cdot\cdot} + \eta_i'd + c'\bar{\xi}_{\cdot} + {\eta_i}'\bar{\xi}_{\cdot} + \bar{\varepsilon}_{i\cdot} \qquad (15)\\
\overline{y}_{\cdot t} &= (\beta_0 + c'd) + \overline{x}_{\cdot t}'\beta + \overline{x}_{\cdot t}'\bar{\rho}_{\cdot} + \overline{x}_{\cdot\cdot}'\delta_t + \overline{x}_{\cdot\cdot}'A'B\overline{x}_{\cdot t} + \bar{\eta}_{\cdot}'d + c'\xi_t + \bar{\eta}_{\cdot}'\xi_t + \bar{\varepsilon}_{\cdot t} \qquad (16)\\
\overline{y}_{\cdot\cdot} &= \frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N} y_{js} = (\beta_0 + c'd) + \overline{x}_{\cdot\cdot}'\beta + \overline{x}_{\cdot\cdot}'\bar{\rho}_{\cdot} + \overline{x}_{\cdot\cdot}'\bar{\delta}_{\cdot} + \overline{x}_{\cdot\cdot}'A'B\overline{x}_{\cdot\cdot}\\
&\quad + \bar{\eta}_{\cdot}'d + c'\bar{\xi}_{\cdot} + \bar{\eta}_{\cdot}'\bar{\xi}_{\cdot} + \bar{\varepsilon}_{\cdot\cdot},
\end{aligned}$$

where $\bar{\delta}_{\cdot} = \frac{1}{T}\sum_{s=1}^{T}\delta_s = A'd + \bar{\xi}$ and $\bar{\rho}_{\cdot} = \frac{1}{N}\sum_{j=1}^{N}\rho_j = B'c + \bar{\eta}$. For large T (or N), (or assuming) averaged $\bar{\xi}_{\cdot} = 0$, equation (15) becomes

$$\begin{aligned}
\overline{y}_{i\cdot} &= (\beta_0 + c'd + c'B\overline{x}_{\cdot\cdot}) + \overline{x}_{i\cdot}'(\beta + A'd + A'B\overline{x}_{\cdot\cdot}) \qquad (17)\\
&\quad + ({\eta_i}'(d + B\overline{x}_{\cdot\cdot}) + \bar{\varepsilon}_{i\cdot}),\\
\overline{y}_{\cdot t} &= (\beta_0 + c'd + \overline{x}_{\cdot\cdot}'A'd) + \overline{x}_{\cdot t}'(\beta + B'c + B'A\overline{x}_{\cdot\cdot}) \qquad (18)\\
&\quad + ((\overline{x}_{\cdot\cdot}'A + c)\xi_t + \bar{\varepsilon}_{\cdot t}),
\end{aligned}$$

Thus, the ordinary least square estimation using time averaged data (17) is inconsistent, as $\beta + A'd + A'B\overline{x}_{\cdot\cdot}$ is different from β, due to the additional terms $\overline{x}_{i\cdot}'\delta_t + \overline{x}_{i\cdot}'A'B\overline{x}_{\cdot t}$

in equation (8). Similarly, the ordinary least square estimation using cross-section averages (18) is inconsistent, due to the additional terms $\overline{x}'_{\cdot t}\rho_i + \overline{x}'_{i\cdot}A'B\overline{x}_{\cdot t}$ in equation (8). It is therefore different from the case in equations (2.4) and (2.5) of Hsiao (2018) on p.647.

3.1 The General Interactive Form

Let g_t be the $r' \times 1$ dimensional unobservable common factors in the data generating process of the regressors x_{it}, similar to the setup of CCE. Specifically, we explicitly write the interactive form between the regressors x_{it} and the factor structure[1] as follows:

$$x_{it} = \Gamma'_{ix} g_t + h_{it}, \tag{19}$$

where Γ_{ix} denotes the factor loadings of x_{it}, **which is correlated with λ_i**, and h_{it} denotes the idiosyncratic error, which is also called the de-factored version of the regressors x_{it} in the literature; e.g., see Cui et al., (2022), Cui et al. (2023), Cao et al. (2023) and others.

Intuitively, we consider the following two cases. For simplicity, let $f_t = (f'_{t1}, f'_{t2})'$. First, if x_{it} is correlated with certain part of the common factors such as f_{t1} and uncorrelated with f_{t2}, then we expect that the space of g_t in model (19) is related with the space of f_{t1}. Thus, we could control the correlation between the regressors and the factor structure by controlling for g_t. **Similar to Juodis, Karabiyik, Westerlund (2021), we allow for the possibility that some of the factors that enter the equation for x_{it}, do not enter the equation for y_{it}.** Second, if x_{it} is correlated with the whole common factor f_t in y_{it}, then we expect that the space of g_t in model (19) should at least cover the space of f_t given $r' \geq r$.

Furthermore, we need to distinguish between three cases, regarding the rank of the $r' \times p$ dimensional matrix Γ_{ix} and the factor number r. First, in the case that $r \leq r' \leq p$, Γ_{ix} is a full row rank matrix and we could transform equation (19) into the setup of equations (3) and (7). The full row rank of Γ_{ix} ensure that Assumption 1 is valid and the two-way Mundlak projection method in section 2 remains valid even under interactive forms. Specifically,

[1]Similar to Freeman and Weidner (2023), the relationship between the regressor and the factor structure can be written as the nonparametric format, such that for $j = \{1, \cdots, p\}$, $x_{it,j} = \sum_{s=1}^{r'} \sigma_s \varphi_x(\alpha_i)\psi_x(\gamma_t) + h_{it,j}$,where $h_{it,j}$ is the idiosyncratic errors for each regressor $x_{it,j}$. Thus, it can be represented as the interactive form where $g_t = \psi_x(\gamma_t)$ be the $r' \times 1$ dimensional unobserved common factors and $\Gamma_i = \sigma_s \varphi_x(\alpha_i)$ is its factor loadings for each regressors,

according equation (1), $\bar{x}_{\cdot t} = \bar{\Gamma}'_x g_t + \overline{h}_{\cdot t}$, **with** $\overline{h}_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N} h_{it} \xrightarrow{p} 0_{p\times 1}$ **as** $N \to \infty$. **Thus,** $(\bar{\Gamma}_x \bar{\Gamma}'_x)^{-1}\bar{\Gamma}'_x \bar{x}_{\cdot t} \xrightarrow{p} g_t$, **and then** $M_{\bar{X}} G \xrightarrow{p} 0_{T\times r'}$, **as** $N \to \infty$.

Second, in the case that $r' > p \geq r$,**the number of regressors is not smaller than the number of common factor that enter the equation for** y_{it}. **We can divide** g_t **into two parts** $g_t = (f'_t, g'_{t,-y})'$, **and then we can transform equation (19)** into

$$f_t = \underset{r\times p}{B_1}\overline{x}_{\cdot t} + \underset{r\times(r'-r)}{B_2} g_{t,-y} + \xi^{\dagger}_t.$$

where $\xi^{\dagger}_t$ **is the projection's error. The Mundlak projection introduces the additional correlation between regressor** x_{it} **and** $g_{t,-y}$, **thus the pooled least squared is still invalid.**

Last, for the case that $r' \geq r > p$, **such that the rank condition is not satisfied, we can not project all the common factors and factor loadings into the space of the cross-sectional** x_{it} **as the two-way Mundlak's format of equations (3) and (7). Thus, the two-way Mundlak projection in Section 2 can be invalid and thus we propose the following methods of Section 3.2, which combine the previous two-way Mundlak projection method with instrumental variables.**

3.2 Two-way Mundlak Projection with Instrumental Variables

In the economic forecasting, the unobserved common factors can be estimated by many observed macroeconomic variable (Stock and Watson, 2002). Bai and Ng (2010) used the exogenous variable in the data rich environment to estimated the common factors. In line with those spirit, we use the exogenous variables to fill the rank deficiency of the equation (19).

A large panel of macroeconomic variables z_{it} are also driven by the common factors g_t as equation (19), such that

$$\underset{q\times 1}{z_{it}} = \Gamma'_{iz} g_t + \zeta_{it}, \tag{20}$$

where Γ_{iz} is the $r' \times q$ factor loadings and ζ_{it} is the error. Thus, we combine the (19) with above (20), obtaining the augmented regressors x^*_{it}

$$\begin{aligned}\underset{(p+q)\times 1}{x^*_{it}} &= \begin{pmatrix} x_{it} \\ z_{it} \end{pmatrix} = \begin{pmatrix} \Gamma'_{ix} \\ \Gamma'_{iz} \end{pmatrix} g_t + \begin{pmatrix} h_{it} \\ \zeta_{it} \end{pmatrix} \\ &= \Gamma'_i g_t + \vartheta_{it} \end{aligned} \tag{21}$$

where $\Gamma_i = (\Gamma_{ix}, \Gamma_{iz})$ and $\vartheta_{it} = (h'_{it}, \zeta'_{it})'$. The equation (20) makes sure $p + q > r' \geq r$.

Let $\overline{x}^*_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N} x^*_{it}$ and $\overline{x}^*_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x^*_{it}$. Thus, we can project the common factor f_t and the factor loadings λ_i into the spaces of the cross-sectional and time average of the augmented regressor x^*_{it} **as the following two-way Mundlak's projection,**

$$f_t = B\overline{x}^*_{\cdot t} + \xi^*_t,$$

and

$$\lambda_i = A\overline{x}^*_{i\cdot} + \eta^*_i.$$

Next, we repeat the procedure in the equations (8)-(10) and then obtain

$$M_{\bar{X}^*} Y M_{\underline{X}^*} = \beta_1 \cdot M_{\bar{X}^*} X^1 M_{\underline{X}^*} + ... + \beta_p \cdot M_{\bar{X}^*} X^p M_{\underline{X}^*} + M_{\bar{X}^*} U^* M_{\underline{X}^*}. \quad (22)$$

where $\underset{N\times(p+1)}{\underline{X}^*} = (\overline{x}^*_{1\cdot}, \overline{x}^*_{2\cdot}, ..., \overline{x}^*_{N\cdot})'$, $\underset{T\times(p+1)}{\bar{X}^*} = (l_T, \bar{X}^*_{\cdot,1}, \cdots, \bar{X}^*_{\cdot,p})$ and $M_{\bar{X}^*} = I_T - \bar{X}^*(\bar{X}^{*\prime}\bar{X}^*)^{-1}\bar{X}^{*\prime}$,$M_{\underline{X}^*} = I_T - \underline{X}^*(\underline{X}^{*\prime}\underline{X}^*)^{-1}\underline{X}^{*\prime}$. The composite error $u^*_{it} = \eta^{*\prime}_i \xi^*_t + \varepsilon_{it}$ and $u^*_i = (u^*_{i1}, u^*_{i2}, \cdots, u^*_{iT})'$, $\underset{T\times N}{U^*} = (u^*_1, u^*_2, ..., u^*_N)$.

Last, we collect all the transformed regressors for individual i at period t, $\underset{NT\times p}{\tilde{X}^*} = [vec(M_{\bar{X}^*} X^1 M_{\underline{X}^*}), \cdots, vec(M_{\bar{X}^*} X^p M_{\underline{X}^*})]$. Similarly, let $\underset{NT\times 1}{\tilde{Y}^*} = vec(M_{\bar{X}^*} Y M_{\underline{X}^*})$, and $\underset{NT\times 1}{\tilde{U}^*} = vec(M_{\bar{X}^*} U^* M_{\underline{X}^*})$. Thus, equation (22) can be further parameterized by

$$\tilde{Y}^* = \tilde{X}^*\beta + \tilde{U}^*, \quad (23)$$

and the least squared estimator is also defined as

$$\widehat{\beta}_{M4} = (\tilde{X}^{*\prime}\tilde{X}^*)^{-1}\tilde{X}^{*\prime}\tilde{Y}^*. \quad (24)$$

We give the similar Assumptions of Section 2.

Assumption 4 *(i) $E||\tilde{x}^*_{it}|| < \infty$; (ii) the matrix $\frac{1}{NT}\tilde{X}^{*\prime}\tilde{X}^*$ converges to non-singular matrix, as $(T, N) \to \infty$.*

Assumption 5 *(i) u^*_{it} is independent of $\tilde{x}^*_{it}$; (ii) Given $\tilde{x}^*_{it}$, let $U^*_t = (u^*_{1t,\cdots,} u^*_{Nt})'$, and $\bar{U}^*_t = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u^*_{it}$, in which U^*_t follows a process $U^*_t = g(\theta_t, \theta_{t-1}, \cdots)$ with $\theta_t = (\theta_{1t}, \cdots, \theta_{Nt})'$ being a sequence of independent and identically distributed random vectors, $E(\theta_t|\tilde{x}^*_{it}) = 0_N$, and $g(\cdot)$ is a measurable function. In addition, let $\bar{U}^\star_t = \frac{1}{\sqrt{N}} U^{*\prime}_t 1_N$, where $U^\star_t = g(\theta_t, \cdots, \theta_1, \theta'_0, \theta'_{-1}, \cdots)$ is the coupled version of U_t, and θ'_t is an independent copy of θ_t. Suppose that $\sum_{t=0}^{\infty} t^2 \delta^U_{t,\kappa} < \infty$, for $\kappa \geq 4$, where $\delta^U_{t,\kappa} = ||\bar{U}_t - \bar{U}^\star_t||_\kappa$.*

Under above Assumptions and let $\underset{NT\times p}{\mathbb{X}^*} = (M_{\underline{X}^*} \otimes M_{\bar{X}^*})\tilde{X}^*$, the $(i-1)T+t$ row of $\mathbb{X}$ is denoted by $\underset{p\times 1}{\mathbb{X}^*_{ti}}$, we obtain the following Proposition.

Theorem 2 *Under Assumptions 4-5, as $(T,N) \to \infty$, then*

$$\sqrt{NT}(\widehat{\beta}_{M4} - \beta) \xrightarrow{d} N(0, V_{\beta 4}).$$

where $V_{\beta 4} = \Psi^{*-1}_{NT}\Phi^*_{NT}\Psi^{*-1}_{NT}$ *with* $\Psi^*_{NT} = \underset{(N,T)\to\infty}{plim} \frac{1}{NT}\tilde{X}^{*\prime}\tilde{X}^*$ *and*

$$\Phi^*_{NT} = \underset{(T,N)\to\infty}{plim} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} E(u^*_{it}u^{*\prime}_{js})\mathbb{X}^*_{ti}\mathbb{X}^{*\prime}_{sj},$$

where $u^*_{it} = \eta^{*\prime}_i\xi^*_t + \varepsilon_{it}$.

We can also estimate $V_{\beta 4}$ by the panel HAC estimation. Let $\underset{p\times 1}{\hat{\nu}^*_{it}} = \tilde{x}^*_{it}\hat{u}^*_{it}$ with $\hat{u}^*_{it} = \tilde{y}^*_{it} - \tilde{x}^{*\prime}_{it}\hat{\beta}_{M4}$. Define $\hat{\nu}^*_t = \sum_{i=1}^{N}\tilde{x}^*_{it}\hat{u}^*_{it}$ and compute a HAC estimator as

$$\hat{\hat{\Omega}}^* = \hat{\hat{\Gamma}}^*_0 + \sum_{s=1}^{T-1}k(s/M)(\hat{\hat{\Gamma}}^*_s + \hat{\hat{\Gamma}}^{*\prime}_s),$$

where $\hat{\hat{\Gamma}}^*_s = \frac{1}{T}\sum_{t=s+1}^{T}\hat{\nu}^*_t\hat{\nu}^{*\prime}_{t-s}$. Thus, the variance-covariance matrix of $\hat{\beta}_{M4}$ is estimated by

$$\hat{V}_{\beta 4} = T(\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}^*_{it}\tilde{x}^{*\prime}_{it})^{-1}\hat{\hat{\Omega}}^*(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}^*_{it}\tilde{x}^{*\prime}_{it})^{-1}.$$

4 Endogenous Regressors

In empirical application, the endogeneity of data popularly exits. In this section, $x_{it} = (x'_{1it}, x'_{2it})'$, in which x_{1it} is a $p_1 \times 1$ vector of exogenous variables x_{1it} and x_{2it} is $p_2 \times 1$ vector of endogenous variables. In particular, $cov(x_{1it}, \varepsilon_{it}) = 0$, and x_{2it} are correlated with ε_{it}, such that $cov(x_{2it}, \varepsilon_{it}) \neq 0$, leading to the inconsistent estimator by the estimator in Section 2.

4.1 The Mundlak-Control Function approach

Let z_{2it} is $m \times 1$ vector of additional exogenous variables or instrument variables with $m \geq p_2$ and $z_{it} = (x'_{1it}, z'_{2it})'$. The endogenous regressors x_{2it} has linear reduced form,

$$x_{2it} = \alpha' z_{it} + q_{it}. \quad (25)$$

$$\hat{\hat{\Omega}}^{*} = \hat{\hat{\Gamma}}_0^{*} + \sum_{s=1}^{T-1} k(s/M)(\hat{\hat{\Gamma}}_s^{*} + \hat{\hat{\Gamma}}_s^{*\prime}),$$

where α is the coefficient and q_{it} is the error. The control function approach (CF; Wooldridge, 2015) assume the error term ε_{it} is expressed by the error term q_{it} in equation (25)

$$\varepsilon_{it} = q_{it}' \underset{p_2\times 1}{\pi} + \varpi_{it}. \tag{26}$$

Let $\omega_{it} = \eta_i' \xi_t + \varpi_{it}$, $\omega_{i\cdot} = (\omega_{i1}, \omega_{i2}, ..., \omega_{iT})'$, and $\underset{T\times N}{\omega} = (\omega_{1\cdot}, \omega_{2\cdot}, ..., \omega_{N\cdot})$. $\underset{T\times N}{Q^j}$ are $T\times N$ matrix being the j^{th} additional error matrix $[q_{it}^j]_{t=1,i=1}^{T,N}$, associated with parameter π_j for $j = \{1, 2, \cdots, p_2\}$. Plugging equation (26) into equation (10) gives

$$M_{\bar{X}} Y M_{\underline{X}} \;=\; \sum_{j=1}^{p} \beta_j \cdot M_{\bar{X}} X^j M_{\underline{X}} + \sum_{j=1}^{p_2} \pi_j \cdot M_{\bar{X}} Q^j M_{\underline{X}} + M_{\bar{X}} \omega M_{\underline{X}} \tag{27}$$

Compared with equation (10), controlling for the correlation between x_{it} and factor structure, equation (27) add additional term $\sum_{j=1}^{p_2} \pi_j \cdot M_{\bar{X}} Q^j M_{\underline{X}}$ to control for the endogeneity between between x_{it} and ε_{it}. If Q^j is observed, the least square estimators of (27) is consistent. We only interest in the slope β and then partial out nuisance parameters.

Furthermore, we vector the variables in equation (27) as Section 2. Similarly, $\underset{NT\times p}{\mathbb{Q}} = [vec(M_{\bar{X}} Q^1 M_{\underline{X}}), \cdots, vec(M_{\bar{X}} Q^{p_2} M_{\underline{X}})]$, and $\tilde{\omega} = vec(M_{\bar{X}} \omega M_{\underline{X}})$. Thus, equation (27) can be further parameterized by

$$\tilde{Y} = \tilde{X}\beta + \mathbb{Q}\pi + \tilde{\omega}. \tag{28}$$

Since q_{it} is not observed in practice, we then follows two-step procedure to estimate the parameters in the spirit of the control function approach. In the first step, run regression (25) to obtain residuals

$$\hat{q}_{it} = x_{2it} - (\sum_{i=1}^{N} \sum_{t=1}^{T} x_{2it} z_{it}')(\sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} z_{it}')^{-1} z_{it},$$

and then $\hat{\mathbb{Q}}$ obtained as above definition. In the second step, after plugging $\hat{\mathbb{Q}}$ into transformed variables, we lastly obtain the interested coefficients β, by the pooled estimator

$$\hat{\beta}_{M4} = (\tilde{X}' M_{\hat{\mathbb{Q}}} \tilde{X})^{-1} \tilde{X}' M_{\hat{\mathbb{Q}}} \tilde{Y}. \tag{29}$$

4.2 Asymptotic Properties

Assumption 6 *(i) $E(q_{it})=0$, and $E(||q_{it}||^4)<\infty$ for all i,t; (ii) $E(q_{it}q_{js}')=\Sigma_{q,ijts}$ with constant $||\Sigma_{q,ijts}||<\infty$ for all i,j,t,s. (iii) for $j=\{1,\cdots,p_2\}$,$E(x_{1it}q_{it}^j)=E(z_{it}q_{it}^j)=0$.*

Assumption 7 *(i) $E||z_{it}||^4<\infty$; (ii) the matrices $(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T z_{it}z_{it}'$ and $(NT)^{-1}\tilde{X}'M_{\mathbb{Q}}\tilde{X}$ converges to non-singular matrix, as $(T,N)\to\infty$.*

Assumption 8 *(i) ω_{it} is independent of $\tilde{x}_{it}$, q_{it} and z_{it} (ii) Given $\tilde{x}_{it}$ and z_{it}, let $W_t=(\omega_{1t,\cdots},\omega_{Nt})'$, and $\bar{W}_t=\frac{1}{\sqrt{N}}\sum_{i=1}^N\omega_{it}$, in which W_t follows a process $W_t=g(\theta_t,\theta_{t-1},\cdots)$ with $\theta_t=(\theta_{1t},\cdots,\theta_{Nt})'$ being a sequence of independent and identically distributed random vectors, $E(\theta_t|\tilde{x}_{it},z_{it})=0_N$, and $g(\cdot)$ is a measurable function. In addition, let $\bar{W}_t^\star=\frac{1}{\sqrt{N}}W_t^{\star\prime}1_N$, where $W_t^\star=g(\theta_t,\cdots,\theta_1,\theta_0',\theta_{-1}',\cdots)$ is the coupled version of W_t, and θ_t' is an independent copy of θ_t. Suppose that $\sum_{t=0}^\infty t^2\delta_{t,\kappa}^U<\infty$, for $\kappa\geq 4$, where $\delta_{t,\kappa}^U=||\bar{W}_t-\bar{W}_t^\star||_\kappa$.*

Similar to Section 2.3, according to Assumptions 2, 6, 7 and 8, we obtain:

(i) Let $\underset{NT\times p}{\tilde{\mathbb{X}}}=(M_{\underline{X}}\otimes M_{\bar{X}})M_{\mathbb{Q}}\tilde{X}$, and its the $(i-1)T+t$ row of $\tilde{\mathbb{X}}$ is $\underset{p\times 1}{\tilde{\mathbb{X}}_{ti}}$,

$$(NT)^{-1/2}\sum_{i=1}^N\sum_{t=1}^T\tilde{\mathbb{X}}_{ti}\omega_{it}\xrightarrow{d}N(0,\Theta_{NT}^{(1)}),$$

where $\Theta_{NT}^{(1)}=\underset{(T,N)\to\infty}{plim}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\sum_{i'=1}^N\sum_{s=1}^T E(\omega_{it}\omega_{i's}')\tilde{\mathbb{X}}_{ti}\tilde{\mathbb{X}}_{si'}'$. is $p\times p$ dimensional nonsingular positive matrix.

(ii) $\underset{NT\times p}{\tilde{\mathbb{Z}}}=(M_{\underline{X}}\otimes M_{\bar{X}})M_{\mathbb{Q}}\tilde{X}M_Z$, and the $(i-1)T+t$ row of $\tilde{\mathbb{Z}}$ is $\underset{p\times 1}{\tilde{\mathbb{Z}}_{ti}}$,

$$(NT)^{-1/2}\sum_{i=1}^N\sum_{t=1}^T\tilde{\mathbb{Z}}_{ti}\pi' q_{it}\xrightarrow{d}N(0,\Theta_{NT}^{(2)}),$$

where $\Theta_{NT}^{(2)}=\underset{(T,N)\to\infty}{plim}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\sum_{i'=1}^N\sum_{s=1}^T E(q_{it}q_{i's}')\tilde{\mathbb{Z}}_{ti}\pi'\pi\tilde{\mathbb{Z}}_{si'}'$ is $p\times p$ dimensional nonsingular positive matrix.

Under above Assumptions, we can show the MLS estimator, combined with control function approach is also consistent.

Proposition 1 *Under Assumptions 1, 2, and 6-8, as $(T,N)\to\infty$, then*

$$\sqrt{NT}(\hat{\beta}_{M4}-\beta)\xrightarrow{d}N(0,V_{\beta 4}).$$

where $V_{\beta 4} = \tilde{\Psi}_{NT}^{-1}\Theta_{NT}\tilde{\Psi}_{NT}^{-1}$ *with* $\tilde{\Psi}_{NT}^{-1} = \underset{(N,T)\to\infty}{plim} \frac{1}{NT}\tilde{X}'M_{\mathbb{Q}}\tilde{X}$ *and* $\Theta_{NT} = \Theta_{NT}^{(1)} + \Theta_{NT}^{(2)}$.

5 The Dependent Wild Bootstrap (DWB) Tests

(Need to adjust) As stated above, for $s = \{2, 3, 4\}$, **the variance of** $\sqrt{NT}(\widehat{\beta}_{Ms} - \beta)$ **is not feasible to be estimated by traditional Panel HAC estimation.** In this section, we apply the DWB procedure with B repetitions for obtaining the sequence $\widehat{\beta}^*_{Ms,b}, b = \{1, \cdots, B\}$, and then estimating $V_{\beta s}$. Specifically, we illustrate the procedures by estimating $V_{\beta 2}$ and show its asymptotic properties. The estimating of $V_{\beta 3}$ and $V_{\beta 4}$ can be derived similarly and then are omitted. An l-dependent time series ϵ_t satisfying the following condition:

Assumption 9 *Let* $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = 1$, $E(\epsilon_t^4) < \infty$,$E(\epsilon_t\epsilon_s) = k(\frac{t-s}{l})$, *where* $\frac{1}{l}$ *and* $\frac{l}{T} \to 0$, *as* $(l, T) \to \infty$, *and* $k(\cdot)$ *is a symmetric kernel function defined on* $[-1, 1]$ *satisfying that* $k(\cdot)$ *is Lipschitz continuous on* $[-1, 1]$, $k(0) = 1$, *and* $K(d) = \int_{-\infty}^{\infty} k(u)e^{-iud}du \geq 0$ *for all* $d \in \mathbb{R}$.

The DWB estimator of $V_{\beta 2}$ is given by the bootstrap population variance–covariance matrix of $(\widehat{\beta}^*_{M2} - \hat{\beta}_{M2})$, conditional on the original data, that is $\hat{V}^{boot}_{M2} = Var^*(\widehat{\beta}^*_{M2} - \hat{\beta}_{M2})$, implemented under the following steps:

1. After get the consistent estimator $\widehat{\beta}_{M2}$, we get $\hat{\tilde{u}}_{it} = \tilde{y}_{it} - \tilde{x}_{it}\widehat{\beta}_{M2}$,
2. DWB procedure (which is repeated B times)
   (a) Generating $\tilde{u}^*_{it,b} = \hat{\tilde{u}}_{it}\epsilon_t$, for $t = \{1, \cdots, T\}$;
   (b) Generating $\tilde{y}^*_{it,b} = \tilde{x}_{tt}\hat{\beta}_{M2} + \tilde{u}^*_{it,b}$, for $t = \{1, \cdots, T\}$;
   (c) obtaining Mundlak type estimator $\hat{\beta}^*_{M2}(b) = (\sum_{i=1}^N \sum_{t=1}^T \tilde{x}_{it}\tilde{x}'_{it})^{-1}(\sum_{i=1}^N \sum_{t=1}^T \tilde{x}_{it}\tilde{y}^*_{it,b})$;
3. Calculate the bootstrap standard error based on the series $\hat{\beta}^*_M(b), b = \{1, 2, \cdots, B\}$,

$$\hat{V}^{boot}_{M2} = \frac{1}{B-1}\sum_{b=1}^B(\hat{\beta}^*_{M2}(b) - \bar{\hat{\beta}}^*_{M2})(\hat{\beta}^*_{M2}(b) - \bar{\hat{\beta}}^*_{M2})'$$

   with $\bar{\hat{\beta}}^*_{M2} = \frac{1}{B}\sum_{b=1}^B \widehat{\beta}^*_{M2,b}$.

We test the null hypothesis H_a^0: $\beta_j = \beta_j^0, j \in \{1, \cdots, p\}$, with scalar β_j^0, against the alternative hypothesis $H_a^1 : \beta \neq \beta_0$. Traditional T test is adaptive, such as $T_{M2,j} = V_{\beta,jj}^{-1/2}(\widehat{\beta}_{M2,j} - \beta_j^0)$ with $V_{\beta 2,jj}$ is the j^{th}diagonal element of $V_{\beta 2}$, if $V_{\beta 2}$ is consistent estimated or known in prior. Thus, according to Theorem 1, then $T_{M2,j} \xrightarrow{d} N(0,1)$. The bootstrap T statistic is $T_{M2,j}^* = (\hat{V}_{M2,jj}^{boot})^{-1/2}(\widehat{\beta}_{M2,j}^*(b) - \widehat{\beta}_{M2,j})$ with $\hat{V}_{M2,jj}^{boot}$ is the j^{th}diagonal element of $\hat{V}_{M2}^{boot}$ and its p-value $p_T^* = \frac{1}{B}\sum_{b=1}^{B} 1\{|\widehat{\beta}_{M2,j}^*(b) - \widehat{\beta}_{M2,j}| > |\widehat{\beta}_{M2,j} - \beta_j^0|\}$.

For the multivariate null hypothesis $H_b^0 : \underset{l \times p}{R_1}\beta = \theta_1$, **with** θ_1 **being an** $l \times 1$ **constant vector, against the alternative hypothesis** $H_b^1 : R_1\beta \neq \theta_1$, **the Wald-type statistics is** $W_{M2} = NT(R_1\beta - \theta_1)'(R_1 V_{\beta 2} R_1')^{-1}(R_1\beta - \theta_1)'$ **and its asymptotic distribution is** χ_p^2, **if** $V_{\beta 2}$ **is available. Since** $V_{\beta 2}$ **is infeasible, instead of** $\hat{V}_{M2}^{boot}$**, the bootstrap Wald test is applied,** $W_{M2}^* = (\widehat{\beta}_{M2}^* - \widehat{\beta}_{M2})'R_1'(R_1\hat{V}_{M2}^{boot}R_1')^{-1}R_1(\widehat{\beta}_{M2}^* - \widehat{\beta}_{M2})$. **(**$W_{M2}^*$ **is compared with which statistic?)**

We give another Assumption 10, same as Assumption 3 in Gao et al. (2023), let $\lfloor q \rfloor$ denotes the largest integer not larger than q,

Assumption 10 *For* $q \in \lfloor q \rfloor$, *suppose that* $\lim_{|x| \to 0} \frac{1-k(x)}{|x|^q} = b_q$ *for some real number* $0 < b_q < \infty$.

Let P^* is the probability measure induced by the wild bootstrap conditional on the observed data, and then we have following Theorem 3.

Theorem 3 *Under Assumptions 1-3, 9, 10, as* $(T, N) \to \infty$, *then*

(i) If $B \to \infty$,

$$(NT)^{-1}\hat{V}_{M2}^{boot} \xrightarrow{p^*} V_{\beta 2};$$

(ii)

$$\sup_{\tau} |P^*(T_{M2}^* \leq \tau) - P(T_{M2} \leq \tau)| \xrightarrow{p} 0,$$

and

$$\sup_{\tau} |P^*(W_{M2}^* \leq \tau) - P(W_{M2} \leq \tau)| \xrightarrow{p} 0.$$

To select the optimal l above, Gao et al. (2023) minimized the mean squared error and they suggest that if $q = 1$, $l_{opt} = O(T^{1/3})$ and if $q = 2$, $l_{opt} = O(T^{1/5})$.

6 The Robust Tests

Following the spirit of robust inference in Vogelsang (2012), we propose a robust testing statistics, which is robust for heteroskedasticity, serial correlation and cross-sectional dependence in the error term $\tilde{U}$. Here, we also illustrate the testing via the Mundlak estimator $\hat{\beta}_{M2}$. Let $\underset{p\times 1}{\hat{\nu}_{it}} = \tilde{x}_{it}\hat{\hat{u}}_{it}$ with $\hat{\hat{u}}_{it} = \tilde{y}_{it} - \tilde{x}'_{it}\hat{\beta}_{M2}$. Define $\hat{\nu}_t = \sum_{i=1}^{N} \tilde{x}_{it}\hat{\hat{u}}_{it}$ and compute a HAC estimator as

$$\hat{\hat{\Omega}} = \hat{\hat{\Gamma}}_0 + \textstyle\sum_{s=1}^{T-1} k(s/M)(\hat{\hat{\Gamma}}_s + \hat{\hat{\Gamma}}'_s),$$

where $\hat{\hat{\Gamma}}_s = \frac{1}{T}\sum_{t=s+1}^{T} \hat{\nu}_t\hat{\nu}'_{t-s}$. $\hat{\hat{\Omega}}$ is equivalent to be expressed as

$$\hat{\hat{\Omega}} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} K_{ts}\hat{\nu}_t\hat{\nu}'_s,$$

where $K_{ts} = k(|t-s|/M)$ is the Bartlett kernel with bandwidth M. According to the definition of u_{it}, we select the bandwidth equal to the sample size $M = T$, as in Kiefer and Vogelsang (2002). Thus, the estimation of variance-covariance matrix of $\hat{\beta}_{M2}$ has the following sandwich form,

$$\hat{V}_{HACSC} = T(\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}'_{it})^{-1}\hat{\hat{\Omega}}(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}'_{it})^{-1}.$$

Based on above state, we consider the test for $R_2\beta = \theta_2$, against $R_2\beta \neq \theta_2$ with R_2 is a $l \times (p+1)$ matrix and θ_2 is a $l \times 1$ dimensional constant. Thus, we define the robust Wald test,

$$Wald_{HACSC} = (R_2\hat{\beta}_{M2} - \theta_2)'(R_2\hat{V}_{HACSC}R'_2)^{-1}(R_2\hat{\beta}_{M2} - \theta_2).$$

and in the case with $l = 1$,the robust T type statistics is

$$t_{HACSC} = \frac{R_2\hat{\beta}_{M2} - \theta_2}{\sqrt{R_2\hat{V}_{HACSC}R'_2}}.$$

For $a \in (0,1]$, $W_l(1,a)$ denotes a $l \times 1$ dimensional vector of standard Brownian motion and B_l denote a $l \times 1$ dimensional standard Brownian bridges. In addition,

Assumption 11 *For $a \in (0,1]$, the matrix $plim_{(N,T)\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}\tilde{x}'_{it} = a\Psi$ and Ψ are assumed to be non-singular.*

Assumption 12 *For $a \in (0,1]$ and $W_p(1,a)$ denotes a $p \times 1$ vector of standard Brownian sheets, $(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}\tilde{u}_{it} \Rightarrow \Lambda W_p(1,a)$ and Q are assumed to be non-singular.*

Assumption 13 *The process $\tilde{x}_{it}\tilde{u}_{it}$ is a mean zero vector of three dimensional stationary random fields indexed by i, t, s.*

Under the fixed b asymptotic framework, in our particular setup $b = 1$, the test statistics weakly convergence to random matrix under above Assumptions, which are borrowed from Kiefer and Vogelsang (2005) ,Vogelsang (2012) and others.

Theorem 4 *Under Assumptions 1, 11-13, as $(T,N) \to \infty$, then*

$$Wald_{HACSC} \Rightarrow W_l(1)'[2\int_0^1 B_l(r)B_l(s)'dr]^{-1}W_l(1),$$

and

$$t_{HACSC} \Rightarrow \frac{W_1(1)}{\sqrt{2\int_0^1 B_1(r)^2 dr}}.$$

To simplify, we use the Bartlett kernel with bandwidth M equal to the sample size T, such that $b = 1$ in the fixed b asymptotic theorem. The critical values of the T test are reported in Table 1 of Kiefer and Vogelsang (2002) and the critical value of the Wald Testing for $q = 1, 2, \cdots, 30$ can be obtain by multiplying $0.5q$ the critical value in Table II in Kiefer, Vogelsang and Bunzel (2000). For the fixed b asymptotic distribution of the testing statistics with $0 < b \leq 1$ and other kernel, such that Case (i) twice continuously differential kernel or Case (ii) twice continuously differential kernel with continuity in Kiefer and Vogelsang (2005) and Vogelsang (2012), are also appropriate in our paper. The critical values for the asymptotic distributions of the Wald and T tests refer to Table B in Vogelsang (2012).

7 Monte Carlo Simulations

This section provides Monte Carlo simulations to examine finite sample performance of our proposed the Mundlak least squared estimators of Section 2 ($MLS2$), Section 3 ($MLS3$). We compared those estimators with the CCE approach (Pesaran, 2006), IFE (Bai, 2009), MLE (Bai and Li, 2014), ATE (Hsiao et al., 2021). For the case of endogeneity, we compare the results of our proposed Mundlak-Control methods (*MLS2-CF* and *MLS3-CF*), with the common correlated estimator with instrumental variables (IVCCE; Harding and Lamarche, 2011); the profile GMM estimator (PGMM; Hong et al., 2023) and the average transformed GMM estimator (TGMM; Hsiao et al., 2023).

For all the simulations, the data $\{y_{it}, x_{it}\}$ are generated as,

$$y_{it} = x'_{it}\beta + \lambda'_i f_t + \varepsilon_{it},$$

where the true slopes $\beta = (\beta_1, \beta_2)'$ with $\beta_1 = 1$ and $\beta_2 = 2$. The factor loadings $\lambda_i = (\lambda_{i1}, \lambda_{i2})'$ are set as $\lambda_{i1} \sim iid\chi^2(1)$ and $\lambda_{i2} \overset{iid}{\sim} N(0.2, 0.2)$. For $t = \{-49, ...0, ..., T\}$, let $v_{ft} \sim iid\chi^2(3) - 3$ with $\rho_f = 0.5$, $f_{i,-50} = 0$. the common factors $f_t = (f_{t1}, f_{t2})'$ follows AR(1) process,

$$f_t = \rho_f f_{t-1} + v_{ft}.$$

For the errors ε_{it}, we consider two cases: (i) ε_{it} follows stationary AR(1) process with heteroskedasticity across each i. Let $\rho_{i\varepsilon} \sim iidU[0.5, 0.9], \sigma_i \sim iidU[0.8, 1.8], \zeta_{it} \sim iid\chi^2(3) - 3$ for both i and $t = \{-49, ...0, ..., T\}$,

$$\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho^2_{i\varepsilon})^{0.5}\zeta_{it},$$

where $\varepsilon_{i,-50} = 0$. (ii) ε_{it} follows stationary AR(1) process with cross-sectional dependence and heteroskedasticity across each i. $f^*_t \sim iidN(0, 1)$ is one additional factor, with loadings $\lambda^*_i \sim iid\chi^2(2) - 2$, which are uncorrelated with x_{it}. The errors

$$\varepsilon_{it} = \lambda^*_i f^*_t + \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho^2_{i\varepsilon})^{0.5}\zeta_{it}.$$

We replicate each experiments 500 replications and report the root mean squared errors ($RMSE$) and mean bias ($Bias$) of $\hat{\beta}_1$, under different combination of samples, such that $N = \{20, 50, 100, 200\}$ and $T = \{20, 50, 100, 200\}$. Let $v_{it,1} \sim iidN(0, 1)$ and $v_{it,2} \sim iid\chi^2(3) - 3$, the regressors x_{it} are driven by various combination of factor structure and $v_{it} = (v_{it,1}, v_{it,2})'$, stated below. In the iterative estimation by the approach of IFE, MLE and ATE, we allow for the maximum number of iterations reaches 2000, until $||\hat{\beta}^{s+1} - \hat{\beta}^s|| < 0.0001$,at the $s + 1$ step. In our simulations, we found the ATE has the fast converge rate of iterative estimation and MLE has the lowest converge rate of iterative estimation. The initial values of slope in all the iterative algorithm is the least squared estimator.

7.1 Data generating process of x_{it}

We consider various correlations between regressors x_{it} and factor structure stated below DGP1-DGP4, another DGP are show in the Appendix.

DGP 1(CLF): the regressors x_{it} are driven by linear combination of factors and its loadings,

$$x_{it,1} = 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + v_{it,1},$$
$$x_{it,2} = f_{t,1} + 2\lambda_{i,1} + 2f_{t,2} + 3\lambda_{i,2} + v_{it,2}.$$

DGP 2 (**CIF1**; Bai and Li, 2014): Let $\epsilon_{i1} \sim iidN(0.5, 1), \epsilon_{i2} \sim iidN(0.2, 1), \epsilon_{i3} \sim iidN(0, 1), \epsilon_{i4} \sim iidN(0.7, 1)$. We set three regressors and two factor, in which the regressors are driven by

$$x_{it} = \begin{pmatrix} x_{it,1} \\ x_{it,2} \\ x_{it,3} \end{pmatrix} = \begin{bmatrix} \lambda_{i1}, & \lambda_{i2}, \\ \lambda_{i1} + \epsilon_{i1}, & \lambda_{i2} + \epsilon_{i2}, \\ \lambda_{i1} + \epsilon_{i3}, & \lambda_{i2} + \epsilon_{i4}, \end{bmatrix} f_t + v_{it},$$

such that $\overline{\Gamma}' = \begin{bmatrix} 0.5 & 0.3 \\ 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ is of full row and full column rank.

DGP 3 (**CIF2**; Bai, 2009; Hsiao et al., 2021): the regressors x_{it} are driven by interactive effects and linear combination of factors and its loadings,

$$x_{it,1} = 1 + \lambda_i' f_t + f_{t,1} + \lambda_{i,1} + f_{t,2} + \lambda_{i,2} + v_{it,1},$$
$$x_{it,2} = 1 + \lambda_i' f_t + f_{t,1} + \lambda_{i,1} + f_{t,2} + \lambda_{i,2} + v_{it,2}.$$

DGP 4 (CLF, CIF1 with Additive effects): additive two-way fixed effects, where the x_{it} is same as that of DGP 1, DGP 2 and

$$y_{it} = x_{it}'\beta + \lambda_i' f_t + \alpha_i + \tau_t + \varepsilon_{it},$$

where $\alpha_i = (\bar{x}_{i,1} + \bar{x}_{i,2})/2 - \frac{1}{N}\sum_{i=1}^{N}((\bar{x}_{i,1} + \bar{x}_{i,2})/2) + \nu_i$ with $\nu_i \sim iidN(0, 1)$, and $\tau_t = (\bar{x}_{t,1} + \bar{x}_{t,2})/2 - \frac{1}{T}\sum_{t=1}^{T}((\bar{x}_{t,1} + \bar{x}_{t,2})/2) + \nu_t$ with $\nu_t \sim iidN(0, 1)$. As note by Bai (2009), while the additive fixed effects and interactive fixed effects both exist, the data should be handle firstly by the two-way transformation to eliminate the additive fixed effects. Then, in DGP 3, the approach of CCE, IFE, MLE and ATE are conducted on the two-way transformed data.

DGP 5 (Endogeneity): the generating of data is same as Design of Hong et al. (2023), exception that one-dimensional regressor x_{it} are generated as linearly correlated with f_t,λ_i and ε_{it},

$$x_{it} = 1 + 3f_{t,1} + \lambda_{i,1} + f_{t,2} + 2\lambda_{i,2} + v_{it},$$

or in the interactive format

$$x_{it} = 1 + f_t'\lambda_i + f_{t,1} + \lambda_{i,1} + f_{t,2} + \lambda_{i,2} + v_{it}.$$

Let $z_{it} = (z_{it,1}, z_{it,2})'$ be the instrumental variables. For the errors in the model,

$$\varepsilon_{it} = \epsilon_{it} + \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}.$$

The variables are generated as

$$\begin{pmatrix} v_{it} \\ z_{it,1} \\ z_{it,2} \\ \epsilon_{it} \end{pmatrix} \sim iidN\left(0_{4\times 1}, \begin{bmatrix} 1 & 0.5 & 0.5 & \eta \\ 0.5 & 1 & 0 & 0 \\ 0.5 & 0 & 1 & 0 \\ \eta & 0 & 0 & 1 \end{bmatrix}\right).$$

The endogeneity is depicted by the correlation efficient η between ϵ_{it} and v_{it}. We set $\eta = 0.5$ in our simulations.

7.2 Results

The results of the root of mean squared error (RMSE) and Bias of all experiments are attached in Appendix B and additional simulations are show in Appendix C. In the DGP 1, the regressors x_{it} is linear correlated with factors and loadings, making the CCE approach has larger RMSE and is inconsistent, show in Table 2. The RMSE of our MLS2 and MLS3 estimators are similar to that of ATE and is smaller than that of IFE and the MLE is most efficient estimator. From the view of Bias, our MLE has better performance in the finite sample as MLE, other approaches have relative bigger bias. Table 15 in the Appendix C reports the results of the DGP 1 with cross sectional dependence in the errors. The dependence errors are generated by one additional unobserved factor f_t^* with loading λ_i^*. The cross-sectional dependence can also be modelled by the spatial model. It shows that our MLE2 is robust. Table 16 in the Appendix C extend the simulation of DGP 1, in which $\hat{r} = 6$. Table 17 in the Appendix C extends the DGP 1 with the true factor number $r = 4$ and the estimated number of factors

$\hat{r} = 2$. They all show our MLS are robust and consistent, thus those results all verify Theorem 1.

Table 3 show the RMSE and Bias of DGP 2 with the interactive effects, which is similar to the setup of Bai and Li (2014). It show that the CCE approach has large RMSE and Bias, while the correlations between λ_i and Γ_i exists. Due to the correlations, the IFE has largest RMSE and Bias. Our MLS2 perform better that the CCE and a litter worse that MLE and ATE. If the errors have larger heteroskedasticity across individuals, our MLS has robust results than the MLE and ATE, as show in Table 4. In addition, we also consider the case of two regressors and one common factors in the Appendix C.

Table 5 and 6 show the RMSE and Bias of DGP 2 with the interactive effects, for the independent and identically distributed errors (iid) and auto-correlated errors respectively. For all the case, we select $\hat{r}' = 4$. In the case of independent and identically distribution, the IFE has the lowest RMSE and bias in the case of auto-correlated errors. Our MLS3 has robust results as the MLE and ATE.

Table 7 consider the estimation of interactive panel data model with additional additive effects[2]. As show in tables, the CCE approach can not allow for additive additive fixed effects, with large RMSE and Bias. Regardless of transformation, the RMSE and Bias of our MLS remains the same magnitude, as those of DGP 1. However, the IFE and MLE3 are consistent after transformation. Without transformation, the ATE is s consistent and the RMSE and Bias are all larger than those of MLS and after transformation, they are a litter larger than those of MLS3. **Table 8 consider the case of interactive form with additional additive effects, such that x_{it} are interactive correlated with $\lambda_i' f_t$ in DGP 2. We can conclude that the RMSE and Bias of our MLS remains the same magnitude as Table 3.**

Table 9 and 10 consider the case of endogeneity, for the linear and interactive formats between the factor structures and regressors. It shows that the RMSE and Bias of all the estimators all decrease as the sample increase. The RMSE of MLS2-CF, MLS3-CF, and TGMM has advantage in the smaller sample. For the large sample with $N = T = 200$, the RMSE of PGMM and TGMM is similar. In all, the performance of PGMM and TGMM is a litter better than our proposed estimator and the IV-CCE has the worst performance.

[2]As noted in Bai (2009), the IFE needs the two-way transformation to eliminate the additive effects firstly, show in Table 20 in the Appendix C.

Table 12 reports the sample size and power of the wild dependent bootstrap estimation, showing that the bootstrap test works well. Table 12 reports the sample size and power of T test, plugging the estimated standard deviation of β_1 by the wild dependent bootstrap procedure. The good performance implied the estimated standard deviation of β_1 is consistent. Last, Table 14 show the sample size and power of the robust T test. For the larger correlation of errors, it encounters more size distortion as show in Vogelsang (2012).

8 Empirical Analysis

In this section, we apply our approach to empirically investigate the output elasticity with respect to public infrastructure in an aggregate production function of China (Feng and Wu, 2018; Feng, 2020). we compare the estimation of MLS with the Mean Group (MG) estimates without considering unobserved factors, CCE mean group estimator (Pesaran, 2006), and IFE (Bai, 2009), and MLE (Bai and Li, 2014).

The empirical model comes from the aggregate production function,

$$g_{it} = \beta_0 + \beta_b b_{it} + \beta_k k_{it} + e_{it}.$$

where g_{it} denotes the logarithm of the gross domestic product (GDP) per labor of the province i at year t, and b_{it} is the logarithm of public infrastructure stock per labor, and k_{it} is the logarithm of non-infrastructure capital stock per labor. Thus, β_k and β_b are the estimated elasticizes of public infrastructure and non-infrastructure capital respectively. Let λ_i denotes the unobserved provincial effect and γ_t is the unobserved year's effect. For the panels with the two-way fixed effect, $e_{it} = \lambda_i + \gamma_t + \epsilon_{it}$ and for the interactive model, $e_{it} = \lambda_i' f_t + \epsilon_{it}$. To eliminate the nonstationality of data, the first order difference of model is done firstly, such as $\Delta g_{it} = g_{it} - g_{i,t-1}$, and similar for Δb_{it}, and Δk_{it}. Thus, the model becomes

$$\Delta g_{it} = \beta_0 + \beta_b \Delta b_{it} + \beta_k \Delta k_{it} + \Delta e_{it}.$$

The panel data set consists of the China's 30 provincial infrastructure investments over the period 1996–2015. This data set is collected from the website of National Bureau of Statistics of China, used in Feng and Wu (2018). The summary statistics of data refer to Table 1 in Feng (2020). The results by those methods are reported in Table 1. The contents of the second column correspond to our proposed estimators. From the results, we see that the estimated $\hat{\beta}_b$ and $\hat{\beta}_k$ by the MLS are only a lit different from the results of other methods.

Table 1: Output Elasticizes: Common Factors

Dependent Variable:						
	MG	MLS	IFE	MLE	CCEMG	ATE
Independent variables:	(1)	(2)	(3)	(4)	(5)	(6)
β_b	0.205***	0.164***	0.197***	0.193***	0.194***	0.193***
	(0.025)	-	(0.017)	(0.018)	(0.023)	-
	[0.029]	[0.021]	[0.021]	[0.025]	[0.021]	[0.033]
β_k	0.361***	0.372***	0.349***	0.354***	0.407***	0.407***
	(0.031)	-	(0.018)	(0.019)	(0.037)	-
	[0.039]	[0.024]	[0.020]	[0.014]	[0.039]	[0.042]
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	569	569	569	569	569	569
Overall R^2	0.65	-	-	0.67	0.72	-
Empirical features:						
slope heterogeneity	Yes	No	No	No	Yes	No
cross-sectional dependence	No	Yes	Yes	Yes	Yes	Yes

Note: (1) For the IFE and MLE, two factors are assumed in the estimation and the time effects are added. (2) the standard error of estimators are reported in parentheses and the wild Bootstrapping standard error of MLS are reported in brackets. Specifically, for the MG, the standard errors are adjusted for 30 clusters or provinces. Bai (2009) in Section 6 show that IFE is a result of least squares with the effects being estimated in the Section 6. Thus, the standard error of IFE is computed as the *regife* command of the Stata. The standard error of MLE is computed as the Remark 2.6 of Bai and Li (2014). The standard error of CCEMG is computed as equation (58) of Pesaran (2006). (3) The stars, *, ** and *** indicate the significance level at 10%, 5% and 1%, respectively.

9 Conclusion and Discussions

In this paper, we research the one-way and two-way Mundlak projection estimators of the panel data model with the interactive fixed effects in detail, allowing the linear and interactive correlation between the regressors and factor structure. In addition, we also combined the CF approach to allow for the case of endogeneity. Those estimators need not the iterative estimation procedure and have good theoretical and finite sample performances, compared with many other estimators.

The framework of the Mundlak projection estimation can be extended to others interactive effects panels, for example, the nonlinear panels with the interactive fixed effects, dynamic panels with the interactive fixed effects, the panel quantile regression with the interactive fixed effects. We will explore the estimation of those models by the

Mundlak projection estimation in the future.

References

[1] Ahn, S. C., Lee, Y. H., & Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2), 219-255.

[2] Ahn, S. C., Lee, Y. H., & Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174(1), 1-14.

[3] Bai, J. (2009). Panel data models with interactive fixed effects. *Econometric*a, 77(4), 1229-1279.

[4] Bai, J. (2024). Likelihood approach to dynamic panel models with interactive effects. *Journal of Econometrics*, 240(1), 105636.

[5] Bai, J., Choi, S. H., & Liao, Y. (2021). Feasible generalized least squares for panel data with cross-sectional and serial correlations. *Empirical Economics*, 60, 309-326.

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

[7] Bai, J., & Liao, Y. (2017). Inferences in panel data with interactive effects using large covariance matrices. *Journal of Econometrics*, 200(1), 59-78.

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

[9] Bai, J., & Ng, S. (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory*, 26(6), 1577-1606.

[10] Cao, Y., Jin, S., Lu, X., & Su, L. (2024). Oracle Efficient Estimation of Heterogeneous Dynamic Panel Data Models with Interactive Fixed Effects. *Journal of Business & Economic Statistics*, 1-16.

[11] Cui, G., Hayakawa, K., Nagata, S., & Yamagata, T. (2023). A Robust Approach to Heteroscedasticity, Error Serial Correlation and Slope Heterogeneity in Linear Models with Interactive Effects for Large Panel Data. *Journal of Business & Economic Statistics*, 41(3), 862-875.

[12] Cui, G., Norkutė, M., Sarafidis, V., & Yamagata, T. (2022). Two-stage instrumental variable estimation of linear panel data models with interactive effects. *The Econometrics Journal*, 25(2), 340-361.

[13] Feng, Q. (2020). Common factors and common breaks in panels: An empirical investigation. *Economics Letters*, 187, 108897.

[14] Feng, Q., & Wu, G. L. (2015). Bubble or riddle? An asset-pricing approach evaluation on China's housing market. *Economic Modelling*, 46, 376-383.

[15] Freeman, H., & Weidner, M. (2023). Linear panel regressions with two-way unobserved heterogeneity. *Journal of Econometrics*, 237(1), 105498.

[16] Gao, J., Peng, B., & Yan, Y. (2023). Higher-order Expansions and Inference for Panel Data Models. *Journal of the American Statistical Association*, 1-12.

[17] Harding, M., & Lamarche, C. (2011). Least squares estimation of a panel data model with multi-factor error structure and endogenous covariates. *Economics Letters*, 111(3), 197-199.

[18] Hayakawa, K. (2013). GMM estimation of short dynamic panel data models with interactive fixed effects. *Journal of the Japan Statistical Society*, 42(2), 109-123.

[19] Holtz-Eakin, D., Newey, W., & Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 1371-1395.

[20] Hong, S., Su, L., & Jiang, T. (2023). Profile GMM estimation of panel data models with interactive fixed effects. *Journal of Econometrics*, 235(2), 927-948.

[21] Hsiao, C. (2018). Panel models with interactive effects. *Journal of Econometrics*, 206(2), 645-673.

[22] Hsiao, C., Shi, Z., & Zhou, Q. (2022). Transformed Estimation for Panel Interactive Effects Models. *Journal of Business & Economic Statistics*, 40(4), 1831-1848.

[23] Hsiao, C., Zhou, Q., & Kong, J. (2023). Panel Interactive Effects Models with Endogenous Regressors: A Transformed GMM Estimator. *Working Paper.*

[24] Jiang, B., Yang, Y., Gao, J., & Hsiao, C. (2021). Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics*, 224(2), 439-465.

[25] Juodis, A., and Sarafidis, V. (2018). Fixed T Dynamic Panel Data Estimators With Multifactor Errors. *Econometric Reviews*, 37, 893–929.

[26] Juodis, A., and Sarafidis, V. (2022). A Linear Estimator for Factor- Augmented Fixed-T Panels With Endogenous Regressors, *Journal of Business & Economic Statistics*, 40:1, 1-15,

[27] Juodis, A., Karabiyik, H., & Westerlund, J. (2021). On the robustness of the pooled CCE estimator. *Journal of Econometrics*, 220(2), 325-348.

[28] Karabiyik, H., Reese, S., & Westerlund, J. (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics*, 197(1), 60-64.

[29] Karabiyik, H., Urbain, J. P., & Westerlund, J. (2019). CCE estimation of factor augmented regression models with more factors than observable. *Journal of Applied Econometrics*, 34(2), 268-284.

[30] Keilbar, G., Rodriguez-Poo, J. M., Soberon, A., & Wang, W. (2022). A semi-parametric approach for interactive fixed effects panel data models. *arXiv e-prints, arXiv-2201.*

[31] Kiefer, N. M., & Vogelsang, T. J. (2002). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, 70(5), 2093-2095.

[32] Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(6), 1130-1164.

[33] Kiefer, N. M., Vogelsang, T. J., & Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, 68(3), 695-714.

[34] Kneip, A., Sickles, R. C., & Song, W. (2012). A new panel data treatment for heterogeneity in time trends. *Econometric theory*, 28(3), 590-628.

[35] Moon, H. R., & Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4), 1543-1579.

[36] Moon, H. R., & Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1), 158-195.

[37] Moon, H. R., & Weidner, M. (2019). Nuclear norm regularized estimation of panel regression models. *arXiv preprint arXiv:1810.10987.*

[38] Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 69-85.

[39] Norkutė, M., Sarafidis, V., Yamagata, T., & Cui, G. (2021). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multi-factor error structure. *Journal of Econometrics*, 220(2), 416-446.

[40] Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multi-factor error structure. *Econometrica*, 74(4), 967-1012.

[41] Phillips, P. C. B., & Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal*, 6(1), 217-259.

[42] Sarafidis, V., & Robertson, D. (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *The Econometrics Journal*, 12(1), 62-81.

[43] Sarafidis, V., & Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews*, 31(5), 483-531.

[44] Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460), 1167-1179.

[45] Robertson, D., & Sarafidis, V. (2015). IV estimation of panels with factor residuals. *Journal of Econometrics*, 185(2), 526-541.

[46] Westerlund, J. (2019). On estimation and inference in heterogeneous panel regressions with interactive effects. *Journal of Time Series Analysis*, 40(5), 852-857.

[47] Westerlund, J., & Urbain, J. P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics*, 185(2), 372-377.

[48] Vogelsang, T. J. (2012). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics*, 166(2), 303-319.

[49] Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2), 420-445.

[50] Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.