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

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

February 1, 2024

Abstract

In this paper, we propose a simple estimation method for panel data models with common factors by generalizing Mundlak projection to the form of interactive fixed effects. It can also be considered as an generalization of Pesaran's CCE approach. We further consider the case of endogenous regressors by combining the Mundlak projection with the control function approach. The advantage of our estimation procedures is that no iterations are needed. In addition, 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 fixed effects panel data models. Monte Carlo simulations are conducted to verify our theoretical results and compare our approach with existing estimators in finite samples. Finally, the proposed method is illustrated in an empirical example of the productivity effect of infrastructure investment.

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 Jiao Tong University(2023), CUHK-Shenzhen(2023), CUHK(2023), the 2023 AMES Singapore. Constructive comments from the audiences at these conferences are highly acknowledged.

†Economics Division, Nanyang Technological University, Singapore. Email: qfeng@ntu.edu.sg.

‡Zhongtai Securities Institute for Financial Studies, Shandong University, China. Email: wangwei_0115@outlook.com.

§Economics Division, Nanyang Technological University, Singapore. Email: wang.wj@ntu.edu.sg.

¶School of Economics and Finance, South China University of Technology. Email: 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. Morkut81c8, 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 (2013) 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 (2013), 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. Morkut81c8 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

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-sectionally correlated, modelled by a multi-factor structure,

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

where f_t is a $r \times 1$ vector of unobservable common factors with $r \times 1$ vector of factor loadings λ_i, in which r is finite and ε_{it} is the idiosyncratic error.

Although the common factors f_t and factor loadings λ_i are strictly exogenous with respect to ε_{it} under the commonly used assumptions, the regressors x_{it} could be correlated with e_{it}, by 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 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.

2.1 The One-way Mundlak Estimator

For time t, let $\overline{x}_{\cdot t} = \frac{1}{N}\sum_{i=1}^N x_{it} = (\frac{1}{N}\sum_{i=1}^N x_{it,1}, \cdots, \frac{1}{N}\sum_{i=1}^N x_{it,p})'$ denote the cross-sectional average of the regressors. If λ_i is uncorrelated with x_{it}, to control the correlation between x_{it} and f_t, we can project the common factor f_t onto the space of the cross-sectional average $\overline{x}_{\cdot t}$ as in the Mundlak's approach (Mundlak, 1978), i.e.,

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

where B is the coefficient matrix, d is the intercept, 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 $\overline{x}_{\cdot t}$. 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 is still affected by the idiosyncratic error ξ_t even as $N \to \infty$ (i.e., the variance of ξ_t is fixed).

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

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

where $u_{it} = \lambda_i' d + \lambda_i' \xi_t + \varepsilon_{it}$. We note under the assumption that λ_i is uncorrelated with x_{it}, $\lambda_i'(d + \xi_t)$ 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 partitioned regression.

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 Mundlak least squared estimator is defined as

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

As the composite error $u_{it} = \lambda_i' d + \lambda_i' \xi_t + \varepsilon_{it}$ has serial correlation and cross-sectional dependence, a GLS estimator may more efficient. What's more important, if λ_i is in fact correlated with x_{it}, the above one-way Mundlak estimator $\widehat{\beta}_{M1}$ is not consistent. Under such a more general setting, we propose the following two-way Mundlak projection estimation procedure.

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

To control the correlation between x_{it} and λ_i, we further project λ_i onto the space of time average of x_{it}. For individual i, denote the time average of regressors as $\overline{x}_{i\cdot} = \frac{1}{T}\sum_{i=1}^{T} x_{it}$, then we have

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

where, A is a constant coefficient matrix, c is the intercept, and η_i is projection error. As noted in Bai (2009, 2013), when λ_i is correlated with the regressors, it can be projected onto the regressors, similar to Keilbara et al. (2023), such that $\lambda_i = g(\overline{x}_{i\cdot}) + \eta_i$ with $g(\overline{x}_{i\cdot})$ is a $K \times 1$ vector of unknown functions, leading to a semi-parametric factor model.

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

$$\begin{aligned} y_{it} &= \beta_0 + x_{it}'\beta + \overline{x}_{\cdot t}'B'\eta_i + \overline{x}_{\cdot t}'B'c + \overline{x}_{i\cdot}'A'\xi_t + \overline{x}_{i\cdot}'A'd \\ &\quad + c'\xi_t + \eta_i'd + \overline{x}_{i\cdot}'A'B\overline{x}_{\cdot t} + c'd + \eta_i'\xi_t + \varepsilon_{it} \\ &= (\beta_0 + c'd) + x_{it}'\beta + \overline{x}_{\cdot t}'\rho_i + \overline{x}_{i\cdot}'\delta_t + \overline{x}_{i\cdot}'A'B\overline{x}_{\cdot t} + u_{it}, \end{aligned} \tag{8}$$

where $\underset{p\times 1}{\rho_i} = B'\eta_i + B'c$, $\underset{p\times 1}{\delta_t} = A'\xi_t + A'd$, $u_{it} = c'\xi_t + \eta_i'd + \eta_i'\xi_t + \varepsilon_{it}$.

Similar to those in Section 2.1, we implement the following steps to remove the nuisance parameters and obtain a direct estimation of the parameter of interest β. Stacking time t, the regression (8) can be written as

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

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

Pre-multiple the equation (9) by $M_{\bar{X}} = I_T - \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'$,

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

Stacking individual i, 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 $\underset{T\times N}{Y} = (Y_1, Y_2, ..., Y_N)$, $\underset{T\times N}{X^j}$ are $T\times N$ matrix being the j^{th} regressor matrix associated with parameter β_j, $\underset{N\times p}{\underline{X}} = (\overline{x}_{1\cdot}, \overline{x}_{2\cdot}, ..., \overline{x}_{N\cdot})'$, $\underset{T\times N}{U} = (u_1, u_2, ..., u_N) = \xi(l_N'\otimes c) + (l_T\otimes d')\eta' + \xi\eta' + \varepsilon$.

Post-multiple the equation by $M_{\underline{X}} = I_N - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$,

$$M_{\bar{X}}YM_{\underline{X}} = \beta_1\cdot M_{\bar{X}}X^1M_{\underline{X}} + \beta_2\cdot M_{\bar{X}}X^2M_{\underline{X}} + ... + \beta_p\cdot M_{\bar{X}}X^pM_{\underline{X}} + M_{\bar{X}}UM_{\underline{X}} \tag{10}$$

Furthermore, 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 reparameterized by

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

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

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

Remark 1: *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.,

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

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

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

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

2.3 Asymptotic Normality

To show the asymptotic properties of the two-way Mundlak type estimator $\widehat{\beta}_{M2}$, we give the following assumptions.

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 2: *(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. When the matrices* *of A and B are not full column rank, we could use parts of regressors to satisfy the full column rank of the matrices of A and B. (2) Assumption 2 directly set the identification condition on the transformed regressors $\tilde{x}_{it}$, which is common in the literature. (3) Since $u_{it} = c'\xi_t + \eta_i' d + \eta_i'\xi_t + \varepsilon_{it}$, 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 $u_{it} = c'\xi_t + \eta_i' d + \eta_i'\xi_t + \varepsilon_{it}$, 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 $u_{it} = \eta_i' d + c'\xi_t + \eta_i'\xi_t + \varepsilon_{it}$.

Theorem 1 states the asymptotic distribution of $\widehat{\beta}_{M2}$ as $(T, N) \to \infty$. 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 Assumption that $N^{-1/2}\sum_{i=1}^{N} \mathbb{X}_{ti} u_{it} \xrightarrow{d} N(0, \Phi_N)$, thus, as $N \to \infty$,

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

with $\Psi_N = \underset{N\to\infty}{plim} \frac{1}{N}\tilde{X}'\tilde{X}$ and $V_{\beta 2} = \Psi_N^{-1}\Phi_N\Psi_N^{-1}$. The case of fixed N and $T \to \infty$ can be derived similarly.

The consistent estimation of covariance matrix V_β is difficult by the traditional panel HAC estimation, since the serial correlation and cross-sectional dependence both exist in u_{it}. Thus, the traditional Newey-West type variance estimation is not feasible. This circumstance is similar to the estimation of D_Z in Bai (2009). He suggested using the partial sample method together with the Newey–West procedure. In addition, Bai and Liao (2017), Bai, Choi and Liao (2021) applied the eigenvalue decomposition or impose the sparsity to estimate V_β.

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.

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

In the cases where x_{it} is correlated with the factor structure $\lambda_i' f_t$, the approach described in Section 2 is not applicable. Specifically, under the model

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

if $\lambda_i' f_t$ is correlated with x_{it}, then $\eta_i'\xi_t$ in the composite errors u_{it} may be correlated with x_{it}, resulting in the inconsistency of the MLS estimators of β in Section 2

Next, we propose a valid approach under such a general setup by combining the two-way Mundlak projection method and defactored regressors.

Remark 3: *In addition, the averaging method in p.647 of Hsiao (2018) and other one-way or two-way transformed estimation are also not applicable. 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*

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

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*

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

Thus, the ordinary least square estimation using time averaged data (17) is inconsistent, since $\beta + A'd + A'B\overline{x}_{\cdot\cdot}$ *instead of* β*, due to 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, since* $\beta + B'c + B'A\overline{x}_{\cdot\cdot}$*, due to terms* $\overline{x}'_{\cdot t}\rho_i + \overline{x}'_{i\cdot}A'B\overline{x}_{\cdot t}$ *in equation (8). It's different from the case in equations (2.4) and (2.5) of Hsiao (2018) on p.647.*

Let $m(\cdot,\cdot)$ denotes an unknown real-valued function, depends on the two way fixed effects α_i and γ_t. Freeman and Weidner (2023) considered the following model,

$$
y_{it} = x'_{it}\beta + m(\alpha_i, \gamma_t) + \varepsilon_{it}
$$

There exists, under weak regularity conditions, the singular value decomposition of a function $m(\cdot,\cdot)$,

$$
m(\alpha_i, \gamma_t) = \sum_{r=1}^{\infty} \sigma_r \varphi_y(\alpha_i)\psi_y(\gamma_t),
$$

where σ_r are some functional singular values and $\varphi_y(\alpha_i)$, $\psi_y(\gamma_t)$ are appropriate normalized functions. Let $\lambda_i = \sigma_r\varphi_y(\alpha_i)$, and $f_t = \psi_y(\gamma_t)$. Thus, the model can be rewritten as the interactive fixed models with an truncated finite sequence of factors. Similar to the relationship between the regressor and the factor structure, let $x_{it,j} = \sum_{s=1}^{r'} \sigma_s\varphi_x(\alpha_i)\psi_x(\gamma_t) + h_{it,j}$,where $h_{it,j}$ is the idiosyncratic errors for each regressor $x_{it,j}$,$j = \{1, \cdots, p\}$. Thus, it can be represented as the interactive form

$$x_{it} = \Gamma_i' g_t + h_{it}, \tag{19}$$

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, h_{it} is called de-factored version of regressors x_{it} in the literature, such as Cui et al., (2022), Cui et al. (2023), Cao et al. (2023) and others. Thus, it can be regarded as the extension of the setup in Keilbara et al. (2023). Intuitively, if x_{it} is correlated with the full common factors f_t, the space of g_t in model (19) will cover the space of f_t, if $r' \geq r$. Although in the case of x_{it} is affected by f_1, not f_2,with $f_t = (f_{t1}', f_{t2}')'$, the space of g_t in model (19) will cover the related space of f_{t1}, if $r' \geq r$. Since f_{t2} is not correlated with the regressor x_{it},the estimating of the space of g_t is enough. We give the following Assumptions 4, 5 and 6:

Assumption 4 *$E||g_t||^4 \leq 0$ and $\frac{1}{T}\sum_{t=1}^{T} g_t g_t'$ converges to an $r' \times r'$ positive definite matrix.*

Assumption 5 *(i) For each $i = \{1, \cdots, N\}$, the $r' \times p$ dimensional random loadings $\Gamma_i = \bar{\Gamma} + \kappa_i$, $\kappa_i \sim IID(0, \Omega_\xi)$, $i = 1, ...N$, where Γ are non-zero and fixed and the variances Ω_η, Ω_ξ are finite.(ii) $E||\kappa_i||^4 \leq 0$ and $\frac{1}{N}\sum_{i=1}^{N} \Gamma_i \Gamma_i'$ converges to an $r' \times r'$ positive definite matrix.*

Assumption 6 *(i) For $j = \{1, \cdots, p\}$, $\tilde{h}_{it,j}$ are mutual independent and $\tilde{h}_{it,j}$ is independent of (ε_{it},f_t^*, λ_i^*, λ_i, f_t); (ii) $E(h_{it,j}) = 0$, $E||h_{it,j}||^8 < \infty$;(iii) $E(h_{\cdot t,j}' h_{\cdot s,j}/N) = \gamma_{N,j}(t,s)$,$|\gamma_{N,j}(t,s)| < \infty$ and $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} |\gamma_{N,j}(t,s)| < \infty$; (iv) $E(h_{it,j}' h_{i't,j}/N) = \tau_{t,j}(i,i')$,$|\tau_{t,j}(i,i')| < \infty$ and $\frac{1}{N}\sum_{i=1}^{N}\sum_{i'=1}^{N} |\tau_{t,j}(i,i')| < \infty$;(v) $E(h_{it,j}' h_{i'ts,j}/N) = \tau_{ts,j}(i,i')$, and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |\tau_{ts,j}(i,i')| < \infty$; (vi) For every (t,s), $E|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[h_{it,j}h_{is,j} - E(h_{it,j}h_{is,j})]|^4 < \infty$; (vii) the matrix $\frac{1}{NT}\tilde{H}'\tilde{H}$ converges to non-singular matrix, as $(T,N) \to \infty$.*

Assumptions 4, 5 and 6 are similarly borrowed from Bai and Ng (2002) and Bai (2009), exception that we setup the random factor loadings Γ_i, in Assumtion 5 (i). They ensure the principal component analysis works for equation (19) under $(T,N) \to \infty$, and then $\hat{h}_{it} \xrightarrow{p} h_{it}$, $\overline{h}_{\cdot t} \xrightarrow{p} 0$ and $\overline{h}_{i\cdot} \xrightarrow{p} 0$.

Plugging equations (19) into the model (1), the equation (9) becomes

$$Y_i = H_i\beta + G\Gamma_i\beta + \bar{X}\rho_i + \delta\overline{x}_{i\cdot} + \bar{X}B'A\overline{x}_{i\cdot} + (\beta_0 + c'd) \cdot l_T + u_i \tag{20}$$

where $H_i = (h_{i1}, \cdots, h_{iT})'$, and $G = (g_1, \cdots, g_T)'$. Pre-multipling the equation (20) by $M_{\bar{X}}$ gives

$$M_{\bar{X}} Y_i = M_{\bar{X}} H_i \beta + M_{\bar{X}} G \Gamma_i \beta + M_{\bar{X}} \delta \overline{x}_{i\cdot} + M_{\bar{X}} u_i. \tag{21}$$

If $r' = p$, Γ_i is full rank matrix and then the rank condtion is satisfied as the CCE appraoch (Pesaran, 2006). According to auxiliary factor model, $\bar{x}_{\cdot t} = \bar{\Gamma}' g_t + \overline{h}_{\cdot t}$, 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, $(\bar{\Gamma}\bar{\Gamma}')^{-1}\bar{\Gamma}'\bar{x}_{\cdot t} \xrightarrow{p} g_t$, and then $M_{\bar{X}} G \xrightarrow{p} 0_{T\times r'}$,as $N \to \infty$. However, $r' > p$ in general [1], such that the rank condition is not satisfied as the case considered in Karabiyik, Reese and Westerlund (2017), Karabiyik, Urbain and Westerlund (2019), Juodis, Karabiyik and Westerlund (2021) and others. While the rank condition is not satisfied, as $N \to \infty$, $M_{\bar{X}} \xrightarrow{p} M_{G\bar{\Gamma}} = I_T - G\bar{\Gamma}(\bar{\Gamma}'G'G\bar{\Gamma})^{\dagger}\bar{\Gamma}'G'$,with $(\bar{\Gamma}'G'G\bar{\Gamma})^{\dagger}$ is the Moore-Penrose inverse of singular matrix $\bar{\Gamma}'G'G\bar{\Gamma}$. Thus, for the second term of the right side of equation (21), $M_{\bar{X}} G \Gamma_i \beta \xrightarrow{p} M_{G\bar{\Gamma}} G\bar{\Gamma}\beta + M_{G\bar{\Gamma}} G\kappa_i\beta = M_{G\bar{\Gamma}} G\kappa_i\beta$, due to Assumtion 5 (i) and the fact that $M_{G\bar{\Gamma}} G\bar{\Gamma} = 0_{T\times p}$. In addition, we can also follow the framework of mean factors in Juodis, Karabiyik and Westerlund (2021) to analyze it.

Thus, equation (21) asymptotically equals

$$M_{\bar{X}} Y_i = M_{\bar{X}} H_i \beta + M_{\bar{X}} \delta \overline{x}_{i\cdot} + M_{G\bar{\Gamma}} G\kappa_i \beta + M_{\bar{X}} u_i. \tag{22}$$

Furthermore, stacking individual i, rewrite above model as

$$M_{\bar{X}} Y = \beta_1 M_{\bar{X}} H^1 + \cdots + \beta_p M_{\bar{X}} H^p + M_{\bar{X}} \delta \underline{X} + M_{G\bar{\Gamma}} G[\kappa_1\beta, \cdots, \kappa_N\beta] + M_{\bar{X}} U,$$

where $\underset{T\times N}{H^j}$ are $T \times N$ matrix being the j^{th} regressor matrix associated with parameter β_j.

Post-multiple the equation by $M_{\underline{X}}$,the model becomes

$$M_{\bar{X}} Y M_{\underline{X}} = \beta_1 M_{\bar{X}} H^1 M_{\underline{X}} + \cdots + \beta_p M_{\bar{X}} H^p M_{\underline{X}} + M_{G\bar{\Gamma}} G[\kappa_1\beta, \cdots, \kappa_N\beta] M_{\underline{X}} + M_{\bar{X}} U M_{\underline{X}} \tag{23}$$

Last, we collect all transformed regressors for individual i at period t, $\underset{NT\times p}{\tilde{H}} = [vec(M_{\bar{X}} H^1 M_{\underline{X}}), \cdots, vec(M_{\bar{X}} H^p M_{\underline{X}})]$, and $\tilde{U}^* = vec(M_{G\bar{\Gamma}} G[\kappa_1\beta, \cdots, \kappa_N\beta] M_{\underline{X}} + M_{\bar{X}} U M_{\underline{X}})$.

[1] Specifically, for each regressor $j = \{1, \cdots, p\}$, $x_{it,j} = \Gamma'_{i,j} g_{t,j} + h_{it,j}$, with the $s\times 1$ dimensional unobserved common factors $g_{t,j}$ and factor loadings $\Gamma_{i,j}$. Stacking the regressors $x_{it} = (x_{it,1}, \cdots, x_{it,p})'$,gives $x_{it} = \Gamma'_i g_t + h_{it}$, with $\underset{ps\times p}{\Gamma_i} = diag(\Gamma_{i,1}, \cdots, \Gamma_{i,p})$ and $\underset{ps\times 1}{g_t} = (g'_{t,1}, \cdots, g'_{t,p})'$. If $s = 1$, Γ_i is full rank matrix under Assumtion 5. If $s > 1$,we obtain $ps > p$ and then Γ_i is not full row rank matrix. In general, according to equation (3), we allow the case of $r > p$ and the space of g_t covers the space of f_t, thus, $ps > p$ in general.

Thus, equation (23) can be further reparameterized by

$$\tilde{Y} = \tilde{H}\beta + \tilde{U}^*, \tag{24}$$

and then the Mundlak least squared estimator is defined as

$$\widehat{\beta}_{M3} = (\tilde{H}'\tilde{H})^{-1}\tilde{H}'\tilde{Y}. \tag{25}$$

Let $\tilde{h}_{it}$ and u_{it}^* denote the $(i-1)T+t$ row of $\tilde{H}$ and $\tilde{U}^*$ respectively. To show the asymptotic properties of $\widehat{\beta}_{M3}$, we give another Assumption 4-7.

Assumption 7 *(i) u_{it}^* is independent of $\tilde{h}_{it}$; (ii) Given $\tilde{h}_{it}$, 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, $E(\theta_t|\tilde{h}_{it}) = 0_N$, and $g(\cdot)$ is a measurable function. In addition, let $\bar{U}_t^\star = \frac{1}{\sqrt{N}}U_t^{\star\prime}1_N$, 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 $\delta_{t,\kappa}^U = ||\bar{U}_t^* - \bar{U}_t^\star||_\kappa$.*

Assumption 7 is similar to Assumption 3 in Section 2, ensuring that the central limit theorem exists. Under the above assumptions, we obtain Proposition 1.

Proposition 1 *Under Assumptions 1, 4-6, 7, as $(T, N) \to \infty$, then*

$$\sqrt{NT}(\widehat{\beta}_{M3} - \beta) \xrightarrow{d} N(0, V_{\beta 3}).$$

where $V_{\beta 3} = \Psi_{NT}^{*-1}\Psi_{NT}^*\Psi_{NT}^{*-1}$ with $\Psi_{NT}^* = \underset{(N,T)\to\infty}{plim} \frac{1}{NT}\tilde{H}'\tilde{H}$ and

$$\Psi_{NT}^* = \underset{(T,N)\to\infty}{plim} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\sum_{j=1}^N\sum_{s=1}^T E(u_{it}^* u_{js}^{*\prime})\tilde{h}_{it}\tilde{h}_{js}'.$$

Since h_{it} is not observed in practice, the principal component analysis for each regressors gives the estimated $\hat{r}' \times 1$ dimensional factor $\hat{g}_{t,j}$, $\hat{r}' \times 1$ dimensional loadings $\hat{\Gamma}_{i,j}$, and $\hat{h}_{it,j}$. Let $\hat{h}_{it} = (\hat{h}_{it,1} \cdots, \hat{h}_{it,p})'$, and $\underset{T\times N}{\hat{H}^j}$ are $T\times N$ matrix being the j^{th} regressor matrix in model (24). The information criterion approach of Bai and Ng (2002) and other methods are applied to determine $\hat{r}$ and $\hat{r}'$. In practice, we use the $\max(\hat{r}', r)$ in the principal component analysis of equation (19) or the estimated number of factors using the principal component analysis of y_{it}. Last, since the principal component analysis works in the approximate factor model under the case of N is fixed and $T \to \infty$, or the case of T is fixed and $N \to \infty$, the above estimation adapts to those cases, if and only if h_{it} is consistent estimated.

In the CCE's approach, due to the existing of equation (19), the rank condition is important for the analysis of the asymptotic distribution . As show in Karabiyik, Reese and Westerlund (2017), Juodis, Karabiyik and Westerlund (2021), the asymptotic distribution of the pooled CCE estimator (CCEP) is not normal distribution in the case of $r' > p$ or $r' < p$. In our Mundlak projection method, the estimators are still asymptotically normal distribution when the rank condition is not satisfied.

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. Similar to the Section 3, the defactored regressors $h_{it} = (h'_{1it}, h'_{2it})'$, $cov(h_{1it}, \varepsilon_{it}) = 0$, and h_{2it} are correlated with ε_{it}, such that $cov(h_{2it}, \varepsilon_{it}) \neq 0$.

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,

$$x_{2it} = \alpha' z_{it} + q_{it}. \tag{26}$$

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 (26)

$$\varepsilon_{it} = q'_{it} \underset{p_2 \times 1}{\pi} + \varpi_{it}. \tag{27}$$

Let $\omega_{it} = c'\xi_t + \eta'_i d + \eta'_i \xi_t + \varpi_{it}$, $\omega_{i\cdot} = (\omega_{i1}, \omega_{i2}, ..., \omega_{iT})'$, 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 $[q^j_{it}]^{T,N}_{t=1,i=1}$, associated with parameter π_j for $j = \{1, 2, \cdots, p_2\}$. Plugging equation (27) into equation (10) gives

$$M_{\bar{X}} Y M_{\underline{X}} = \textstyle\sum_{j=1}^{p} \beta_j \cdot M_{\bar{X}} X^j M_{\underline{X}} + \sum_{j=1}^{p_2} \pi_j \cdot M_{\bar{X}} Q^j M_{\underline{X}} + M_{\bar{X}} \omega M_{\underline{X}} \tag{28}$$

Compared with equation (10), controlling for the correlation between x_{it} and factor structure, equation (28) 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 (28) is consistent. We only interest in the slope β and then partial out nuisance parameters.

Furthermore, we vectorize the variables in equation (28) 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 (28) can be further reparameterized by

$$\tilde{Y} = \tilde{X}\beta + \mathbb{Q}\pi + \tilde{\omega}. \tag{29}$$

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 (26) to obtain residuals

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

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

$$\hat{\beta}_{M4} = (\tilde{X}'M_{\hat{\mathbb{Q}}}\tilde{X})^{-1}\tilde{X}'M_{\hat{\mathbb{Q}}}\tilde{Y}. \tag{30}$$

4.2 Asymptotic Properties

Assumption 8 *(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 9 *(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 10 *(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, 8, 9 and 10, 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 2 *Under Assumptions 1, 2, and 8-10, 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

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 11 *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^{0}_{a}: $\beta_j=\beta^{0}_{j}, j\in\{1,\cdots,p\}$, with scalar β^{0}_{j}, against the alternative hypothesis $H^{1}_{a}:\beta\neq\beta_0$. Traditional T test is adaptive, such as $T_{M2,j}=V^{-1/2}_{\beta,jj}(\widehat{\beta}_{M2,j}-\beta^{0}_{j})$ 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 $T^{*}_{M2,j}=(\hat{V}^{boot}_{M2,jj})^{-1/2}(\widehat{\beta}^{*}_{M2,j}(b)-\widehat{\beta}_{M2,j})$ with $\hat{V}^{boot}_{M2,jj}$ is the j^{th}diagonal element of $\hat{V}^{boot}_{M2}$ and its p-value $p^{*}_{T}=\frac{1}{B}\sum_{b=1}^{B}1\{|\widehat{\beta}^{*}_{M2,j}(b)-\widehat{\beta}_{M2,j}|>|\widehat{\beta}_{M2,j}-\beta^{0}_{j}|\}$.

For the multivariate null hypothesis $H^{0}_{b}:\underset{l\times p}{R_1}\beta=\theta_1$, with θ_1 being an $l\times 1$ constant vector, against the alternative hypothesis $H^{1}_{b}:R_1\beta\neq\theta_1$, the Wald-type statistics is $W_{M2}=NT(R_1\beta-\theta_1)'(R_1V_{\beta 2}R'_1)^{-1}(R_1\beta-\theta_1)'$ and its asymptotic distribution is χ^{2}_{p}, if $V_{\beta 2}$ is available. Since $V_{\beta 2}$ is infeasible, instead of $\hat{V}^{boot}_{M2}$, the bootstrap Wald test is applied, $W^{*}_{M2}=(\widehat{\beta}^{*}_{M2}-\widehat{\beta}_{M2})'R'_1(R_1\hat{V}^{boot}_{M2}R'_1)^{-1}R_1(\widehat{\beta}^{*}_{M2}-\widehat{\beta}_{M2})$.

We give another Assumption 12, same as Assumption 3 in Gao et al. (2023), let $\lfloor q\rfloor$ denotes the largest integer not larger than q,

Assumption 12 *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 2.

Theorem 2 *Under Assumptions 1-3, 11, 12, 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 $\hat{\hat{\nu}}_t = \sum_{i=1}^{N} \tilde{x}_{it}\hat{\hat{u}}_{it}$ and compute a HAC estimator as

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

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

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

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,

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

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 13 *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 14 *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 15 *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 3 *Under Assumptions 1, 13-15, 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 $t = \{-49, ...0, ..., T\}$,

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

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

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

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 (CIF; 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 (CIF; 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 (CIF with Additive effects): additive two-way fixed effects, where the x_{it} is same as that of DGP 1 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},

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

or in the interactive format

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

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

$$\begin{pmatrix} v_{it} \\ z_{it,1} \\ z_{it,2} \\ \epsilon_{it} \end{pmatrix} \sim iidN\left(0_{4\times1}, \begin{bmatrix} 1 & 0.5 & 0.5 & \eta \\ 0.5 & 1 & 0 & 0 \\ 0.5 & 0 & 1 & 0 \\ \eta & 0 & 0 & 1 \end{bmatrix}\right).$$

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 13 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 14 in the Appendix C extend the simulation of DGP 1, in which $\hat{r} = 6$. Table 15 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 acorss 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 autocorrelated errors respectively. For all the case, we select $\hat{r}' = 4$. In the case of iid, the IFE has the lowest RMSE and bias in the case of autocorrelated 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.

[2]As noted in Bai (2009), the IFE needs the two-way transformation to eliminate the additive effects firstly, show in Table 18 in the Appendix C.

Table 8 and 9 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.

Table 10 reports the sample size and power of the wild dependent bootstrap estimation, showing that the bootstrap test works well. Table 10 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 12 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 Example

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 estimated results with the two-way fixed effect estimator, Mean Group estimators, 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 China92256a9a 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 MOLS are only a lit different from the results of other methods. However, the estimated variance by the DWB procedures are obvious larger than those of other methods.

Table 1: Output Elasticities: Common Factors

Dependent Variable:					
	MG	MOLS	IFE	MLE	CCEMG
Independent variables:	(1)	(2)	(3)	(4)	(5)
β_b	0.205***	0.164	0.197***	0.193***	0.194***
	(0.025)	[0.118]	(0.017)	(0.018)	(0.023)
β_k	0.361***	0.372***	0.349***	0.354***	0.407***
	(0.031)	[0.153]	(0.018)	(0.019)	(0.037)
Year effects	Yes	Yes	Yes	Yes	Yes
No. of observations	569	569	569	569	569
Overall R^2	0.65			0.67	0.72
Empirical features:					
slope heterogeneity	Yes	No	No	No	Yes
cross-sectional dependence	No	Yes	Yes	Yes	Yes

Note: (1) Standard errors are reported in parentheses. (2) Bootstrapping standard errors are reported in brackets. (3) The stars, *, ** and *** indicate the significance level at 10%, 5% and 1%, respectively. (4) Standard errors are adjusted for 30 clusters for column 1.

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. In addition, we will relax the rank condition in the Section 3 to justify the consistency of the estimator.

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. (2013). Likelihood approach to dynamic panel models with interactive effects. *Available at SSRN 2332992.*

[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] Cao, Y., Jin, S., Lu, X., & Su, L. (2023). Oracle Efficient Estimation of Heterogeneous Dynamic Panel Data Models with Interactive Fixed Effects, *Working paper.*

[10] 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.

[11] 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.

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

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

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

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

[16] 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.

[17] 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.

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

[19] 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.

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

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

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

[23] 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.

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

[25] 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,

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

[27] 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.

[28] 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.

[29] 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.*

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

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

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

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

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

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

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

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

[38] 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.

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

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

[41] 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.

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

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

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

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

[46] 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.

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

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

A The Tables of Simulations and Empirical Analysis

A.1 The RMSE and Bias

A.2 Bootstrap and Robust Tests

Table 2: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with f_t and λ_i in DGP 1

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	14.98	8.92	6.61	4.30	-0.69	0.98	0.13	-0.14
	50	9.02	5.51	4.00	2.72	-0.35	0.14	0.13	-0.13
	100	6.64	4.00	2.78	1.92	0.10	-0.22	-0.11	-0.05
	200	4.67	2.82	1.83	1.33	-0.18	-0.08	-0.06	-0.01
$MLS3$	20	13.99	8.40	6.22	4.07	-0.49	0.95	0.34	-0.17
	50	9.05	5.54	3.96	2.64	-0.24	0.12	0.12	-0.11
	100	6.71	3.94	2.73	1.90	0.06	-0.16	-0.13	-0.06
	200	4.68	2.81	1.81	1.32	-0.21	-0.06	-0.06	-0.01
CCE	20	20.94	12.97	9.21	5.93	1.86	2.26	1.27	0.58
	50	17.29	11.89	8.71	4.98	2.83	2.99	2.58	0.48
	100	15.45	11.71	7.52	4.77	3.23	3.43	2.34	0.97
	200	17.20	10.33	7.05	4.15	5.22	3.05	2.91	0.73
IFE	20	16.86	11.39	9.43	8.28	7.87	4.87	3.71	3.50
	50	14.42	8.75	6.93	6.45	6.48	2.40	1.94	2.31
	100	11.04	6.83	4.85	3.76	2.39	1.57	0.84	0.94
	200	9.48	4.89	3.27	2.13	2.02	0.73	0.50	0.32
MLE	20	11.89	7.72	5.20	3.53	0.42	0.54	0.20	-0.11
	50	8.15	5.08	3.51	2.29	-0.30	0.15	-0.01	-0.03
	100	6.29	3.74	2.53	1.80	-0.17	-0.11	-0.16	-0.02
	200	4.39	2.67	1.71	1.25	-0.09	-0.11	-0.02	-0.01
ATE	20	14.77	8.12	6.06	4.16	2.42	1.27	0.69	0.60
	50	16.10	5.70	3.70	2.45	6.04	0.61	0.34	0.21
	100	13.12	9.01	2.83	1.78	4.28	3.36	0.08	0.12
	200	9.75	6.90	5.00	1.32	1.49	1.48	0.83	0.05

Note:(1)(i)$x_{it,1} = 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + v_{it,1}$, $x_{it,2} = f_{t,1} + 2\lambda_{i,1} + 2f_{t,2} + 3\lambda_{i,2} + v_{it,2}$;(ii) The errors ε_{it} follow 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 $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.

(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; $MLS3$: the Mundlak estimator β_{M3} in Section 3 of this paper; CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 3: The $RMSE$ and $Bias$ of β_1 with x_{it} are are correlated with $\lambda_i' f_t$ in DGP 2

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	18.44	9.92	5.97	4.17	10.01	4.84	2.22	1.55
	50	14.55	7.18	4.53	2.79	10.65	4.91	2.34	1.36
	100	12.90	6.56	3.83	2.16	9.67	4.77	2.58	1.27
	200	11.93	6.75	3.22	1.96	9.39	4.99	2.42	1.38
CCE	20	20.53	11.32	6.96	4.92	14.28	6.98	3.63	2.21
	50	17.27	9.13	5.26	3.41	14.08	7.33	3.56	2.13
	100	15.41	8.33	4.79	2.66	13.02	6.98	3.87	1.98
	200	15.53	8.35	4.30	2.53	13.47	7.14	3.79	2.13
IFE	20	44.71	45.66	46.35	46.27	42.76	44.73	45.74	45.71
	50	46.23	47.66	47.57	48.24	45.15	47.23	47.23	47.98
	100	45.80	47.27	47.95	47.87	45.00	46.65	47.58	47.59
	200	44.32	43.45	40.30	32.76	43.06	41.77	35.97	24.13
MLE	20	10.31	6.76	4.42	3.13	0.84	0.07	-0.47	0.22
	50	6.46	4.02	2.91	2.04	0.34	0.28	0.003	-0.04
	100	5.05	3.08	2.17	1.54	-0.01	0.06	0.10	0.08
	200	3.63	2.29	1.53	1.11	-0.17	-0.09	-0.02	0.08
ATE	20	32.98	26.67	22.84	20.67	29.42	24.52	21.55	20.02
	50	19.53	13.51	8.71	7.06	16.53	11.10	7.32	6.30
	100	16.45	5.63	3.45	2.54	13.09	3.13	2.37	1.80
	200	10.90	2.97	1.78	1.37	7.18	1.06	0.55	0.67

Note: (i) the data $\{y_{it}, x_{it}\}$ are generated as,

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

where the true slopes $\beta = (\beta_1, \beta_2, \beta_3)'$ with $\beta_1 = \beta_3 = 1$ and $\beta_2 = 2$. The factor loadings $\lambda_i = (\lambda_{i1}, \lambda_{i2})'$ are set as $\lambda_{i1} \sim iidN(0.5, 1)$ and $\lambda_{i2} \overset{iid}{\sim} N(0.3, 1)$. 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,

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

(ii) The errors ε_{it} follows stationary AR(1) process with heteroskedasticity across each i. Let $\rho_{i\varepsilon} \sim iidU[0.5, 0.9]$, $\sigma_i \sim iidU[0.5, 1.5], \zeta_{it} \sim iid\chi^2(3) - 3$ for both i and $t = \{-49, ...0, ..., T\}$,

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

where $\varepsilon_{i,-50} = 0$.

(iii) Let $\epsilon_{i1} \sim iidN(0.5, 1), \epsilon_{i2} \sim iidN(0.2, 1), \epsilon_{i3} \sim iidN(0, 1), \epsilon_{i4} \sim iidN(0.7, 1)$. The regressors

$$x_{it} = \begin{pmatrix} x_{it,1} \\ x_{it,2} \\ x_{it,3} \end{pmatrix} = \begin{bmatrix} \lambda_{i1}, & \lambda_{i2}, \\ \lambda_{i1} + \epsilon_{i1}, & \lambda_{i2} + \epsilon_{i2}, \\ \lambda_{i1} + \epsilon_{i3}, & \lambda_{i2} + \epsilon_{i4}, \end{bmatrix} f_t + v_{it}, \overline{\Gamma} = \begin{bmatrix} 0.5, & 0.3, \\ 1, & 0.5, \\ 0.5, & 1, \end{bmatrix}$$

Table 4: The $RMSE$ and $Bias$ of β_1 with x_{it} are correlated with $\lambda_i' f_t$ in DGP 2 with larger heteroskedasticity

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	22.40	12.86	8.26	6.26	9.92	5.39	1.66	1.59
	50	16.41	9.82	5.94	4.17	8.70	4.89	2.04	1.24
	100	14.44	7.70	4.89	3.01	9.28	4.39	2.55	1.41
	200	12.81	7.23	3.81	2.29	9.42	4.96	2.39	1.24
CCE	20	24.88	15.17	9.07	6.40	14.52	8.13	2.86	2.39
	50	18.66	11.70	6.81	4.47	13.27	7.66	3.44	2.00
	100	16.55	9.49	5.84	3.41	12.66	7.00	3.89	2.10
	200	15.62	9.29	4.70	2.65	12.99	7.61	3.65	1.93
IFE	20	45.80	46.77	47.48	47.13	43.12	45.69	46.66	46.59
	50	48.01	50.23	49.76	50.61	46.34	49.65	49.41	50.36
	100	49.66	50.56	50.76	50.50	48.73	50.15	50.53	50.34
	200	48.83	50.53	50.42	50.93	48.13	50.23	50.29	50.83
MLE	20	16.34	9.41	6.48	4.71	0.10	0.57	-0.02	0.21
	50	11.46	6.52	4.55	3.30	-0.03	-0.11	-0.15	-0.09
	100	7.80	4.76	3.30	2.38	-0.45	-0.04	-0.07	0.11
	200	5.52	3.70	2.37	1.64	0.12	0.02	-0.02	0.01
ATE	20	37.60	35.49	34.09	32.56	33.45	34.13	33.27	32.04
	50	25.92	32.94	28.22	24.58	22.63	31.66	27.36	24.01
	100	25.49	19.83	18.05	11.88	23.45	18.27	16.32	11.23
	200	23.84	16.92	6.33	3.98	22.17	14.80	4.57	3.41

Note: this DGP is similar to Table 3, exception that $\sigma_i \sim iidU[1,4]$.

Table 5: The $RMSE$ and $Bias$ of β_1 with x_{it} are correlated with $\lambda_i' f_t$ in DGP 2

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS3$	20	6.31	3.54	2.64	1.81	0.004	-0.06	-0.28	-0.23
	50	3.61	2.08	1.51	1.09	-0.11	-0.19	-0.10	-0.12
	100	2.52	1.52	1.01	0.76	-0.10	-0.10	-0.001	-0.03
	200	1.71	1.06	0.71	0.51	-0.13	0.02	-0.04	-0.03
CCE	20	16.67	16.19	15.36	14.97	11.20	11.71	10.87	10.74
	50	16.29	15.56	15.06	15.73	11.75	12.00	11.62	12.71
	100	16.29	15.33	15.53	15.78	12.63	11.82	12.50	12.98
	200	16.47	15.96	15.07	14.81	12.58	13.19	12.09	12.18
IFE	20	4.94	2.78	1.98	1.40	0.14	0.005	-0.003	0.04
	50	2.95	1.68	1.16	0.86	0.18	-0.03	0.05	-0.01
	100	1.96	1.25	0.82	0.60	-0.03	-0.005	0.03	-0.02
	200	1.31	0.82	0.56	0.43	0.06	0.04	0.02	-0.002
MLE	20	6.11	3.62	2.58	1.92	0.12	0.15	-0.02	-0.05
	50	3.41	2.16	1.48	1.10	0.12	-0.09	0.02	-0.04
	100	2.36	1.49	1.00	0.75	0.05	-0.02	0.05	0.01
	200	1.61	1.02	0.70	0.51	0.06	0.07	0.01	0.01
ATE	20	6.04	3.36	2.38	1.65	0.20	0.21	-0.05	0.03
	50	3.40	2.05	1.40	1.01	0.26	0.01	0.10	-0.03
	100	2.26	1.45	0.94	0.73	0.01	0.02	0.07	-0.01
	200	1.55	0.96	0.66	0.49	0.06	0.06	0.01	-0.002

Note: (i) The factor and loadings $\lambda_{i1} \sim iidN(0,1)$ and $\lambda_{i2} \sim iidN(0,1)$;$f_{t1} \sim iidN(0,1)$ and $f_{t2} \sim iidN(0,1)$; (ii) $x_{it,1} = 1 + \lambda_i' f_t + f_{t,1} + \lambda_{i,1} + f_{t,2} + \lambda_{i,2} + v_{it,1}$, $x_{it,2} = 1 + \lambda_i' f_t + f_{t,1} + \lambda_{i,1} + f_{t,2} + \lambda_{i,2} + v_{it,2}$; (iii) The errors $\varepsilon_{it} \sim iidN(0,1)$.
(2) $MLS3$: the Mundlak estimator β_{M3} in Section 3 of this paper ($\hat{r}' = 4$); CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 6: The $RMSE$ and $Bias$ of β_1 with x_{it} are correlated with $\lambda_i' f_t$ in DGP 3

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS3$	20	16.54	9.61	6.96	4.94	0.74	-0.50	-0.02	-0.01
	50	10.17	5.72	4.01	2.82	-0.27	-0.03	0.06	0.19
	100	7.03	4.09	2.95	2.10	0.12	-0.21	0.11	0.06
	200	4.93	3.03	2.11	1.35	0.23	0.13	0.17	0.06
CCE	20	60.58	58.05	57.34	56.92	56.88	55.97	55.96	56.23
	50	60.56	60.25	58.80	58.64	58.12	58.90	57.94	58.19
	100	61.69	60.13	59.90	58.62	59.71	58.79	59.20	58.23
	200	62.64	60.42	60.32	59.08	61.05	59.26	59.69	58.69
IFE	20	59.37	64.97	66.72	69.88	52.51	58.84	61.09	65.92
	50	56.41	58.09	60.74	61.42	44.46	44.69	48.71	49.27
	100	50.88	47.08	51.13	54.14	35.15	29.25	33.66	37.68
	200	45.42	44.90	42.38	39.67	27.45	25.67	22.66	20.08
MLE	20	12.29	7.80	5.37	4.20	0.76	-0.55	-0.21	0.16
	50	8.11	4.96	3.59	2.88	0.24	-0.14	0.26	0.15
	100	6.14	3.86	2.77	2.04	0.13	-0.15	0.12	0.07
	200	4.74	3.02	2.20	1.45	0.21	0.15	0.15	0.05
ATE	20	26.20	10.48	6.97	5.12	14.24	3.27	2.94	2.51
	50	23.32	5.75	3.46	2.39	11.03	0.35	0.31	0.61
	100	13.83	6.48	2.61	1.74	4.69	1.38	0.23	0.09
	200	7.71	4.55	2.88	1.29	2.24	0.66	0.47	0.05

Note: (1) The factor loadings $\lambda_{i1} \sim iid\chi^2(1)$ and $\lambda_{i2} \overset{iid}{\sim} N(0.2, 1)$; Let $v_{ft} \sim iid\chi^2(3) - 3$, the common factors $f_t = 0.5 f_{t-1} + v_{ft}$; (ii) $x_{it,1} = 1 + \lambda_i' f_t + f_{t,1} + \lambda_{i,1} + f_{t,2} + \lambda_{i,2} + v_{it,1}$, $x_{it,2} = 1 + \lambda_i' f_t + f_{t,1} + \lambda_{i,1} + f_{t,2} + \lambda_{i,2} + v_{it,2}$; (iii) the errors ε_{it} follow 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 $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.
(2) $MLS3$: the Mundlak estimator β_{M3} in Section 3 of this paper ($\hat{r}' = 4$);; CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 7: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with $f_t'\lambda_i$ in DGP 4 with additive fixed effects

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	16.25	9.31	6.41	4.36	0.37	0.29	0.19	0.16
	50	9.55	5.63	4.12	2.80	-0.06	0.17	-0.25	0.02
	100	6.82	4.04	2.79	2.01	-0.19	-0.02	0.27	0.10
	200	4.75	2.90	1.979	1.44	0.04	-0.03	0.07	-0.06
CCE	20	53.03	50.84	52.84	52.53	42.52	43.25	46.35	46.42
	50	60.91	59.15	57.60	54.65	52.99	54.07	53.17	50.36
	100	60.15	58.77	57.88	56.38	53.93	55.19	54.83	53.87
	200	59.22	56.58	56.74	56.79	53.78	53.84	54.75	55.48
IFE	20	96.13	100.30	97.47	96.77	76.64	81.66	80.46	79.78
	50	106.73	108.82	113.49	115.20	92.36	96.33	102.07	104.12
	100	113.66	121.72	120.26	125.23	103.10	115.82	113.39	119.98
	200	120.93	129.45	130.21	129.01	113.74	127.17	128.42	127.11
MLE	20	18.04	14.44	14.43	13.83	-6.95	-8.80	-10.6	-10.8
	50	13.13	11.44	10.97	10.32	-8.17	-9.19	-9.33	-9.09
	100	10.88	9.73	9.30	9.33	-7.92	-8.48	-8.46	-8.78
	200	9.24	8.80	8.90	8.78	-7.72	-8.13	-8.48	-8.49
ATE	20	21.62	12.54	9.81	9.37	4.05	1.83	2.13	2.08
	50	13.22	7.26	5.60	4.53	1.01	0.86	0.89	1.01
	100	8.53	4.73	3.52	2.67	-1.30	-1.20	1.09	0.69
	200	5.84	3.53	2.46	1.87	-1.36	-1.52	-1.00	0.24

Note: (1)(i) $x_{it,1} = 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + v_{it,1}$, $x_{it,2} = f_{t,1} + 2\lambda_{i,1} + 2f_{t,2} + 3\lambda_{i,2} + v_{it,2}$; (ii)$y_{it} = x_{it}'\beta + \lambda_i' f_t + \alpha_i + \tau_t + \varepsilon_{it}$,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 $\nu_t \sim iidN(0,1)$. (iii) the errors ε_{it} follow 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 $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.

(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 8: $RMSE$ and $Bias$ of $\hat{\beta}_1$ with x_{it} are correlated with f_t, λ_i and ε_{it} in DGP 5

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
MLS2-CF	20	23.36	14.05	9.74	7.06	-0.66	-0.04	-0.17	-0.15
	50	14.60	8.20	6.54	4.43	0.82	-0.17	0.02	0.07
	100	9.88	7.07	4.13	2.97	-0.46	-0.29	-0.05	0.15
	200	7.24	4.62	3.18	2.33	0.02	0.32	0.11	0.10
$IVCEE$	20	29.85	21.41	18.64	17.38	4.12	4.43	3.55	4.54
	50	18.94	14.55	12.03	11.37	1.67	2.66	1.37	1.39
	100	14.92	10.73	8.48	7.87	1.26	0.28	0.47	0.67
	200	10.12	7.46	6.52	5.55	-0.54	-0.13	0.26	0.13
$PGMM$	20	36.48	15.62	9.28	6.22	4.71	-0.35	0.08	0.27
	50	18.42	10.87	5.79	3.56	1.81	-0.21	0.26	0.08
	100	12.54	6.77	4.21	2.44	-0.18	-0.13	0.15	-0.02
	200	8.55	4.72	2.81	1.99	0.38	-0.53	-0.05	0.03
$TGMM$	20	22.98	13.92	8.95	6.65	1.12	0.68	0.25	-0.01
	50	13.57	8.73	5.96	4.18	0.76	0.18	0.02	0.29
	100	10.09	5.91	4.53	3.03	0.37	-0.005	0.02	-0.10
	200	6.88	4.46	3.13	2.00	0.19	-0.30	0.005	0.05

Note: (1) $\lambda_{i1} \overset{iid}{\sim} N(1,2), \lambda_{i2} \overset{iid}{\sim} N(0.2, 0.2), f_t \sim AR(1)$; (ii) $x_{it} = 1 + 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + v_{it}$. Let $z_{it} = (z_{it,1}, z_{it,2})'$ is instrumental variables. For the errors, $\varepsilon_{it} = \epsilon_{it} + \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1-\rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$. Let $\eta = 0.6$,and the variables are generated as

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

(2) MLS2-CF: the Mundlak estimator β_{M2}, combined with the Control Function method in Section 4 of this paper; IVCCE: the common correlated estimator with instrumental variables (Harding and Lamarche, 2011); PGMM: the profile GMM estimator (Hong et al., 2023); TGMM: The average transformed GMM estimator (Hsiao et al., 2023).

Table 9: $RMSE$ and $Bias$ of $\hat{\beta}_1$ with x_{it} are correlated with f_t, λ_i and ε_{it} in DGP 5

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
MLS3-CF	20	17.07	10.19	6.78	5.00	-1.40	0.10	0.05	0.11
	50	10.08	5.90	4.26	2.70	-1.34	-0.72	-0.27	-0.06
	100	7.79	4.10	2.64	1.78	-2.35	-0.91	-0.41	-0.18
	200	5.36	2.78	1.87	1.37	-1.97	-0.80	-0.48	-0.35
$IVCEE$	20	19.00	17.54	17.52	17.73	13.01	12.44	12.84	13.19
	50	11.47	10.94	10.29	10.28	5.60	5.93	4.49	5.60
	100	7.62	7.39	7.44	7.37	2.91	2.70	2.87	2.60
	200	5.97	5.05	5.11	5.03	1.62	1.52	1.55	1.09
$PGMM$	20	35.59	16.97	9.68	5.78	17.12	2.74	0.52	0.20
	50	14.92	14.60	6.55	3.56	1.02	5.67	0.51	0.26
	100	8.33	5.11	6.63	2.21	-0.53	0.28	0.57	0.23
	200	5.37	3.27	1.87	1.60	0.17	0.12	0.02	-0.03
$TGMM$	20	11.77	6.67	4.83	3.42	-0.30	0.06	-0.12	-0.09
	50	6.90	4.05	2.92	2.04	0.01	-0.03	0.02	0.07
	100	4.99	3.01	2.10	1.40	-0.14	-0.01	0.09	0.01
	200	3.39	2.05	1.34	1.05	0.07	0.09	-0.03	-0.15

Note: (1) $\lambda_{i1} \overset{iid}{\sim} N(1,2), \lambda_{i2} \overset{iid}{\sim} N(0.2, 0.2), f_t \sim AR(1)$; (ii) $x_{it} = 1 + f_t'\lambda_i + f_{t,1} + \lambda_{i,1} + f_{t,2} + \lambda_{i,2} + v_{it}$. Let $z_{it} = (z_{it,1}, z_{it,2})'$ is instrumental variables. For the errors, $\varepsilon_{it} = \epsilon_{it} + \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1-\rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$. Let $\eta = 0.6$,and the variables are generated as

$$\begin{pmatrix} v_{it} \\ z_{it,1} \\ z_{it,2} \\ \epsilon_{it} \end{pmatrix} \sim iidN \left(0_{4\times1}, \begin{bmatrix} 1 & 0.5 & 0.5 & \eta \\ 0.5 & 1 & 0 & 0 \\ 0.5 & 0 & 1 & 0 \\ \eta & 0 & 0 & 1 \end{bmatrix} \right).$$

(2) MLS3-CF: the Mundlak estimator β_{M2}, combined with the Control Function method in Section 4 of this paper ($\hat{r}' = 4$);; IVCCE: the common correlated estimator with instrumental variables (Harding and Lamarche, 2011); PGMM: the profile GMM estimator (Hong et al., 2023); TGMM: The average transformed GMM estimator (Hsiao et al., 2023).

Table 10: The Size and Power of Bootstrap T test (p_T^*,5% level) of DGP 1

		Size				Power($\beta_1=1.1$)				Power($\beta_1=1.3$)			
ρ_ε	N\T	20	50	100	200	20	50	100	200	20	50	100	200
0	20	0.080	0.050	0.086	0.054	0.17	0.40	0.65	0.91	0.86	1.00	1.00	1.00
	50	0.070	0.050	0.058	0.054	0.41	0.79	0.96	0.99	1.00	1.00	1.00	1.00
	100	0.068	0.080	0.050	0.046	0.67	0.95	1.00	1.00	1.00	1.00	1.00	1.00
	200	0.082	0.048	0.068	0.054	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.3	20	0.074	0.070	0.056	0.054	0.20	0.32	0.47	0.80	0.75	0.97	1.00	1.00
	50	0.074	0.058	0.064	0.066	0.35	0.66	0.88	1.00	0.97	1.00	1.00	1.00
	100	0.064	0.038	0.062	0.070	0.56	0.88	0.99	1.00	1.00	1.00	1.00	1.00
	200	0.056	0.062	0.058	0.040	0.84	0.99	1.00	1.00	1.00	1.00	1.00	1.00
0.6	20	0.080	0.064	0.048	0.046	0.14	0.22	0.35	0.57	0.60	0.85	0.98	1.00
	50	0.068	0.044	0.072	0.046	0.25	0.43	0.68	0.89	0.90	0.99	1.00	1.00
	100	0.072	0.070	0.050	0.052	0.41	0.67	0.93	1.00	1.00	1.00	1.00	1.00
	200	0.052	0.072	0.038	0.044	0.66	0.91	1.00	1.00	1.00	1.00	1.00	1.00
0.9	20	0.080	0.064	0.048	0.046	0.09	0.12	0.13	0.20	0.33	0.46	0.65	0.87
	50	0.068	0.044	0.072	0.046	0.11	0.18	0.25	0.34	0.60	0.81	0.94	0.99
	100	0.072	0.070	0.050	0.052	0.21	0.27	0.44	0.60	0.85	0.97	0.99	1.00
	200	0.052	0.072	0.038	0.044	0.34	0.50	0.65	0.91	0.94	1.00	1.00	1.00

Note: (1)$p_T^* = \frac{1}{B}\sum_{b=1}^{B} 1\{|\widehat{\beta}_{M,j}^*(b) - \widehat{\beta}_{M,j}| > |\widehat{\beta}_{M,j} - \beta_j^0|\}$; (2)(i) $\lambda_{i1} \overset{iid}{\sim} N(1,2), \lambda_{i2} \overset{iid}{\sim} N(0.2,0.2), f_t \sim AR(1), v_{it,1} \overset{iid}{\sim} N(0,1), v_{it,2} \overset{iid}{\sim} N(0,1)$, with $\rho_f = 0.5$. (ii) $x_{it,1} = 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + v_{it,1}$, $x_{it,2} = f_{t,1} + 2\lambda_{i,1} + 2f_{t,2} + 3\lambda_{i,2} + v_{it,2}$; (iii) Let $\lambda_i^* = (\lambda_{i1}^*, \lambda_{i2}^*)' \overset{iid}{\sim} N(0,1)$ and $f_t^* = (f_{t1}^*, f_{t2}^*)' \overset{iid}{\sim} N(0,1)$,the errors $\varepsilon_{it} = \lambda_i^{*\prime} f_t^* + \rho_\varepsilon \varepsilon_{i,t-1} + \sigma_i(1-\rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$ with $\zeta_{it} \overset{iid}{\sim} N(0,1)$ and $\sigma_i^2 \overset{iid}{\sim} U[0.5,1.5]$. (3) Replications = 500; B = 500.

Table 11: The Size and Power of Bootstrap T test $T_M = (\widehat{\beta}_{M,j} - \beta^0_{M,j})/\sqrt{\hat{V}^{boot}_{M,jj}}$ of DGP 1

		Size				Power($\beta_1 = 1.1$)				Power($\beta_1 = 1.3$)			
ρ_ε	N\T	20	50	100	200	20	50	100	200	20	50	100	200
0	20	0.074	0.058	0.060	0.060	0.22	0.41	0.66	0.88	0.84	1.00	1.00	1.00
	50	0.054	0.052	0.040	0.048	0.39	0.71	0.96	1.00	0.99	1.00	1.00	1.00
	100	0.062	0.044	0.052	0.060	0.69	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	200	0.064	0.048	0.032	0.056	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.3	20	0.064	0.062	0.068	0.066	0.21	0.32	0.52	0.80	0.80	0.97	1.00	1.00
	50	0.084	0.054	0.052	0.066	0.31	0.63	0.87	0.98	0.98	1.00	1.00	1.00
	100	0.054	0.060	0.050	0.052	0.57	0.89	0.99	1.00	1.00	1.00	1.00	1.00
	200	0.062	0.060	0.044	0.052	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.6	20	0.054	0.046	0.060	0.070	0.16	0.21	0.34	0.57	0.64	0.87	0.99	1.00
	50	0.064	0.056	0.066	0.054	0.23	0.43	0.67	0.94	0.91	1.00	1.00	1.00
	100	0.056	0.042	0.040	0.068	0.43	0.69	0.91	1.00	0.99	1.00	1.00	1.00
	200	0.032	0.032	0.054	0.066	0.67	0.91	0.99	1.00	1.00	1.00	1.00	1.00
0.9	20	0.068	0.052	0.056	0.074	0.11	0.10	0.12	0.23	0.34	0.49	0.66	0.89
	50	0.036	0.044	0.038	0.044	0.16	0.15	0.22	0.38	0.64	0.80	0.95	1.00
	100	0.040	0.042	0.042	0.048	0.20	0.30	0.42	0.64	0.86	0.96	1.00	1.00
	200	0.036	0.026	0.024	0.046	0.35	0.50	0.71	0.89	0.96	0.99	1.00	1.00

Note: (1)(i) $\lambda_{i1} \overset{iid}{\sim} N(1,2), \lambda_{i2} \overset{iid}{\sim} N(0.2,0.2), f_t \sim AR(1), v_{it,1} \overset{iid}{\sim} N(0,1), v_{it,2} \overset{iid}{\sim} N(0,1)$; (ii) $x_{it,1} = 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + v_{it,1},\ x_{it,2} = f_{t,1} + 2\lambda_{i,1} + 2f_{t,2} + 3\lambda_{i,2} + v_{it,2}$; (iii) Let $\lambda^*_i = (\lambda^*_{i1}, \lambda^*_{i2})' \overset{iid}{\sim} N(0,1)$ and $f^*_t = (f^*_{t1}, f^*_{t2})' \overset{iid}{\sim} N(0,1)$,the errors $\varepsilon_{it} = \lambda^{*\prime}_i f^*_t + \rho_\varepsilon \varepsilon_{i,t-1} + \sigma_i(1-\rho^2_{i\varepsilon})^{0.5}\zeta_{it}$ with $\zeta_{it} \overset{iid}{\sim} N(0,1)$ and $\sigma^2_i \overset{iid}{\sim} U[0.5,1.5]$. (2) Replications = 500; B = 500.

Table 12: The size and power of robust T test of DGP 1 under Fixed b theory(5% level)

		$\rho_\varepsilon = \rho_f = 0.3$				$\rho_\varepsilon = \rho_f = 0.6$				$\rho_\varepsilon = \rho_f = 0.9$			
Stats	T\N	20	50	100	200	20	50	100	200	20	50	100	200
Size	20	0.076	0.056	0.053	0.055	0.108	0.068	0.065	0.043	0.166	0.140	0.112	0.080
	50	0.055	0.055	0.064	0.052	0.121	0.098	0.076	0.052	0.192	0.138	0.123	0.080
	100	0.088	0.058	0.046	0.038	0.105	0.087	0.066	0.058	0.196	0.156	0.122	0.088
	200	0.058	0.065	0.041	0.041	0.087	0.070	0.072	0.056	0.190	0.148	0.104	0.090
Power	20	1	1	1	1	0.995	1	1	1	0.984	0.995	1	1
	50	1	1	1	1	1	1	1	1	1	1	1	1
	100	1	1	1	1	1	1	1	1	1	1	1	1
	200	1	1	1	1	1	1	1	1	1	1	1	1

Note: (i) The common factors $f_t = \rho_f f_{t-1} + v_{ft}$ with $v_{ft} = (v_{ft,1}, v_{ft,2})' \overset{iid}{\sim} N((0,0)', I_2)$; (ii) The factor loadings $\lambda_i = (\lambda_{i1}, \lambda_{i2})'$ are set as $\lambda_{i1} \overset{iid}{\sim} N(1,2)$ and $\lambda_{i2} \overset{iid}{\sim} N(0.2, 0.2)$;

(iii) $v_{it,1} = \rho_f v_{i,t-1,1} + \eta_{it,1}$ and $v_{it,2} = \rho_f v_{i,t-1,2} + \eta_{it,2}$; (iv) Let $\lambda_i^* = (\lambda_{i1}^*, \lambda_{i2}^*)' \overset{iid}{\sim} N(0,1)$ and and $f_t^* = \rho_f f_{t-1}^* + v_{ft}^*$. The errors $\varepsilon_{it} = \lambda_i^{*\prime} f_t^* + \rho_\varepsilon \varepsilon_{i,t-1} + \sigma_i (1 - \rho_{i\varepsilon}^2)^{0.5} \zeta_{it}$ with $\zeta_{it} \overset{iid}{\sim} N(0,1)$ and $\sigma_i^2 \overset{iid}{\sim} U[0.5, 1.5]$.

B Proof of Theorems

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 $u_{it}=\eta_i'd+c'\xi_t+\eta_i'\xi_t+\varepsilon_{it}$.

Proof of Theorem 1. According to the two-way Mundlak projection estimator in equations (11) and (12),

$$\begin{aligned}\widehat{\beta}_{M2}-\beta&=(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{U}\\&=(\tilde{X}'\tilde{X})^{-1}\tilde{X}'vec(M_{\bar{X}}\mu M_{\underline{X}})\\&=(\tilde{X}'\tilde{X})^{-1}\tilde{X}'(M_{\underline{X}}\otimes M_{\bar{X}})vec(\mu).\end{aligned}$$

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

$$\begin{aligned}\sqrt{NT}(\widehat{\beta}_{M2}-\beta)&=(\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}\frac{1}{\sqrt{NT}}\tilde{X}'(M_{\underline{X}}\otimes M_{\bar{X}})vec(\mu)\\&=(\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}(\frac{1}{\sqrt{NT}}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}\mu_{it}).\end{aligned}\tag{31}$$

In addition, as $(N,T)\to\infty$, $\Psi_{NT}=plim_{(T,N)\to\infty}(NT)^{-1}\tilde{X}'M_{\tilde{G}}\tilde{X}$. Lastly, according to Assumptions 1-3 and the part 1 of Theorem 2.1 in Gao et al.(2023), we obtain the results.

In addition, we could decompose the asymptotic variance Φ_{NT} into four terms. Plugging $\mu_{it}=c'\xi_t+d'\eta_i+\eta_i'\xi_t+\varepsilon_{it}$ into equation (31) and then scaled by $\sqrt{NT}$ on both side of the equation. Thus,

$$
\begin{aligned}
\sqrt{NT}(\widehat{\beta}_{M2}-\beta) &= (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}c'\xi_t) \\
&+ (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}d'\eta_i) \\
&+ (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}\eta_i'\xi_t) \\
&+ (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}\varepsilon_{it}).
\end{aligned}
$$

Under Assumptions 1-3, let $C = diag(c_1', \cdots, c_T')$, $D = diag(d_1', \cdots, d_N')$ and $\tilde{\mathbb{X}}_{t\cdot} = (\tilde{\mathbb{X}}_{t1}, \cdots, \tilde{\mathbb{X}}_{tN})$,$\tilde{\mathbb{X}}_{\cdot i} = (\tilde{\mathbb{X}}_{1i}, \cdots, \tilde{\mathbb{X}}_{Ti})$,

$$
\begin{aligned}
\Phi_{NT,1} =& plim_{(T,N)\to\infty}Var((NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}c'\xi_t) = plim_{(T,N)\to\infty}(NT)^{-1}Var(\sum_{i=1}^{N}\tilde{\mathbb{X}}_{\cdot i}C\xi) \\
=& plim_{(T,N)\to\infty}(NT)^{-1}\sum_{i=1}^{N}E(\tilde{\mathbb{X}}_{\cdot i}C\xi\xi'C'\tilde{\mathbb{X}}_{\cdot i}'),
\end{aligned}
$$

and

$$
\begin{aligned}
\Phi_{NT,2} =& plim_{(T,N)\to\infty}Var((NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}d'\eta_i) = plim_{(T,N)\to\infty}(NT)^{-1}Var(\sum_{t=1}^{T}\tilde{\mathbb{X}}_{t\cdot}D\eta) \\
=& plim_{(T,N)\to\infty}(NT)^{-1}\sum_{t=1}^{T}E(\tilde{\mathbb{X}}_{t\cdot}D\eta\eta'D'\tilde{\mathbb{X}}_{t\cdot}'),
\end{aligned}
$$

and

$$
\begin{aligned}
\Phi_{NT,3} =& plim_{(T,N)\to\infty}Var((NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}\eta_i'\xi_t) \\
=& plim_{(T,N)\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}E(\tilde{\mathbb{X}}_{ti}\eta_i'\xi_t\eta_j'\xi_s\tilde{\mathbb{X}}_{sj}').
\end{aligned}
$$

Similarly, according to Assumptions 1-3,

$$
\begin{aligned}
\Phi_{NT,4} =& plim_{(T,N)\to\infty}Var((NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}\varepsilon_{it}) \\
=& plim_{(T,N)\to\infty}E[(NT)^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}\varepsilon_{it})(\sum_{j=1}^{N}\sum_{s=1}^{T}\varepsilon_{js}\tilde{\mathbb{X}}_{sj}')] \\
=& plim_{(T,N)\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}E(\tilde{\mathbb{X}}_{ti}\varepsilon_{it}\varepsilon_{js}\tilde{\mathbb{X}}_{sj}').
\end{aligned}
$$

Lastly, according to Assumptions 1-3 and the part 1 of Theorem 2.1 in Gao et al. (2023), we obtain

$$\sqrt{NT}(\widehat{\beta}_{M2}-\beta) \sim N(0, \Psi_{NT}^{-1}\Phi_{NT}\Psi_{NT}^{-1}).$$

with $\Phi_{NT} = \Phi_{NT,1} + \Phi_{NT,2} + \Phi_{NT,3} + \Phi_{NT,4}$.

Theorem 2. Under Assumptions 1-3, 11, 12, as $(T, N) \to \infty$, then

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

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

(ii)

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

and

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

Proof of Theorem 2: (i) Let $S_{NT}^* = \sqrt{NT}(\widehat{\beta}_{M2}^* - \widehat{\beta}_{M2})$ and $S_{NT} = \sqrt{NT}(\widehat{\beta}_{M2} - \beta_0)$. For the definition of DWB,

$$\begin{aligned}
S_{NT}^* &= (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}\frac{1}{\sqrt{NT}}\tilde{X}'\tilde{U}^* = (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{x}_{it}\hat{\tilde{u}}_{it}\epsilon_t \\
&= (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{x}_{it}(\tilde{x}_{it}'\beta + \tilde{u}_{it} - \tilde{x}_{it}'\hat{\beta}_{M2})\epsilon_t \\
&= (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{x}_{it}\tilde{u}_{it}\epsilon_t \\
&+ (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{x}_{it}\tilde{x}_{it}'(\beta - \hat{\beta}_{M})\epsilon_t = S_{1,NT}^* + S_{2,NT}^*.
\end{aligned}$$

According to Theorem 1, such that $\beta - \hat{\beta}_{M2} = O_p(\frac{1}{\sqrt{NT}})$ and Assumption 2, the above second term $S_{2,NT}^* = o_p(1)$. Thus, $S_{NT}^* = S_{1,NT}^* + o_p(1)$.

Thus, to show (1), we need to proof that $E(E^*(S_{1,NT}^{*2})) - E(S_{NT}^2) = o_p(1)$. This proof follows the proof of Theorem 2.3 in Gao et al. (2023).

Let $w_{it} = \tilde{x}_{it}\tilde{u}_{it}$, $W_t = (w_{1t}, w_{2t}, \cdots, w_{Nt})'$, and $\bar{W}_t = \frac{1}{\sqrt{N}}W_t'1_N$, thus, according to equation (2.5) in in Gao et al. (2023),

$$E^*(S_{1,NT}^{*2}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\bar{W}_t\bar{W}_s k(\frac{t-s}{l}),$$

Thus,

$$\begin{aligned}
l^q[E(E^*(S_{1,NT}^{*2}) - E(S_{NT}^2)] &= l^q\sum_{s=-l}^{l}[k(s/l) - 1]E(\bar{W}_s\bar{W}_0) - 2l^q\sum_{s=1}^{l}\frac{s}{T}[k(\frac{s}{l}) - 1]E(\bar{W}_s\bar{W}_0) \\
&- 2l^q\sum_{s=l+1}^{T-1}\frac{T-s}{T}E(\bar{W}_s\bar{W}_0) = I_1 + I_2 + I_3.
\end{aligned}$$

As the proof of Proposition A.1 in Gao et al. (2023), it's obvious that $\sum_{t=1}^{\infty} t^2|E(\bar{W}_t\bar{W}_0)| < \infty$. Next, Same arguments as the proof of Theorem 2.3 in Gao et al. (2023), show that $I_1 = -c_q \sum_{s=-\infty}^{\infty} |k|^q E(\bar{U}_s\bar{U}_0) + o_p(1)$, and $I_2 = o_p(1)$,$I_3 = o_p(1)$. Last, we obtain the result (i).

(ii) It's sufficient to show that

$$S^*_{NT} = (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}\frac{1}{\sqrt{NT}}\tilde{X}'\tilde{U}^* \xrightarrow{d^*} N(0, V_{\beta 2}).$$

Thus, combined with Theorem 1, we obtain $\sup_{\tau\in\mathbb{R}} |P^*(S^*_{NT} < \tau) - P(S_{NT} < \tau)| = o_p(1)$ and get the results.

Next, we show $\sqrt{NT}(\widehat{\beta}^*_{M2} - \widehat{\beta}_{M2}) \xrightarrow{d^*} N(0, V_{\beta 2})$. According to the results of (i), $E(E^*(S^{*2}_{NT})) - E(S^2_{NT}) = o_p(1)$, with $E(\tilde{u}_{it}) = 0, E(\epsilon_t) = 0$ in Assumption . Since the variables $\tilde{x}_{it}, \tilde{u}_{it}$ and ϵ_t are mutual independent and according to the proof of Theorem in Gao et al. (2023), we conclude that,

$$S^*_{1,NT} = (\frac{1}{NT}\tilde{X}'\tilde{X})^{-1}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{x}_{it}\tilde{u}_{it}\epsilon_t \xrightarrow{d^*} N(0, V_{\beta 2}).$$

Last, we obtain the results.

Theorem 3. Under Assumptions 13-15, as $(T, N) \to \infty$, then

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

and

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

Proof of Theorem 3: The Results for t_{HACSC} and $Wald_{HACSC}$ require the limit of $(NT)^{-1/2}\hat{\hat{S}}_{[aT]}$,

$$\begin{aligned}
(NT)^{-1/2}\hat{\hat{S}}_{[aT]} &= (NT)^{-1/2}\textstyle\sum_{t=1}^{[aT]}\hat{\hat{v}}_t = (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\displaystyle\sum_{i=1}^{N}\tilde{x}_{it}\hat{\tilde{u}}_{it} \\
&= (NT)^{-1/2}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}(\tilde{x}'_{it}\beta + \tilde{u}_{it} - \tilde{x}'_{it}\hat{\beta}_M) \\
&= (NT)^{-1/2}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}u_{it} - (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}\tilde{x}'_{it}(\hat{\beta}_M - \beta) \\
&= B_1 - B_2.
\end{aligned}$$

For B_1, under Assumptions 14,

$$(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}u_{it} \Rightarrow \Lambda W_{p+1}(1,a)$$

with $W_{p+1}(1,a)$ is a $(p+1)\times 1$ vector of standard Brownian sheets.

For B_2, according to (31)

$$\begin{aligned}\sqrt{NT}(\hat{\beta}_{aug}-\beta_{aug}) &= ((NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}')^{-1}[(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}u_{it}] \\ &\Rightarrow \mathbf{\Psi}^{-1}\Lambda W_{p+1}(1,1) \equiv N(0,\mathbf{\Psi}^{-1}\Omega\mathbf{\Psi}^{-1}) \qquad (32)\end{aligned}$$

Lastly,

$$(NT)^{-1/2}\hat{\hat{S}}_{[aT]} \Rightarrow \Lambda[W_{p+1}(1,a)-aW_{p+1}(1,1)]. \qquad (33)$$

Let $\hat{\hat{\nu}}_t = \sum_{i=1}^{N}\hat{\nu}_{it} = \sum_{i=1}^{N}\tilde{x}_{it}\hat{\hat{u}}_{it}$, $\hat{\hat{S}}_{[rT]} = \sum_{t=1}^{[aT]}\hat{\hat{\nu}}_t$,

$$\begin{aligned}(NT)^{-1/2}\hat{\hat{S}}_t &= (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}\hat{\hat{u}}_{it} \\ &= (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}(\mu_{it}-\tilde{x}_{it}'(\hat{\beta}_M-\beta)) \\ &= (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}\mu_{it} - (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[rT]}\tilde{x}_{it}\tilde{x}_{it}'(\hat{\beta}_M-\beta) \\ &= (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}\mu_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{[aT]}\tilde{x}_{it}\tilde{x}_{it}'((NT)^{-1}\tilde{X}'\tilde{X})^{-1}(NT)^{-1/2}\tilde{X}'U \\ &\Rightarrow \Lambda W_{p+1}(1,a) - rW_{p+1}(1,a) = B_{p+1}(a).\end{aligned}$$

Following Kiefer and Vogelsang (2002), for the Bartlett kernel with bandwidth equal to the sample size,

$$\begin{aligned}\hat{\Omega} &= Var[(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\hat{\hat{u}}_{it}] \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{x}_{it}\hat{\hat{u}}_{it})(1-|t-s|/T)(\hat{\hat{u}}_{js}\tilde{x}_{js}') \\ &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}(\sum_{i=1}^{N}\tilde{x}_{it}\hat{\hat{u}}_{it})(1-|t-s|/T)(\sum_{j=1}^{N}\hat{\hat{u}}_{js}\tilde{x}_{js}') \\ &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\hat{\nu}}_t(1-|t-s|/T)\hat{\hat{\nu}}_s.\end{aligned}$$

Following equation (1) in Kiefer and Vogelsang (2002),

$$\begin{aligned}\hat{\Omega} &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\hat{\nu}}_t\,(1-|t-s|/T)\,\hat{\hat{\nu}}_s \\ &= \frac{2}{T}\sum_{t=1}^{T}(NT)^{-1/2}\hat{\hat{S}}_t(NT)^{-1/2}\hat{\hat{S}}_t' \\ &\Rightarrow \Lambda[2\int_0^1 B_l(r)B_l(r)'dr]\Lambda'\end{aligned}$$

$$
\begin{aligned}
Wald_{HACSC} &= T(R_2\hat{\beta}_M-\theta_2)'(R\mathbf{\Psi}^{-1}\hat{\Omega}\mathbf{\Psi}^{-1}R')^{-1}(R_2\hat{\beta}_M-\theta_2)/l \\
&\Rightarrow W_l(1)'[2\int_0^1 B_l(r)B_l(r)'dr]^{-1}W_l(1)
\end{aligned}
$$

and in the case with $l=1$,

$$
t_{HACSC}=\frac{\sqrt{T}\left(R_2\hat{\beta}_M-\theta_2\right)}{\sqrt{R_2\mathbf{\Psi}^{-1}\hat{\Omega}\mathbf{\Psi}^{-1}R_2'}}\Rightarrow\frac{W_q(1)}{2\int_0^1 B_1(r)^2dr}.
$$

Proposition 2. Under Assumptions 1, 2, and 8-10, as $(T,N)\to\infty$, then

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

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

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

Proof of Proposition 2. According to the Mundlak estimator in(12) and equations (27),

$$
\tilde{Y}=\tilde{X}\beta+\hat{\mathbb{Q}}\alpha+(\mathbb{Q}-\hat{\mathbb{Q}})\alpha+\tilde{W}.
$$

Thus, let $\underset{NT\times p}{\tilde{\mathbb{X}}}=(M_{\underline{X}}\otimes M_{\bar{X}})\,M_{\hat{\mathbb{Q}}}\tilde{X}$, and the $(i-1)T+t$ row is denoted by $\underset{p\times 1}{\tilde{\mathbb{X}}_{ti}}$, and $\underset{NT\times p}{\tilde{\mathbb{Z}}}=M_Z(M_{\underline{X}}\otimes M_{\bar{X}})M_{\hat{\mathbb{Q}}}\tilde{X}$, and the $(i-1)T+t$ row is denoted by $\underset{p\times 1}{\tilde{\mathbb{Z}}_{ti}}$. Let $\underset{NT\times p_2}{V}=[V^1,V^2,\cdots,V^{p_2}]$ and $V-\hat{V}=M_ZV$, thus

$$
\begin{aligned}
\hat{\beta}_{M4}-\beta &= (\tilde{X}'M_{\hat{\mathbb{Q}}}\tilde{X})^{-1}\tilde{X}'M_{\hat{\mathbb{Q}}}[(\mathbb{Q}-\hat{\mathbb{Q}})\alpha+\tilde{W}] \\
&= (\tilde{X}'M_{\hat{\mathbb{Q}}}\tilde{X})^{-1}\tilde{X}'M_{\hat{\mathbb{Q}}}(M_{\underline{X}}\otimes M_{\bar{X}})vec(\omega) \\
&+ (\tilde{X}'M_{\hat{\mathbb{Q}}}\tilde{X})^{-1}\tilde{X}'M_{\hat{\mathbb{Q}}}(M_{\underline{X}}\otimes M_{\bar{X}})M_ZQ\alpha \\
&= (\tilde{X}'M_{\hat{\mathbb{Q}}}\tilde{X})^{-1}(\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{X}}_{ti}\omega_{it})+(\tilde{X}'M_{\hat{\mathbb{Q}}}\tilde{X})^{-1}(\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbb{Z}}_{ti}\alpha'q_{it}).
\end{aligned}
$$

Similar to Section 2.3, according to Assumptions 9 and 10 and Theorem 1 in Gao et al.(2023), 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 $\underset{p\times 1}{\tilde{\mathbb{X}}_{ti}}$,

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

where $\Theta_{NT,1}^{(1)}$ is $p \times p$ dimensional non-singular 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 $\Theta_{NT}^{(2)}$ is $p \times p$ dimensional non-singular positive matrix.

Thus, as $(N,T) \to \infty$, $\hat{\mathbb{Q}} \xrightarrow{p} \mathbb{Q}$ and$\tilde{\Psi}_{NT} = plim_{(N,T)\to\infty}\frac{1}{NT}\tilde{X}' M_{\mathbb{Q}}\tilde{X}$. Lastly, according to Assumptions 10, we obtain

$$\sqrt{NT}(\hat{\beta}_{M4} - \beta) \sim N(0, \tilde{\Psi}_{NT}^{-1}\Phi_{NT}\tilde{\Psi}_{NT}^{-1}).$$

where $\Phi_{NT} = \Theta_{NT}^{(1)} + \Theta_{NT}^{(2)}$.

C Additional Simulations and Results

C.1 Addtional Simulations of Section 7

DGP 1: Table 13 reports the results of the DGP 1 with cross sectional dependence in the errors. Table 14 extend the simulation of DGP 1, in which $\hat{r} = 6$. Table 15 extends the DGP 1 with the true factor number $r = 4$ and the estimated number of factors $\hat{r} = 2$.

DGP 2(CIF; Pesaran, 2006; Bai and Li, 2014): Table 16 reports the results of the DGP 2 with two regressors and one factor. Table 17 report the results of this case with larger heteroskedasticity.

DGP 4: Table 18 reports the another results of the DGP 4 with the additive fixed effects. For each estimators, we first do the two-way transformation to eliminate the additive fixed effects.

C.2 Additional DGP

In addition, we also consider the following additional three DGP6-8:

DGP 6 (Simplified Case) : the regressors x_{it} are driven by linear combination of common factors,

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

DGP 7 (Pesaran, 2006) : the regressors are mainly driven by common factors,

$$x_{it} = \Gamma_i' f_t + v_{it},$$

where all the elements of Γ_i follows $iidN(0, 1)$.

DGP 8 : the regressors x_{it} are driven by linear combination of factor loadings,

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

Table 19 reports the RMSE and Bias of DGP 6. The RMSE and Bias of the five estimators all decrease, as N or T become larger. In this simplified case, the estimators of all the approaches are all consistent. The performance of MLS1 is very similar to that of the CCE approach. The RMSE and Bias of MLS1 is a litter better than those

Table 13: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with f_t and λ_i in DGP 1 with Cross-sectional Dependence in ε_{it}

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	16.37	9.72	6.62	4.67	-0.62	-0.05	0.083	0.41
	50	9.17	5.84	4.22	3.02	0.11	0.39	-0.031	0.09
	100	6.72	4.55	2.95	2.06	0.36	0.05	0.079	-0.02
	200	5.03	3.06	2.09	1.52	0.07	0.11	0.215	0.01
CCE	20	27.55	19.56	15.79	14.54	5.88	7.1	8.09	9.97
	50	23.88	18.26	15.91	14.76	5.88	8.501	10.23	11.39
	100	23.32	18.71	15.59	14.02	5.91	9.83	10.28	10.99
	200	23.17	16.78	14.42	13.75	5.83	8.36	9.87	11.29
IFE	20	29.23	24.38	20.84	19.40	3.63	-1.02	-2.32	-4.25
	50	26.53	20.75	18.33	16.38	1.47	-5.20	-6.47	-10.07
	100	25.32	20.78	16.37	16.26	-1.03	-6.31	-8.69	-10.54
	200	24.41	19.01	17.31	15.72	-2.74	-5.72	-11.56	-12.58
MLE	20	16.83	10.92	8.97	7.52	-6.35	-5.56	-5.85	-5.77
	50	10.12	7.35	6.48	5.78	-4.49	-4.52	-4.86	-4.78
	100	7.67	6.01	5.20	4.88	-3.65	-4.11	-4.22	-4.42
	200	6.44	5.04	4.36	4.39	-3.96	-3.85	-3.84	-4.10
ATE	20	21.98	16.69	14.39	13.43	4.09	-9.29	-10.43	-11.10
	50	14.88	9.13	9.463	8.04	2.01	2.12	-7.91	-7.09
	100	13.84	6.94	5.153	4.18	0.66	0.26	0.88	-3.51
	200	10.89	5.60	3.908	2.67	-0.31	0.43	0.51	0.16

Note: (1)The DGP is similar to DGP 4, except that the errors ε_{it} follow AR(1) process with heteroskedasticity and cross-sectional dependence across each i. Let $\rho_{i\varepsilon} \sim iidU[0.5, 0.9]$,$\sigma_i \sim iidU[0.8, 1.8]$,$\zeta_{it} \sim iid\chi^2(3) - 3$. $f_t^* \sim iidN(0, 1)$ is one additional factor, with loadings $\lambda_i^* \sim iid\chi^2(2) - 2$, which are uncorrelated with x_{it}. The errors $\varepsilon_{it} = \lambda_i^* f_t^* + \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.
(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 14: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with f_t and λ_i in DGP 1 with $\hat{r} = 6$

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	14.98	8.92	6.61	4.30	-0.69	0.98	0.13	-0.14
	50	9.02	5.51	4.00	2.72	-0.35	0.14	0.13	-0.13
	100	6.64	4.00	2.78	1.92	0.10	-0.22	-0.11	-0.05
	200	4.67	2.82	1.83	1.33	-0.18	-0.08	-0.06	-0.01
CCE	20	20.94	12.97	9.21	5.93	1.86	2.26	1.27	0.58
	50	17.29	11.89	8.71	4.98	2.83	2.99	2.58	0.48
	100	15.45	11.71	7.52	4.77	3.23	3.43	2.34	0.97
	200	17.20	10.33	7.05	4.15	5.22	3.05	2.91	0.73
IFE	20	16.68	11.35	9.59	8.38	7.79	4.91	3.67	3.67
	50	14.42	8.78	7.03	6.45	6.48	2.60	2.25	2.31
	100	11.04	6.83	4.85	3.76	2.39	1.57	0.84	0.94
	200	9.48	4.89	3.27	2.13	2.02	0.73	0.50	0.32
MLE	20	18.19	11.75	6.16	2.97	13.83	7.13	1.76	0.34
	50	10.15	5.00	2.98	1.94	4.48	0.46	0.05	0.03
	100	6.60	3.59	2.34	1.55	1.48	-0.14	-0.13	-0.03
	200	4.98	2.65	1.72	1.26	0.43	-0.07	0.02	0.03
ATE	20	14.77	8.12	6.06	4.16	2.42	1.27	0.69	0.60
	50	16.10	5.70	3.70	2.45	6.04	0.61	0.34	0.21
	100	13.12	9.01	2.83	1.78	4.28	3.36	0.08	0.12
	200	9.75	6.90	5.00	1.32	1.49	1.48	0.83	0.05

Note:(1) the DGP and estimators is same as DGP 4, except that in the estimation of IFE, MLE and ATE, we set the estimated number of common factors $\hat{r} = 6$.
(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 15: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with f_t and λ_i in DGP 1 with $r = 4$ and $\hat{r} = 2$

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	24.33	16.10	11.22	7.46	-0.17	1.66	0.61	0.11
	50	16.29	9.88	7.11	4.77	0.54	0.26	0.30	-0.14
	100	11.13	6.91	4.66	3.19	-0.54	-0.45	0.12	-0.06
	200	7.37	4.70	3.34	2.28	0.02	0.05	-0.14	-0.21
CCE	20	25.80	19.23	14.40	12.38	-7.39	-5.75	-6.31	-6.80
	50	23.22	15.27	12.15	9.63	-6.94	-5.97	-5.86	-5.75
	100	19.97	13.28	10.19	8.97	-6.80	-6.13	-4.86	-5.86
	200	18.10	12.81	9.57	8.04	-7.29	-6.32	-5.44	-4.88
IFE	20	31.77	28.29	29.14	28.19	16.37	15.04	15.39	12.58
	50	28.94	29.11	25.08	24.15	17.04	16.12	13.30	13.47
	100	27.90	24.49	23.23	20.38	17.32	13.59	13.11	10.89
	200	30.33	24.98	19.46	15.82	18.36	13.91	10.77	8.81
MLE	20	19.85	11.20	8.82	6.10	1.60	0.64	0.78	0.77
	50	13.15	8.10	5.85	4.08	0.84	-0.05	-0.06	-0.61
	100	9.49	6.05	4.22	3.02	-0.96	-0.61	-0.52	-0.56
	200	7.50	4.37	3.27	2.25	0.01	-0.62	-0.92	-0.80
ATE	20	20.62	13.77	10.37	9.75	4.54	2.07	0.54	0.74
	50	23.65	9.49	7.54	6.17	10.84	1.03	0.28	0.50
	100	20.79	15.60	5.75	4.43	9.65	7.21	0.33	0.16
	200	21.01	15.20	12.13	3.68	8.62	6.70	5.59	0.21

Note:(1)(i)$x_{it,1} = 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + 3f_{t,3} + \lambda_{i,3} + 2f_{t,4} + \lambda_{i,4} + v_{it,1}$, $x_{it,2} = f_{t,1} + 2\lambda_{i,1} + 2f_{t,2} + 3\lambda_{i,2} + f_{t,3} + 2\lambda_{i,3} + 2f_{t,4} + 3\lambda_{i,4} + v_{it,2}$;(ii) The errors ε_{it} follow 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 $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.
(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 16: Two Regressors and One Factor

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	15.26	9.09	6.24	4.38	1.40	0.45	0.52	0.28
	50	10.28	5.90	4.25	2.83	3.02	0.62	0.15	0.21
	100	7.48	4.22	2.97	1.93	2.23	0.52	0.32	0.11
	200	6.43	3.30	2.16	1.42	2.28	0.99	0.29	0.18
CCE	20	15.14	9.26	6.67	4.67	3.95	1.46	0.99	0.67
	50	11.80	6.02	4.36	2.92	5.80	1.66	0.74	0.50
	100	9.58	4.60	3.05	1.93	5.21	1.73	0.94	0.36
	200	8.49	3.85	2.30	1.48	5.27	2.15	0.88	0.49
IFE	20	37.87	38.43	39.02	38.85	33.41	36.47	37.36	37.46
	50	38.91	38.73	39.14	39.22	35.72	36.77	37.39	37.56
	100	37.04	34.87	32.32	30.44	33.09	29.56	26.12	23.16
	200	30.80	23.54	13.02	8.51	24.97	14.58	4.76	2.37
MLE	20	11.08	7.27	4.82	3.64	-0.96	-0.20	-0.04	0.22
	50	7.98	4.89	3.63	2.62	0.29	-0.23	0.02	0.07
	100	5.62	3.84	2.65	1.81	0.08	-0.31	-0.02	0.001
	200	4.09	2.57	1.90	1.32	-0.05	0.08	-0.09	0.04
ATE	20	23.76	18.07	16.43	15.34	17.38	15.44	14.88	14.37
	50	15.42	9.10	6.98	5.86	10.23	6.02	4.98	4.74
	100	11.51	4.89	3.47	2.40	6.60	1.84	1.63	1.38
	200	6.60	3.19	2.20	1.46	3.36	1.20	0.43	0.50

Note: the DGP is similar to Table 1, exception that the AR(1) errors, The errors 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 $t = \{-49, ...0, ..., T\}$,

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

where $\varepsilon_{i,-50} = 0$.

of CCE. For example, when $N = T = 200$, the RMSE of MLS1 and CCE are 0.0147 and 0.0149, and the Bias of MLS1 and CCE are -0.0014 and -0.0019. From view of RMSE, the MLE has the best performance and the IFE has the largest RMSE, the ATE is closed to our MLS1. From view of Bias, the IFE and ATE estimators have more accuracy and the Bias of MLE are similar to that of MLS1 and CCE.

In Table 20, we consider the case of common correlated estimation (DGP 7), where the regressor are driven by factors with heterogeneous coefficients. The RMSE of MLE is the smallest and most efficient. Compared with the RMSE of the CCE approach, the ATE and IFE has similar efficient and our MLS is slightly lower than them. From the

Table 17: Two Regressors and One Factor with larger heteroskedasticity

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	21.59	12.48	8.30	5.76	2.77	1.62	0.78	0.11
	50	15.33	8.22	5.44	3.99	3.60	1.42	0.67	0.24
	100	11.72	6.41	4.09	2.90	4.31	1.78	0.51	0.44
	200	9.69	4.45	3.04	2.02	4.60	1.58	0.78	0.21
CCE	20	23.63	14.00	9.24	6.20	9.22	4.33	2.09	1.00
	50	17.60	9.36	5.94	4.10	9.29	3.84	2.08	0.89
	100	16.44	7.99	4.48	3.09	11.19	4.50	1.98	1.08
	200	14.76	6.14	3.72	2.18	10.86	4.24	2.17	0.92
IFE	20	35.99	36.71	35.33	34.78	31.89	35.10	34.47	34.24
	50	36.83	36.64	36.82	37.40	35.15	36.00	36.46	37.19
	100	37.75	37.46	37.73	37.84	36.83	37.02	37.54	37.72
	200	37.38	37.58	37.55	37.46	36.57	37.25	37.31	37.31
MLE	20	16.90	9.60	6.55	4.52	1.05	-0.24	0.36	-0.04
	50	10.91	6.70	4.80	3.24	-0.68	0.004	0.11	-0.12
	100	7.30	4.97	3.43	2.54	0.60	0.07	-0.16	0.08
	200	5.69	3.41	2.50	1.72	0.27	0.04	0.15	-0.06
ATE	20	27.91	22.15	19.53	17.99	19.79	18.92	17.84	16.90
	50	24.57	15.62	12.22	10.16	20.16	12.46	10.32	9.16
	100	23.43	12.74	6.57	5.18	20.16	9.63	4.68	4.09
	200	21.71	7.49	3.90	2.44	18.70	5.56	2.46	1.43

Note: the DGP is similar to Table 4, exception that the AR(1) errors, The errors ε_{it} follows stationary AR(1) process with heteroskedasticity across each i. Let $\rho_{i\varepsilon} \sim iidU[0.5, 0.9]$, $\sigma_i \sim iidU[1, 4], \zeta_{it} \sim iid\chi^2(3) - 3$ for both i and $t = \{-49, ...0, ..., T\}$,

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

where $\varepsilon_{i,-50} = 0$.

view of Bias, our MLS1 has more advantages, when the sample is smaller and the Bias of MLS is a litter bigger than that of other estimators. For example, while $N = T = 20$, the Bias of MLS is -0.0016 and others are more bigger than it. Table 21 reports the results of DGP 8. From the view of RMSE, the MLE and ATE have the best efficient and Our MLS is better than that of IFE and CCE. Since this case is not em in CCE, the CCE has larger RMSE and Bias.

Table 22 reports the sample size and power of the wild dependent bootstrap under the sparsity induced weaker factor(Uematsu and Yamagata, 2022 JBES).

Table 23 reports the sample size of the wild dependent bootstrap and robust test in

Table 18: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with $f_t'\lambda_i$ in DGP 3 with the additive fixed effects and transformation

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	16.25	9.31	6.41	4.36	0.37	0.29	0.19	0.16
	50	9.55	5.63	4.12	2.80	-0.06	0.17	-0.25	0.02
	100	6.82	4.04	2.79	2.01	-0.19	-0.02	0.27	0.10
	200	4.75	2.90	1.979	1.44	0.04	-0.03	0.07	-0.06
CCE	20	53.03	50.84	52.84	52.53	42.52	43.25	46.35	46.42
	50	60.91	59.15	57.60	54.65	52.99	54.07	53.17	50.36
	100	60.15	58.77	57.88	56.38	53.93	55.19	54.83	53.87
	200	59.22	56.58	56.96	56.86	53.78	53.84	52.19	50.46
IFE	20	14.90	8.47	5.76	4.01	0.57	0.42	0.17	0.04
	50	9.43	5.56	3.89	2.65	0.14	0.10	-0.33	0.02
	100	6.67	3.96	2.61	1.92	-0.24	-0.03	0.27	0.07
	200	4.63	2.81	1.91	1.51	0.17	-0.02	-0.06	0.001
MLE	20	13.82	7.68	5.05	3.44	0.57	0.28	0.11	0.05
	50	8.91	5.06	3.40	2.46	0.36	0.03	-0.21	0.08
	100	6.48	3.69	2.50	1.79	-0.003	0.02	0.24	0.002
	200	4.51	2.72	1.79	1.34	0.11	0.08	0.001	0.001
ATE	20	15.82	8.57	5.67	3.91	0.64	0.35	0.14	-0.07
	50	9.36	5.69	3.88	2.64	-0.02	0.10	-0.33	0.004
	100	6.73	3.99	2.65	1.92	-0.88	-0.10	0.25	0.07
	200	4.86	2.84	1.91	1.45	-0.61	-0.03	0.07	0.01

Note: (1)(i) $x_{it,1} = 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + v_{it,1}$, $x_{it,2} = f_{t,1} + 2\lambda_{i,1} + 2f_{t,2} + 3\lambda_{i,2} + v_{it,2}$; (ii)$y_{it} = x_{it}'\beta + \lambda_i' f_t + \alpha_i + \tau_t + \varepsilon_{it}$,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 $\nu_t \sim iidN(0,1)$. (iii) the errors ε_{it} follow 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 $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.
(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).
(3) For the estimations of CCE, IFE, MLE and ATE, we firstly handle the data by the two-way within transformation, as the suggestion of Bai (2009) to eliminate the additive fixed effects.

Table 19: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with f_t in DGP 1

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	14.31	8.48	6.18	4.22	-0.04	-0.43	-0.49	-0.16
	50	9.52	5.38	3.94	2.80	-0.19	-0.27	-0.05	-0.11
	100	6.34	3.99	2.91	1.94	0.09	-0.09	-0.10	0.06
	200	5.08	2.82	1.94	1.47	-0.24	-0.13	0.16	-0.14
$MLS3$	20	21.22	11.15	7.67	4.94	-0.60	-0.60	-0.18	-0.08
	50	11.85	6.45	4.30	3.06	-0.03	-0.13	0.05	-0.12
	100	7.51	4.41	3.12	2.05	-0.11	-0.09	0.01	0.05
	200	5.90	3.07	2.06	1.52	-0.33	-0.11	0.21	-0.12
CCE	20	14.78	8.72	6.77	4.55	-0.07	-0.73	-0.47	-0.13
	50	9.11	5.68	4.07	2.90	-0.28	-0.26	-0.05	-0.19
	100	6.45	4.03	2.94	1.97	0.18	-0.04	-0.18	0.14
	200	4.89	2.78	1.93	1.49	-0.23	-0.12	0.14	-0.16
IFE	20	21.10	14.85	10.81	7.93	8.87	6.73	4.23	4.05
	50	18.89	10.73	6.20	3.83	7.82	3.40	1.86	0.86
	100	16.72	7.66	4.20	2.51	7.26	1.61	0.55	0.52
	200	13.42	5.41	2.89	1.76	4.19	0.54	0.51	-0.02
MLE	20	12.38	7.59	5.01	3.58	0.21	-0.19	-0.29	-0.10
	50	8.36	4.97	3.44	2.50	-0.20	-0.22	0.14	-0.20
	100	6.11	3.77	2.68	1.84	-0.17	-0.11	-0.11	0.03
	200	4.62	2.67	1.86	1.39	-0.14	-0.09	0.11	-0.13
ATE	20	16.37	9.10	6.60	4.38	3.85	1.71	0.70	1.27
	50	16.96	6.37	4.12	2.76	5.47	0.56	0.49	0.16
	100	12.93	8.45	3.07	1.95	4.30	1.33	0.16	0.14
	200	10.34	6.17	3.89	1.48	2.03	0.15	0.27	-0.06

Note: (1)(i) $x_{it,1} = 3f_{t,1} + 2f_{t,2} + v_{it,1}$, $x_{it,2} = f_{t,1} + 2f_{t,2} + v_{it,2}$; (ii) The errors ε_{it} follow AR(1) process with heteroskedasticity across each i. Let $\rho_{i\varepsilon} \sim iidU[0.5, 0.9]$,$\sigma_i \sim iid[0.8, 1.8]$,$\zeta_{it} \sim iid\chi^2(3) - 3$ for both i and $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1-\rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.
(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; $MLS3$: the Mundlak estimator β_{M3} in Section 3 of this paper; CCE: the Common Correlated Estimator (Pesaran,2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 20: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with f_t in DGP 2

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	20.21	10.24	6.37	4.50	-0.16	0.11	0.12	-0.04
	50	17.68	7.79	4.23	2.84	1.96	0.11	0.07	-0.16
	100	17.23	6.27	3.42	2.15	-0.26	0.26	0.03	0.10
	200	15.55	5.17	2.68	1.56	-1.14	-0.14	-0.17	-0.17
$MOLS3$	20	19.35	10.54	6.40	4.39	0.22	-0.40	-0.30	-0.15
	50	11.99	6.38	4.01	2.80	0.49	0.10	-0.09	-0.18
	100	8.70	4.74	3.25	2.09	0.63	0.29	-0.04	0.09
	200	6.13	3.53	2.08	1.51	-0.54	-0.20	0.04	-0.16
CCE	20	16.86	9.33	6.44	4.64	-1.13	-0.20	0.16	-0.15
	50	11.58	5.93	3.85	2.84	0.31	0.17	-0.15	-0.12
	100	10.68	4.55	3.08	2.06	-0.13	0.18	0.08	0.09
	200	8.81	3.50	2.11	1.44	-0.67	-0.04	0.06	-0.14
IFE	20	21.86	10.62	7.55	5.14	6.01	3.03	1.41	1.10
	50	13.43	6.43	3.97	2.84	2.73	0.78	0.33	0.12
	100	8.92	4.57	3.17	2.07	1.04	0.37	0.26	0.15
	200	6.01	3.10	2.04	1.44	0.49	0.06	0.13	-0.10
MLE	20	11.74	6.86	4.73	3.56	-0.92	0.12	-0.23	-0.17
	50	7.85	5.04	3.35	2.44	0.05	0.05	0.02	-0.17
	100	6.50	3.66	2.76	1.89	0.11	0.09	0.05	0.05
	200	4.21	2.70	1.77	1.33	-0.03	-0.15	0.02	-0.14
ATE	20	15.72	8.71	6.42	4.50	2.09	1.44	0.83	0.71
	50	12.56	5.89	3.79	2.74	0.61	0.37	0.26	0.07
	100	9.10	5.28	3.09	2.03	0.52	-0.02	0.16	0.14
	200	6.23	3.69	2.57	1.43	-0.09	-0.04	0.15	-0.12

Note: (1)(i) $x_{it} = \Gamma_i' f_t + v_{it}$; (ii)The errors ε_{it} follow 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 $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.
(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; $MLS3$: the Mundlak estimator β_{M3} in Section 3 of this paper; CCE: the Common Correlated Estimator (Pesaran,2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 21: The $RMSE$ and $Bias$ of β_1 with x_{it} are linear correlated with λ_i in DGP 3

		$RMSE \times 100$				$Bias \times 100$			
Method	T\N	20	50	100	200	20	50	100	200
$MLS2$	20	19.15	11.96	7.73	5.55	-1.29	0.54	0.59	0.05
	50	10.74	6.68	4.44	3.07	-0.36	-0.36	-0.20	-0.15
	100	7.07	4.57	3.19	2.18	0.05	0.31	-0.10	0.16
	200	4.91	3.03	2.15	1.52	0.03	-0.06	0.03	0.01
$MLS3$	20	25.61	14.16	9.54	6.38	-1.82	0.71	0.65	-0.09
	50	13.60	7.61	4.86	3.34	-0.10	-0.29	-0.21	-0.10
	100	8.38	5.17	3.34	2.29	0.10	0.40	-0.09	0.18
	200	5.61	3.21	2.31	1.57	-0.01	-0.09	0.04	-0.01
CCE	20	21.49	13.65	8.76	6.30	-0.85	0.37	0.40	-0.09
	50	13.20	8.73	6.11	4.04	0.31	-0.44	-0.24	-0.19
	100	9.55	5.96	4.12	2.87	-0.60	0.16	0.14	0.17
	200	6.85	4.56	3.04	2.10	-0.32	0.04	0.02	0.07
IFE	20	32.95	30.11	27.87	27.36	1.02	2.16	2.35	3.31
	50	18.37	13.56	11.18	10.62	2.20	1.29	1.64	2.22
	100	9.30	6.12	4.68	3.67	0.44	0.49	0.15	0.65
	200	5.77	3.31	2.39	1.72	0.19	0.09	0.10	0.09
MLE	20	12.49	7.00	4.96	3.57	-0.56	0.30	0.04	0.06
	50	8.04	4.99	3.65	2.45	-0.12	-0.28	-0.22	-0.11
	100	5.95	3.86	2.59	1.86	-0.01	0.17	-0.09	0.05
	200	4.48	2.66	1.94	1.37	0.14	0.02	-0.10	0.03
ATE	20	15.82	8.36	5.83	4.31	-0.52	0.80	0.32	0.06
	50	12.87	5.55	3.63	2.45	1.38	-0.27	0.001	-0.07
	100	8.24	5.00	2.62	1.75	0.31	0.35	0.003	0.07
	200	5.61	3.21	2.24	1.40	0.14	0.06	0.08	0.03

Note: (1)(i) $x_{it,1} = \lambda_{i1} + \lambda_{i2} + v_{it,1}$, $x_{it,2} = 2\lambda_{i1} + 3\lambda_{i2} + v_{it,2}$; (ii)The errors ε_{it} follow 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 $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i(1 - \rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$.
(2) $MLS2$: the Mundlak estimator β_{M2} in Section 2 of this paper; $MLS3$: the Mundlak estimator β_{M3} in Section 3 of this paper; CCE: the Common Correlated Estimator (Pesaran, 2006); IFE: the iterative PCA estimator (Bai, 2009); MLE: the maximum likelihood estimation of Bai and Li (2014); ATE: the average transformed estimator (Hsiao et al., 2021).

Table 22: The size of Bootstrap T testing with x_{it} are linear correlated with f_t and λ_i under weaker loadings

		Size				Power($\beta_1 = 1.1$)				Power($\beta_1 = 1.3$)			
ρ_ε	N\T	20	50	100	200	20	50	100	200	20	50	100	200
0	20	0.044	0.058	0.030	0.042	0.18	0.40	0.65	0.91	0.86	1.00	1.00	1.00
	50	0.046	0.036	0.050	0.054	0.42	0.78	0.96	0.99	1.00	1.00	1.00	1.00
	100	0.048	0.060	0.052	0.040	0.67	0.95	1.00	1.00	1.00	1.00	1.00	1.00
	200	0.042	0.052	0.052	0.042	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.3	20	0.044	0.064	0.066	0.040	0.20	0.33	0.47	0.81	0.75	0.97	1.00	1.00
	50	0.050	0.062	0.042	0.042	0.35	0.65	0.88	1.00	0.97	1.00	1.00	1.00
	100	0.060	0.060	0.062	0.046	0.56	0.87	0.99	1.00	1.00	1.00	1.00	1.00
	200	0.050	0.046	0.066	0.058	0.83	0.99	1.00	1.00	1.00	1.00	1.00	1.00
0.6	20	0.042	0.040	0.034	0.058	0.14	0.23	0.34	0.57	0.61	0.85	0.98	1.00
	50	0.046	0.050	0.056	0.062	0.25	0.42	0.68	0.89	0.90	1.00	1.00	1.00
	100	0.048	0.052	0.064	0.042	0.42	0.67	0.93	1.00	1.00	1.00	1.00	1.00
	200	0.030	0.046	0.054	0.022	0.66	0.91	1.00	1.00	1.00	1.00	1.00	1.00
0.9	20	0.054	0.046	0.058	0.032	0.09	0.12	0.14	0.21	0.34	0.46	0.65	0.87
	50	0.042	0.062	0.064	0.034	0.10	0.18	0.24	0.35	0.60	0.82	0.94	0.99
	100	0.056	0.050	0.044	0.046	0.22	0.27	0.44	0.61	0.85	0.98	0.99	1.00
	200	0.054	0.056	0.048	0.052	0.34	0.50	0.65	0.90	0.94	1.00	1.00	1.00

Note: (1)(i) $\lambda_{i1} \overset{iid}{\sim} N(1,2), \lambda_{i2} \overset{iid}{\sim} N(0.2,0.2)$,$f_t \sim AR(1), v_{it,1} \overset{iid}{\sim} N(0,1), v_{it,2} \overset{iid}{\sim} N(0,1)$; (ii) $x_{it,1} = 3f_{t,1} + \lambda_{i,1} + 2f_{t,2} + \lambda_{i,2} + v_{it,1},\ x_{it,2} = f_{t,1} + 2\lambda_{i,1} + 2f_{t,2} + 3\lambda_{i,2} + v_{it,2}$; (iii) Let $\lambda_i^* = (\lambda_{i1}^*, \lambda_{i2}^*)' \overset{iid}{\sim} N(0,1)$ and $f_t^* = (f_{t1}^*, f_{t2}^*)' \overset{iid}{\sim} N(0,1)$,the errors $\varepsilon_{it} = \lambda_i^{*\prime} f_t^* + \rho_\varepsilon \varepsilon_{i,t-1} + \sigma_i(1-\rho_{i\varepsilon}^2)^{0.5}\zeta_{it}$ with $\zeta_{it} \overset{iid}{\sim} N(0,1)$ and $\sigma_i^2 \overset{iid}{\sim} U[0.5,1.5]$. (2) Replications = 500; B = 500; (c) Sparsity Induced Weaker Factor(Uematsu and Yamagata, 2022 JBES) : $N^{0.7}$ elements in loadings are nonzero, eg. $200^{0.7} = 40.8$.

Table 23: The Bootstrap and Robust Tests of endogenous regressor

	Bootstrap				Robust Inference			
T\N	20	50	100	200	20	50	100	200
20	0.060	0.046	0.046	0.060	0.063	0.053	0.055	0.043
50	0.057	0.046	0.059	0.051	0.061	0.058	0.050	0.051
100	0.054	0.062	0.044	0.050	0.058	0.057	0.043	0.052
200	0.053	0.050	0.050	0.042	0.057	0.050	0.057	0.048

Note: (1) The size of Bootstrap T test $T_M^* = (\widehat{\beta}^*_{M,j} - \widehat{\beta}_{M,j})/\sqrt{\hat{V}^{boot}_{M,jj}}$, while x_{it} are linear correlated with f_t, λ_i and ε_{it}.The size of robust T test of Mundlak-CF estimator under Fixed b theory(5% level), while x_{it} are linear correlated with f_t, λ_i and ε_{it}
(2) $\lambda_{i1} \overset{iid}{\sim} N(1,2), \lambda_{i2} \overset{iid}{\sim} N(0.2, 0.2), f_t \sim AR(1), v_{it,1} \overset{iid}{\sim} N(0,1), v_{it,2} \overset{iid}{\sim} N(0,1)$; (ii) Let $z_{it} \overset{iid}{\sim} N(1,1)$, $x_{it,1} = f_{t,1} - 1.5\lambda_{i,1} + 2f_{t,1} + \lambda_{i,1} + z_{it} + v_{it,1}$, $x_{it,2} = f_{t,1} + 3\lambda_{i,1} - f_{t,1} + 2\lambda_{i,1} + v_{it,2}$; (3) the errors $\varepsilon_{it} = 0.9v_{it,1} + \rho_\varepsilon \varepsilon_{i,t-1} + \sigma_i(1-\rho^2_{i\varepsilon})^{0.5}\zeta_{it}$ with $\zeta_{it} \overset{iid}{\sim} N(0,1)$ and $\sigma_i^2 \overset{iid}{\sim} U[0.5, 1.5]$. (2) Replications = 500; B = 500.

the case of endogeneity.