FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

MURILO SEPULVIDA CARDOSO

INSTRUMENTAL VARIABLE WITH INTERACTIVE FIXED EFFECTS

São Paulo
2020

MURILO SEPULVIDA CARDOSO

Instrumental variable with interactive fixed effects

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas como requisito para obtenção do título de Mestre em Economia de Empresas.

Campo de Conhecimento:
Econometria Teórica

Orientador: Profa. Cristine Pinto

São Paulo

2020

Cardoso, Murilo Sepulvida.
Instrumental variable with interactive fixed effects / Murilo Sepulvida Cardoso. - 2020.
62 f.

Orientador: Cristine Pinto.
Dissertação (mestrado CMEE) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Análise de painel. 2. Variáveis instrumentais (Estatística). 3. Análise fatorial. 4. Modelos lineares (Estatística). I. Pinto, Cristine. II. Dissertação (mestrado CMEE) – Escola de Economia de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 330.115

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

MURILO SEPULVIDA CARDOSO

Instrumental variable with interactive fixed-effects

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas como requisito para obtenção do título de Mestre em Economia de Empresas.

Campo de Conhecimento:
Economia Política

Data de Aprovação:
___/___/______

Banca Examinadora:

____
Profa. Dra. Cristine Pinto (Orientador)
FGV-EESP

____
Prof. Dr. Bruno Ferman
FGV-EESP

____
Prof. Dr. Marcelo Fernandes
FGV-EESP

____
Prof. Dr. Marcelo J Moreira
FGV-EPGE

AGRADECIMENTOS

A trajetória que tive até aqui e a conclusão de mais esse passo importante na minha vida não poderiam ter ocorrido sem algumas pessoas de suma importância na minha vida. E é graça ao esforço, sacrifício e incentivo de uma dessas pessoas que hoje posso concluir mais um dos meus objetivos na vida que é a conclusão desse mestrado. Mãe, este feito não é só meu, é seu também. Sofrendo na pele o quão difícil é uma vida com poucas oportunidades você se esforçou ao máximo para me dar aquelas oportunidades que você não pôde ter. E hoje, tudo que eu consegui e consigo é possível graças a você. Sou muito grato por tudo que fez por mim. Obrigado, Rita de Cassia Sepulvida Cardoso, continue sendo essa mulher e mãe guerreira que sempre foi.

Não menos importante, agradeço também a minha irmã, Tamires Sepulvida Cardoso, que não só foi uma irmã mas, em muitos momentos, uma segunda mãe para mim. Sem você, eu sei que não teria chegado aonde cheguei. Sou e serei sempre grato a você por tudo que fez por mim. Ao meu pai, que infelizmente não pode estar aqui para ver este feito, mas que enquanto esteve aqui, nunca negou nada ao seu alcance para que pudéssemos ser felizes. Ao meu padrinho, Ricardo Pieroni, que sempre me incentivou a estudar, aos meus sogros Antonio Albino e Roseane Silva de Moura Albino e aos meus cunhados Richard Selma Carvalho e João Pedro de Moura Albino.

Além disso, durante o mestrado tive amigos que espero levar para vida toda e que foram de grande ajuda sempre. Seja com matérias, materiais e discussões sobre projetos ou com boas conversas pós aula na biblioteca e no bar a cada fim de trimestre. Obrigado, Angélica Brum, Antonio Leon, Kelly Gonçalvez, Jessé Pizzino, Luiz Felipe Fontes, Bruno Tebaldi, Otávio Conceição, Matheus Antonhy e Renan Alvez.

Tive oportunidade, também, de trabalhar com profissionais incríveis que me ajudaram a crescer como pesquisador. Obrigado professores Bruno Ferman e Marcelo Fernandes, sem o auxílio de vocês este trabalho não teria acontecido. Obrigado pelas orientações, críticas e dicas, tanto nesse, como em outros projetos e, também, para minha carreira. Agradeço ao CNPq pela bolsa durante o mestrado.

Por fim, mas não menos importante, pelo contrário, gostaria de agradecer à minha futura esposa, Carolina de Moura Albino, que sempre esteve ao meu lado nesses últimos quase 8 anos. Obrigado pelo companheirismo, pelo apoio, pela compreensão, e pelo amor. Você faz parte dessa conquista. Espero conseguir retornar todo o apoio que você me deu e me dá, muito obrigado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

ABSTRACT

The biggest challenge of the instrumental variable model is finding a valid instrument, one that satisfies the assumptions of relevance and, especially, exogeneity. To impose a less restrictive assumption, some estimators combine unobservable effects with instrumental variables, as example IV-FE. In this paper, we study a large-panel instrumental variable model where the instrument may be correlated with unobservable variables, even when they vary across both dimensions. This variation implies that standard approaches in the literature, such as IV, IV-FE, and linear factor models (Pesaran, 2006; Bai, 2009), are inconsistent. We construct two and show their $\sqrt{NT}$ convergence under both large N and large T. Our exogeneity assumption is less restrictive than the standard instrumental variable model assumption since the instrument shall be exogenous given the factor structure. We show their asymptotic normality distribution for some rate of N/T even when the errors have autocorrelation and heteroskedasticity in both dimensions. We also study the trade-off between our estimators and a standard IV estimator when the error has an interactive fixed-effect structure, and the instrument is valid. In this case, we show that our estimator is more efficient than IV if the variance of the factor structure is sufficiently large than the error variance and less efficient if the variance of factor structure is sufficiently smaller.

Keywords: Instrumental variable, Linear factor model, interactive fixed-effect model, endogenous regressors, large panel.

JEL Codes: C23, C26

RESUMO

O maior desafio do modelo de variável instrumental é encontrar um instrumento valido, que satisfaça as hipóteses de relevância e, especialmente, exogeneidade. Para impor uma hipótese menos restritiva, alguns estimadores combinam efeitos não observáveis com variável instrumental, como por exemplo IV-FE. Neste artigo, nós estudamos um modelo de variável instrumental em *large panel* onde o instrumento pode ser correlacionado com variáveis não observadas, mesmo se elas variam nas duas dimensões. Essa variação implica que os métodos padrões na literatura, como IV, IV-FE e modelo de fatores lineares (Pesaran, 2006; Bai, 2009), são inconsistentes. Nós construímos dois estimadores e mostramos suas $\sqrt{NT}$-convergências quando N e T tendem ao infinito. Nossa hipótese de exogeneidade é menos restritiva que a hipótese do modelo de variável instrumental padrão dado que o instrumento deve ser exógeno dado a estrutura de fatores. Nós mostramos suas normalidades assintóticas para algumas taxas de N/T mesmo para o caso onde os erros são auto-correlacionados e heterocedásticos nas duas dimensões. Nos também estudamos o *trade-off* entre nossos estimadores e os estimador padrão de IV quando os erros destes apresentam uma estrutura de *interactive fixed-effect* e o instrumento é valido. Neste caso, mostramos que nosso estimador é mais eficiente que o IV se a variância da estrutura de fatores for suficientemente maior do que a variância do erro e menos eficiente se a variância da estrutura de fatores for suficientemente pequena.

Palavras-Chave: Variável instrumental, Linear factor model, modelo de efeitos fixos interativos, regressores endógenos, dados em painel

Códigos JEL: C23, C26

List of Figures

List of Tables

TABLE OF CONTENTS

1 Introduction

Consider estimating the causal effect of X_{it} on Y_{it} in a linear panel regression model. If the regressors X_{it} are orthogonal to the error term, then pooled OLS is consistent. However, this orthogonality assumption is very restrictive and often not met. To resolve this issue, there are two alternative approaches.

The first approach involves controlling for unobserved effects. It assumes that one can estimate, or somehow eliminate, the unobserved variables that correlate with X_{it} (Wooldridge, 2002). The principal is the fixed-effects model (FE), that decomposes errors into individual fixed effect, time fixed effect, and idiosyncratic error, where individual fixed effect and time fixed effect are constant across i and t, respectively. The key assumption is the strict exogeneity between the regressors (X_{it}) and the idiosyncratic error. Some papers relax the need for unobserved factors to be invariant in at least one dimension by using a linear factor model or latent factor models, requiring less restrictive assumptions to estimate β consistently (Holtz-Eakin et al., 1988; Murtazashvili and Wooldridge, 2008; Abadie et al., 2010; Bai, 2009; Pesaran, 2006). The structure they use is less restrictive and gives more flexibility for the error structure than the fixed effect models. The unobserved factor structure can vary across i and t, which is not possible in the fixed effect models. In general, the cost is that the panel data must be large in two dimensions (i.e., $N, T \to \infty$) or the data generation process must be restricted.

The second approach relies on instrumental variables (IV). The difficulty of this approach is to find a valid and strong instrument since, in this case, the instrument must be exogenous to all unobserved variables. There are a large number of estimators that combine instrumental variables and unobserved effects. The most popular is the fixed-effect instrumental variable (FE-IV), that relaxes the exogeneity hypothesis. It allows individual- and time-invariant unobserved variables to be correlated with the instrument.

Consider the model:

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + \varepsilon_{it} \tag{1}$$

In sum, to estimate β consistently with fewer restrictive assumptions, the literature suggests two ways: the first is estimating the factor structure without restriction, as used by Bai (2009), but assuming the strict exogeneity between X_{it} and ε_{it}. The second assumes that exists an instrumental variable (Z_{it}) which is orthogonal to the error (ε_{it}) and the factor structure ($\lambda_i' F_t$), or puts some restriction on factor structure when assuming only the exogeneity of the instrument with the error.

In this paper, we take a different route by assuming that the instrument is orthogonal to the error in Equation (1) without any restriction on the factor structure.[1] In particular, we propose

[1] To estimate the factor structure it is necessary to put restrictions on the estimated factors. However, these

an instrumental variable approach within a linear factor model with unobservable factors. Our estimator is consistent when N and T are large, requiring similar assumptions to the standard instrumental variable estimator. The only difference is that one must condition the exogeneity and relevance of the instrumental variables on the factor structure.

The first assumption is stronger than assumptions for a standard IV estimator. The correlation of instrument and covariate needs to be different from zero, after conditioning on the factor structure. The second assumption, on the other hand, is less restrictive, since the instrument needs not be exogenous to all non-observable variables in the model. It only needs to be uncorrelated with the unobservable variables not captured by the factor structure, introducing several new valid instruments. We can think about this as a control in a standard IV approach; in fact, we are controlling for the factor structure.

Other researchers have constructed instrumental estimators together with interactive fixed effects. However, they often impose more restrictive exogeneity assumptions on the instrument or more restrictions about the data generation process and the convergence rate. Harding and Lamarche (2011) utilize the estimator constructed by Pesaran (2006) to create an IV estimator. They need to assume that the instrument must be exogenous to loading factors that affect the dependent variable.[2] In this case, if T is finite, $N \to \infty$, and under some conditions about the factors generation process, Murtazashvili and Wooldridge (2008) show that FE-IV is also consistent. However, our estimators is more broadly applicable. In our approach, we put any restriction on the correlation between the instrument and the factor structure. That is, the instrument can be correlated with the loading factor, the common factor or both. The closest approaches to ours are the ones proposed by Lee et al. (2012), Moon et al. (2018) and Moon and Weidner (2017). Nevertheless, they only show asymptotic normality for $T/N \to \rho^2$ where $0 < \rho < \infty$ and in absence of error correlation in both dimensions.[3] On the contrary, we show the asymptotic normality for some cases, even when $T/N \to 0$, $N/T \to 0$ or in the presence of correlation.

This paper is divided as follows. Section 2 proposes two estimators: one estimated in two steps, called Two-Step, and the other in one step, called Joint. In Section 3, we discuss the assumptions necessary for consistency. We establish consistency and asymptotic distribution in Section 4. Section 5 examines the finite sample properties and compares the bias of our estimators to other extant estimators in the literature through Monte Carlo simulations. Section 6 compares our estimators and standard IV estimator when IV exogeneity assumption holds. Section 7 concludes. We relegate all technical proofs to Appendix B.

---

restrictions representing no restriction for the model since the restricted factors span the same space as the unrestricted factors.

[2] Pesaran (2006) assumes that λ is uncorrelated to the covariate.

[3] Moon and Weidner (2017) does argue that correlation in the cross-section represents no problem for inference. However, correlation in both dimensions is never allowed.

2 Estimation

We examine the panel data model

$$Y_i = X_i\beta + F\lambda_i + \varepsilon_i \tag{2}$$

where

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix}, X_i = \begin{bmatrix} X'_{i1} \\ X'_{i2} \\ \vdots \\ X'_{iT} \end{bmatrix}, F = \begin{bmatrix} F'_1 \\ F'_2 \\ \vdots \\ F'_T \end{bmatrix}, \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix},$$

Y_{it} is a scalar, and X_{it} is a $p \times 1$ matrix. F_t and λ_i are $r \times 1$ matrix.

Define $\Lambda = (\lambda_1, \lambda_2, \cdots, \lambda_N)'$. The equation above in matrix form is

$$Y = X\beta + F\Lambda' + \varepsilon \tag{3}$$

where $Y = (Y_1, Y_2, \ldots, Y_N)$ is a $T \times N$ matrix, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)$ is a $T \times N$ matrix, X is a three-dimensional matrix with p sheets of $T \times N$ (i.e. $T \times N \times p$), and β is a vector with p coordinates when the jth element of β is associated with the jth sheet of X. F is $T \times r$ and Λ is an $N \times r$ matrix.

Assume that there exists a vector of instruments, Z, such that

$$Z_{jit} = \lambda'_{z,jit}F + e_{it} \qquad j \in \{1, 2, ..., K\}$$

where $Z_{it} = [Z_{1it}, Z_{2it}, \ldots, Z_{Kit}]'$ is a $K \times 1$ matrix, $Z_i = [Z_{1i}, Z_{2i}, \ldots, Z_{Ki}]'$ is a $T \times K$ matrix and $\lambda'_{z,jit}$ is the loading factor for equation of Z.

REMARK 1. Each one of the variables X_{it}, Z_{it} and Y_{it} may have common factors that affect more than one variable and idiosyncratic common factors that affect only itself. For example, $Y_{it} = \beta X_{it} + \tilde{\lambda}'_{y,i} f_t + \tilde{\lambda}'_{yy,i} f_{y,t} + \varepsilon_{it}$, $X_{it} = \tilde{\lambda}'_{x,i} f_t + \tilde{\lambda}'_{xx,i} f_{x,t} + e_{x,it}$ and $Z_{it} = \tilde{\lambda}'_{z,i} f_t + \tilde{\lambda}'_{zz,i} f_{z,t} + e_{z,it}$, where f_t are the common factors that affect all variables and $f_{K,t}$ are the idiosyncratic common factors of the variable K. We can define $F_t = (f, f_{y,t}, f_{x,t}, f_{z,t})$ and Λ_k so that $\lambda'_{k,i}F_t$ is equal to the relevant factor structure for the equation of k. In the example above, for Y_{it} we define $\lambda_i = (\tilde{\lambda}_{y,i}, \tilde{\lambda}_{yy,i}, \mathbf{0}, \mathbf{0})$, for X_{it}, $\lambda_{x,i} = (\tilde{\lambda}_{x,i}, \mathbf{0}, \tilde{\lambda}_{xx,i}, \mathbf{0})$ and, for Z_{it}, $\lambda_{z,i} = (\tilde{\lambda}_{z,i}, \mathbf{0}, \mathbf{0}, \tilde{\lambda}_{zz,i})$.

REMARK 2. Correlation in the remark above comes from the common factor structure, although it is straightforward to entertain correlation from loading factors.

Let $S = \{1, 2, ..., p\}$. Rewrite X_i and Y_i in reduced form

$$X_{1i} = Z_i\pi_{11} + F\lambda_{1i} + \nu_{11i}$$
$$X_{2i} = Z_i\pi_{12} + F\lambda_{2i} + \nu_{12i}$$
$$\vdots$$
$$X_{pi} = Z_i\pi_{1k} + F\lambda_{pi} + \nu_{1ki}$$
$$Y_i = Z_i\pi_2 + F\lambda_i + \nu_{2i}$$

where Y_i, X_{ji} and ν_{ji} for every $j \in S$ is $T \times 1$. Z_i is $T \times K$, π_{1j} is a $K \times 1$ vector for every $j \in S$ and π_2 is $K \times 1$ vector.

The system of equations above has (p + 1) * T equations, with a unique solution if $F'F/T = I$ and $\Lambda'\Lambda =$ diagonal, and $E[Z_i'M_F X_i]$ has a full rank. The first two assumptions do not represent any effective restriction for the model; this is only a rotation (normalization). They are necessary for the estimation of the factor structure. Otherwise, since the common factor and loading factor are estimated parameters, we would have more parameters to estimate than equations. For more detailed analysis, see Bai and Ng (2002) and Bai (2009). The last condition is a standard rank condition in an instrumental variable approach. However, this condition needs to hold not for the moments of $Z_i'X_i$, but in the space generated by this variable that is orthogonal to the factor structure.

We construct two estimators. One is called Two-Step, that is similar to the two least square estimators. The other, called Joint, will estimate all p + 1 equations jointly.

2.1 Two-Step Estimation

The Two-Step procedure is very similar to the two least squares estimators. However, in the first and the second stages, we also estimate the loading and common factors. Our first stage is defined by the p first equations of the system composed by the reduced-form equations.

Define
$$\Pi_1 = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{p1} \end{bmatrix}.$$

Therefore our procedure consists of the following steps.

Estimation Procedure I - Two-Step

1. Estimate π_{j1}, $F_{1s,j}$ and $\lambda_{1s,j}$ for all $j \in S$ with the p first equation in the reduced-form system using Bai (2009) and construct

$$\hat{X}_i = M_{\hat{F}} Z_i \hat{\Pi}_1$$

where $M_{\hat{F}} = I - \hat{F}\hat{F}'/T$.

2. Estimate β using Bai (2009) by the equation below:

$$Y_{it} = \hat{X}_{it}'\beta + \lambda_{2i}'F_t + \nu_{2it}. \tag{4}$$

The first step procedure is generated by an interactive process where $\hat{\pi}$ is a function of F, and $\hat{F}$ is a function of π:

$$\hat{\pi}_{1j}(F, \Lambda) = \left(\sum_{i=1}^{N} Z_i'Z_i\right)^{-1} \sum_{i=1}^{N} Z_i'(X_{ji} - F\lambda_i)\ , \forall j \in S. \tag{5}$$

For each F and Λ, we estimate π_{1j} using (5) and for each π_{1j} we estimate F and λ by the pure factor model $W_{ji}(\pi_j) = X_{ji} - Z_i\pi_j = F\lambda_{ji} + \varepsilon_{ji}$, see Bai (2009) for more details.

In the second step, the procedure is the same, but uses the covariate $\hat{X}_i$ to do the same interactive process. By construction, $\hat{X}_{it}$ a vector of variables that represent all variation in X_i driven only by the variation on the instrumental variable (Z_i) such that is orthogonal to the factor structure. Under assumptions explained in Section 3, the instrument is orthogonal to the error in Equation (2). Thus, the estimator for β can be given by the system below:

$$\hat{\beta} = \left(\sum_{i=1}^{N} \hat{X}_i'M_{\hat{F}}\hat{X}_i\right)^{-1} \sum_{i=1}^{N} \hat{X}_i'M_{\hat{F}}Y_i \tag{6}$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\hat{\beta}\right)\left(Y_i - X_i\hat{\beta}\right)'\right]\hat{F} = \hat{F}V_{NT} \tag{7}$$

where V_{NT} is a diagonal matrix that consists of the r largest eigenvalues of Matrix 5 given by Bai (2009), arranged in decreasing order (Bai, 2009).

REMARK 3. In practice, it is unnecessary, to re-estimate the factors in the second stage to maintain consistency. In fact, we can use the estimated factors from the first stage in the second stage. All non-estimated factors in the first stage are not relevant to identification. If such factors exist, they are correlated neither with X_{it} nor with Z_{it}. However, if Y_i has specific factors, controlling for these factors in most cases leads to a more precise estimation.

2.1.1 Alternative estimation

Bai's (2009) procedure is more efficient when the mean of factors is zero. He shows that when the factors have non-zero mean, or if the model has time and (or) an individual fixed effect,[4] it is more efficient to utilize a fixed-effect model together with the factor model:

$$Y_i = X_i\beta + \mu + \alpha_i + \xi_t + F\lambda_i + \varepsilon_i. \tag{8}$$

A better estimation is to transform this model as above before performing the procedure with Bai's (2009) algorithm.

$$\dot{Y}_i = \dot{X}_i\beta + F\lambda_i + \dot{\varepsilon}_i \tag{9}$$

where

$$\dot{u}_{it} = u_{it} - \overline{u}_{i.} - \overline{u}_{.t} + \overline{u}_{..}$$

$$\overline{u}_{i.} = \frac{1}{T}\sum_{t=1}^{T} u_{it}, \quad \overline{u}_{.t} = \frac{1}{N}\sum_{i=1}^{N} u_{it}, \quad \overline{u}_{..} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}.$$

We can repeat the same procedure above; now, in Equation (9). With this procedure, we impose that the factor structure has zero mean; consequently, the method of Bai (2009) is more efficient in this environment. One point that emerges here is that if the factor structure has zero mean, the non-transformed model is more efficient than the transformed model.

Therefore, our estimator in the first and second stage is now equal to

$$\hat{\pi}_{1j}(F,\Lambda) = \left(\sum_{i=1}^{N} \dot{Z}_i'\dot{Z}_i\right)^{-1} \sum_{i=1}^{N} \dot{Z}_i'(\dot{X}_{ji} - F\lambda_i)\,, \forall j \in S \tag{10}$$

$$\hat{\beta} = \left(\sum_{i=1}^{N} \dot{\hat{X}}_i' M_{\hat{F}} \dot{\hat{X}}_i\right)^{-1} \sum_{i=1}^{N} \dot{\hat{X}}_i' M_{\hat{F}} \dot{Y}_i. \tag{11}$$

Another alternative estimation is possible. Before the second stage we multiply Equation (4) by $M_{\hat{F}}$ and continue to implement the Estimation Procedure I - Two-Step. This algorithm is the same except for equations in the second stage that are defined as

$$M_{\hat{F}} Y_{it} = M_{\hat{F}} \hat{X}_{it}'\beta + M_{\hat{F}} \lambda_{2i}' F_t + M_{\hat{F}} \nu_{2it}$$

$$Y_{it}^{\oint} = \hat{X}_{it}'\beta + \lambda_{2i}^{\oint\prime} F_t + \nu_{2it}^{\oint}.$$

This procedure is more economical in terms of computational cost. Unlike the first method, here we do not re-estimate all the factors in the second step. We only estimate the

[4] It is easy to show that the model $Y_i = X_i\beta + \mu + \alpha_i + \xi_t + F\lambda_i + \varepsilon_i$ is a homeomorphism to the equation $Y_i = X_i\beta + F\lambda_i + \varepsilon_i$ when $E[F_t] \neq 0$ and $E[\lambda_i] \neq 0$ and X_{it} has a constant.

idiosyncratic factors of Y in the second step. In terms of asymptotic variance and consistency, however, there is no difference.[5]

The two algorithms above can be used together to achieve greater efficiency and a less computationally costly algorithm in most cases.

The reduced forms are written as

$$\begin{aligned}
\dot{X}^{\oint}_{1i} &= \dot{Z}^{\oint}_{i}\pi_{11} + F\lambda^{\oint}_{1i} + \dot{\nu}^{\oint}_{11i} \\
\dot{X}^{\oint}_{2i} &= \dot{Z}^{\oint}_{i}\pi_{12} + F\lambda^{\oint}_{2i} + \dot{\nu}^{\oint}_{12i} \\
&\vdots \\
\dot{X}^{\oint}_{pi} &= \dot{Z}^{\oint}_{i}\pi_{1k} + F\lambda^{\oint}_{pi} + \dot{\nu}^{\oint}_{1ki} \\
\dot{Y}^{\oint}_{i} &= \dot{Z}_{i}\pi_{2} + F\lambda^{\oint}_{i} + \dot{\nu}^{*}_{2i}.
\end{aligned}$$

Estimation Procedure II - Two-Step alternative estimation

1. Estimate π_{j1}, $F_{1s,j}$ and $\lambda_{1s,j}$ for all $j \in S$ with the p first equation in the reduced-form system using Bai (2009) and construct

$$\hat{\dot{X}}_i = M_{\hat{F}}\dot{Z}_i\hat{\Pi}_1$$

where $M_{\hat{F}} = I - \hat{F}\hat{F}'/T$.

2. Estimate β using Bai (2009) by the equation below:

$$\dot{Y}^{\oint}_{it} = \hat{\dot{X}}^{\oint\prime}_{it}\beta + \lambda^{\oint\prime}_{2i}F_t + \dot{\nu}^{\oint}_{2it}.$$

REMARK 4. The procedure above uses $\dot{Z}$ as an instrument for $\dot{X}$. In other words, we are using the between transformation model with Bai (2009). Therefore, in a dynamic panel data model we cannot use Procedure II - Two-Step alternative estimation. Nickell (1981) shows that in a model with dynamic panel data, we cannot use the between transformation. It generates a bias since, necessarily, the instrumental variable will be correlated with the error. In these cases, therefore, we cannot use the second procedure. However, Procedure I - Two-Step remains consistent.

2.2 Joint Estimation

Given the necessary assumptions, our model can consistently estimate all the reduced-form equations using a linear factor procedure created by Bai (2009). Because our estimation issues come from latent factors that affect all variables, or a non-singular subgroup of them that

[5] For more details see Bai and Li (2017).

contains Y, when we jointly estimate all equations, we gain variation to better estimate these latent factors. The cost here is that each new variable adds more specific factors to our estimation. For each of them an additional $2p(N+T)$ factors need to be estimated, with most equal to zero.[6]

Define

$$Q = \begin{bmatrix} Y \\ X_1 \\ \vdots \\ X_p \end{bmatrix}, D_j = \begin{bmatrix} \mathbf{0}_{j-1} \\ 1 \\ \mathbf{0}_{p-j} \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} Z \\ \vdots \\ Z \end{bmatrix}, \nu = \begin{bmatrix} \nu_2 \\ \nu_{11} \\ \vdots \\ \nu_{1p} \end{bmatrix}$$

and $\mathbf{0}_k$ is a $k \times 1$ matrix of zeros. Given all reduced form, we can jointly estimate them as

$$Q = 1_{(p+1)T\times N}\alpha + \mathbf{Z}\gamma_1 + \sum_{j\in S}(D_j \otimes Z)\gamma_{2j} + F\Lambda + \nu \tag{12}$$

$$\beta_j = f(\gamma_1, \gamma_{2j})$$

where $f(.)$ is a minimum distance function continuous at all points of $\gamma_1 \neq 0$.

For example, if there are one covariate and one instrument, then $f(\gamma_1, \gamma_2) = \frac{\gamma_1+\gamma_2}{\gamma_1}$ which is equivalent to $\frac{\pi_2}{\pi_1}$. Where π_1 and π_2 are defined on reduced-form equations.

2.2.1 The asymmetric property

Equation (12) is not symmetric due to the construction of $Y, X_1, \ldots, X_p$. If we define these as $T \times N$ matrices then Q, D_j, Z are $(p+1)T \times N$ matrices. More importantly, F and Λ are $(p+1)T \times k$ and $k \times N$ matrices, respectively. This provides more variance to estimate the loading factors (Λ). If we re-define the variables $Y, X_1, \ldots, X_p$ as $N \times T$ such that Q, D_j, Z are $(p+1)N \times T$, then F and Λ are $T \times k$ and $k \times (p+1)N$ matrices, respectively. Thus, we gain more variance to estimate the common factors (F). Here it is not clear which method is better for bias reduction. The decision depends on T, N, T/N, $cov(Z_{it}, F)$, $cov(Z_{it}, \Lambda)$ and the number of specific loading and common factors. However, the linear factors estimate has a small sample bias term that depends negatively on the $min(T, N)$.[7] Therefore, for cases when T (or N) is small, it is better to define the variables as a $(p+1)N \times T$ (or $(p+1)T \times N$) matrix.

3 Assumptions

In this section, we show all assumptions necessary for consistency of our estimators. The first four assumptions come from Bai (2009) and are required to consistently estimate the

[6] Not all parameters are estimated together. Usually, and in the Bai (2009) procedure, we estimate only one of the common factors and loading factors. The other is estimated as a residual of the model.

[7] See Bai and Ng (2002) and Bai (2009).

equations in reduced forms. The last assumption is the necessary condition for the instrumental variable estimator to be consistent and plays a central role in the identification of the model.

The important matrices in our limiting theory are defined above:[8]

$$W_i(Z,F) = M_F Z_i - \frac{1}{N}\sum_{k=1}^{N} M_F Z_k \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k \tag{13}$$

$$D_Z(F) = \frac{1}{NT}\sum_{i=1}^{N} W_i'(Z,F)W_i(Z,F). \tag{14}$$

Observing the factor F_t (or λ_i), a condition for consistency is that $D_Z(F)$ is positive definite. However, we estimate both common and loading factors in this framework. Therefore, we need a stronger assumption (Bai, 2009). This leads to Assumption 1.

Let $||A|| = (tr(A'A))^{1/2}$ and $M \in \Re$.

Assumption 1. *$E||Z_{it}||^4 < M$ and $E||X_{it}||^4 < M$. Let $\mathcal{F} = \{F : F'F/T = I\}$. We assume*

$$\underset{F\in\mathcal{F}}{Inf}\, D_Z(F) > 0.$$

Assumption 2.

(i) $E||F_t||^4 < M$ *and* $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F > 0$ *for some* $r \times r$ *matrix* Σ_F, *as* $T \to \infty$

(ii) $E||\lambda_i||^4 < M$ *and* $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\lambda > 0$ *for some* $r \times r$ *matrix* Σ_Λ, *as* $N \to \infty$

Assumption 2 relates to the existence of r factors. That is, there exists a finite number of factors such that all variance of unobserved factor structure may be represented by them. This assumption is not strongly restrictive. For example, Bai (2009) argues that the common factor can be a dynamic process such that $F_t = \sum_{i=1}^{\infty} C_i e_{t-i}$, where e_t is i.i.d. with zero mean. The same reasoning may be extended to the loading factor (λ_i).

Assumption 3. *Serial and Cross-Sectional Weak Dependence and Heteroskedasticity:*

For all $a_{it} \in \{\nu_{2it}, \nu_{11i}, \nu_{12i}, ..., \nu_{1Ki}\}$

(i) $E(a_{it}) = 0$ *and* $E|a_{it}|^8 \leq M$.

(ii) Let $E(a_{it}a_{js}) = \sigma_{a,ij,ts}$.

[8] Bai (2009) argues these matrices are equal to a deviation of $M_F Z_i$ from its weighted average.

$|\sigma_{a,ij,ts}| \leq \bar{\sigma}_{a,ij}$ for all (t,s) and $|\sigma_{a,ij,ts}| \leq \bar{\sigma}_{a,ts}$ for all (i,j), such that

$$\frac{1}{N}\sum_{i,j=1}^{N}\bar{\sigma}_{a,ij} \leq M, \frac{1}{T}\sum_{t,s=1}^{T}\bar{\sigma}_{a,ts} \leq M, \frac{1}{NT}\sum_{i,j,t,s=1}|\sigma_{a,ij,ts}| \leq M.$$

The largest eigenvalue of $\Omega_i = E(a_i a_i')$ $(T \times T)$ is bounded uniformly in i and T.

(iii) For every (t,s), $E\left|N^{-1/2}\sum_{i=1}^{N}\left[\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})\right]\right|^4 \leq M$.

(iv) Moreover

$$T^{-2}N^{-1}\sum_{t,s,u,v}\sum_{i,j}|\text{cov}(a_{it}a_{is}, a_{ju}a_{jv})| \leq M.$$

$$T^{-1}N^{-2}\sum_{t,s}\sum_{i,j,k,\ell}|\text{cov}(a_{it}a_{jt}, a_{ks}a_{\ell s})| \leq M.$$

Assumption 4. *For all $a_{it} \in \{\nu_{2it}, \nu_{11i}, \nu_{12i}, ..., \nu_{1Ki}\}$. a_{it} is independent of λ_j and F_s for all i, t, j and s.*

Assumptions 3 and 4 come directly from Bai (2009) except that here, this procedure works for all reduced-form equations. Therefore, we need to impose these assumptions for all errors in reduced forms of X_i, that is for $\nu_{11i}, ...\nu_{1Ki}$, and for the error in reduced form of Y_i (ν_{2it}). Since the error in the reduction form of Y_i is a function of ε_i and the errors $\nu_{11i}, ...\nu_{1Ki}$, it is sufficient to impose the restriction on $\varepsilon_i, \nu_{11i}, \nu_{12i}, ..., \nu_{1Ki}$. This implies that F_t and λ_i are independent of ν_{2it} for all i and t.

However, for Assumption 3, especially item iv, since in general $cov(\varepsilon_{it}\nu_{h1js}) \neq 0$ for some (i,t), (j,s) and $h \in \{1, ..., K\}$,[9] this assumption is a little more restrictive than the assumption for the set ε_{it}, $\nu_{11i}, ...\nu_{1Ki}$.

Assumption 5. *Instrumental variable assumption*

*(i) **(Rank condition)** $\frac{1}{NT}\sum_{i=1}^{N}E(X_i M_{F^0} Z_i)$ has full rank.*

*(ii) **(Exogeneity)** ε_{it} is independent of Z_{js} for all i, t, j and s.*

Assumption 5 is the principal assumption; it gives the necessary condition for the instrumental variable to be consistent in this framework. The rank condition here is very similar to the instrumental variable approach.[10] However, a stronger assumption is needed, that Z_{it} is correlated to X_{it} given the factor structure. In contrast, the second part of the assumption is less restrictive. The instruments needs exogeneity, given the factor structure. That is, the instrumental variable can be correlated with the unobserved (and observed) factor structure.

9 Otherwise Bai's (2009) method works and the instrumental approach is not necessary.
10 In a standard IV model, $E[X_i Z_i]$ must have full rank.

We can think about this as an IV model with control variables. In our case, the factor structure represents our control variables. The benefit here is that even when the unobserved factor structure is correlated with the instrument, we can construct a consistent estimator for the causal effect of X on Y. Thus, in many cases, we have a new world of instrumental variables that had not previously been valid. Now, they are, and with less-restrictive assumptions.

REMARK 5. Assumption 5 rules out dynamic panel data models (Blundell and Bond, 1998; Bond, 2002). The procedure works well even with lagged dependent variables. However, when X_{it} has lagged variables we need to assume that ε_{it} is not serially correlated (Bai, 2009).

REMARK 6. Assumption 4 also rules out shift-share models (Bartik, 1991; Blanchard et al., 1992; David et al., 2013; Jaravel, 2019). The instrumental variable in this model is a factor structure. The common factor is the exogenous shock and the loading factor is the country, city or region shares of the industry, sector or other elements affected by the shock. Thus, when the non-observable factor structure is a part of the shift-share's instrumental variable, Assumption 4 does not hold. However, when the standard assumptions of the shift-share model hold, our estimators will be consistent. Moreover, in some situations, our estimators have less variance than the shift-share estimator, see Section 6 for more details.

4 Limit Theory

Denote $(\beta^0, F^0, \Lambda^0)$ the true parameters of the model in Equation (3), where F^0, Λ^0 satisfies Assumption 2. In this section, we show the consistency and asymptotic theory of our estimators. We will treat the dimension of F^0 and Λ^0 as known. We can adapt the method of Bada and Kneip (2014), who show that the distribution is the same whether we know the dimension of factors, or if we do not know the dimension of factors and estimate them. Moon and Weidner (2015) also show that, under some conditions, the asymptotic distribution is independent of the number of factors if you do not underestimate the real number. All proofs can be found in Appendix B.

4.1 Two-Step - Limiting theory

Let $F_{1s,j}$ be the common factor for the first stage equation of X_j, for all $j \in S$. Therefore, in the first stage we are minimizing Equation (15).[11]

For all $j \in S$

$$S_{1S-j,NT}(\pi_{1j}, F) = \frac{1}{NT}\sum_{i=1}^{N}(X_{ji} - Z_i\pi_{1j})' M_F (X_{ji} - Z_i\pi_{1j}) - \frac{1}{NT}\sum_{i=1}^{N}\nu'_{ji}M_{F^0}\nu ji \quad (15)$$

[11] See Bai (2009) for more details.

and

$$(\hat{\pi}_{1j}, \hat{F}_{1s,j}) = \underset{\pi_{1j}, F}{\text{arg min}}\; S_{1S-j,NT}(\pi_{1j}, F) \tag{16}$$

where the consistency comes directly from Bai (2009) under Assumptions 1 through 5. Each equation in the first stage, under these assumptions, satisfies all assumptions of Bai (2009).

The second stage is similar; however, $\hat{X}_i$ is used as a covariate of Y. Therefore,

$$S_{\hat{X},NT}(\beta, F) = \frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - \hat{X}_i\beta\right)' M_F \left(Y_i - \hat{X}_i\beta\right) - \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' M_{F^0}\varepsilon_i \tag{17}$$

and

$$(\hat{\beta}, \hat{F}_{2s}) = \underset{\beta, F}{\text{arg min}}\; S_{\hat{X},NT}(\beta, F) \tag{18}$$

where F_{2s} is the common factor of the second-stage equation.

The first point that could be problematic is that for each equation in the first and second stages we are estimating different factors.[12] In two-stage least squares, it is necessary to use the same exogenous variables (generate the same space) for all equations in the first and second stages. Indeed, we define the same common factor (that we called F) for all reduced-form equations. However, in our procedure, we estimate different factors in each step. This should not be a problem because of Proposition 1.

Proposition 1. *Let $S = \{1, 2, ..., p\}$, $F^0_{1s,j}$ for all $j \in S$ and F^0_{2s} be the factors in the first- and second-stage equations, respectively. Under Assumptions 1 to 5 then, for all $j, j' \in S$,*

$$span(M_{F^0_{1s,j}} Z_i) = span(M_{F^0_{1s,j'}} Z_i).$$

Moreover,

$$span(M_{F^0_{1s,j}} Z_i) = span(M_{F^0_{2s}} Z_i) = span(M_{F^0} Z_i).$$

Proposition 1 implies that in the models defined by the reduced-form equations with F or $F_{1s,j}$ and F_{2s} there are no differences for the estimator's asymptotic properties. In other words, if and only if $span(F\Lambda') \xrightarrow{p} span(F^0\Lambda^0)$ then $\forall j \in S$, $span(F_{1s,j}\Lambda'_{1s,j}) \xrightarrow{p} span(F^0\Lambda^0_j)$ and $span(F_{2s}\Lambda'_{2s}) \xrightarrow{p} span(F^0\Lambda^0)$. The intuition of this result is that if a factor belongs to any first

[12] See Appendix A for an example.

equation but does not belong to another equation of the reduced-form system, the jth equation for example, then this factor does not affect the instruments and the covariate X_{jit}. Thus, the true loading factor that corresponds to this common factor is equal to zero in the jth equation.

Therefore, directly from Bai (2009) under Assumptions 1 to 5 where $N, T \to \infty$,

$$\forall j \in S, \hat{\pi}_{1j} \xrightarrow{p} \pi_{1j} \text{ and } ||P_{\hat{F}_{1s,j}} - P_{F^0_{1s,j}}|| \xrightarrow{p} 0$$

and by Proposition 1,

$$\forall j \in S, ||M_{\hat{F}_{1s,j}} Z_i - M_{F^0} Z_i|| \xrightarrow{p} 0.$$

Proposition 2. *Let* $X_i^* = M_{F^0} Z_i \Pi_1$ *and* $\hat{X}_i = M_{\hat{F}} Z_i \hat{\Pi}_1$. *Under Assumptions 1 to 5*

$$N, T \to \infty \implies \hat{X}_i \xrightarrow{p} X_i^*.$$

Corollary 1. *Under Assumptions 1 to 5*

$$N, T \to \infty \implies D_{\hat{X}}(F) \xrightarrow{p} D_{X^*}(F)$$

where $D_{\hat{X}}(F) = \frac{1}{NT} \sum_{i=1}^{N} W_i'(\hat{X}, F) W_i(\hat{X}, F)$,
$D_{X^*}(F) = \frac{1}{NT} \sum_{i=1}^{N} W_i'(X^*, F) W_i(X^*, F)$ *and* $W_i(.)$ *is defined as* (13).

Therefore by Proposition 2 and Corollary 1, it is sufficient to show that, under the assumptions in Section 3, all assumptions necessary in Bai's (2009) method hold for X^*. This implies that this assumption is asymptotically true for $\hat{X}_i$ as well. Proposition 3 proves exactly this for Assumption 1. The other assumptions do not depend on X thus, no changes are necessary.

Proposition 3. *Let* $X_i^* = M_{F^0} Z_i \Pi_1$. *If Assumptions 1 to 5 hold, then*

(i) $E||X_{it}^*||^4 < M$

(ii) Let $\mathcal{F} = \{F : F'F/T = I\}$, $\underset{F \in \mathcal{F}}{Inf} D_{X^*}(F) > 0$.

Under Assumptions 1 to 5 and Proposition 3 we can show that when $N, T \to \infty$, $(\beta^0, HF^0) = \underset{\beta, F}{\arg\min}\, S_{\hat{X},NT}(\beta, F)$. This fact comes from Lemma A2 in Appendix B that shows

$$S_{\hat{X},NT}(\beta, F) = \tilde{S}_{\hat{X},NT}(\beta, F) + o_p(1) \tag{19}$$

where

$$\tilde{S}_{\hat{X},NT}(\beta, F) = \beta' \left(\frac{1}{NT} \sum_{i=1}^{N} \hat{X}_i' M_F \hat{X}_i \right) \beta + tr \left[\left(\frac{F^{0\prime} M_F F^0}{T} \right) \left(\frac{\Lambda'\Lambda}{N} \right) \right] + 2\beta' \frac{1}{NT} \sum_{i=1}^{N} \hat{X}_i' M_F F^0 \lambda_i.$$

Under assumptions in Section 3 and Proposition 3 we can also show that $\tilde{S}_{\hat{X},NT}(\beta, F) \xrightarrow{p} \tilde{S}_{X^*,NT}(\beta, F)$ uniformly in β and F, such that $\tilde{S}_{X^*,NT}(\beta, F)$ has a unique minimum in (β^0, HF^0). This fact is enough to prove that $\hat{\beta} - \beta \xrightarrow{p} 0$. The consistency of $P_{\hat{F}}$ to P_{F^0} does not depend on $\hat{X}$ only by the consistency of β, therefore under the assumptions in Section 3 and following Bai (2009), $||P_{\hat{F}} - P_{F^0}|| \xrightarrow{p} 0$.

Theorem 1. *Consistency: Under Assumptions 1 to 5, as $N, T \to \infty$, the following statements hold:*

1. *The estimator $\hat{\beta}$ is consistent such that $\hat{\beta} \xrightarrow{p} \beta$*
2. *The space generated by $\hat{F}$ is asymptotically equal to the space generated by F^0. In other words, $||P_{\hat{F}} - P_{F^0}|| \xrightarrow{p} 0$.*

By Lemma B3 if $T/N^2 \to 0$ we can write $\sqrt{NT}(\hat{\beta} - \beta)$ as

$$\sqrt{NT}(\hat{\beta} - \beta) = \left(\frac{1}{NT} \sum_{i=1}^{N} W_i'(X^*, \hat{F}) W_i(X^*, \hat{F}) \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} W_i'(X^*, \hat{F}) \varepsilon_i + O_p(1)$$

Without heteroskedasticity or auto-correlation, we have no asymptotic bias, and the term $O_p(1)$ becomes $o_p(1)$. However, when heteroskedasticity and auto-correlation exist in both dimensions (cross-sectional and serial), we can estimate the bias term and construct a consistent estimator (Bai, 2009).

Therefore with Proposition B4, an adaptation of Proposition B3 of Bai (2009) for X^*, if $T/N^2 \to 0$ and $N/T^2 \to 0$ we can write

$$\sqrt{NT}(\hat{\beta} - \beta^0) = \left(D_{X^*}(F^0) \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} W_i'(X^*, F^0) \varepsilon_i + O_p(1). \tag{20}$$

The first term converges by the law of large numbers. For the second term, however, we need to utilize the central limit theory. Thus the following assumption is necessary.

Assumption 6. *There exists a nonnegative matrix D_W such that*

$$\text{plim}\,\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_{ij,ts}W_{it}(X^*,F^0)W_{it}'(X^*,F^0)=D_W,$$

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}W_i'(X^*,F^0)\varepsilon_i \xrightarrow{d} N\left(0,D_W\right).$$

Let

$$D_0 = plimD_{X^*}(F^0) = plim\frac{1}{NT}\sum_{i=1}^{N}W_i'(X^*,F^0)W_i(X^*,F^0). \tag{21}$$

Theorem 2. *Under Assumptions 1 to 6 and $N,T\to\infty$.*

(i) In the absence of serial correlation and heteroskedasticity and when $N/T\to 0$

$$\sqrt{NT}(\hat{\beta}-\beta^0)\xrightarrow{d} N(0,D_0^{-1}D_WD_0^{-1\prime})$$

where $D_W = D_1 = plim\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}\sum_{t=1}^{T}W_{it}(X^*,F^0)W_{it}'(X^*,F^0)$ and $\sigma_{ij} = E(\varepsilon_{it}\varepsilon_{jt})$.

(ii) In the absence of cross-sectional correlation and heteroskedasticity and when $T/N\to 0$

$$\sqrt{NT}(\hat{\beta}-\beta^0)\xrightarrow{d} N(0,D_0^{-1}D_WD_0^{-1\prime})$$

where $D_W = D_2 = plim\frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\omega_{ts}\sum_{i=1}^{N}W_{it}(X^*,F^0)W_{it}'(X^*,F^0)$ and $\omega_{ts} = E(\varepsilon_{it}\varepsilon_{is})$.

If ε_{it} are i.i.d. with distribution equal to $N(0,\sigma^2)$ then $\sqrt{NT}(\hat{\beta}-\beta^0)\xrightarrow{p} N(0,\sigma^2D_0^{-1})$. Theorem 2 follows from the fact that in the absence of cross-sectional correlation and heteroskedasticity, or in the absence of correlation and heteroskedasticity, the last term in Equation (20) is $o_p(1)$.

In the case of cross-sectional correlation and heteroskedasticity, for each i the number of periods need to be large enough. To estimate a robust variance when heteroskedasticity and correlation exist, it is necessary that T increases faster than N ($N/T\to 0$). At the other extreme, in the case of serial correlation and heteroskedasticity, N must increase at a rate faster than T ($N/T\to 0$).

In the case of both serial and cross-sectional correlation and heteroskedasticity, we have an asymptotic bias. This bias depends on both correlation and heteroskedasticity; however, it does not depend on β. Therefore, we can estimate it and correct the bias.[13]

[13] See Bai (2009) for more details.

We can show that if $T/N^2 \to 0$

$$\sqrt{NT}\left(\hat{\beta}-\beta^0\right) = D_{X^*}(\hat{F})^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i^{*\prime}M_{F^0}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k^{*\prime}M_{F^0}\right]\varepsilon_i + \sqrt{\frac{T}{N}}\xi_{NT}+\sqrt{\frac{N}{T}}\zeta_{NT}+o_p(1)$$

where ξ_{NT} is the term that represents the bias generated by cross-section and ζ_{NT} is the term related to serial correlation and heteroskedasticity.[14]

However, we do not use X^* in the second step. Instead, we use $\hat{X}$. Thus, to prove the equality above we need show that $\xi_{NT}(\hat{X}) \xrightarrow{p} \xi_{NT}$ and $\zeta_{NT}(\hat{X}) \xrightarrow{p} \zeta_{NT}$. These proofs are in Lemma B4 in Appendix B. Moreover, we can show that $\sqrt{T/N}(\xi_{NT}-B) = o_p(1)$ and $\sqrt{N/T}(\zeta_{NT}-C) = o_p(1)$, $D_{x^*}(\hat{F})^{-1} - D_{X^*}(F)^1 = o_p(1)$, $\sqrt{N/T}(D_{x^*}(\hat{F})^{-1}-D_{X^*}(F)^{-1}) = o_p(1)$ and $\sqrt{T/N}(D_{x^*}(\hat{F})^{-1}-D_{X^*}(F)^{-1}) = o_p(1)$ with an appropriate convergence rate of T and N. This implies that

$$\sqrt{NT}\left(\hat{\beta}-\beta^0\right) = D_{X^*}(F^0))^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i^{*\prime}M_{F^0}-\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k^{*\prime}M_{F^0}\right]\varepsilon_i + \sqrt{\frac{T}{N}}B+\sqrt{\frac{N}{T}}C+o_p(1)$$

where

$$B = -D\left(F^0\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{\left(X_i^*-V_i\right)^{\prime}F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1} \times\left(\frac{\Lambda^{\prime}\Lambda}{N}\right)^{-1}\lambda_k\left(\frac{1}{T}\sum_{t=1}^{T}\sigma_{ik,tt}\right) \tag{22}$$

and

$$C = -D\left(F^0\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}X_i^{*\prime}M_{F^0}\Omega F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda^{\prime}\Lambda}{N}\right)^{-1}\lambda_i \tag{23}$$

$V_i = \frac{1}{N}\sum_{j=1}^{N}a_{ij}X_j^*$, $a_{ij} = \lambda_i^{\prime}\left(\Lambda^{\prime}\Lambda/N\right)^{-1}\lambda_j$, and $\Omega = \frac{1}{N}\sum_{k=1}^{N}\Omega_k$ with $\Omega_k = E\left(\varepsilon_k\varepsilon_k^{\prime}\right)$.

Therefore, the following theorem comes straight from Bai (2009).

Theorem 3. *Under Assumptions 1-6 and when $T/N \to \rho > 0$*

$$\sqrt{NT}\left(\hat{\beta}-\beta^0\right) \xrightarrow{d} N\left(\rho^{1/2}B_0+\rho^{-1/2}C_0, D_0^{-1}D_WD_0^{-1}\right)$$

where B_0 is the probability limit of B defined by Equation (22), C_0 is the probability limit of C defined by Equation (23) and D_W is defined in Assumption 6.

[14] See Propositions A2 and 3 for definitions of ζ_{NT} and ξ_{NT}.

The bias is given by $\rho^{1/2}B_0 + \rho^{-1/2}C_0$. Bai (2009) shows that without cross-sectional correlation and heteroskedasticity, then $B_0 = 0$, and without serial correlation and heteroskedasticity, then $C_0 = 0$. Therefore, without cross-sectional and serial correlation and heteroskedasticity when $T/N \to \rho$ we have no bias and the asymptotic distribution of β is given by $\sqrt{NT}\left(\hat{\beta} - \beta^0\right) \sim N(0, D_0^{-1}D_W D_0^{-1})$. If B_0 or C_0 are different from zero, we can consistently estimate them and construct a consistent estimator of β^0.

4.1.1 Bias correction

In the case of serial correlation and heteroskedasticity, the asymptotic bias depends on ρ, B, and C. Therefore, we only need consistent estimators for them.

Let

$$\hat{B} = -D_{\hat{X}}(\hat{F})^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{X}_i - \hat{V}_i\right)'\hat{F}\left(\frac{1}{T}\hat{F}'\hat{F}\right)^{-1}\left(\hat{\Lambda}'\hat{\Lambda}\right)^{-1}\hat{\lambda}_i'\hat{\Psi}_{ik}$$

$$\hat{C} = -D_{\hat{X}}(\hat{F})^{-1}\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i'M_{\hat{F}}\hat{\Omega}\hat{F}\left(\frac{1}{T}\hat{F}'\hat{F}\right)^{-1}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i'$$

where $\hat{V}_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij}\hat{X}_j$, $\hat{\Psi}_{ik} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{\epsilon}_{it}\hat{\epsilon}_{kt}$, and $\hat{\Omega} = \frac{1}{N}\sum_{k=1}^{N}\hat{\Omega}_k = \frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_k\hat{\varepsilon}_k'$.

If there is no serial and cross-sectional correlation $\hat{\Omega}_k$ simplifies to a $T \times T$ diagonal matrix with elements $\hat{\omega}_{ik} = \frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_k^2$ and $\hat{\Psi} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$.

Proposition 4. *Under Assumptions 1 to 6 and in the absence of serial and cross-sectional correlation*

(i) $\hat{B} - B = o_p(1)$

(ii) $\hat{C} - C = o_p(1)$.

Therefore, let

$$\hat{\beta}^{\dagger} = \hat{\beta} - \frac{1}{N}\hat{B} - \frac{1}{T}\hat{C}. \tag{24}$$

Thus, $\beta^{\dagger} \xrightarrow{p} \beta^0$ under Assumptions 1 to 6 with an appropriate convergence rate of T and N.

In the absence of cross-sectional and serial correlation, by Proposition 4, $\hat{\beta}^{\dagger} - \beta^0 = o_p(1)$. Moreover, the asymptotic distribution of $\sqrt{NT}\left(\hat{\beta}^{\dagger} - \beta^0\right)$ can be defined even if $T/N \to 0$ or $N/T \to 0$. We need only assume that $T/N^2 \to 0$ and $N/T^2 \to 0$ which is less restrictive than $T/N \to \rho$.

Theorem 4. *Under Assumptions 1 to 6, in the absence of serial and cross-sectional correlation and when* $T/N^2 \to 0$ *and* $N/T^2 \to 0$

$$\sqrt{NT}\left(\hat{\beta}^{\dagger} - \beta^0\right) \xrightarrow{d} N\left(0, D_0^{-1} D_W D_0^{-1}\right)$$

where $\hat{\beta}^{\dagger}$ *is defined as in Equation (24) and*

$$D_W = D_3 = plim \tfrac{1}{NT} \textstyle\sum_{i=1}^{N} \sum_{i=1}^{T} W_{it}(F^0, X_i^*) W_{it}'(F^0, X_i^*) \sigma_{it}^2.$$

When we also have serial correlation, Bai (2009) argues that we need to estimate $T^{-1} X_i' \Omega_k F^0$ and $T^{-1} F'^0 \Omega_k F^0$ to consistently estimate C. We can estimate it using the heteroskedasticity and autocorrelation (HAC) robust limiting covariance.[15]

4.1.2 Estimating covariance matrices

To estimate the covariance matrix, define

$$\hat{D}_0 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it}(\hat{F}, \hat{X}_i) W_{it}'(\hat{F}, \hat{X}_i) \tag{25}$$

$$\hat{D}_1 = \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_i^2 \left(\frac{1}{T} \sum_{i=1}^{T} W_{it}(\hat{F}, \hat{X}_i) W_{it}'(\hat{F}, \hat{X}_i) \right) \tag{26}$$

$$\hat{D}_2 = \frac{1}{T} \sum_{t=1}^{T} \hat{\omega}_t^2 \left(\frac{1}{N} \sum_{i=1}^{N} W_{it}(\hat{F}, \hat{X}_i) W_{it}'(\hat{F}, \hat{X}_i) \right) \tag{27}$$

$$\hat{D}_3 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{i=1}^{T} W_{it}(\hat{F}, \hat{X}_i) W_{it}'(\hat{F}, \hat{X}_i) \hat{\varepsilon}_{it}^2 \tag{28}$$

where $\hat{\sigma}_i^2 = \frac{1}{T} \sum_{i=1}^{T} \hat{\varepsilon}_{it}^2, \hat{\omega}_t^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_{it}^2$, and $W_{it}(\hat{F}, \hat{X}_i)$ is defined in Equation (13) but replacing Λ by $\hat{\Lambda}$. For the construction of $\hat{\varepsilon}_{it}$ we need to be careful, if we use the second-step residuals, since in the second step we are not using X_{it} but $\hat{X}_{it}$. The residuals do not converge to ε_{it} (Wooldridge, 2002). Therefore, a consistent estimator for $\hat{\varepsilon}_{it}$ is

$$\hat{\varepsilon}_{it} = Y_{it} - \hat{\beta} X_{it} - \hat{\lambda}_i' \hat{F}_t. \tag{29}$$

Proposition 5. *Under Assumptions 1 to 6*

(i) $\hat{D}_0 \xrightarrow{p} D_0$*;*

moreover, in the absence of serial and cross-sectional correlation

[15] See Bai (2009) Remark 6.

(ii) $\hat{D}_1 \xrightarrow{p} D_1$;

(iii) $\hat{D}_2 \xrightarrow{p} D_2$ *and*

(iv) $\hat{D}_3 \xrightarrow{p} D_3$

where $\hat{D}_0, \hat{D}_1, \hat{D}_2$, *and* $\hat{D}_3$ *are defined in Equations* (25), (26), (27), *and* (28), *respectively.* D_0, D_1, D_2, *and* D_3 *are defined in Theorems 2 and 4.*

In the case of serial correlation, Bai (2009) proposes that we use the HAC robust limiting covariance for estimating D_W.[16]

4.1.3 Alternative estimation

Let $\hat{F}_1$ be the factor estimated in the first stage and $\hat{F}_{-1}$ be all factors not estimated in the first stage that are estimated in the second stage.

Estimation Procedure II uses $\dot{X}_{it}$ and $\dot{Z}_{it}$ instead of X_{it} and Z_{it} in the first step. And it uses $\dot{\hat{X}}_{it}^{\oint}$ and $\dot{Z}_{it}^{\oint}$ instead of $\hat{X}_{it}$ and Z_{it} in the second step. Thus, in the first step we are minimizing

$$S_{1S-j,NT}(\pi_{1j}, F_1) = \frac{1}{NT}\sum_{i=1}^{N}\left(\dot{X}_{ji} - \dot{Z}_{ji}\pi_{1j}\right)' M_{F_1}\left(\dot{X}_{ji} - \dot{Z}_{ji}\pi_{1j}\right) - \frac{1}{NT}\sum_{i=1}^{N}\dot{\nu}'_{ji}M_{F_1^0}\dot{\nu}_{ji}. \tag{30}$$

In the second stage we estimate the equation

$$\dot{Y}_i^{\oint} = \dot{\hat{X}}_i^{\oint}\beta + \lambda'_{-1i}F_{-1} + \varepsilon_i^{\oint}.$$

Since $\dot{Y}_i^{\oint} = M_{\hat{F}_1}\dot{Y}_i$ and $\dot{\hat{X}}_{it}^{\oint} = M_{\hat{F}_1}\dot{\hat{X}}_i$ then, by the Frisch-Waugh-Lovell theorem proved by Frisch and Waugh (1933), the model above is equivalent to estimating the following model:

$$\begin{aligned} \dot{Y} &= \dot{\hat{X}}\beta + \lambda'_{1i}F_1 + \lambda'_{-1i}F_{-1} + \varepsilon_i \\ &= \dot{\hat{\mathcal{X}}} + \lambda'_{-1i}F_{-1} + \varepsilon_i \end{aligned}$$

where $\dot{\hat{\mathcal{X}}} = [\dot{\hat{X}}, F_1]$.[17] However, we do not observe F_{-1}. Let $\dot{\hat{\hat{\mathcal{X}}}}_{it} = [\dot{\hat{X}}, \hat{F}_1]$, since $\hat{F}_1 \xrightarrow{p} F_1$ then $\dot{\hat{\hat{\mathcal{X}}}}_{it} \xrightarrow{p} \dot{\hat{\mathcal{X}}}$.[18]

Thus, we minimize

[16] See Remark 8 of Bai (2009).
[17] For more details see Bai and Li (2017).
[18] This convergence does not depend on β.

$$S_{\tilde{X},NT}(\beta, F_{-1}) = \frac{1}{NT}\sum_{i=1}^{N}\left(\dot{Y}_i - \dot{\hat{X}}_i\beta\right)' M_F \left(\dot{Y}_i - \dot{\hat{X}}_i\beta\right)$$
$$-\frac{1}{NT}\sum_{i=1}^{N}\dot{\varepsilon}_i' M_{F^0}\dot{\varepsilon}_i + o_p(1)$$

For consistency, Assumption 1 must be changed to

$$\underset{F\in\mathcal{F}}{Inf}\,\dot{D}_Z(F) > 0 \tag{31}$$

where $\dot{W}_i(Z,F) = M_F\dot{Z}_i - \frac{1}{N}\sum_{k=1}^{N} M_F\dot{Z}_k\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$ and $\dot{D}_Z(F) = \frac{1}{NT}\sum_{i=1}^{N}\dot{W}_i'(Z,F)\dot{W}_i(Z,F)$.

Therefore, the minimizing function can be written as

$$S_{\hat{\dot{X}},NT}(\beta, F) = \tilde{S}_{\hat{\dot{X}},NT}(\beta, F) + o_p(1) \tag{32}$$

where

$$\tilde{S}_{\hat{\dot{X}},NT}(\beta, F) \;=\; \beta'\frac{1}{NT}\sum_{i=1}^{N}\hat{\dot{X}}_i' M_F\hat{\dot{X}}_i\beta + tr\frac{F^{0\prime}M_F F^0}{T}\frac{\Lambda'\Lambda}{N}$$
$$+2\beta'\frac{1}{NT}\sum_{i=1}^{N}\hat{\dot{X}}_i' M_F F^0\lambda_i.$$

Under assumptions in Section 3 we can also show that $\tilde{S}_{\hat{\dot{X}},NT}(\beta, F) \xrightarrow{p} \tilde{S}_{\dot{X}^*,NT}(\beta, F)$ uniformly in β and F, such that $\tilde{S}_{X^*,NT}(\beta, F)$ has a unique minimum in (β^0, HF^0) as in Theorem 1. This implies that $\beta \xrightarrow{p} \beta^0$.

The asymptotic distribution is equal to the derivative for the standard case when we use X_{it}, Y_{it} and Z_{it}. The only difference now is that we change these variables to $\dot{X}_{it}, \dot{Y}_{it}$ and $\dot{Z}_{it}$. For example, when the error is i.i.d with mean 0 and variance equal to σ^2 then, $\sqrt{NT}(\beta - \beta^0) \xrightarrow{d} N(0, \sigma^2\dot{D}_Z^{-1}(F))$.

4.2 Joint - Limiting theory

In the joint estimator we use Bai's (2009) procedure in Equation (12). Let

$$H_i = \begin{bmatrix} Z & D_1\otimes\mathbf{Z}_i & D_2\otimes\mathbf{Z}_i & \cdots & D_p\otimes\mathbf{Z}_i \end{bmatrix}$$

$$\theta = \begin{bmatrix} \gamma_1' & \gamma_{21}' & \gamma_{22}' & \cdots & \gamma_{2p}' \end{bmatrix}'.$$

Therefore Equation (12) can be written as

$$Q_i = H_i\theta + F\lambda_i + \varepsilon_i. \tag{33}$$

We minimize the equation below with respect to θ:

$$S_{H,NT}(\theta, F) = \frac{1}{NT}\sum_{i=1}^{N}(Q_i - H_i\theta)' M_F (Q_i - H_i\theta) - \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' M_{F^0}\varepsilon_i. \tag{34}$$

Under Assumptions 1 to 5, all assumptions necessary for the consistency of the estimator in Equation (33) hold. Therefore, directly from Bai (2009) we have that $\hat{\theta} \xrightarrow{p} \theta^0$ and $||P_F - P_{F^0}|| \xrightarrow{p} 0$. Since $f(.)$ is continuous in θ^0 if Assumption 5 holds, the consistency of $\hat{\beta}$ comes directly from the continuous mapping theorem.

Theorem 5. *Consistency: Under Assumptions 1 to 5, as $N, T \to \infty$, the following statements hold*

(i) The estimator $\hat{\beta}$ is consistent such that $\hat{\beta} \xrightarrow{p} \beta^0$; and

(ii) The space generated by $\hat{F}$ is asymptotically equal to the space generated by F^0. That is, $||P_{\hat{F}} - P_{F^0}|| \xrightarrow{p} 0$.

For the asymptotic distribution, since we know the distribution of θ as $N, T \to \infty$, let

$$W_{it}(H, F^0) = M_F H_i - \frac{1}{N}\sum_{k=1}^{N} M_F H_k \lambda_i', \tag{35}$$

and

$$D_0 = plim D_H(F^0) = plim \frac{1}{NT}\sum_{i=1}^{N} W_i'(H, F^0) W_i(H, F^0). \tag{36}$$

Assumption 7. *There exists a non-negative matrix D_W such that*

$$\text{plim}\,\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_{ij,ts} W_{it}(H_i, F^0) W_{it}'(H_i, F^0) = D_W,$$

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} W_i'(H_i, F^0)\varepsilon_i \xrightarrow{d} N(0, D_W).$$

In the absence of serial heteroskedasticity and correlation and when $N/T \to 0$, under Assumptions 1 to 5 and 7, the asymptotic distribution of θ is given by

$$\sqrt{NT}(\hat{\theta} - \theta^0) \xrightarrow{d} N(0, D_0^{-1} D_W D_0^{-1\prime})$$

where $D_W = D_1 = plim\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}\sum_{t=1}^{T}W_{it}(H, F^0)W_{it}'(H, F^0)$ and $\sigma_{ij} = E(\varepsilon_{it}\varepsilon_{jt})$.

Therefore, since β is a function of θ, $\sqrt{NT}\left(\hat{\beta} - \beta^0\right) = \sqrt{NT}\left(f(\hat{\theta}) - f(\theta^0)\right)$. By the delta method $\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, \nabla f(\beta)' D_0^{-1} D_W D_0^{-1\prime} \nabla f(\beta))$. The same reasoning applies to the case without cross-sectional heteroskedasticity and correlation and $T/N \to 0$. The following theorem gives these results.

Theorem 6. *Under Assumptions 1 to 5 and 7,*

(i) In the absence of serial correlation and heteroskedasticity and when $N/T \to 0$

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, \nabla f(\beta)' \Sigma \nabla f(\beta))$$

where $\Sigma = D_0^{-1} D_W D_0^{-1\prime}$,
$D_W = D_1 = plim\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}\sum_{t=1}^{T}W_{it}(H, F^0)W_{it}'(H, F^0)$ and $\sigma_{ij} = E(\varepsilon_{it}\varepsilon_{jt})$

(ii) In the absence of cross-sectional correlation and heteroskedasticity and when $T/N \to 0$

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, \nabla f(\beta)' \Sigma \nabla f(\beta))$$

where $\Sigma = D_0^{-1} D_W D_0^{-1\prime}$,
$D_W = D_2 = plim\frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\omega_{ts}\sum_{i=1}^{N}W_{it}(H, F^0)W_{it}'(H, F^0)$ and $\omega_{ts} = E(\varepsilon_{it}\varepsilon_{is})$.

As in the Two-Step method from Section 2.1, if we have heteroskedasticity and correlation in both dimensions, the asymptotic distribution of β has a bias. This leads to Bai's (2009) procedures where θ has an asymptotic bias in this case. However, as in Two-Step we can construct an unbiased estimator for θ. If we have heteroskedasticity and correlation in both dimensions and when $T/N \to \rho > 0$ then $\sqrt{NT}\left(f(\hat{\theta}) - f(\theta^0)\right) \xrightarrow{d} N\left(\rho^{1/2}B_0 + \rho^{-1/2}C_0, D_0^{-1}D_W D_0^{-1}\right)$, where B_0 and C_0 are the probability limits of B and C defined in Equations (22) and (23), respectively; however, changing X_i^* for H_i. That is,

$$\hat{B} = -D_H(\hat{F})^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(H-\hat{V}_i\right)'\hat{F}\left(\frac{1}{T}\hat{F}'\hat{F}\right)^{-1}\left(\hat{\Lambda}'\hat{\Lambda}\right)^{-1}\hat{\lambda}_i'\hat{\Psi}_{ik} \tag{37}$$

$$\hat{C} = -D_H(\hat{F})^{-1}\frac{1}{NT}\sum_{i=1}^{N}H_i'M_{\hat{F}}\hat{\Omega}\hat{F}\left(\frac{1}{T}\hat{F}'\hat{F}\right)^{-1}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i' \tag{38}$$

$$\tag{39}$$

where $\hat{V}_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij}H_j$, $\hat{\Psi}_{ik} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{\epsilon}_{it}\hat{\epsilon}_{kt}$, and, $\hat{\Omega} = \frac{1}{N}\sum_{k=1}^{N}\hat{\Omega}_k = \frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_k\hat{\varepsilon}_k'$.

If there is no serial and cross-sectional correlation, $\hat{\Omega}_k$ simplifies to a $T \times T$ diagonal matrix with elements $\hat{\omega}_{ik} = \frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_k^2$ and $\hat{\Psi} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$.

Therefore, we can construct

$$\hat{\theta}^{\dagger} = \hat{\theta} - \frac{1}{N}\hat{B} - \frac{1}{T}\hat{C} \tag{40}$$

which, by Bai (2009), leads to $\sqrt{NT}\left(\hat{\theta}^{\dagger} - \theta^0\right) \xrightarrow{d} N\left(0, D_0^{-1}D_W D_0^{-1}\right)$.[19]

Thus, let $\beta^{\dagger} = f(\theta^{\dagger})$.

Theorem 7. *Under Assumptions 1 to 5 and 7, when* $T/N \to \rho > 0$

$$\sqrt{NT}\left(\hat{\beta}^{\dagger} - \beta^0\right) \xrightarrow{d} N\left(0, \nabla f(\beta^0)'\Sigma\nabla f(\beta^0)\right)$$

where $\Sigma = D_0^{-1}D_W D_0^{-1}$, B_0 *and* C_0 *is the probability limit of* $\hat{B}$ *and* $\hat{C}$, *respectively.* $\hat{B}$ *and* $\hat{C}$ *are defined by Equation* (37) *and* D_W *is defined in Assumption 7.*

In absence of serial and cross-sectional correlation we also determine the consistency if $T/N \to 0$.

Theorem 8. *Under Assumptions 1 to 5 and 7,* $T/N^2 \to 0$ *and* $N/T^2 \to 0$ *and in absence of serial and cross-sectional correlation*

$$\sqrt{NT}\left(\beta^{\dagger} - \beta^0\right) \xrightarrow{d} N\left(0, \nabla f(\beta^0)'\Sigma\nabla f(\beta^0)\right)$$

where $\Sigma = D_0^{-1}D_W D_0^{-1}$ B_0 *is the probability limit of* B, C_0 *is the probability limit of* C *defined in Equation* (23) *and* D_W *is defined in Assumption 7.*

[19] For more details see Bai (2009).

5 Finite Sample Properties - Simulation

We now present some estimators' properties in finite samples through Monte Carlo simulations. First, we show the data generation process utilized in the simulations. Then, we show the estimators' consistency and variance results. And finally, we compare our estimators with some estimators in the literature: instrumental variable, interactive fixed-effect (Bai, 2009) and FE-IV.

5.1 Data Generation Process

In this section, we show the finite sample properties of our estimators. The data generation process (DGP) is constructed as follows.

Let Λ_j be the loading factor, with respect to F for $j \in \{x, y, z\}$.

$$
\begin{aligned}
Y &= \beta_1 X_1 + \beta_2 X_2 + F\Lambda_y + E_y \\
X_1 &= F\Lambda_{x_1} + E_{x_1} \\
X_2 &= F\Lambda_{x_2} + E_{x_2} \\
Z &= F\Lambda_z + E_z
\end{aligned}
$$

where Y, X, Z and E are the matrix $T \times N$, F and Λ_j are defined as

$$F = (f', f_y', f_{x1}', f_{x1}', f_z')' \tag{41}$$

and

$$\Lambda_j = (\lambda, \mathbf{0}, ..., \mathbf{0}, \lambda_{jj}', \mathbf{0}, ..., \mathbf{0})'. \tag{42}$$

For example, $\Lambda_y = (\lambda', \lambda_{yy}', \mathbf{0}, \mathbf{0}, \mathbf{0})'$ and $F\Lambda_y = f\lambda + f_y\lambda_{yy}$. The loading and common factor $f\lambda$ is the structure that affects all variables and $f_y\lambda_{yy}$ is the idiosyncratic factor structure.

The loading and common factors are generated by a multivariate normal distribution with mean and variance equal to 1 and covariance equal to 0,

$$F \sim N(\mathbf{1}_{R_F}, \mathbf{I}_{R_F}) \quad \Lambda \sim N(\mathbf{1}_{R_\Lambda}, \mathbf{I}_{R_\Lambda})$$

where r is the dimension of factors (f and λ_j), r_j the dimension of the idiosyncratic factors of the variable j (f_j and λ_{jj}),[20] $\mathbf{1}_k$ a vector $k \times 1$ of 1, $\mathbf{I}_k$ an identity matrix with dimension $k \times k$, $R_F = r + r_y + r_x + r_z$ and $R_\Lambda = 3r + r_y + r_x + r_z$.

[20] f is a $T \times r$ matrix, f_y is a $T \times r_y$ matrix, λ_y is an $r \times N$ matrix and λ_{yy} is $r_y \times N$.

The errors are distributed by the multivariate normal distribution such that

$$E \sim N(\mathbf{0}_4, V)$$

where $E = (E_y, E_{x_1}, E_{x_2}, E_z)$ and

$$\mathbf{V} = \begin{pmatrix} 1 & 1/2 & 0 & 0 \\ 1/2 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 1/2 \\ 0 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

That is, $cov(e_{y,it}, e_{x_1,it}) = 1/2$, $cov(e_{x_1,it}, e_{z,it}) = 1/2$ and $cov(e_{y,it}, e_{z,it}) = 0$. The first covariance is responsible for the bias of the standard estimators in this framework as in the fixed effects estimator and Bai's (2009) method. The last two covariances ensure that Assumption 5 holds. The bias in the instrumental variable estimator comes from the fact that f and λ affect the instrument and the dependent variable.

For simulation with heteroskedasticity we use the same DGP above. However we construct a new error such that

$$\tilde{e}_{y,it} = \sigma_{it} * e_{y,it} \tag{43}$$

where $\sigma_{it} \sim U(0.5, 1.5)$. Thus,

$$Y = \beta_1 X_1 + \beta_2 X_2 + F\Lambda_y + \tilde{E}_y.$$

5.2 Convergence

The entire Monte Carlo simulation was performed with 1000 interactions. Table 2 shows the mean and standard deviation of the Two-Step estimator, and Table 1 shows them for the Joint estimator.

The simulations were performed with the errors (ε_{it}) with heteroskedasticity across t and i as defined in Equation (43); and in the absence of correlation in both dimensions. The factor structure has dimension equal to 4. That is, $F = (f, f_y, f_{x_1}, f_{x_2}, f_z)$ is $T \times 5$ where one of these factors affects all variables (f), and f_y, f_{x_1}, f_{x_2}, and f_z are idiosyncratic factors of Y, X_1, X_2, and Z, respectively. All of them are $T \times 1$ matrix.

In Appendix C, we show the results for DGPs when there is no heteroskedasticity and correlation in either dimension, when there is no idiosyncratic common factor for any variables, and when Z has an idiosyncratic factor. The results show no significant changes.

Table 1 – β_1 Convergence - Two-Step estimator with heteroskedasticity

N \T	10	15	30	50	100	200	500	1000
10	0.845 (6.8)	1.052 (0.61)	1.042 (0.24)	1.052 (0.17)	1.047 (0.11)	1.042 (0.09)	1.039 (0.07)	1.039 (0.06)
15	1.076 (0.66)	1.081 (0.3)	1.056 (0.15)	1.054 (0.11)	1.035 (0.07)	1.019 (0.05)	1.008 (0.03)	1.005 (0.02)
30	1.062 (0.22)	1.069 (0.15)	1.068 (0.09)	1.047 (0.06)	1.021 (0.04)	1.011 (0.03)	1.004 (0.02)	1.002 (0.01)
50	1.047 (0.16)	1.058 (0.11)	1.04 (0.06)	1.034 (0.04)	1.017 (0.03)	1.008 (0.02)	1.003 (0.01)	1.002 (0.01)
100	1.047 (0.12)	1.038 (0.07)	1.02 (0.04)	1.017 (0.03)	1.016 (0.02)	1.008 (0.01)	1.003 (0.01)	1.001 (0.01)
200	1.044 (0.09)	1.019 (0.05)	1.01 (0.03)	1.009 (0.02)	1.009 (0.01)	1.008 (0.01)	1.003 (0.01)	1.002 (0.00)
500	1.041 (0.07)	1.007 (0.03)	1.003 (0.02)	1.003 (0.01)	1.003 (0.01)	1.003 (0.01)		
1000	1.042 (0.06)	1.005 (0.02)	1.002 (0.01)	1.002 (0.01)	1.001 (0.01)	1.001 (0.00)		

Notes: This table shows the mean of 1000 simulations for the estimator Joint. The standard deviation are in parentheses. $\beta = (\beta_1, \beta_2) = (1, 2)$. f is $T \times 1$. The idiosyncratic common factor of Y (f_y), X_1 (f_{x_1}), X_2 (f_{x_2}) and Z (f_z) is $T \times 1$ matrix. The errors are heteroskedastic as described in section 5.1.

Table 2 – β_1 Convergence - Joint estimator with heteroskedasticity

N \T	10	15	30	50	100	200	500	1000
10	1.629 (9.12)	1.171 (1.08)	1.061 (0.32)	1.061 (0.25)	1.021 (0.17)	1.017 (0.14)	1.015 (0.11)	1.003 (0.09)
15	1.205 (2.62)	1.13 (0.39)	1.084 (0.19)	1.065 (0.12)	1.039 (0.07)	1.021 (0.05)	1.008 (0.03)	1.006 (0.03)
30	1.08 (0.32)	1.088 (0.18)	1.114 (0.1)	1.079 (0.07)	1.045 (0.04)	1.025 (0.03)	1.01 (0.02)	1.006 (0.01)
50	1.051 (0.24)	1.063 (0.13)	1.075 (0.07)	1.081 (0.05)	1.047 (0.03)	1.027 (0.02)	1.011 (0.01)	1.007 (0.01)
100	1.027 (0.17)	1.041 (0.07)	1.042 (0.04)	1.047 (0.03)	1.055 (0.02)	1.031 (0.01)	1.013 (0.01)	1.006 (0.01)
200	1.028 (0.14)	1.021 (0.05)	1.024 (0.03)	1.028 (0.02)	1.031 (0.01)	1.031 (0.01)	1.013 (0.01)	1.007 (0.00)
500	1.016 (0.11)	1.008 (0.03)	1.009 (0.02)	1.012 (0.01)	1.013 (0.01)	1.013 (0.01)		
1000	1.013 (0.1)	1.004 (0.02)	1.005 (0.01)	1.006 (0.01)	1.006 (0.01)	1.006 (0.00)		

Notes: This table shows the mean of 1000 simulations for the estimator Two-Step. The standard deviation are in parentheses. $\beta = (\beta_1, \beta_2) = (1, 2)$. f is $T \times 1$. The idiosyncratic common factor of Y (f_y), X_1 (f_{x_1}), X_2 (f_{x_2}) and Z (f_z) is $T \times 1$ matrix. The errors are heteroskedastic as described in section 5.1.

In this DGP, we have heteroskedasticity in both dimensions, and we do not have correlation in either dimension. Therefore, by Theorems 4 and 8, when $N/T^2 \to 0$ and $T/N^2 \to 0$, the estimators are consistent and the simulations show it. Even with N and T small, the convergence is fast. This is more notable since here we are estimating a factor structure with dimension equal to four, which is a relatively high number of factors.

In general, the Two-Step procedure is better than Joint in terms of convergence and standard deviation, especially for small numbers of N and T. However, when N is large and T very small (as $N = 1000$ and $T = 5$), or T is large and N small (as $N = 5$ and $T = 1000$), the

Joint performs better. This is expected, since Bai's (2009) procedure and the Two-Step estimator has a bias term which is inversely proportional to $min\{N, T\}$. Whiles, the Joint estimator has a bias term inversely proportional to $min\{2N, T\}$ (or $min\{N, 2T\}$).

5.3 Bias comparison

In this section, we compare our estimators with the estimators in the literature, Instrumental Variable (IV), Instrumental Variable and Fixed Effects (IV-FE), and linear factor model (Bai, 2009). Figure 1 shows the 3D graphic of the bias in all estimators. The x-axis is the individual numbers in panel data (N), the y-axis is the period numbers (T), and the z-axis is the bias in absolute value.

The surface is estimated using a bootstrap procedure where each pair (n,t) is withdrawn from a uniform distribution between 3 and 150.[21] Each point corresponds to the bias means in 1000 interactions with the pair (n, t). The DGP is the same as above, with five loading and five common factor. The surface was estimated using a polynomial fit with 4 degrees.

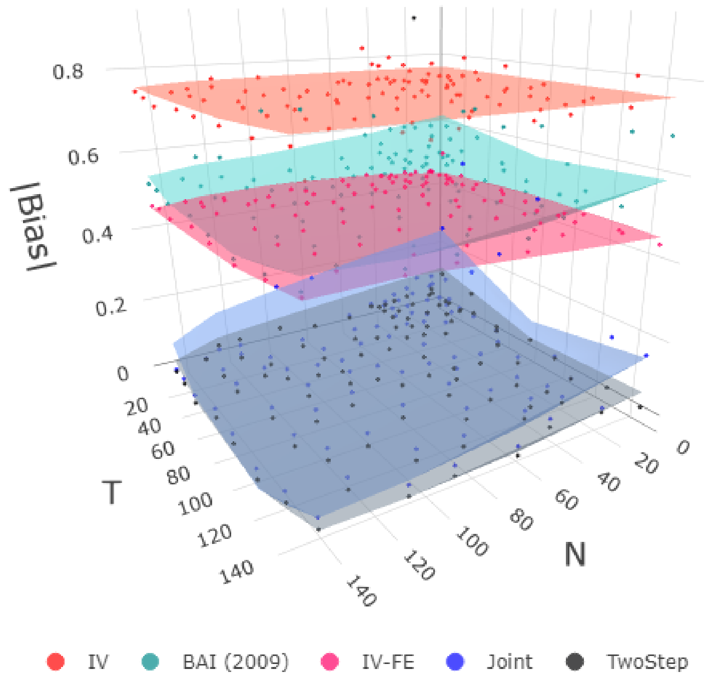

Figure 1 – Bias comparison

The pooled IV estimator has the highest bias of all estimators utilized and is given by the correlation of Z_{it} and the factor structure. Its asymptotic bias is around 0.70. Bai's (2009) estimator is asymptotically biased also, with an bias equal to the covariance of X_{it} and ε_{it}, which in our DGP is equal to 0.5. The instrumental variable fixed effect estimator (IV-FE) reduces the bias to around 0.45. Our two estimators, when $N, T \to \infty$, have no asymptotic bias. Moreover,

[21] We use a weight in the uniform distribution to have a greater mass of points in the subset $n < 50$ and $t < 50$.

our estimators, for any pair of (n, t), have less bias than the other estimators in our DGP and the convergence is fast.

Table 3 – Bias comparison

N,T	Estimadores	10		50		100		1000	
		$Bias$	Sd	$Bias$	Sd	$Bias$	Sd	$Bias$	Sd
10	Two-Step	-0.155	6.801	0.052	0.168	0.047	0.112	0.039	0.062
	Joint	0.629	9.124	0.061	0.248	0.021	0.169	0.003	0.09
	IV	0.746	1.072	0.821	2.477	0.769	0.43	0.751	0.3
	IV-FE	0.46	0.511	0.459	0.152	0.47	0.122	0.463	0.097
	Bai (2009)	0.58	0.222	0.599	0.127	0.579	0.11	0.566	0.082
50	Two-Step	0.047	0.155	0.034	0.041	0.017	0.029	0.002	0.009
	Joint	0.051	0.239	0.081	0.045	0.047	0.032	0.007	0.009
	IV	0.745	0.357	0.747	0.171	0.755	0.152	0.75	0.118
	IV-FE	0.463	0.153	0.49	0.071	0.494	0.06	0.492	0.041
	Bai (2009)	0.601	0.13	0.54	0.042	0.536	0.036	0.525	0.031
100	Two-Step	0.047	0.119	0.017	0.028	0.016	0.019	0.001	0.006
	Joint	0.027	0.174	0.047	0.031	0.055	0.02	0.006	0.006
	IV	0.756	0.334	0.755	0.144	0.752	0.115	0.751	0.084
	IV-FE	0.464	0.131	0.49	0.058	0.492	0.044	0.498	0.029
	Bai (2009)	0.586	0.105	0.535	0.036	0.522	0.03	0.503	0.013
1000	Two-Step	0.042	0.064	0.002	0.009	0.001	0.006		
	Joint	0.013	0.102	0.006	0.009	0.006	0.006		
	IV	0.742	0.312	0.748	0.119	0.753	0.083		
	IV-FE	0.466	0.094	0.494	0.041	0.496	0.029		
	Bai (2009)	0.571	0.084	0.524	0.03	0.502	0.01		

Notes: This table shows the mean and the standard deviation (sd) of 1000 simulations for the estimators: Two-Step, Joint, IV, IV-FE, Bai (2009). $\beta = (\beta_1, \beta_2) = (1, 2)$. f is $T \times 1$. The idiosyncratic common factor of Y (f_y), X_1 (f_{x_1}), X_2 (f_{x_2}) and Z (f_z) is $T \times 1$ matrix. There is no factors that affect Y, X_1, X_2 and Z $(f = NULL)$. The errors are i.i.d. with normal distribution.

6 More Efficient IV Estimator

Consider the model in Equation (2) and an instrumental variable (Z_{it}) that satisfies all assumptions in the standard instrumental model. That is, the Z_{it} is orthogonal to the factor structure and the error for all i and t. In this case, both standard IV and our estimators are consistent. This fact can be tested using, for example, the Hausman test (Hausman, 1978).

However, since our method estimates the factor structure, the error variance is reduced. Thus, a trade-off exists. Our estimators have less error variance but estimate a higher number of parameters. The standard IV estimator estimates a lower number of parameters but has a high error variance. The best choice between the IV and one of our estimators depends, especially, on factor structure variance. If the factor structure has a high variance in comparison to error variance, it is better to estimate it and reduce the error variance. Otherwise, if the factor structure has a low variance in comparison to error variance, it is better not to estimate the factor structure, and instead utilize the standard IV approach.

When the variance of this factor structure is high, we construct a more efficient estimator than the standard IV estimator. Table 4 shows the variance and the point estimate in the case when the standard IV assumptions hold. That is, the instrument is orthogonal to the factor structure and the error in Equation (1). Unless the pair (n,t) is equal to (10,10), our estimations have lower variance than the IV estimator. Also, our estimators, when N and T are small (which implies that N*T are small too), show a better finite-sample convergence in this DGP.

Table 4 – Standard Deviation Comparison - β_1

N \T	Estimator	10		15		50		100		500		1000	
		$\beta = 1$	Sd	$\beta = 1$	Sd	$\beta = 1$	Sd	$\beta = 1$	Sd	$\beta = 1$	Sd	$\beta = 1$	Sd
	IV	1.642	8.78	0.725	7.414	0.977	4.88	0.762	6.745	2.306	46.035	1.863	28.755
10	Two-Step	1.007	5.508	0.935	1.059	1.035	0.185	1.021	0.113	1.008	0.038	1.005	0.027
	Joint	0.79	10.099	0.963	1.002	1.052	0.17	1.039	0.105	1.011	0.042	1.006	0.03
	IV	2.15	24.265	1.237	5.064	0.948	3.912	0.831	8.125	1.678	11.391	0.158	22.678
15	Two-Step	0.968	0.848	1.063	0.577	1.045	0.115	1.032	0.067	1.007	0.03	1.005	0.022
	Joint	1.056	1.31	1.056	0.613	1.051	0.109	1.033	0.067	1.007	0.03	1.007	0.023
	IV	0.919	12.201	1.188	3.673	1.085	1.377	0.943	1.506	0.988	0.395	1.011	0.368
50	Two-Step	1.018	0.184	1.042	0.114	1.061	0.042	1.039	0.028	1.011	0.013	1.005	0.009
	Joint	1.049	0.149	1.05	0.105	1.034	0.04	1.016	0.026	1.004	0.012	1.001	0.009
	IV	0.773	7.154	1.238	9.847	0.992	0.443	0.762	7.247	1.006	0.231	1.000	0.231
100	Two-Step	1.022	0.107	1.031	0.067	1.04	0.03	1.046	0.018	1.011	0.008	1.006	0.006
	Joint	1.039	0.099	1.031	0.066	1.017	0.027	1.015	0.017	1.003	0.008	1.002	0.006
	IV	0.77	17.344	0.845	6.137	0.995	0.369	0.988	0.248				
500	Two-Step	1.006	0.039	1.007	0.031	1.01	0.011	1.012	0.007				
	Joint	1.011	0.043	1.007	0.03	1.003	0.011	1.003	0.008				
	IV	1.311	9.63	1.035	4.829	1.004	0.355	1.003	0.228				
1000	Two-Step	1.003	0.03	1.003	0.023	1.006	0.008	1.006	0.006				
	Joint	1.004	0.031	1.005	0.023	1.002	0.008	1.001	0.006				

Notes: This table shows the mean and the standard deviation (sd) of 1000 simulations for the estimators: IV, Two-Step and Joint. $\beta = (\beta_1, \beta_2) = (1, 2)$. f is $T \times 1$. The idiosyncratic common factor of Y (f_y), X_1 (f_{x_1}), X_2 (f_{x_2}) and Z (f_z) is $T \times 1$ matrix. There is no factors that affect Y, X_1, X_2 and Z ($f = NULL$). The errors are i.i.d. with normal distribution.

7 Conclusion

In this paper, we examine the case when we have a “bad"instrument, that is, an instrument for which the exogeneity assumption does not hold. In this case, the literature in panel data uses the IV-FE model and, in some cases, grouped fixed effects to obtain consistency. We extend these models using the literature of interactive fixed effects; specifically, we utilize the estimator of Bai (2009). We construct two estimators with instrumental variables together with the interactive fixed-effect model, one called Two-Step, which is similar to the two stage least squares estimator (2SLS), and another called Joint with which we estimate all reduced-form equations jointly using Bai’s (2009) method.

Our estimators require a stronger relevance assumption, since the instrumental variable needs to be correlated with the covariate given the factor structure ($E[X_i M_F Z_i] \neq 0$). However, the exogeneity assumption is less restrictive than the assumption commonly used, since Z_{it} is orthogonal to ε_{it}, which is the error orthogonal to the factor structure.

We show the consistency of our estimators when $N, T \to \infty$ and their asymptotic distribution for some rates of T, N. In the absence of correlation and heteroskedasticity in both dimensions, we need only require that $N, T \to \infty$. If heteroskedasticity exists in both dimensions, but no correlation exists, it is necessary that $N/T^2 \to 0$ and $T/N^2 \to 0$. If correlation is also present, we need that $N/T \to \rho$. The construction of estimators for the covariance matrix in all of these cases is explained as well.

Analysis of finite sample properties through Monte Carlos simulations shows that the two estimators have a faster convergence and small standard deviations, even when N, T is not large. When we compare their bias with the bias of other estimators in the literature: Bai (2009), IV, and IV-FE, we verify that our estimators have less bias than all other estimators for all pairs of (n, t) with N and T greater than 3.

Finally, we discuss the trade-off between our estimation and standard IV estimation when this last one is consistent. We show that when the variance of the factor structure is high enough in comparison to error variance, our estimators are more efficient.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505.

Bada, O. and Kneip, A. (2014). Parameter cascading for panel models with unknown number of unobserved factors: An application to the credit spread puzzle. *Computational Statistics & Data Analysis*, 76:95–115.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Li, K. (2017). Practical notes on panel data models with interactive effects.

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

Bartik, T. J. (1991). Who benefits from state and local economic development policies?

Blanchard, O. J., Katz, L. F., and Robert, E. (1992). Hall, and barry eichengreen. 1992."regional evolutions.". *Brookings papers on economic activity*, 1(1).

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143.

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese economic journal*, 1(2):141–162.

David, H., Dorn, D., and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–68.

Frisch, R. and Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, pages 387–401.

Harding, M. and Lamarche, C. (2011). Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. *Economics Letters*, 111(3):197–199.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, pages 1251–1271.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society*, pages 1371–1395.

Jaravel, X. (2019). The unequal gains from product innovations: Evidence from the us retail sector. *The Quarterly Journal of Economics*, 134(2):715–783.

Lee, N., Moon, H. R., and Weidner, M. (2012). Analysis of interactive fixed effects dynamic linear panel regression with measurement error. *Economics Letters*, 117(1):239–242.

Moon, H. R., Shum, M., and Weidner, M. (2018). Estimation of random coefficients logit demand models with interactive fixed effects. *Journal of Econometrics*, 206(2):613–644.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579.

Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195.

Murtazashvili, I. and Wooldridge, J. M. (2008). Fixed effects instrumental variables estimation in correlated random coefficient panel data models. *Journal of Econometrics*, 142(1):539–552.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, pages 1417–1426.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data mit press. *Cambridge, MA*, 108.

8 APPENDIX

A Factors estimated in each step

Consider first the simple case with one covariate and one instrument.

$$Y_{it} = \beta X_{it} + \lambda_y' f_t + \lambda_{yy}' f_{y,t} + \epsilon_{it}$$
$$X_{1it} = \lambda_x' f_t + \lambda_{xx}' f_{x,t} + e_{x,it}$$
$$Z_{it} = \lambda_z' f_t + \lambda_{zz}' f_{z,t} + e_{z,it}$$

Therefore, the first and second steps are defined as

$$X_{1it} = Z_{it}'\pi_{11} + \lambda_{1i}' F_t + \nu_{1it}$$
$$Y_{it} = Z_{it}'\pi_{21} + \lambda_{2i}' F_t + \nu_{2it}$$

where $\lambda_{1i} = [\lambda_x + \pi_{11}\lambda_z, \mathbf{0}, \lambda_{xx}, \pi_{11}\lambda_{zz}]$ and $\lambda_{2i} = [\lambda_y + \beta\lambda_x + \beta\pi_{11}\lambda_z, \lambda_{yy}, \beta\lambda_{xx}, \beta\pi_{11}\lambda_{zz}]$.

We estimate only the factors when $\lambda_k > 0$, where the subscript k is referred to as the k factor.

- All factors of X, Y, and Z will be estimated in the second stage (and we can define $F_t = F_{2t}$ with no loss of generality).
- In the first stage we estimate only factors that affect X and Z.

B Proofs

B.1 Two-Step

Demonstração. Proposition 1:

Let $j, j' \in S$ be arbitrarily chosen. Define $M_F = I - P_F$, where P_F is the projection matrix of the vector F, that is, $P_F = F(F'F)^{-1}F' = FF'/T$ with the restrictions $F'F/T = I$.

First, we will prove that $span(M_{F_{1s,j}} Z_i) = span(M_{F_{1s,j'}} Z_i)$.

By definition of M_F,

$$M_{F_{1s,j}} Z_i = (I - P_{F_{1s,j}})Z_i = Z_i - P_{F_{1s,j}} Z_i$$

and

$$M_{F_{1s,j'}} Z_i = (I - P_{F_{1s,j'}})Z_i = Z_i - P_{F_{1s,j'}} Z_i.$$

Claim 1: $M_{F_{1s,j}} Z_i = M_{F_{1s,j'}} Z_i$

Proof of Claim 1: Assume, by way of contradiction, that $M_{F_{1s,j}} Z_i \neq M_{F_{1s,j'}} Z_i$. We have that

$$\begin{gathered} Z_i - P_{F_{1s,j}} Z_i \neq Z_i - P_{F_{1s,j'}} Z_i \\ P_{F_{1s,j}} Z_i \neq P_{F_{1s,j'}} Z_i. \end{gathered} \tag{44}$$

Let $F_{1s,-j}$ be the factors of $F_{1s,j'}$ that are not estimated in the jth equation of the first stage,[22] i.e., all the elements of $F_{1s,j'}$ that are not in $F_{1s,j}$.

Since all the elements in $F_{1s,j'}$ belong to either $F_{1s,j}$ or $F_{1s,-j}$ we can write it as

$$F_{1s,j'} = [F_{1s,j}\ F_{1s,-j}] \text{ for all } j$$

where $F_{1s,j}$ and $F_{1s,-j}$ are mutually exclusive.

Therefore, since $P_F = FF'/T$ under the restriction $F'F/T = I$

$$P_{F_{1s,j'}} = P_{F_{1s,j}} + P_{F_{1s,-j}} \tag{45}$$

by Equations (44) and (45)

[22] See Appendix A1 for more details.

$$P_{F_{1s,j}} Z_i \neq P_{F_{1s,j}} Z_i + P_{F_{1s,-j}} Z_i$$
$$P_{F_{1s,-j}} Z_i \neq \mathbf{0}.$$

This implies that there exists at least one factor in $F_{1s,j'}$ that does not belong to $F_{1s,j}$ and $P_{F_{1s,-j}} Z_i \neq \mathbf{0}$. Let this factor be f^*, since Z_i is covariate in the jth equation and f^* is not in $F_{1s,j}$. This implies that $f^* \perp Z_i$. Therefore, $P_{F_{1s,-j}} Z_i = \mathbf{0}$. Contradiction.

We have proved that $M_{F_{1s,j}} Z_i = M_{F_{1s,j'}} Z_i$. Therefore,

$$span(M_{F_{1s,j}} Z_i) = span(M_{F_{1s,j'}} Z_i).$$

Now we shall prove that $span(M_{F_{1s,j}} Z_i) = span(M_F Z_i)$. Without any loss of generality, we can assume that $F = F_{2s}$. Therefore, consider a decomposition of factors as $F = [F_{1s,j} F_{1s,-j}]$, where $F_{1s,-j}$ is now equal to all the factors in the second step (equal to all factors in F) that are not in $F_{1s,j}$. Again, we can define F and P_F as

$$F = [F_{1s,j} \; F_{1s,-j}] \text{ and}$$
$$P_F = P_{F_{1s,j}} + P_{F_{1s,-j}}.$$

This implies

$$P_F Z_i \quad = \quad P_{F_{1s,j}} Z_i + P_{F_{1s,-j}} Z_i = P_{F_1} Z_i.$$

Since Z and $F_{1s,-j}$ are orthogonal, then $P_{F_{1s,-j}} Z_i = 0$. Therefore, $M_F Z_i = M_{F_2} Z_i$ and

$$span(M_{F_{1s,j}} Z_i) = span(M_F Z_i) = span(M_{F_{2s}} Z_i).$$

Since j and j' are arbitrary, we have proved the desired result. □

Demonstração. Proof of Proposition 2

Under Assumptions 1 to 5 we have that if $N, T \to \infty$

$$\text{for all } j \in S, \hat{\pi}_{j1} \xrightarrow{p} \pi_{j1} \text{ and } ||P_{F_{1s,j}} - P_{F^0_{1s,j}}|| \xrightarrow{p} 0.$$

Directly from Bai (2009) and by Proposition 1, if $||P_{F_{1s,j}} - P_{F^0_{1s,j}}|| \xrightarrow{p} 0$ then $||M_{F_{1s,j}} Z_i - M_{F^0} Z_i|| \xrightarrow{p} 0$.

Define $A = (a_{ij})$ to be a matrix. If $\hat{a}_{ij} \xrightarrow{p} a_{ij}$ then the matrix $\hat{A} = (\hat{a}_{ij}) \xrightarrow{p} A$. Therefore, if for all $j \in S, \hat{\pi}_{j1} \xrightarrow{p} \pi_{j1}$ then $\hat{\Pi}_1 \xrightarrow{p} \Pi_1$.

Thus,

$$\begin{aligned}\hat{X}_i &= M_F Z_i \Pi_1 + M_F Z_i(\hat{\Pi}_1 - \Pi_1) + (M_{\hat{F}} - M_F) Z_i(\hat{\Pi}_1) \\ &= X_i^* + O_p(1) o_p(1) + o_p(1) O_p(1) = X_i^* + o_p(1).\end{aligned}$$

□

Demonstração. of COROLLARY 1

First, we prove that $W_i(\hat{X}, F)$ converges to $W_i(X^*, F)$. After this, the result is direct.

$$\begin{aligned}W_i(\hat{X}, F) &= M_F \hat{X}_i - \frac{1}{N}\sum_{k=1}^{N} M_F \hat{X}_k \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k \\ &= M_F X_i^* - \frac{1}{N}\sum_{k=1}^{N} M_F X *_k \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k \\ &\quad + M_F(\hat{X}_i - X_i^*) + \frac{1}{N}\sum_{k=1}^{N} M_F \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k(\hat{X}_k - X_k^*).\end{aligned}$$

By Proposition 2, $X_i \xrightarrow{p} X^*$ and, by Assumption 2, $M_F = O_p(1)$ and $M_F\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k = O_p(1)$. Therefore,

$$\begin{aligned}W_i(\hat{X}, F) &= M_F X_i^* - \frac{1}{N}\sum_{k=1}^{N} M_F X *_k \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k + o_p(1) \\ &= W_i(X^*, F) + o_p(1).\end{aligned}$$

This implies that

$$\begin{aligned}D_{\hat{X}}(F) &= \frac{1}{NT}\sum_{i=1}^{N} W_i'(\hat{X}, F) W_i(\hat{X}, F) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\Big(W_i'(X^*, F) W_i(X^*, F) + o_p(1) W_i(\hat{X}, F) + W_i'(\hat{X}, F) o_p(1) + o_p(1) o_p(1)\Big).\end{aligned}$$

Since under Assumptions 1 and 2, $W_i(\hat{X}, F) = O_p(1)$.[23] Thus,

$$D_{\hat{X}}(F) = W_i'(X^*, F) W_i(X^*, F) + o_p(1).$$

□

[23] Π_1 is finite because all their entries are less than $max\{0, \pi_{j1}\}$ and π_{j1} is the minimum of a convex function.

Demonstração. of PROPOSITION 3:

Let $i \in \{1, 2, ..., N\}$ and $t \in \{1, 2, ..., T\}$ be arbitrarily chosen.

Proof part **(i):** $E||X^*_{it}||^4 < M$

$$X^*_i \;=\; Z_i\Pi_1 + P_F Z_i \Pi_1. \tag{46}$$

Therefore,

$$X^*_{it} = Z_{it}\Pi_1 - \frac{1}{T}\sum_{s=1}^{T} F_t F'_s Z_{is}\Pi_1.$$

Insofar as for all j in S, $(\pi_{j1}, F_{j,1s}) = argminS_{N,T}(\beta, F)$ and $S_{N,T}$ is convex which implies that Π is finite, that is, $||\Pi_1|| < M'$. Let M = max{M, M'}.

$$E||Z_{it}||^4 < M \text{ and } ||\Pi_1||^4 < M.$$

Thus, the first term of Equation 46 is bounded. We need to show now that $E||\frac{1}{T}\sum_{s=1}^{T} F_t F'_s Z_{is}\pi_1||^4$ is bounded.

By the Cauchy-Schwarz inequality

$$E\left|\left|\frac{1}{T}\sum_{s=1}^{T} F_t F'_s Z_{is}\pi_1\right|\right| \;\leq\; \left(\frac{1}{T}\sum_{s=1}^{T} E||F_t||^2.E||F_s||^2.E||Z_{is}||^2.||\pi_{1j}||^2\right)^{1/2}.$$

Squaring both sides

$$E\left|\left|\frac{1}{T}\sum_{s=1}^{T} F_t F'_s Z_{is}\pi_1\right|\right|^2 \;\leq\; \frac{1}{T}\sum_{s=1}^{T} E||F_t||^2.E||F_s||^2.E||Z_{is}||^2.||\pi_{1j}||^2$$

Again using the Cauchy-Schwarz inequality

$$E\left|\left|\frac{1}{T}\sum_{s=1}^{T} F_t F'_s Z_{is}\pi_1\right|\right|^4 \;\leq\; \frac{1}{T}\sum_{s=1}^{T} E||F_t||^4.E||F_s||^4.E||Z_{is}||^4.||\pi_{1j}||^4.$$

With Assumptions 2 and 4, $E||F_t||^4 < M$, $E||F_s||^4$, $E||Z_{is}||^4 < M$ and $||\pi_{1j}||^4$ then

$$E\left|\left|\frac{1}{T}\sum_{s=1}^{T}F_tF_s'Z_{is}\pi_1\right|\right|^4 \leq \frac{1}{T}\sum_{s=1}^{T}E||F_t||^4.E||F_s||^4.E||Z_{is}||^4.||\pi_{1j}||^4$$
$$\leq \frac{1}{T}\sum_{s=1}^{T}MMMM \leq \frac{1}{T}\sum_{s=1}^{T}M^4 \leq M^4.$$

Thus, by Equation (46)

$$E||X_{it}^*||^4 = (Z_i\Pi_1 + P_F Z_i\Pi_1)^4 .$$

Since $Z_i\Pi_1$ and $P_F Z_i\Pi_1$ are finite, there exists a K such that $E||X_{it}^*||^4 < K$. Let M := max{M,K}. Then

$$E||X_{it}^*||^4 < M$$

(ii) $\underset{F\in\mathcal{F}}{Inf} D_{X^*}(F) > 0$

$$W_i(X^*,F) = M_F X_i^* - \frac{1}{N}\sum_{k=1}^{N}M_F X_k^*\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$$
$$= M_F M_{F_{1s,j}} Z_i\Pi_1 - \frac{1}{N}\sum_{k=1}^{N}M_F M_{F_{1s,j}} Z_i\Pi_1\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$$

By Proposition 1 we have that $M_{F_{1s,j}}Z_i = M_{F_{1s,j'}}Z_i = M_F Z_i$

$$W_i(X^*,F) = M_F M_F Z_i\Pi_1 - \frac{1}{N}\sum_{k=1}^{N}M_F M_F Z_i\Pi_1\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$$
$$= M_F Z_i\Pi_1 - \frac{1}{N}\sum_{k=1}^{N}M_F Z_i\Pi_1\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k.$$

Observe that $\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$ is a scalar. Therefore,

$$
\begin{aligned}
W_i(X^*, F) &= M_F Z_i \Pi_1 - \frac{1}{N}\sum_{k=1}^{N} \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k M_F Z_i \Pi_1 \\
&= \left(M_F Z_i - \frac{1}{N}\sum_{k=1}^{N} \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k M_F Z_i \right) \Pi_1 \\
&= W_i(Z_i, F)\Pi_1.
\end{aligned}
$$

which implies that

$$
\begin{aligned}
D_{X^*}(F) &= W_i(X^*, F)' W_i(X^*, F) \\
&= (W_i(Z_i, F)\Pi_1)' (W_i(Z_i, F)\Pi_1) \\
&= \Pi_1' W_i(Z_i, F)' W_i(Z_i, F)\Pi_1 \\
&= \Pi_1' D_z(F)\Pi_1.
\end{aligned}
$$

Observe that Π_1 is not a function of F. Since $\underset{F\in\mathcal{F}}{Inf} D_X(F) > 0$ and the fact that Π_1 has full rank, implied by Assumption 5 (ii): relevance. We have that $\underset{F\in\mathcal{F}}{Inf} D_{X^*}(F) > 0$.

□

The Lemma below is important to the proof of Theorem 1.

Lemma B 1. *Under Assumptions 1-5 and when* $N, T \to \infty$

$$
\underset{F}{sup} \left\| \frac{1}{NT}\sum_{i=1}^{N} \hat{X}_i' M_F \varepsilon_i \right\| = o_p(1).
$$

Demonstração. of Lemma B1

Under Assumptions 1-5 and when $N, T \to \infty$ by Proposition 2 it follows that $\hat{X}_i = X_i^* + o_p(1)$. Since $M_F\varepsilon_i$ is $O_p(1)$

$$
\frac{1}{NT}\sum_{i=1}^{N} \hat{X}_i' M_F \varepsilon_i = \frac{1}{NT}\sum_{i=1}^{N} X_i^{*\prime} M_F \varepsilon_i + o_p(1).
$$

From Assumption 5, $\frac{1}{NT}\sum_{i=1}^{N} Z_i' M_F \varepsilon_i = o_p(1)$ given that Π_1 is bounded by the fact that this is the minimum of a convex function. Thus, under Assumptions 1-5 and when $N, T \to \infty$

$$\frac{1}{NT}\sum_{i=1}^{N} X_i^{*\prime}\varepsilon_i = \frac{1}{NT}\sum_{i=1}^{N}(M_F Z_i \Pi_1)'\varepsilon_i$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\Pi_1' Z_i' M_F \varepsilon_i = o_p(1)$$

and

$$\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F \varepsilon_i = \frac{1}{NT}\sum_{i=1}^{N} X_i^{*\prime} M_F \varepsilon_i + o_p(1).$$

The last equation arises from the fact that M_F and ε_i are bounded ($O_p(1)$) by Assumptions 2 and 3, respectively, and by Proposition 2 (i.e., $\hat{X} \xrightarrow{p} X^*$).

Thus

$$\underset{F}{sup}\left|\left|\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F \varepsilon_i\right|\right| \leq \underset{F}{sup}\left|\left|\frac{1}{NT}\sum_{i=1}^{N} X_i^{*\prime} M_F \varepsilon_i\right|\right| + o_p(1)$$
$$= \underset{F}{sup}\left|\left|\frac{1}{NT}\sum_{i=1}^{N} X_i^{*\prime}\varepsilon_i\right|\right| + o_p(1)$$
$$= o_p(1).$$

The last inequality arises from the fact that under Assumptions 1 to 5, $X_i^* \perp \varepsilon_i$. Since $0 \leq \underset{F}{sup}\left|\left|\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F \varepsilon_i\right|\right| \leq o_p(1)$ we have the desired result. □

Lemma B 2. *Under Assumptions 1 to 5 and when $N, T \to \infty$*

$$S_{\hat{X},NT}(\beta, F) = \tilde{S}_{\hat{X},NT}(\beta, F) + o_p(1)$$

where $o_p(1)$ is uniformly bounded on F and β and

$$\tilde{S}_{NT,\hat{X}}(\beta, F) = \beta'\left(\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F \hat{X}_i\right)\beta + tr\left[\left(\frac{F^{0\prime} M_F F^{0\prime}}{T}\right)\left(\frac{\Lambda'\Lambda}{N}\right)\right] + 2\beta'\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F F^0 \lambda_i.$$

Demonstração. of Lemma B2

$$S_{\hat{X},NT}(\beta, F) = \frac{1}{NT}\sum_{i=1}^{N}(Y_i - \hat{X}_i\beta)' M_F (Y_i - \hat{X}_i\beta) - \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' M_F \varepsilon_i.$$

Expanding $S_{\hat{X},NT}(\beta, F)$,

$$S_{2S,NT}(\beta, F) = \tilde{S}_{2S,NT}(\beta, F) + 2\beta'\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F \varepsilon_i + 2\frac{1}{NT}\sum_{i=1}^{N}\lambda_i' F^{0\prime} M_F \varepsilon_i$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'(P_F - P_F^0)\varepsilon_i$$

where

$$\tilde{S}_{NT,\hat{X}}(\beta, F) = \beta'\left(\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F \hat{X}_i\right)\beta + tr\left[\left(\frac{F^{0\prime} M_F F^{0\prime}}{T}\right)\left(\frac{\Lambda'\Lambda}{N}\right)\right]$$
$$+2\beta'\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F F^0 \lambda_i.$$

Under Assumptions 1-5 and Proposition 3 we have that [24]

$$\sup_F \left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i' F^{0\prime} M_F \varepsilon_i\right\| = o_p(1) \tag{47}$$

$$\sup_F \left\|\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' P_F \varepsilon_i\right\| = o_p(1) \tag{48}$$

$$\sup_F \left\|\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_F \varepsilon_i\right\| = o_p(1). \tag{49}$$

Where (47) and (48) come from Bai (2009) and (49) comes from Lemma 1. Therefore, by Equations (47), (48) and (49),

$$S_{\hat{X},NT}(\beta, F) = \tilde{S}_{\hat{X},NT}(\beta, F) + o_p(1).$$

□

Demonstração. of THEOREM 1

By Lemma 2,

$$S_{\hat{X},NT}(\beta, F) = \tilde{S}_{\hat{X},NT}(\beta, F) + o_p(1).$$

[24] See Bai's (2009) supplemental material for the proofs.

where $o_p(1)$ is uniform in F and β. For consistency of β to β^0 therefore, we only need to prove that $\tilde{S}_{\hat{X},NT}(\beta, F) \xrightarrow{p} \tilde{S}_{X^*,NT}(\beta, F)$ uniformly on F and β and by Proposition 3, $(\beta^0, F^0) = argmin\tilde{S}_{X^*,NT}(\beta, F)$ comes directly from Bai (2009).

CLAIM 1: $\tilde{S}_{\hat{X},NT}(\beta, F) \xrightarrow{p} \tilde{S}_{X^*,NT}(\beta, F)$ *uniformly on* F *and* β*.*

Proof of Claim 1:

By Lemma 2, the fact that $\hat{X}$ does not depend on the F estimated in the second stage (only the F estimated in the first stage) and Assumption 2 implies M_F, F^0 and λ is bounded. Which implies that

$$\tilde{S}_{\hat{X},NT}(\beta, F) = \tilde{S}_{X^*,NT}(\beta, F) + o_p(1)$$

where $o_p(1)$ is uniform in F (estimated in the second stage) and β. Thus, we prove Claim 1.

Therefore,

$$\begin{aligned} S_{\hat{X},NT}(\beta, F) &= \tilde{S}_{\hat{X},NT}(\beta, F) + o_p(1) \\ &= \tilde{S}_{X^*,NT}(\beta, F) + o_p(1) + o_p(1) \\ &= \tilde{S}_{X^*,NT}(\beta, F) + o_p(1) \end{aligned}$$

since under Propositions 3 and Assumptions 1 to 5, all assumptions in Bai (2009) hold for X^*. Therefore, $(\beta^0, F^0) = argmin\tilde{S}_{X^*,NT}(\beta, F)$ comes from Bai (2009) which implies $\hat{\beta} \xrightarrow{p} \beta$.[25] The proof of the second part of the theorem has no changes, given that $\hat{\beta} \xrightarrow{p} \beta$, and is provided by Bai (2009). □

Lemma B 3. *Under Assumptions 1 to 5*

(i) $\hat{X}_i(\beta - \hat{\beta}) = X_i^*(\beta - \hat{\beta}) + o_p(1)$

(ii) $\hat{X}_i'\hat{F} = X_i^{*\prime}\hat{F} + o_p(1)$

(iii) $\hat{X}_i'F^0H = X_i^{*\prime}F^0H + o_p(1)$

(iv) $\hat{X}_i'M_{\hat{F}}F^0H = X_i^{*\prime}M_{\hat{F}}F^0H + o_p(1)$

(v) $\hat{X}_i'M_{\hat{F}}F^0 = X_i^{*\prime}M_{\hat{F}}F^0 + o_p(1)$

(vi) $\hat{X}_i'M_{\hat{F}}\varepsilon_i = X_i^{*\prime}M_{\hat{F}}\varepsilon_i + o_p(1)$.

Demonstração. Lemma 3

All the proofs comes directly from Proposition 2 and Assumptions 1-3 and are omitted. □

25 See Bai (2009) for the proof.

Proposition B 1. *Under Assumptions 1-5 and when $N, T \to \infty$, we can make the following statements:*

1. *V_{NT} is invertible and $V_{NT} \xrightarrow{p} V$, where $V(r \times r)$ is a diagonal matrix consisting of the eigenvalues of $\Sigma_\Lambda \Sigma_F$.*
2. *Let $H = (\Lambda'\Lambda/N)\left(F^0\hat{F}/T\right)V_{NT}^{-1}$. Then H is an $r \times r$ invertible matrix and*

$$\frac{1}{T}\left\|\hat{F} - F^0 H\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - H'F_t^0\right\|^2$$
$$= O_p\left(\|\hat{\beta} - \beta\|^2\right) + O_p\left(\frac{1}{\min[N, T]}\right).$$

Demonstração. PROPOSITION A1

Under Assumptions 1-5 together with Proposition 3, the model with X^* as covariate satisfies all assumptions required by Bai (2009). Therefore, by Lemma B3, as $N, T \to \infty$

$$\hat{F}V_{NT} = I1 + \cdots + I9 + o_p(1).$$

where I1 to I9 are defined as in the proof of Proposition A.1 in Bai (2009) changing X_i to X_i^*. After this, with Proposition 3, the proof has no changes and can be found as Proposition A.1 of Bai (2009). □

Let

$$\xi_{NT}(\hat{X}) = -D_{\hat{X}}(\hat{F})^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{\left(\hat{X}_i - \hat{V}_i\right)' F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}$$
$$\times\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_k\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\right) \quad (50)$$

and

$$\zeta_{NT}(\hat{X}) = -D_{\hat{X}}(\hat{F})^{-1}\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i' M_{\hat{F}}\Omega\hat{F}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i \quad (51)$$

where $\hat{V}_i = \frac{1}{N}\sum_{j=1}^{N} a_{ij}\hat{X}_j$, $a_{ij} = \lambda_i'\left(\Lambda'\Lambda/N\right)^{-1}\lambda_j$, and $\Omega = \frac{1}{N}\sum_{k=1}^{N}\Omega_k$ with $\Omega_k = E\left(\varepsilon_k\varepsilon_k'\right)$.

Lemma B 4. *Under Assumptions 1 to 5 and when $N, T \to \infty$*

(*) $\xi_{NT}(\hat{X}) \xrightarrow{p} \xi_{NT}$

(*) $\zeta_{NT}(\hat{X}) \xrightarrow{p} \zeta_{NT}$

where $\zeta_{NT}(\hat{X})$ *and* $\xi_{NT}(\hat{X})$ *are defined as in Equations* (50) *and* (51)*, respectively. Also,* ζ_{NT}*,* ξ_{NT} *are defined as in Propositions A2 and A3, respectively.*

Demonstração. Lemma B4

First we prove part (i)

Under Assumptions 1 and 2 we have that $D\left(\hat{F}\right)^{-1} = O_p(1)$, by 2 $F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1} = O_p(1)$, $\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_k = O_p(1)$ and by 3 $\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\right) = O_p(1)$. Therefore,

$$\xi_{NT}(\hat{X}) = \xi_{NT} + O_p(1) \times \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{\left(\hat{X}_i - \hat{V}_i\right)' - (X_i^* - V_i)'}{T} \times O_p(1).$$

By Proposition 2 and $a_{ij} = O_p(1)$ then $||(\hat{X}_i - \hat{V}_i)' - (X_i^* - V_i)'|| \xrightarrow{p} 0$ which implies that

$$\xi_{NT}(\hat{X}) = \xi_{NT} + O_p(1).$$

Thus, we have the desired result.

Part (ii) is similar, and we omit the proof.

□

Proposition B 2. *Under Assumptions 1-5. If* $T/N^2 \to 0$*, then*

$$\begin{aligned}\sqrt{NT}\left(\hat{\beta} - \beta^0\right) =& D_{X^*}(\hat{F})^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i^{*\prime}M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k^{*\prime}M_{\hat{F}}\right]\varepsilon_i \\ &+ \sqrt{\frac{N}{T}}\zeta_{NT} + o_p(1)\end{aligned}$$

where $a_{ik} = \lambda_i'\left(\Lambda'\Lambda/N\right)\lambda_k$ *and* $\zeta_{NT} = -D_{X^*}(\hat{F})^{-1}\frac{1}{NT}\sum_{i=1}^{N}X_i^{*\prime}M_{\hat{F}}\Omega\hat{F}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i$ *with* $\Omega = \frac{1}{N}\sum_{k=1}^{N}\Omega_k$ *and* $\Omega_k = E\left(\varepsilon_k\varepsilon_k'\right)$*.*

Demonstração. PROPOSITION A2

By Bai (2009),

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i'M_{\hat{F}}\hat{X}_i\right)(\hat{\beta} - \beta) = \frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i'M_{\hat{F}}F^0\lambda_i + \frac{1}{NT}\sum_{i=1}^{N}\hat{X}_i'M_{\hat{F}}\varepsilon_i.$$

Under Assumptions 1 to 5 and by Proposition 2, Lemma B1 and the fact that $\frac{1}{NT}\sum_{i=1}^N \hat{X}_i' M_{\hat{F}} \hat{X}_i = \frac{1}{NT}\sum_{i=1}^N X_i^{*\prime} M_{\hat{F}} X_i^* + o_p(1)$,[26] then

$$\left(\frac{1}{NT}\sum_{i=1}^N X_i^{*\prime} M_{\hat{F}} X_i^* + o_p(1)\right)(\hat{\beta}-\beta) = \frac{1}{NT}\sum_{i=1}^N \hat{X}_i' M_{\hat{F}} F^0 \lambda_i + \frac{1}{NT}\sum_{i=1}^N X_i^{*\prime} M_{\hat{F}} \varepsilon_i + o_p(1) \qquad (52)$$

where

$$\frac{1}{NT}\sum_{i=1}^N \hat{X}_i' M_{\hat{F}} F^0 \lambda_i = -\frac{1}{NT}\sum_{i=1}^N \hat{X}_i' M_{\hat{F}} [I1 + \cdots + I8 + o_p(1)] \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i.$$

Thus, since $\hat{X}_i$, $M_{\hat{F}}$, $\hat{F}$, F^0 and λ_i are bounded and Proposition 2,I1 to I8 are bounded and

$$\frac{1}{NT}\sum_{i=1}^N \hat{X}_i' M_{\hat{F}} F^0 \lambda_i = -\frac{1}{NT}\sum_{i=1}^N X_i^{*\prime} M_{\hat{F}} [I1 + \cdots + I8] \times \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i + o_p(1).$$

Therefore by (52)

$$\left(\frac{1}{NT}\sum_{i=1}^N X_i^{*\prime} M_{\hat{F}} X_i^*\right)(\hat{\beta}-\beta) = \frac{1}{NT}\sum_{i=1}^N X_i^{*\prime} M_{\hat{F}} F^0 \lambda_i + \frac{1}{NT}\sum_{i=1}^N X_i^{*\prime} M_{\hat{F}} \varepsilon_i + o_p(1).$$

Using Lemma B4, the proof remains the same and follows Bai's (2009) proof of Proposition A.2.

□

Lemma B 5. *Let $\delta_{NT} = min\{T, N\}$. Under Assumptions 1 to 5 the following statements hold:*

(i) $\zeta_{NT} = O_p(1)$, *where ζ_{NT} is given in Proposition A2.*[27]

(ii) $HH' = (F^0F^0/T)^{-1} + O_p(\|\hat{\beta}-\beta\|) + O_p\left(\delta_{NT}^{-2}\right)$.[28]

[26] Since $M_{\hat{F}}$ and $\hat{X}_i$ are bounded.
[27] Lemma A.6 in Bai (2009).
[28] Lemma A.7 (i) in Bai (2009).

(iii) $\|P_{\hat{F}} - P_{F^0}\|^2 = O_p(\|\hat{\beta} - \beta\|) + O_p\left(\delta_{NT}^{-2}\right)$.[29]

(iv)

$$\begin{aligned}\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left[X_i^{*\prime} M_{\hat{F}} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k^{*\prime} M_{\hat{F}} \right] \varepsilon_i \\ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left[X_i^{*\prime} M_{F^0} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k^{*\prime} M_{F^0} \right] \varepsilon_i \\ + (\sqrt{\frac{T}{N}}) \xi_{NT}^{\dagger} + \sqrt{T} O_p \left(\left\| \hat{\beta} - \beta^0 \right\|^2 \right) \\ + O_p \left(\left\| \hat{\beta} - \beta^0 \right\| \right) + \sqrt{T} O_p \left(\delta_{NT}^{-2} \right)\end{aligned}$$

where

$$\begin{aligned}\xi_{NT}^{\dagger} = -\tfrac{1}{N} \textstyle\sum_{i=1}^{N} \sum_{k=1}^{N} \frac{(X_i^* - V_i)' F^0}{T} \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \\ \times \left(\frac{\Lambda' \Lambda}{N} \right)^{-1} \lambda_k \left(\tfrac{1}{T} \textstyle\sum_{t=1}^{T} \varepsilon_{it} \varepsilon_{kt} \right) = O_p(1).\end{aligned}$$

Demonstração. Lemma B5

Under Assumptions 1 to 5 and by Proposition 3 the model with X^* satisfies all assumptions in Bai (2009). Therefore, the proofs of this lemma are the same as those of Bai (2009). □

Proposition B 3. *Under Assumptions 1 to 5 and when $N, T \to \infty$ if $T/N^2 \to 0$ and $T/N^2 \to 0$*

$$\begin{aligned}\sqrt{NT} \left(\hat{\beta} - \beta^0 \right) = D_{X^*}(\hat{F})^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left[X_i^{*\prime} M_{F^0} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k^{*\prime} M_{F^0} \right] \varepsilon_i \\ + \sqrt{\frac{T}{N}} \xi_{NT} + \sqrt{\frac{N}{T}} \zeta_{NT} + o_p(1)\end{aligned}$$

where $\xi_{NT} = D_{X^*}(\hat{F}) \xi_{NT}^{\dagger}$, $\xi_{NT}^{\dagger}$ is defined in Lemma B5 and ζ_{NT} is defined in Proposition B2.

Demonstração. As in the previous lemma, under Assumptions 1 to 5, and since $N, T \to \infty$ and by Proposition 2 the proofs of this lemma are the same as those of Bai's (2009) Corollary 1. □

Demonstração. THEOREM 2

Under Assumptions 1 to 5, X^* satisfies all assumptions in Bai (2009). Together with Propositions 2 and B3, the proof is identical to that of Theorem 2 in Bai (2009).

□

The lemmas and proposition below are an adaptation of Lemma A.9 and Proposition A.3 of Bai (2009).

[29] Lemma A.7 (i) in Bai (2009).

Lemma B 6. *Under Assumptions 1 to 5*

(i) $D_{X^*}(\hat{F}) - D_{X^*}(F^0)) = o_p(1)$

(ii) $\sqrt{T/N}\left[D_{X^*}(\hat{F})^{-1} - D\left(F^0\right)^{-1}\right] = o_p(1)$ *if* $T/N^2 \to 0$

(iii) $\sqrt{N/T}\left[D_{X^*}(\hat{F})^{-1} - D\left(F^0\right)^{-1}\right] = o_p(1)$ *if* $N/T^2 \to 0$

(iv) $\sqrt{T/N}\left(\xi_{NT} - B\right) = o_p(1)$ *if* $T/N^2 \to 0$

(v) $\sqrt{N/T}\left(\zeta_{NT} - C\right) = o_p(1)$ *if* $N/T^2 \to 0.$

Demonstração. All these equations do not depend on $\hat{X}$ and X^*. Thus, under Assumptions 1 to 5, they satisfy all assumptions of Bai (2009). Therefore, the proof is the same as that of Bai (2009). □

Proposition B 4. *Under Assumptions 1 to 5, if* $T/N^2 \to 0$ *and* $N/T^2 \to 0$ *then*

$$\sqrt{NT}\left(\hat{\beta} - \beta^0\right) = D\left(F^0\right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left[X_i^{*\prime} M_{F^0} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k^{*\prime} M_{F^0}\right] \varepsilon_i + \sqrt{\frac{T}{N}}B + \sqrt{\frac{N}{T}}C + o_p(1)$$

where B *and* C *are in* (22) *and* (23) *respectively.*

Demonstração. See Bai's (2009) Proposition A.3. Under Assumptions 1 to 5, X^* satisfies all necessary assumptions. □

Demonstração. THEOREM 3

Under Assumptions 1 to 5 X^* satisfies all assumptions of Bai (2009). Together with Proposition B4 the proof is identical to the proof of Theorem 3 of Bai (2009).

□

Demonstração. THEOREM 4 Under Assumptions 1 to 5, X^* satisfies all assumptions of Bai (2009). Together with Proposition B4 the proof is identical to the proof of Theorem 3 of Bai (2009).

□

B.2 Joint

Demonstração. of THEOREM 5

$$\beta = f(\theta)$$

where $f(.)$ is the minimum distance function. By Bai (2009), under Assumptions 1 to 5, $\hat{\theta} \xrightarrow{p} \theta$. f is continuous for all θ with $\gamma_1 \neq 0$. Under Assumption 5, $\gamma_1 \neq 0$, since $E[X_{it} M_{F^0} Z_{it}] \neq 0$. Thus, f is continuous in θ if 5 holds.

Applying the continuous mapping theorem, we have that

$$\hat{\theta} \xrightarrow{p} \theta \Rightarrow f(\hat{\theta}) \xrightarrow{p} f(\theta).$$

Therefore,

$$\hat{\beta} \xrightarrow{p} \beta.$$

□

Demonstração. THEOREM 6

We know, from Bai (2009), the distribution of θ as $N, T \to \infty$.[30] Assume first that we have no serial correlation and heteroskedasticity and that $N/T \to 0$. Therefore, the distribution of θ is given by

$$\sqrt{NT}(\hat{\theta} - \theta) \xrightarrow{d} N(0, D_0^{-1} D_W D_0'^{-1})$$

where D_W is defined in Theorem 6 (i).

Let $\Sigma = D_0^{-1} D_W D_0'^{-1}$.

Therefore, since β is a function of θ ($\beta = f(\theta)$) then

$$\sqrt{NT}\left(\hat{\beta} - \beta\right) = \sqrt{NT}\left(f(\hat{\theta}) - f(\theta)\right).$$

Under Assumption 5, $\nabla f(\beta^0)$ exist. This implies that, by the delta method

$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \nabla f(\beta)' \Sigma \nabla f(\beta)).$$

The case without cross-sectional correlation and heteroskedasticity is identical and we omit it.

□

[30] See Theorem 2 of Bai (2009).

C Simulations

Without heteroskedasticity and serial correlation in both dimensions

In this DGP, correlation and heteroskedasticity are absent in both dimensions. Therefore, under $T, N \to \infty$ at any rate the estimators are consistent, and the simulations show it. Even so, with N and T small, the convergence is fast. This is more notable since here we are estimating a factor structure with dimension equal to four, which is a relatively high number of factors.

Table 5 – Simulation - Two-Step with idiosyncratic factor in Y and X

N,T	10	15	30	50	100	200	500	1000
10	1.033 (0.72)	1.026 (0.46)	0.995 (0.26)	0.999 (0.19)	1.011 (0.14)	1.018 (0.11)	1.014 (0.1)	1.019 (0.09)
15	1.018 (0.39)	1.041 (0.24)	1.003 (0.16)	1.006 (0.12)	1.009 (0.09)	1.007 (0.06)	1.01 (0.05)	1.008 (0.05)
30	1.025 (0.21)	1.012 (0.14)	1.015 (0.08)	1.006 (0.07)	1.000 (0.05)	1.003 (0.03)	1.002 (0.02)	1.002 (0.02)
50	1.014 (0.15)	1.012 (0.09)	1.013 (0.06)	1.011 (0.04)	1.002 (0.03)	1.002 (0.02)	1.002 (0.01)	1.002 (0.01)
100	1.011 (0.1)	1.012 (0.06)	1.008 (0.04)	1.006 (0.03)	1.006 (0.02)	1.000 (0.01)	1.001 (0.01)	1.001 (0.01)
200	1.011 (0.07)	1.011 (0.05)	1.004 (0.03)	1.004 (0.02)	1.004 (0.01)	1.003 (0.01)	1.000 (0.01)	1.001 (0.00)
500	1.011 (0.04)	1.007 (0.03)	1.004 (0.02)	1.003 (0.01)	1.002 (0.01)	1.002 (0.01)		
1000	1.011 (0.03)	1.006 (0.02)	1.004 (0.01)	1.003 (0.01)	1.001 (0.01)	1.001 (0.00)		

Notes: This table shows the mean of 1000 simulations for the estimator Two-step. The standard deviations are in parentheses. $\beta = (\beta_1, \beta_2) = (1, 1)$. f is $T \times 1$. The idiosyncratic common factor of Y (f_y) is $T \times 1$, of X_1 (f_{x_1}) is $T \times 1$, of X_2 (f_{x_2}) is $T \times 1$ and there are no idiosyncratic common factor for Z (f_z=NULL). The errors are i.i.d. with normal distribution.

Table 6 – Simulation - Joint with idiosyncratic factor in Y and X

	10	15	30	50	100	200	500	1000
10	1.31 (2.77)	1.105 (6.02)	1.361 (1.41)	1.336 (1.53)	1.319 (0.73)	1.349 (0.78)	1.311 (0.81)	1.369 (1.03)
15	1.038 (0.37)	1.152 (0.51)	1.127 (0.33)	1.07 (0.24)	1.049 (0.18)	1.021 (0.12)	1.013 (0.08)	1.011 (0.07)
30	1.019 (0.17)	1.024 (0.13)	1.035 (0.14)	1.028 (0.1)	1.012 (0.07)	1.012 (0.05)	1.006 (0.03)	1.003 (0.03)
50	1.012 (0.12)	1.021 (0.09)	1.024 (0.06)	1.024 (0.06)	1.016 (0.04)	1.01 (0.03)	1.005 (0.02)	1.003 (0.01)
100	1.008 (0.08)	1.013 (0.06)	1.015 (0.04)	1.02 (0.03)	1.023 (0.02)	1.013 (0.02)	1.006 (0.01)	1.003 (0.01)
200	1.005 (0.06)	1.008 (0.04)	1.008 (0.03)	1.013 (0.02)	1.017 (0.01)	1.014 (0.01)	1.006 (0.01)	1.003 (0.00)
500	1.000 (0.03)	1.003 (0.03)	1.005 (0.02)	1.006 (0.01)	1.007 (0.01)	1.008 (0.01)		
1000	1.000 (0.03)	1.002 (0.02)	1.003 (0.01)	1.004 (0.01)	1.004 (0.01)	1.004 (0.00)		

Notes: This table shows the mean of 1000 simulations for the estimator Joint. The standard deviations are in parentheses. $\beta = (\beta_1, \beta_2) = (1, 1)$. f is $T \times 1$. The idiosyncratic common factor of Y (f_y) is $T \times 1$, of X_1 (f_{x_1}) is $T \times 1$, of X_2 (f_{x_2}) is $T \times 1$ and there are no idiosyncratic common factors for Z (f_z=NULL).

Without idiosyncratic factors

Tables 7 and 8 shows the simulations when the DGP is performed with only one factor (f, that is $T \times 1$) and has no idiosyncratic common factors for any variables.

Table 7 – β_1 Convergence - Two-Step estimator without idiosyncratic factors

	10	15	30	50	100	200	500	1000
10	1.02 (0.43)	0.989 (0.23)	0.993 (0.15)	1.000 (0.12)	0.997 (0.09)	0.999 (0.06)	0.998 (0.04)	1.001 (0.04)
15	1.016 (0.2)	1.025 (0.15)	0.995 (0.11)	0.991 (0.09)	0.997 (0.06)	1.001 (0.04)	1.000 (0.03)	1.001 (0.02)
30	1.007 (0.13)	1.009 (0.09)	1.017 (0.06)	0.999 (0.05)	1.001 (0.04)	1.000 (0.03)	1.000 (0.02)	1.000 (0.01)
50	1.01 (0.09)	1.008 (0.07)	1.011 (0.05)	1.01 (0.04)	1.001 (0.03)	1.000 (0.02)	1.000 (0.01)	1.000 (0.01)
100	1.004 (0.06)	1.004 (0.05)	1.004 (0.03)	1.006 (0.02)	1.005 (0.02)	1.000 (0.01)	1.000 (0.01)	1.000 (0.01)
200	1.000 (0.05)	1.003 (0.04)	1.002 (0.02)	1.002 (0.02)	1.003 (0.01)	1.003 (0.01)	1.000 (0.01)	1.000 (0)
500	1.001 (0.03)	1.001 (0.02)	1.001 (0.01)	1.001 (0.01)	1.001 (0.01)	1.001 (0.01)		
1000	1.001 (0.02)	1.000 (0.01)	1.000 (0.01)	1.000 (0.01)	1.000 (0.01)	1.001 (0)		

Notes: This table shows the mean of 1000 simulations for the estimator Two-step. The standard deviations are in parentheses. $\beta = (\beta_1, \beta_2) = (1, 1)$. f is $T \times 1$. There is not idiosyncratic common factor of Y (f_y), X_1 (f_{x_1}), X_2 (f_{x_2}), and Z (f_z). The errors are i.i.d. with normal distribution as in Section 5.1.

Table 8 – β_1 Convergence - Joint estimator without idiosyncratic factors

	10	15	30	50	100	200	500	1000
10	1.22 (2.47)	1.197 (2.72)	1.13 (0.4)	1.107 (0.32)	1.057 (0.23)	1.041 (0.15)	1.011 (0.1)	1.012 (0.09)
15	1.034 (0.34)	1.12 (0.41)	1.074 (0.26)	1.043 (0.19)	1.031 (0.14)	1.015 (0.09)	1.008 (0.06)	1.005 (0.05)
30	1.024 (0.17)	1.026 (0.12)	1.035 (0.12)	1.02 (0.09)	1.01 (0.06)	1.004 (0.04)	1.001 (0.03)	1.000 (0.02)
50	1.016 (0.12)	1.023 (0.08)	1.024 (0.05)	1.01 (0.06)	1.007 (0.04)	1.002 (0.03)	1.001 (0.02)	1.000 (0.01)
100	1.013 (0.07)	1.011 (0.05)	1.011 (0.03)	1.013 (0.02)	1.002 (0.02)	1.001 (0.02)	1.001 (0.01)	1.000 (0.01)
200	1.008 (0.05)	1.006 (0.03)	1.005 (0.02)	1.005 (0.02)	1.006 (0.01)	1.000 (0.01)	1.000 (0.01)	1.000 (0.00)
500	1.002 (0.03)	1.002 (0.02)	1.003 (0.01)	1.002 (0.01)	1.002 (0.01)	1.002 (0.01)		
1000	1.001 (0.02)	1.001 (0.01)	1.001 (0.01)	1.001 (0.01)	1.001 (0.01)	1.001 (0.00)		

Notes: This table shows the mean of 1000 simulations for the estimator Two-step. The standard deviations are in parentheses. $\beta = (\beta_1, \beta_2) = (1, 1)$. f is $T \times 1$. There is not idiosyncratic common factor of Y (f_y), X_1 (f_{x_1}), X_2 (f_{x_2}), and Z (f_z). The errors are i.i.d. with normal distribution as in Section 5.1.

Instrument idiosyncratic factor

Tables 9 and 10 shows the simulation where the DGP is done with one common factor that affect all variables (f, that is $T \times 1$) and one idiosyncratic common factors for the instrument (f_z, that is $T \times 1$).

Table 9 – β_1 Convergence - Two-Step estimator with idiosyncratic factor only on instrument

	10	15	30	50	100	200	500	1000
10	1.02 (0.43)	0.989 (0.23)	0.993 (0.15)	1.000 (0.12)	0.997 (0.09)	0.999 (0.06)	0.998 (0.04)	1.001 (0.04)
15	1.016 (0.2)	1.025 (0.15)	0.995 (0.11)	0.991 (0.09)	0.997 (0.06)	1.001 (0.04)	1.000 (0.03)	1.001 (0.02)
30	1.007 (0.13)	1.009 (0.09)	1.017 (0.06)	0.999 (0.05)	1.001 (0.04)	1.000 (0.03)	1.000 (0.02)	1.000 (0.01)
50	1.01 (0.09)	1.008 (0.07)	1.011 (0.05)	1.01 (0.04)	1.001 (0.03)	1.000 (0.02)	1.000 (0.01)	1.000 (0.01)
100	1.004 (0.06)	1.004 (0.05)	1.004 (0.03)	1.006 (0.02)	1.005 (0.02)	1.000 (0.01)	1.000 (0.01)	1.000 (0.01)
200	1.000 (0.05)	1.003 (0.04)	1.002 (0.02)	1.002 (0.02)	1.003 (0.01)	1.003 (0.01)	1.000 (0.01)	1.000 (0.00)
500	1.001 (0.03)	1.001 (0.02)	1.001 (0.01)	1.001 (0.01)	1.001 (0.01)	1.001 (0.01)		
1000	1.001 (0.02)	1.000 (0.01)	1.000 (0.01)	1.000 (0.01)	1.000 (0.01)	1.001 (0.00)		

Notes: This table shows the mean of 1000 simulations for the estimator Two-Step. The standard deviations are in parentheses. $\beta = (\beta_1, \beta_2) = (1, 1)$. f is $T \times 1$. The idiosyncratic common factor of Z (f_z) is $T \times 1$ and there are no idiosyncratic common factors for Y, X_1 and X_2. The errors are i.i.d. with normal distribution as in Section 5.1.

Table 10 – β_1 Convergence - Joint estimator with idiosyncratic factor only on instrument

	10	15	30	50	100	200	500	1000
10	1.124 (6.9)	1.283 (2.35)	1.162 (0.73)	1.137 (0.59)	1.101 (0.27)	1.054 (0.18)	1.028 (0.12)	1.013 (0.08)
15	1.021 (0.99)	1.323 (7.65)	1.121 (0.31)	1.103 (0.2)	1.065 (0.15)	1.03 (0.11)	1.017 (0.08)	1.009 (0.06)
30	1.029 (0.3)	1.073 (0.16)	1.112 (0.12)	1.07 (0.09)	1.036 (0.06)	1.017 (0.04)	1.008 (0.03)	1.005 (0.02)
50	1.044 (0.14)	1.065 (0.09)	1.071 (0.05)	1.06 (0.06)	1.033 (0.04)	1.018 (0.03)	1.007 (0.02)	1.003 (0.01)
100	1.036 (0.08)	1.035 (0.05)	1.038 (0.03)	1.037 (0.02)	1.03 (0.02)	1.016 (0.02)	1.007 (0.01)	1.003 (0.01)
200	1.022 (0.05)	1.02 (0.04)	1.021 (0.02)	1.019 (0.02)	1.02 (0.01)	1.015 (0.01)	1.006 (0.01)	1.003 (0.00)
500	1.007 (0.03)	1.008 (0.02)	1.008 (0.02)	1.008 (0.01)	1.008 (0.01)	1.008 (0.01)		
1000	1.003 (0.02)	1.004 (0.02)	1.004 (0.01)	1.004 (0.01)	1.004 (0.01)	1.004 (0.00)		

Notes: This table shows the mean of 1000 simulations for the estimator Joint. The standard deviations are in parentheses. $\beta = (\beta_1, \beta_2) = (1, 1)$. f is $T \times 1$. The idiosyncratic common factor of Z (f_z) is $T \times 1$ and there are no idiosyncratic common factors for Y, X_1 and X_2. The errors are i.i.d. with normal distribution as in Section 5.1.