UvA ECONOMETRICS

Discussion Paper: 2014/07

Fixed T Dynamic Panel Data Estimators with Multi-Factor Errors

Artūras Juodis and Vasilis Sarafidis

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Fixed T Dynamic Panel Data Estimators with Multi-Factor Errors[☆]

Artūras Juodis[a,b], Vasilis Sarafidis[c]

[a] *University of Amsterdam*
[b] *Tinbergen Institute*
[c] *Monash University*

Abstract

This paper analyzes a growing group of fixed T dynamic panel data estimators with a multi-factor error structure. We use a unified notational approach to describe these estimators and discuss their properties in terms of deviations from an underlying set of basic assumptions. Furthermore, we consider the extendability of these estimators to practical situations that may frequently arise, such as their ability to accommodate unbalanced panels. Using a large-scale simulation exercise, we consider scenarios that remain largely unexplored in the literature, albeit they are of great empirical relevance. In particular, we examine (i) the effect of the presence of weakly exogenous covariates, (ii) the effect of changing the magnitude of the correlation between the factor loadings of the dependent variable and those of the covariates, (iii) the impact of the number of moment conditions on bias and size for GMM estimators, and finally the effect of sample size. Thus, our study may serve as a useful guide to practitioners who wish to allow for multiplicative sources of unobserved heterogeneity in their model.

Keywords: Dynamic Panel Data, Factor Model, Maximum Likelihood, Fixed T Consistency, Monte Carlo Simulation.
JEL: C13, C15, C23.

1. Introduction

There is a large literature on estimating dynamic panel data models with a two-way error components structure and T fixed. Such models have been used in a wide range of economic and financial applications; e.g. Euler equations for household consumption, adjustment cost models for firms' factor demands and empirical models of economic growth. In all these cases the autoregressive parameter has structural significance and measures state dependence, which is due to the effect of habit formation, technological/regulatory constraints, or imperfect information and uncertainty that often underlie economic behavior and decision making in general.

[☆]Part of this paper was written while the first author enjoyed the hospitality of the Department of Econometrics and Business Statistics at Monash University. Financial support from the NWO MaGW grant "Likelihood-based inference in dynamic panel data models with endogenous covariates" is gratefully acknowledged by the first author. We would also like to thank seminar participants at Monash University and the Tinbergen Institute for useful comments.

[*]Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE, Amsterdam, The Netherlands. E.mail: a.juodis@uva.nl.

[**]Department of Econometrics and Business Statistics, Monash University, 900 Dandenong Road, Caulfield East, Victoria 3145, Australia. E.mail: vasilis.sarafidis@monash.edu

Recently there has been a surge of interest in developing dynamic panel data estimators that allow for richer error structures − mainly factor residuals. In this case standard dynamic panel data estimators fail to provide consistent estimates of the parameters; see e.g. Sarafidis and Robertson (2009), and Sarafidis and Wansbeek (2012) for a recent overview. The multi-factor approach is appealing because it allows for multiple sources of multiplicative unobserved heterogeneity, as opposed to the two-way error components structure that represents additive heterogeneity. For example, in an empirical growth model the factor component may reflect country-specific differences in the rate at which countries absorb time-varying technological advances that are potentially available to all of them. In a partial adjustment model of factor input prices, the factor component may capture common shocks that hit all producers, albeit with different intensities.

The majority of estimators developed in the literature are based on the Generalized Method of Moments (GMM) approach. In particular, Ahn, Lee, and Schmidt (2013) in a seminal paper extend Ahn, Lee, and Schmidt (2001) to the case of multiple factors, and propose a GMM estimator that relies on quasi-long-differencing to eliminate the common factor component. Nauges and Thomas (2003) utilise the quasi-differencing approach of Holtz-Eakin, Newey, and Rosen (1988), which is computationally tractable for the single factor case, and propose similar moment conditions to Ahn et al. (2001) *mutatis mutandis*. Sarafidis, Yamagata, and Robertson (2009) propose using the popular linear first-differenced and System GMM estimators with instruments based solely on strictly exogenous regressors. Robertson and Sarafidis (2013) develop a GMM approach that introduces new parameters to represent the unobserved covariances between the factor component of the error and the instruments. Furthermore, they show that given the model's structure there exist restrictions in the nuisance parameters that lead to a more efficient GMM estimator compared to quasi-differencing approaches. Hayakawa (2012) shows that the moment conditions proposed by Ahn et al. (2013) can be linearized at the expense of introducing extra parameters. Furthermore, following Bai (2013b), he discusses a GMM estimator that approximates the factor loadings using a Chamberlain (1982) type projection approach. Bai (2013b), on the other hand, proposes a maximum likelihood estimator.

This paper analyzes the aforementioned group of estimators. The objective of our study is to serve as a useful guide for practitioners who wish to allow for multiplicative sources of unobserved heterogeneity in their model. To achieve this, we describe all methods using a unified notational approach, to the extent that this is possible of course, and discuss their properties under deviations from a baseline set of assumptions commonly employed. We pay particular attention to computing the number of identifiable parameters correctly, which is a requirement for asymptotically valid inferences and consistent model selection procedures. Furthermore, we consider the extendability of these estimators to practical situations that may frequently arise, such as their ability to accommodate unbalanced panels, estimate models with common *observed* factors and others.

Next, we investigate the finite sample performance of the estimators under a number of different designs. In particular, we examine (i) the effect of the presence of weakly exogenous covariates, (ii) the effect of changing the magnitude of the correlation between the factor loadings of the dependent variable and those of the covariates, (iii) the impact of the number of moment conditions on bias and size for GMM estimators, (iv) the impact of different levels of persistence in the data, and finally the effect of sample size. These are important considerations with high empirical relevance. Notwithstanding, to the best our knowledge they remain largely unexplored. For example, the simulation study in Robertson and Sarafidis (2013) does not consider the effect of using a different number of instruments on the finite sample properties of the estimator. In Ahn, Lee, and Schmidt

(2001) the design focuses on strictly exogenous regressors, while in Bai (2013b) the results reported do not include inference. The practical issue of how to choose initial values for the non-linear algorithms is considered in the Appendix. The results of our simulation study indicate that there are non-negligible differences in the finite sample performance of the estimators, depending on the parameterisation considered. Naturally, no estimator dominates the remaining ones universally, although it is fair to say that some estimators are more robust than others.

The outline of the rest of the paper is as follows. The next section introduces the dynamic panel data model with a multi-factor error structure and discusses some underlying assumptions that are commonly employed in the literature. Section 3 presents a large range of dynamic panel estimators developed for such model when T is small, and discusses several technical points regarding their properties. Section 4 investigates the finite sample performance of the estimators. A final section concludes. The Appendix analyzes in detail the implementation of all these methods.

In what follows we briefly discuss notation. The usual $\text{vec}(\cdot)$ operator denotes column stacking operator, while $\text{vech}(\cdot)$ is the corresponding operator that stacks only the elements on and below the main diagonal. The commutation matrix $\mathbf{K}_{a,b}$ is defined such that for any $[a \times b]$ matrix $\boldsymbol{A}$, $\text{vec}(\boldsymbol{A}') = \mathbf{K}_{a,b}\,\text{vec}(\boldsymbol{A})$. The elimination matrix $\boldsymbol{B}_a$ is defined such that for any $[a \times a]$ matrix (not necessarily symmetric) $\text{vech}(\cdot) = \boldsymbol{B}_a\,\text{vec}(\cdot)$. The lag-operator matrix $\boldsymbol{L}_T$ is defined such that for any $[T \times 1]$ vector $\boldsymbol{x} = (x_1, \ldots, x_T)'$, $\boldsymbol{L}_T\boldsymbol{x} = (0, x_1, \ldots, x_{T-1})'$. The j^{th} column of the $[x \times x]$ identity matrix is denoted by $\boldsymbol{e}_j$. Finally, $\mathbb{I}_{(\cdot)}$ is the usual indicator function. For further details regarding the notation used in this paper see Abadir and Magnus (2002).

2. Model

We consider the following dynamic panel data model with a multi-factor error structure:

$$y_{i,t} = \alpha y_{i,t-1} + \sum_{k=1}^{K} \beta_k x_{i,t}^{(k)} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{i,t}; \quad i = 1, \ldots, N, t = 1, \ldots, T, \tag{1}$$

where the dimension of the unobserved components $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$ is $[L \times 1]$. Stacking the observations over time for each individual i yields

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,T})'$ and similarly for $(\boldsymbol{y}_{i,-1}, \boldsymbol{x}_i^{(k)})$, while $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)'$ and is of dimension $[T \times L]$. In what follows we list some assumptions that are commonly employed in the literature, followed by some preliminary discussion. In Section 3 we provide further discussion with regards to which of these assumptions can be strengthened/relaxed for each estimator analyzed.

Assumption 1: $x_{i,t}^{(k)}$ has finite moments up to fourth order for all k;

Assumption 2: $\varepsilon_{i,t} \sim i.i.d.\,(0, \sigma_\varepsilon^2)$ and has finite moments up to fourth order;

Assumption 3: $\boldsymbol{\lambda}_i \sim i.i.d.\,(\mathbf{0}, \boldsymbol{\Sigma}_\lambda)$ with finite moments up to fourth order, where $\boldsymbol{\Sigma}_\lambda$ is positive definite. $\boldsymbol{F}$ is non-stochastic and uniformly bounded such that $||\boldsymbol{F}|| < b < \infty$;

Assumption 4: $E\left(\varepsilon_{it}|y_{i0},...,y_{it-1},\boldsymbol{\lambda}_i',x_{i1}^{(k)},...x_{i\tau}^{(k)}\right)=0$ for all t and k.

Assumption 1 is a standard regularity condition. Assumptions 2-3 are employed mainly for simplicity and can be relaxed to some extent, details of which will be documented later.[1]

Assumption 4 can be crucial for identification, depending on the estimation approach. To begin with, it implies that the idiosyncratic errors are conditionally serially uncorrelated. This can be relaxed in a relatively straightforward way, particularly for GMM estimators; for example, one could assume instead that either $E\left(\varepsilon_{it}|y_{i0},...,y_{is},\boldsymbol{\lambda}_i',x_{i1}^{(k)},...x_{i\tau}^{(k)}\right)=0$, where $s<t-1$, or $E\left(\varepsilon_{it}|\boldsymbol{\lambda}_i',x_{i1}^{(k)},...x_{i\tau}^{(k)}\right)=0$. In the former case a moving average process of a certain order in ε_{it} is permitted and moment conditions with respect to (lagged values of) y_{is} can be used. In the latter case, an autoregressive process in ε_{it} is permitted and moment conditions with respect to (lagged values of) $x_{i\tau}^{(k)}$ remain valid.
In addition, Assumption 4 implies that the idiosyncratic error is conditionally uncorrelated with the factor loadings. This is required for identification based on internal instruments in levels. Moreover, Assumption 4 characterises the exogeneity properties of the covariates. In particular, we will refer to covariates that satisfy $\tau=T$ as strictly exogenous with respect to the idiosyncratic error component, whereas covariates that satisfy only $\tau=t$ are weakly exogenous. When $\tau<t$ the covariates are endogenous. The exogeneity properties of the covariates play a major role in the analysis of likelihood based estimators because the presence of weakly exogenous or endogenous regressors may lead to inconsistent estimates of the structural parameters, α and β_k. Finally, notice that the set of our assumptions implies that y_{it} has finite fourth-order moments, but it does not imply conditional homoskedasticity for the two error components.

Under Assumptions 1-4, the following set of population moment conditions is valid by construction:

$$\mathrm{E}[\mathrm{vech}(\boldsymbol{\varepsilon}_i\boldsymbol{y}_{i,-1}')]=\mathbf{0}_{T(T+1)/2}. \tag{2}$$

In addition, the following sets of moment conditions are valid, depending on whether $\tau=T$ or $\tau=t$ hold true, respectively:

$$\mathrm{E}[\mathrm{vec}(\boldsymbol{\varepsilon}_i\boldsymbol{x}_i^{(k)\prime})]=\mathbf{0}_{T^2}; \tag{3}$$

$$\mathrm{E}[\mathrm{vech}(\boldsymbol{\varepsilon}_i\boldsymbol{x}_i^{(k)\prime})]=\mathbf{0}_{T(T+1)/2}. \tag{4}$$

For all GMM estimators one can easily modify the above moment conditions to allow for endogenous x's. For example, for (say) $\tau=t-1$ one may redefine $\boldsymbol{x}_i^{(k)}:=(x_{i,0},\dots,x_{i,T-1})'$ and proceed in exactly the same way.
From now on we will use the triangular structure of the moment conditions induced by the vech$(\cdot)$ operator to construct the estimating equations for the GMM estimators. To achieve this we adopt the following matrix notation for the stacked model:

$$\boldsymbol{Y}=\alpha\boldsymbol{Y}_{-1}+\sum_{k=1}^{K}\beta_k\boldsymbol{X}_k+\boldsymbol{\Lambda}\boldsymbol{F}'+\boldsymbol{E};\quad i=1,\dots,N,$$

[1] The zero-mean assumption for $\varepsilon_{i,t}$ is actually implied by Assumption 4.

where $(\boldsymbol{Y}, \boldsymbol{Y}_{-1}, \boldsymbol{X}_k, \boldsymbol{E})$ are $[N \times T]$ matrices with typical rows $(\boldsymbol{y}_i', \boldsymbol{y}_{i,-1}', \boldsymbol{x}_i^{(k)\prime}, \boldsymbol{\varepsilon}_i')$ respectively. Similarly a typical row element of $\boldsymbol{\Lambda}$ is given by $\boldsymbol{\lambda}_i'$.

3. Estimators

Remark 1. For notational symmetry, while describing GMM estimators we assume that $x_{i,0}^{(k)}$ observations are not included in the set of available instruments. Otherwise additional T or $T-1$ (depending on the estimator analyzed) moment conditions are available. The same strategy is used in the Monte Carlo section of this paper.

3.1. ***Holtz-Eakin, Newey, and Rosen (1988)/Nauges and Thomas (2003)***

The finite sample analogues of the population moment conditions in equation (2) are given by

$$\operatorname{vech}\left(\frac{1}{N}(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k - \boldsymbol{\Lambda}\boldsymbol{F}')'\boldsymbol{Y}_{-1}\right);$$
$$\operatorname{vech}\left(\frac{1}{N}(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k - \boldsymbol{\Lambda}\boldsymbol{F}')'\boldsymbol{X}_k\right).$$

These moment conditions depend on the unknown matrices $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$. In the simple fixed effects model where $\boldsymbol{F} = \boldsymbol{\imath}_T$, the first-differencing transformation proposed by Anderson and Hsiao (1982) is the most common approach to eliminate the fixed effects from the equation of interest. Using a similar idea in the model with only one unobserved time varying factor, i.e.

$$y_{i,t} = \alpha y_{i,t-1} + \sum_{k=1}^{K}\beta_k x_{i,t}^{(k)} + \lambda_i f_t + \varepsilon_{i,t}; \quad i = 1,\dots,N, t = 1,\dots,T,$$

Holtz-Eakin, Newey, and Rosen (1988) suggest eliminating the unobserved factor component using the following quasi-differencing (QD) transformation:

$$y_{i,t} - r_t y_{i,t-1} = \alpha(y_{i,t-1} - r_t y_{i,t-2}) + \sum_{k=1}^{K}\beta_k(x_{i,t}^{(k)} - r_t x_{i,t-1}^{(k)}) + \varepsilon_{i,t} - r_t\varepsilon_{i,t-1}; \quad i = 1,\dots,N, t = 2,\dots,T, \tag{5}$$

where $r_t = \frac{f_t}{f_{t-1}}$. By construction equation (5) is free from $\lambda_i f_t$ because

$$\lambda_i f_t - r_t\lambda_i f_{t-1} = \lambda_i f_t - \frac{f_t}{f_{t-1}}\lambda_i f_{t-1} = 0, \quad \forall t = 2,\dots,T.$$

It is easy to see that the QD approach is well defined only if *all* $f_t \neq 0$. Collecting all parameters involved in quasi-differencing we can define the corresponding $[T-1 \times T]$ QD transformation matrix by

$$\boldsymbol{D}(\boldsymbol{r}) = \begin{pmatrix} -r_2 & 1 & 0 & \cdots & 0 \\ 0 & -r_3 & & \vdots & 0 \\ \vdots & \vdots & \vdots & 1 & \vdots \\ 0 & 0 & \dots & -r_T & 1 \end{pmatrix},$$

with the first-differencing (FD) transformation being a special case with $r_2 = \ldots = r_T = 1$. Pre-multiplying the terms inside the vech$(\cdot)$ operator in the sample analogue of the population moment conditions above by $\boldsymbol{D}(\boldsymbol{r})$, and noticing that $\boldsymbol{D}(\boldsymbol{r})\boldsymbol{F} = \boldsymbol{0}$, we can rewrite the estimating equations for the QD estimator as

$$\boldsymbol{m}_l = \text{vech}\left(\frac{1}{N}\boldsymbol{D}(\boldsymbol{r})\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{Y}_{-1}\boldsymbol{J}'\right);$$
$$\boldsymbol{m}_k = \text{vech}\left(\frac{1}{N}\boldsymbol{D}(\boldsymbol{r})\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{X}_k\boldsymbol{J}'\right) \quad \forall k.$$

Here $\boldsymbol{J} = (\boldsymbol{I}_{T-1}, \boldsymbol{0}_{T-1})$ is a selection matrix that appropriately truncates the whole set of instruments in order to ensure that the term inside the vech$(\cdot)$ operator is a square matrix. One can easily see that the total number of moment conditions and parameters under the weak exogeneity assumption for all x is given by

$$\#moments = \frac{(K+1)(T-1)T}{2}; \quad \#parameters = K + 1 + (T-1).$$

Here the total number of parameters consists of two terms. The first term corresponds to $K+1$ parameters of interest (or *structural/model* parameters), while there are $T-1$ nuisance parameters corresponding to time-varying factors.

The approach of Holtz-Eakin et al. (1988) as it stands is tailored for models with one unobserved factor. In principle, it can be extended to multiple factors by removing each factor consecutively based on a $\boldsymbol{D}_{(l)}(\boldsymbol{r}^{(l)})$ matrix, with the final transformation matrix being a product of an L matrix of that type. However, this approach soon becomes computationally very cumbersome as the estimating equations become multiplicative in $\boldsymbol{r}^{(l)}$. On the other hand, if the model involves some observed factors, the corresponding $\boldsymbol{D}_{(\cdot)}(\cdot)$ matrix is known, leading to a simple estimator that involves equations containing $\boldsymbol{r}$ and structural parameters only. For example, Nauges and Thomas (2003) augment the model of Holtz-Eakin et al. (1988) by allowing for time-invariant fixed effects:

$$y_{i,t} = \eta_i + \alpha y_{i,t-1} + \sum_{k=1}^{K}\beta_k x_{i,t}^{(k)} + \lambda_i f_t + \varepsilon_{i,t}; \quad i = 1, \ldots, N, t = 1, \ldots, T,$$

where η_i is eliminated using the FD transformation matrix $\boldsymbol{D}(\boldsymbol{\imath}_{T-1})$, which yields

$$\Delta y_{i,t} = \alpha\Delta y_{i,t-1} + \sum_{k=1}^{K}\beta_k\Delta x_{i,t}^{(k)} + \lambda_i\Delta f_t + \Delta\varepsilon_{i,t}; \quad i = 1, \ldots, N, t = 1, \ldots, T,$$

followed by the QD transformation, albeit operated based on a $[T-2 \times T-1]$ matrix $\boldsymbol{D}(\boldsymbol{r})$. The resulting number of parameters and moment conditions can be modified accordingly from those in Holtz-Eakin et al. (1988).

Remark 2. The FD transformation is by no means the only way to eliminate the fixed effects from the model. Another commonly discussed transformation is *Forward Orthogonal Deviations* (FOD). If one uses FOD instead of FD, the identification of structural parameters would require that all

$f_t^* \neq 0$.[2] Depending on the properties of f's one might prefer to use FOD or FD in the framework of Nauges and Thomas (2003).

Remark 3. Assumption 2 can be easily relaxed. For example, unconditional time-series and cross-sectional heteroskedasticity of the idiosyncratic error component, $\varepsilon_{i,t}$, is allowed in the two-step version of the estimator. Serial correlation can be accommodated by choosing the set of instruments appropriately, as in the discussion provided in Section 2. This is a particular attractive feature, which is common to all GMM estimators discussed in this paper. Unconditional heteroskedasticity in $\boldsymbol{\lambda}_i$ can also be allowed, although this is a less interesting extension for practical purposes since there are no repeated observations over each $\boldsymbol{\lambda}_i$.

The condition in Assumption 4 that implies no conditional correlation between the idiosyncratic error and the factor loadings could be relaxed in principle, although this is far less trivial because the moment conditions in (2) are violated in this case. Using instruments with respect to variables expressed in quasi-differences may provide a valid identification strategy. However, computationally the estimation task becomes far more complex.

Finally, endogeneity of the regressors can be easily allowed. The exogeneity property of the covariates can be determined using an overidentifying restrictions test statistic. The same holds for all GMM estimators discussed in this paper, which is of course a desirable property from the empirical point of view since the issue of endogeneity in panels with T fixed, e.g. microeconometric panels, may frequently arise.

3.2. ***Ahn, Lee, and Schmidt (2013)***

As we have mentioned before, the QD approach in Holtz-Eakin et al. (1988) is difficult to generalise to more than one factor (or one unobserved factor plus observed factors). Rather than eliminating factors using the FD type transformation, Ahn, Lee, and Schmidt (2013) propose using a quasi-long-differencing (QLD) type transformation. To explain this approach we partition $\boldsymbol{F} = (\boldsymbol{F}_A', -\boldsymbol{F}_B')'$ where $\boldsymbol{F}_A$ and $\boldsymbol{F}_B$ are of dimensions $[T - L \times L]$ and $[L \times L]$ respectively. Then assuming that $\boldsymbol{F}_B$ is invertible, one can redefine factors and factor loadings as

$$\boldsymbol{F}\boldsymbol{\lambda}_i = \begin{pmatrix} \boldsymbol{F}^* \\ -\boldsymbol{I}_L \end{pmatrix} \boldsymbol{\lambda}_i^*; \quad \boldsymbol{F}^* = \boldsymbol{F}_A \boldsymbol{F}_B^{-1}; \quad \boldsymbol{\lambda}_i^* = \boldsymbol{F}_B \boldsymbol{\lambda}_i.$$

Using this normalization Ahn et al. (2013) propose eliminating the factors using the following QLD transformation matrix $\boldsymbol{D}(\boldsymbol{F}^*)$:

$$\boldsymbol{D}(\boldsymbol{F}^*) = (\boldsymbol{I}_{T-L}, \boldsymbol{F}^*) = \boldsymbol{J} + \boldsymbol{F}^* \boldsymbol{J}_L; \quad \boldsymbol{J} = (\boldsymbol{I}_{T-L}, \mathbf{O}_{T-L \times L}),$$

where $\boldsymbol{J}_L = (\mathbf{O}_{L \times (T-L)}, \boldsymbol{I}_L)$, an $[L \times T]$ matrix. As a result one can express all available moment conditions for this estimator as

$$\boldsymbol{m}_l = \operatorname{vech}\left(\boldsymbol{D}(\boldsymbol{F}^*) \frac{1}{N} \left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k \right)' \boldsymbol{Y}_{-1} \boldsymbol{J}' \right);$$
$$\boldsymbol{m}_k = \operatorname{vech}\left(\boldsymbol{D}(\boldsymbol{F}^*) \frac{1}{N} \left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k \right)' \boldsymbol{X}_k \boldsymbol{J}' \right) \quad \forall k.$$

[2] Here $f_t^* \equiv c_t(f_t - (f_{t+1} + \ldots + f_T)/(T-t))$ with $c_t^2 = (T-t)/(T-t+1)$.

Counting the number of moment conditions and resulting parameters we have

$$\#moments = \frac{(K+1)(T-L)(T-L+1)}{2}; \quad \#parameters = K+1+(T-L)L.$$

However, we will further argue that the number of identifiable parameters is smaller than $K+1+(T-L)L$. To explain the reason for this, rewrite the equation for $y_{i,1}$ as

$$y_{i,1} + \sum_{l=1}^{L} f_1^{(l)} y_{i,T-l} = \alpha \left(y_{i,0} + \sum_{l=1}^{L} f_1^{(l)} y_{i,T-l-1} \right) + \beta \left(x_{i,1} + \sum_{l=1}^{L} f_1^{(l)} x_{i,T-l} \right) + \dots \tag{6}$$

This equation has $2+L$ unknown parameters in total, while the number of moment conditions is 2 ($y_{i,0}$ and $x_{i,1}$). Thus, L "nuisance parameters" are identified only up to a linear combination, unless $L \leq 2$ (or $K+1$ for the general model), and the total number of identifiable parameters is

$$\#parameters = K+1+(T-L)L - \mathbb{I}_{(L \geq K+1)} \frac{(L-K-1)(L-K)}{2}.$$

Remark 3 regarding Assumptions 2-4, as discussed above, applies identically here as well. Ahn et al. (2013) show that under conditional homoskedasticity in $\varepsilon_{i,t}$ the estimation procedure simplifies considerably because it can be performed through iterations. Furthermore, for the case where the regressors are strictly exogenous, the resulting estimator is invariant to the normalization scheme; see their Appendix A.

3.3. ***Robertson and Sarafidis (2013)***

3.3.1. Unrestricted Estimator FIVU

Rather than removing the incidental parameters $\boldsymbol{\lambda}_i$, Robertson and Sarafidis (2013) propose a GMM estimator that makes use of centered moment conditions of the following form:

$$\boldsymbol{m}_l = \mathrm{vech}\left(\frac{1}{N} \left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k \right)' \boldsymbol{Y}_{-1} - \boldsymbol{F}\boldsymbol{G}' \right);$$
$$\boldsymbol{m}_k = \mathrm{vech}\left(\frac{1}{N} \left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k \right)' \boldsymbol{X}_k - \boldsymbol{F}\boldsymbol{G}_k' \right) \quad \forall k,$$

where the true values of the $(\boldsymbol{G}, \boldsymbol{G}_k)$ matrices are defined as

$$\boldsymbol{G} = \mathrm{E}[\boldsymbol{y}_{i,-1}\boldsymbol{\lambda}_i']; \quad \boldsymbol{G}_k = \mathrm{E}[\boldsymbol{x}_i^{(k)}\boldsymbol{\lambda}_i'],$$

with typical row elements $\boldsymbol{g}_t'$ and $\boldsymbol{g}_t^{(k)'}$ respectively. The $(\boldsymbol{G}, \boldsymbol{G}_k)$ matrices essentially represent the unobserved covariances between the instruments and the factor loadings in the error term. This approach adopts essentially a random effects treatment of the factor loadings, which is natural because N is large and there are no repeated observations over $\boldsymbol{\lambda}_i$. Notice that as in Holtz-Eakin et al. (1988) and Ahn, Lee, and Schmidt (2013), factors corresponding to loadings that are uncorrelated with the regressors can be accommodated through the variance-covariance matrix of the idiosyncratic error component, $\varepsilon_{i,t}$, since the latter is left unrestricted.

The total number of moment conditions is given by

$$\#moments = \frac{(K+1)T(T+1)}{2}.$$

As the model stands right now, $\boldsymbol{G}$ (all $K+1$) and $\boldsymbol{F}$ are not separately identifiable because

$$\boldsymbol{F}\boldsymbol{G}' = \boldsymbol{F}\boldsymbol{U}\boldsymbol{U}^{-1}\boldsymbol{G}'$$

for any invertible $[L \times L]$ matrix $\boldsymbol{U}$. This rotational indeterminacy is typically eliminated in the factor literature by requiring an $[L \times L]$ submatrix of $\boldsymbol{F}$ to be the identity matrix. These restrictions correspond to the L^2 term in the equation below. Furthermore, additional normalizations are required due to the fact that the moment conditions are of a vech$(\cdot)$ type. In particular, the number of identifiable parameters is

$$\#parameters = (K+1)(1+TL) + TL - L^2 - (K+1)\frac{L(L-1)}{2} - \mathbb{I}_{(L \geq K+1)}\frac{(L-K-1)(L-K)}{2}.$$

The $(K+1)L(L-1)/2$ term corresponds to the unobserved "last" $\boldsymbol{g}$, while the last term involving the indicator function corresponds to the unobserved "first" $\boldsymbol{f}$ and is identical to the right-hand side term in the corresponding expression for Ahn, Lee, and Schmidt (2013).

Notwithstanding, as shown in Robertson and Sarafidis (2013) if one is only interested in the structural parameters, α and β_k, it is not essential to impose any identifying normalizations on $\boldsymbol{G}$ and $\boldsymbol{F}$; the resulting unrestricted estimator for structural parameters is consistent and asymptotically normal, while the variance-covariance matrix can be consistently estimated using the corresponding sub-block of the generalized inverse of the unrestricted variance-covariance matrix.[3]

Compared with the QLD estimator of Ahn et al. (2013) this estimator utilises $(K+1)L[T-(L-1)/2])$ extra moment conditions, at the expense of estimating exactly the same number of additional parameters. Hence these estimators are asymptotically equivalent.

3.3.2. Restricted Estimator FIVR

The autoregressive nature of the model suggests that individual rows of the $\boldsymbol{G}$ matrix have also an autoregressive structure, i.e.

$$\boldsymbol{g}_t = \alpha \boldsymbol{g}_{t-1} + \sum_{k=1}^{k} \beta_k \boldsymbol{g}_t^{(k)} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}} \boldsymbol{f}_t.$$

For identification one may impose $L(L+1)/2$ restrictions so that w.l.o.g. $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} = \boldsymbol{I}_L$. Thus, one can express $\boldsymbol{F}$ in terms of other parameters as follows:

$$\boldsymbol{F} = \left(\boldsymbol{L}_T' - \alpha \boldsymbol{I}_T\right) \boldsymbol{G} + \boldsymbol{e}_T \boldsymbol{g}_T' - \sum_{k=1}^{k} \beta_k \boldsymbol{G}_k.$$

Here $\boldsymbol{L}_T$ is the usual lag matrix, while the additional parameter $\boldsymbol{g}_T$ is introduced to take into account the fact that in the original set of moment conditions $\boldsymbol{g}_T = \mathrm{E}[\boldsymbol{\lambda}_i y_{i,T}]$ does not appear as a parameter.

[3]For further details see Theorem 3 in the corresponding paper.

Robertson and Sarafidis (2013) show that FIVR is asymptotically more efficient than FIVU and procedures that involve some form of differencing. Furthermore, the restrictions imposed on a subset of the nuisance parameters provide substantial efficiency gains in finite samples.

Counting the total number of moment conditions and parameters, we have

$$\#moments = \frac{(K+1)T(T+1)}{2}; \quad \#parameters = (K+1)(1+TL) + L - (K+1)\frac{L(L-1)}{2}.$$

Remark 4. In principle we have additional T moment conditions (by the zero mean assumption of $\varepsilon_{i,t}$ for each time period t), given by

$$\boldsymbol{m}_\iota = \mathrm{vec}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{\imath}_N - \boldsymbol{F}\boldsymbol{g}_\iota\right).$$

Here $\boldsymbol{g}_\iota$ represents the mean of λ_i. The same is exactly true for Ahn et al. (2013), although there exist $(T-L)$ moment conditions in that case.

3.4. *Linear Hayakawa (2012)*

Hayakawa (2012) proposes a linearized GMM version of the QLD model in Ahn et al. (2013) under strict exogeneity. The moment conditions can be written as follows:

$$\boldsymbol{m}_l = \mathrm{vech}\left(\frac{1}{N}\left(\boldsymbol{Y}(\boldsymbol{J} + \boldsymbol{F}^*\boldsymbol{J}_L)' - \boldsymbol{Y}_{-1}(\alpha\boldsymbol{J} + \boldsymbol{F}^*_\alpha\boldsymbol{J}_L)' - \sum_{k=1}^{K}\boldsymbol{X}_k\left(\beta\boldsymbol{J} + \boldsymbol{F}^*_{\beta_k}\boldsymbol{J}_L\right)'\right)'\boldsymbol{Y}_{-1}\boldsymbol{J}'\right);$$

$$\boldsymbol{m}_k = \mathrm{vec}\left(\frac{1}{N}\left(\boldsymbol{Y}(\boldsymbol{J} + \boldsymbol{F}^*\boldsymbol{J}_L)' - \boldsymbol{Y}_{-1}(\alpha\boldsymbol{J} + \boldsymbol{F}^*_\alpha\boldsymbol{J}_L)' - \sum_{k=1}^{K}\boldsymbol{X}_k\left(\beta\boldsymbol{J} + \boldsymbol{F}^*_{\beta_k}\boldsymbol{J}_L\right)'\right)'\boldsymbol{X}_k\right) \quad \forall k.$$

The estimator of Ahn et al. (2013) can be obtained directly by noting that

$$\boldsymbol{F}^*_\alpha = \alpha\boldsymbol{F}^*; \quad \boldsymbol{F}^*_{\beta_k} = \beta_k\boldsymbol{F}^*.$$

In total, under strict exogeneity of all $x^{(k)}_{i,t}$ we have

$$\#moments = \frac{(T-L)(T-L+1)}{2} + KT(T-L);$$

$$\#parameters = \underbrace{K+1+(T-L)L}_{ALS} + \underbrace{(T-L)L(K+1)}_{linearization} - \frac{L(L-1)}{2}.$$

Notice that the last term in the equation for the total number of parameters is not present in the original study of Hayakawa (2012). To explain the necessity of this term consider the $T-L$'th equation (for ease of exposition we set $L=2$) *without* exogenous regressors:

$$y_{i,T-2} - f^{(1)}_{T-2}y_{i,T} - f^{(2)}_{T-2}y_{i,T-1} = \alpha y_{i,T-3} + f^{(1)}_{\alpha_{T-2}}y_{i,T-1} + f^{(2)}_{\alpha_{T-2}}y_{i,T-2} + \varepsilon_{T-2,t} - f^{(1)}_{T-2}\varepsilon_{i,T} - f^{(2)}_{T-2}\varepsilon_{i,T-1}.$$

Clearly only $f^{(2)}_{T-2} + f^{(1)}_{\alpha_{T-2}}$ can be identified but not the individual terms separately. As a result $L(L-1)/2$ normalizations need to be imposed. Furthermore, as it can be easily seen this term is unaltered if additional regressors are present in the model so long as they do not contain other lags of $y_{i,t}$ or lags of exogenous regressors.

Remark 5. In principle one can use the same linearisation strategy in the Holtz-Eakin, Newey, and Rosen (1988) approach.

3.4.1. Linearized GMM Hayakawa (2012) under weak exogeneity

For simplicity consider only the case with a single weakly exogenous regressor. Observe that we can rewrite the first equation of the transformed model as

$$y_{i,1} + \sum_{l=1}^{L} f_1^{(l)} y_{i,T-l} = \alpha y_{i,0} + \beta x_{i,1} + \sum_{l=1}^{L} f_{\alpha_1}^{(l)} y_{i,T-l-1} + \sum_{l=1}^{L} f_{\beta_1}^{(l)} x_{i,T-l} + \dots \tag{7}$$

This equation contains $2 + 3L$ unknown parameters, with only two available moment conditions (assuming $x_{i,0}$ is not observed, otherwise 3). Hence the full set of parameters in this equation cannot be identified without further normalizations. It then follows that the minimum value of T required in order to identify the structural parameters of interest is such that (for simplicity assume $L = 1$):

$$2(T-1) = 2 + 3 \Longrightarrow \quad \min\{T\} = 1 + \lceil 2.5 \rceil = 4.$$

For more general models with $K > 1$, the condition $\min\{T\} = 4$ continues to hold as

$$(K+1)(T-1) \geq (K+2) + (K+1) \Longrightarrow \quad \min\{T\} = 1 + \left\lceil \frac{2K+3}{K+1} \right\rceil = 4.$$

Notice that for the non-linear estimator $\min\{T\} = 3$ in the single-factor case. As a result, for $L = 1$ under weak exogeneity the number of identifiable parameters and moment conditions is given by

$$\begin{aligned}
\#moments &= (K+1)\frac{(T-L)(T-L+1)}{2} - (K+1);\\
\#parameters &= \underbrace{K+1+(T-L)L}_{ALS} + \underbrace{(T-L)L(K+1)}_{linearization} - \frac{L(L-1)}{2} - (K+2),
\end{aligned}$$

where $-(K+1)$ and $-(K+2)$ adjustments are made to take into account the fact that for $t = 1$ there are $(K+2)$ nuisance parameters to be estimated with $(K+1)$ available moment conditions. Both expressions can be similarly modified for $L > 1$.

3.5. ***GMM with projection Hayakawa (2012)***

Following Bai (2013b), Hayakawa (2012) suggests approximating $\boldsymbol{\lambda}_i$ using a Mundlak (1978)-Chamberlain (1982) type projection of the following form:

$$\boldsymbol{\lambda}_i = \boldsymbol{\Phi} \boldsymbol{z}_i + \boldsymbol{\nu}_i,$$

where $\boldsymbol{z}_i = (1, \boldsymbol{x}_i^{(1)\prime}, \dots, \boldsymbol{x}_i^{(K)\prime}, y_{i,0})'$. Notice that by construction $\boldsymbol{\nu}_i$ is uncorrelated with $\boldsymbol{z}_i$. As a result, the stacked model for individual i can be written as

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i-} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{z}_i + \boldsymbol{F}\boldsymbol{\nu}_i + \boldsymbol{\varepsilon}_i. \tag{8}$$

While Bai (2013b) proposes maximum likelihood estimation of the above model, Hayakawa (2012) advocates a GMM estimator; in our standard notation the total set of moment conditions is given by

$$\boldsymbol{m}_l = \operatorname{vec}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k - \boldsymbol{Z\Phi}'\boldsymbol{F}'\right)'\boldsymbol{Y}_{-1}\boldsymbol{e}_1\right);$$
$$\boldsymbol{m}_\iota = \left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k - \boldsymbol{Z\Phi}'\boldsymbol{F}'\right)'\boldsymbol{\imath}_N\right);$$
$$\boldsymbol{m}_k = \operatorname{vech}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k - \boldsymbol{Z\Phi}'\boldsymbol{F}'\right)'\boldsymbol{X}_k\right), \quad \forall k.$$

Assuming weak exogeneity we have

$$\#moments = 2T + \frac{KT(T+1)}{2};$$
$$\#parameters = \underbrace{(K+1) + (T-L)L}_{ALS} + \underbrace{L(TK+2)}_{Projection}.$$

Similarly to the FIVU estimator of Robertson and Sarafidis (2013) the number of identifiable parameters is smaller than the nominal one and depends on the projected variables $\boldsymbol{z}_i$.

3.6. Equivalence with FIVU

As described in Bond and Windmeijer (2002), consider a more general projection specification of the following form:

$$\boldsymbol{\lambda}_i = \boldsymbol{\Phi z}_i + \boldsymbol{\nu}_i,$$

where $\boldsymbol{z}_i = (\boldsymbol{x}_i^{(1)'}, \ldots, \boldsymbol{x}_i^{(K)'}, \boldsymbol{y}_{i-}')'$. The true value of $\boldsymbol{\Phi}$ has the usual expression for the projection estimator

$$\boldsymbol{\Phi}_0 := \mathrm{E}\left[\boldsymbol{\lambda}_i\boldsymbol{z}_i'\right]\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]^{-1}.$$

The first term in the notation of Robertson and Sarafidis (2013) is simply

$$\mathrm{E}\left[\boldsymbol{\lambda}_i\boldsymbol{z}_i'\right] = (\boldsymbol{G}_1', \ldots, \boldsymbol{G}_K', \boldsymbol{G}'). \tag{9}$$

This estimator coincides asymptotically with the FIVU estimator of Robertson and Sarafidis (2013), as well as with the QLD estimator of Ahn et al. (2013) if all $T(T+1)(K+1)/2$ moment conditions are used. A proof for the equivalence between FIVU and QLD is given in Robertson and Sarafidis (2013).

3.7. ***Sarafidis, Yamagata, and Robertson (2009)***

In their discussion of the test for cross-sectional dependence, Sarafidis et al. (2009) observe that if one can assume

$$\boldsymbol{x}_{i,t} = \boldsymbol{\Pi}(\boldsymbol{x}_{i,t-1}, \ldots, \boldsymbol{x}_{i,0}) + \boldsymbol{\Gamma}_{xi}\boldsymbol{f}_t + \boldsymbol{\pi}(\varepsilon_{i,t-1}, \ldots, \varepsilon_{i,0}) + \boldsymbol{\varepsilon}_{i,t}^x \tag{10}$$

where $\boldsymbol{\Pi}(\cdot)$ and $\boldsymbol{\pi}(\cdot)$ are measurable functions, and the stochastic components are such that

$$\begin{aligned}\mathrm{E}[\boldsymbol{\varepsilon}^x_{i,s}\varepsilon_{i,l}] &= \mathbf{0}_K, \forall s, l;\\ \mathrm{E}[\mathrm{vec}(\boldsymbol{\Gamma}_{xi})\boldsymbol{\lambda}'_i] &= \mathbf{O}_{KL\times L},\end{aligned}$$

then the following GMM moment conditions are valid even in the presence of unobserved factors in both equations for $y_{i,t}$ and $\boldsymbol{x}_{i,t}$:

$$\begin{aligned}&\mathrm{E}[(y_{i,t} - \alpha y_{i,t-1} - \boldsymbol{\beta}'\boldsymbol{x}_{i,t})\Delta\boldsymbol{x}_{i,s}] = 0, \forall s \leq t;\\ &\mathrm{E}[(\Delta y_{i,t} - \alpha\Delta y_{i,t-1} - \boldsymbol{\beta}'\Delta\boldsymbol{x}_{i,t})\boldsymbol{x}_{i,s}] = 0, \forall s \leq t-1.\end{aligned}$$

The total number of valid (*non-redundant*) moment conditions is given by

$$\#moments = K\left(\frac{(T-1)T}{2} + (T-1)\right),$$

if one does not include $\boldsymbol{x}_{i,0}$ and $\Delta\boldsymbol{x}_{i,1}$ among the instruments. Under mean stationarity additional moment conditions become available in the equations in levels, giving rise to a system GMM estimator.

Identification of the structural parameters crucially depends on the fact that no lagged values of $y_{i,t}$ are present in (10) as well as uncorrelated factor loadings. However, it is important to stress that all exogenous regressors are allowed to be weakly exogenous due to the possible non-zero $\boldsymbol{\pi}(\cdot)$ function, or even endogenous provided that $\varepsilon_{i,t}$ is serially uncorrelated.

3.8. ***Maximum Likelihood estimator of Bai (2013b)***

As in Hayakawa (2012) this estimator uses the projection

$$\boldsymbol{\lambda}_i = \boldsymbol{\Phi}\boldsymbol{z}_i + \boldsymbol{\nu}_i.$$

However instead of relying on covariances, this approach makes use of the following variance estimator:

$$\boldsymbol{S}(\alpha,\boldsymbol{\beta}) = \frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right)'\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right).$$

Evaluated at the true values of the parameters the expected value of $\boldsymbol{S}$ is

$$\mathrm{E}[\boldsymbol{S}(\alpha_0,\boldsymbol{\beta}_0)] = \boldsymbol{\Sigma} = \boldsymbol{I}_T\sigma^2 + \boldsymbol{F}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}\boldsymbol{F}'.$$

One can normalize $\boldsymbol{\Sigma}_{\boldsymbol{\nu}} = \boldsymbol{I}_L$ and redefine $\boldsymbol{F} := \boldsymbol{F}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}^{1/2}$ and $\boldsymbol{\Phi} := \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}^{-1/2}$. To evaluate the distance between $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$ Bai (2013b)[4] suggests maximising the following QML objective function to obtain consistent estimates of the underlying parameters:

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\left(\log|\boldsymbol{\Sigma}| + \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\right),$$

[4]Strictly speaking in the current paper the author solely describes the approach in terms of the likelihood function, while in Bai (2013a) the author describes a QML objective function as just one possibility.

where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}', \sigma^2, \operatorname{vec} \boldsymbol{F}', \operatorname{vec} \boldsymbol{\Phi}')'$. The theoretical and finite sample properties of this estimator without factors are discussed in Alvarez and Arellano (2003), Kruiniger (2013) and Norkutė (2014) among others.
The above version of the estimator requires time series homoskedasticity in $\varepsilon_{i,t}$ for consistency. If this condition holds true and all covariates are strictly exogenous, the estimator provides efficiency gains over the GMM estimators analyzed before since the latter do not make use of moment conditions that exploit homoskedasticity (see e.g. Ahn et al. (2001)). The estimator can be modified in a straightforward manner under time series heteroskedasticity to estimate all σ_t^2. On the other hand, cross-sectional heteroskedasticity cannot be allowed unfortunately.

Furthermore, the estimator generally requires $\tau = T$ in Assumption 4, i.e. strict exogeneity of the regressors. An exception to this is discussed in the following remark.

Remark 6. If one knows that all exogenous regressors have the following dynamic specification:

$$x_{i,t}^{(k)} = \beta_x x_{i,t-1}^{(k)} + \alpha_x y_{i,t-1} + \boldsymbol{f}_t' \boldsymbol{\lambda}_i^{x(k)} + \varepsilon_{i,t}^x, \tag{11}$$

so that all $x_{i,t}^{(k)}$ are possibly weakly exogenous and follow an autoregressive process of first order, then according to Bai (2013b) it is sufficient to project on $(1, x_{i,0}^{(1)}, \ldots, x_{i,0}^{(K)}, y_{i,0})$ only, resulting in a more efficient estimator. A necessary condition for this approach to be valid is that factor loadings $(\boldsymbol{\lambda}_i^{x(k)}, \boldsymbol{\lambda}_i)$ are independent, once conditioned on initial observations $(1, x_{i,0}^{(1)}, \ldots, x_{i,0}^{(K)}, y_{i,0})$.

3.9. ***Some general remarks on the estimators***

3.9.1. Unbalanced samples

As it is discussed in Juodis (2014), for the quasi-long-differencing transformation of Ahn et al. (2013) in the model with weakly exogenous regressors it is necessary that for *all* individuals the last L observations are available to the researcher. Otherwise the $\boldsymbol{D}(\boldsymbol{F}^*)$ transformation matrix becomes individual-specific (or group-specific if one can group observations based on availability). If the model contains only strictly exogenous regressors then it is sufficient that there exist L time indices $t^{(1)}, \ldots, t^{(L)}$ where observations for all individuals are available.

The extension of FIVU and FIVR to unbalanced samples follows trivially by simply introducing indicators, depending on whether a particular moment condition is available for individual i or not (as for the standard fixed effects estimator). Similarly, the quasi-differencing estimator of Nauges and Thomas (2003) can be trivially modified as in the standard Arellano and Bond (1991) procedure.

The projection estimator of Hayakawa (2012) requires further modifications in order to take into account that projection variables $\boldsymbol{z}_i$ are not fully observed for each individual. We conjecture that the modification could be performed in a similar way as in the model without a factor structure, as discussed by Abrevaya (2013). For maximum likelihood based estimators, such extendability appears to be a more challenging task.

Remark 7. The above discussion relies on that there exists a large enough number of consecutive time periods for each individual in the sample. For example, FIVU requires at least two consecutive periods and quasi-differencing type procedures require at least three. Under these circumstances, we note that estimators in their existing form may not be fully efficient. For example, if one observes *only* $y_{i,T}$ and $y_{i,T-2}$ for a substantial group of individuals, assuming exogenous covariates are available at all time periods, then one could in principle use backward substitution and consider moment conditions within the FIVU framework, which are quadratic in the autoregressive parameter and result in efficiency gains. For projection type methodologies, however, such substantial unbalancedness may affect the consistency of the estimators as one cannot substitute unobserved quantities for zeros in the projection term. This issue is discussed in detail by Abrevaya (2013).

3.9.2. Observed factors

In some situations of practical importance researchers might want to estimate models with both observed and unobserved factors at the same time. Taking the structure of observed factors into account may improve the efficiency of the estimators, although one can still consistently estimate the model by treating the observed factors as unobserved. One such possibility has been already discussed in Nauges and Thomas (2003) for models with an individual-specific, time-invariant effect. In this section we will briefly summarize implementability issues for all estimators when observed factors are present in the model alongside their unobserved counterparts.[5]

For the GMM estimators that involve some form of differencing, e.g. Holtz-Eakin et al. (1988) and Ahn et al. (2013), one can deal with observed factors using a similar procedure as in Nauges and Thomas (2003), that is, by removing the observed factors first (one-by-one) and then proceeding to remove the unobserved factors from the model. The first step can be most easily implemented using a quasi-differencing matrix $\boldsymbol{D}(\boldsymbol{r})$ with known weights. For the class of GMM estimators of Robertson and Sarafidis (2013) (FIVU) and Hayakawa (2012), since the unobserved factors are not removed from the model, the treatment of the observed factors is somewhat easier. One merely needs to split the $\boldsymbol{F}\boldsymbol{G}'$ terms into two parts, observed and unobserved factors, and then proceed as in the case of unobserved factors. In this case the number of identified parameters will be smaller than in the case where one treats the observed factors as unobserved. As a result, one gains in efficiency, at the expense, however, of robustness.

For FIVR one needs to take care when solving for $\boldsymbol{F}$ in terms of the remaining parameters, because in the model with observed factors one estimates the variance-covariance matrix of the factor loadings for the observed factors, while for those which are unobserved their variance-covariance matrix is normalized. The extension of the likelihood estimator of Bai (2013b) to observed factors can be implemented in a similar way to the projection GMM estimator. As in FIVR, one would have to estimate the variance-covariance matrix of the factor loadings for the observed factors, while the covariances of unobserved factors can be w.l.o.g. normalized as before.

4. Finite Sample Performance

This section investigates the finite sample performance of the estimators analyzed above using simulated data. Our focus lies on examining the effect of the presence of weakly exogenous covariates, the effect of changing the magnitude of the correlation between the factor loadings of the dependent variable and those of the covariates, as well as the impact of changing the number of moment conditions on bias and size for GMM estimators. We also investigate the effect of changing the level of persistence in the data, as well as the sample size in terms of both N and T.

[5] We assume that certain regularity conditions hold, which prohibit perfect collinearity between the observed and unobserved factors.

4.1. MC Design

We consider model (1) with $K = 1$, i.e.

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + u_{i,t}; \quad u_{i,t} = \sum_{\ell=1}^{L} \lambda_{\ell,i} f_{\ell,t} + \varepsilon_{i,t}^{y}.$$

The process for $x_{i,t}$ and for f_t is given, respectively, by

$$x_{i,t} = \delta y_{i,t-1} + \alpha_x x_{i,t-1} + \sum_{\ell=1}^{L} \gamma_{\ell,i} f_{\ell,t} + \varepsilon_{i,t}^{x};$$
$$f_{\ell,t} = \alpha_f f_{\ell,t-1} + \sqrt{1-\alpha_f^2}\varepsilon_{\ell,t}^{f}; \quad \varepsilon_{\ell,t}^{f} \sim \mathcal{N}(0,1), \forall \ell.$$

The factor loadings are generated by $\lambda_{\ell,i} \sim \mathcal{N}(0,1)$ and

$$\gamma_{\ell,i} = \rho \lambda_{\ell,i} + \sqrt{1-\rho^2}\upsilon_{\ell,i}^{f}; \quad \upsilon_{\ell,i}^{f} \sim \mathcal{N}(0,1) \text{ for all } \ell,$$

where ρ denotes the correlation between the factor loadings of the y and x processes. Furthermore, the idiosyncratic errors are drawn as

$$\varepsilon_{i,t}^{y} \sim \mathcal{N}(0,1); \quad \varepsilon_{i,t}^{x} \sim \mathcal{N}\left(0, \sigma_x^2\right).$$

The starting period for the model is $t = -S$ and the initial observations are generated as

$$y_{i,-S} = \sum_{\ell=1}^{L} \lambda_{\ell,i} f_{\ell,-S} + \varepsilon_{i,-S}^{y}; \quad x_{i,-S} = \sum_{\ell=1}^{L} \gamma_{\ell,i} f_{\ell,-S} + \varepsilon_{i,-S}^{x};$$
$$f_{-S} \sim \mathcal{N}(0,1).$$

The signal-to-noise ratio of the model is defined as follows:

$$SNR \equiv \frac{1}{T}\sum_{t=1}^{T} \frac{\operatorname{var}\left(y_{i,t} | \lambda_{\ell,i}, \gamma_{\ell,i}, \{f_{\ell,s}\}_{s=-S}^{t}\right)}{\operatorname{var} \varepsilon_{i,t}^{y}} - 1.$$

σ_x^2 is set such that the signal-to-noise ratio is equal to $SNR = 5$ in all designs.[6] This particular value of SNR is chosen so that it is possible to control this measure across all designs. Lower values of SNR (e.g. 3 as in Bun and Kiviet (2006)) would require $\sigma_x^2 < 0$ ceteris paribus in order to satisfy the desired equality for all designs.

We set $\beta = 1 - \alpha$ such that the long run parameter is equal to 1, $\alpha_x = 0.6$, $\alpha_f = 0.5$ and $L = 1$.[7] We consider $N = \{200; 800\}$ and $T = \{4; 8\}$. Furthermore, $\alpha = \{0.4; 0.8\}$, $\rho = \{0; 0.6\}$

[6] To ensure this, we also set $S = 5$.

[7] Similar results have been obtained for $L = 2$. To avoid repeating similar conclusions we refrain from reporting these results. We note that the number of factors can be estimated for all GMM estimators based on the model information criteria developed by Ahn et al. (2013). The performance of these procedures appears to be more than satisfactory; the interested reader may refer to the aforementioned paper, as well as to the Monte Carlo study in Robertson, Sarafidis, and Westerlund (2014). The size of L is treated as known in this paper because there is currently no equivalent methodology proposed for testing the number of factors within the likelihood framework.

and $\delta = \{0; 0.3\}$. The minimum number of replications performed equals $2,000$ for each design and the factors are drawn in each replication. The choice of the initial values of the parameters for the nonlinear algorithms is discussed in Appendix A.1. When at least one of the estimators fails to converge in a particular replication, that replication is discarded.[8]

Note that for the likelihood methods we use standard errors based on a "sandwich" variance-covariance matrix, as opposed to the simple inverse of the Hessian variance matrix. First order conditions as well as Hessian matrices for likelihood estimators are obtained using analytical derivatives to speed-up the computations.[9]

Although feasible, in this paper we do not implement the linearized GMM estimator of Hayakawa (2012) adapted to weakly exogenous regressors. This is mainly due to the fact that this estimator merely provides an easy way to obtain reasonable starting values for the remaining estimators, which involve non-linear optimization algorithms. Motivated from our theoretical discussion regarding the estimators considered in this paper, some implications can be discussed a priori, based on our Monte Carlo design.

1. When $\delta \neq 0$, likelihood based estimators are inconsistent, with the exception of the modified estimator of Bai (2013b) conditional on $(y_{i,0}, x_{i,0})$.
2. For $\rho \neq 0$ the projection likelihood estimator conditional on $(y_{i,0}, x_{i,0})$ is inconsistent because the conditional independence assumption is violated.
3. For $\alpha = 0.8, \rho = 0, \delta = 0$ the projection GMM estimator might suffer from weak instruments because $y_{i,0}$ remains the only relevant instrument.

4.2. MC Results

The results are reported in the Appendix in terms of median bias and root median square error. The latter is defined as

$$RMSE = \sqrt{\text{med}\left[(\widehat{\alpha}_r - \alpha)^2\right]},$$

where $\widehat{\alpha}_r$ denotes the value of α obtained in the r^{th} replication using a particular estimator (and similarly for β). As an additional measure of dispersion we report the radius of the interval centered on the median containing 80% of the observations, divided by *1.28*. This statistic, which we shall refer to as 'quasi-standard deviation' (denoted qStd) provides an estimate of the population standard deviation if the distribution were normal, with the advantage that it is more robust to the occurrence of outliers compared to the usual expression for the standard deviation. The reason we report this statistic is that, on the one hand, the root mean square error is extremely sensitive to outliers, and on the other hand it is fair to say that the root median square error does not depend on outliers pretty much at all. Therefore, the former could be unduly misleading given that in principle, for

[8]For the numerical maximization we used the BFGS method as implemented in the OxMetrics statistical software. Convergence is achieved when the difference in the value of the given objective function between two consecutive iterations is less than 10^{-4}. Other values of this criterion were considered in the preliminary study with similar qualitative conclusions, although the number of times particular estimators fail to converge varies. For further details on OxMetrics see Doornik (2009).

[9]In the preliminary study, results based on analytical and numerical derivatives were compared. Since the results were quantitatively and qualitatively almost identical (for designs where estimators were consistent), we prefer the use of analytical derivatives solely for practical reasons.

any given data set, one could estimate the model using a large set of different initial values in an attempt to avoid local minima, or lack of convergence in some cases (which we deal with in our experiments by discarding those particular replications). In a large-scale simulation experiment as ours, however, the set of initial values naturally needs to be restricted in some sensible/feasible way. The quasi-standard deviation lies in-between in that, while it provides a measure of dispersion that is less sensitive to outliers compared to the root mean square error, it is still more informative about the variability of the estimators relative to the root median square error. Finally, we report size, where nominal size is set at 5%. For the GMM estimators we also report size of the overidentifying restrictions (J) test statistic.

Initially we discuss results for the OLS estimator, the GMM estimator proposed by Sarafidis, Yamagata, and Robertson (2009) and the linearized GMM estimator of Hayakawa (2012); these estimators have been used to obtain initial values for the parameters for the non-linear estimators, among other (random) choices. As we can see in Table A.1, in many circumstances the OLS estimator exhibits large median bias, while the size of the estimator is most often not far from unity. On the other hand, the linear GMM estimator proposed by Sarafidis, Yamagata, and Robertson (2009) does fairly well both in terms of bias and RMSE when $\delta = 0$ and $\rho = 0$, i.e. when the covariate is strictly exogenous with respect to the total error term, $u_{i,t}$. The size of the estimator appears to be somewhat upwardly distorted, especially for T large, but one expects that this would substantially improve if one made use of the finite-sample correction proposed by Windmeijer (2005). On the other hand, the estimator is not consistent for the remaining parameterisations of our design and this is well reflected in its finite sample performance. Notably, the J statistic appears to have high power to detect violations of the null, even if N is small.

With regards to the linearized GMM estimator of Hayakawa (2012), both median bias and RMSE are reasonably small, even for $N = 200$, so long as $\delta = 0$, i.e. under strict exogeneity of x with respect to the idiosyncratic error. However, the estimator appears to be quite sensitive to high values of α, especially in terms of qStd, an outcome that may be partially related to the fact that the value of β is small in this case, which implies that a many-weak instruments type problem might arise. Naturally, the performance of the estimator deteriorates for $\delta = 0.3$ as the moment conditions are invalidated in this case. While the size of the J statistic appears to be distorted upwards when the estimator is consistent, it has in general quite large power to detect violations of strict exogeneity, and for high values of α this holds true even with a relatively small size of N.

Tables A.3 and A.4 report results for the quasi-long-differenced GMM estimator proposed by Ahn, Lee, and Schmidt (2013). The only difference between the two tables is that A.3 is based on the "pseudo-full" set of moment conditions, i.e. $T(T-1)$, obtained by always treating x as weakly exogenous, while A.4 is based on the 4 most recent lags of the variables. In the latter case the number of instruments is of order $\mathcal{O}(T)$. This strategy is possible to implement only for $T = 8$, as for $T = 4$ there are not enough degrees of freedom to identify the model when truncating the moment conditions to such extent.[10] The estimator appears to have small median bias under all designs. This is expected given that the estimator is consistent. The qStd results indicate that the estimator has large dispersion in some designs, especially when T is small. We have explored further the underlying reason for this result. We found that this is often the case when the value of the

[10]To be more precise, the total number of moment conditions for the subset estimator is $q(2(T-1)+1-q)$, where in our case $q = 4$.

factor at the last time period, i.e. f_T, is close to zero. Thus, the estimator appears to be potentially sensitive to this issue, because the normalization scheme sets $f_T = 1$.[11] The two-step version improves on these results. On the other hand, inferences based on one-step estimates seem to be relatively more reliable. This outcome may be attributed to the standard argument provided for linear GMM estimators, which is that two-step estimators rely on an estimate of the variance-covariance matrix of the moment conditions, which, in samples where N is small, can lead to conservative standard errors. Notice here that a Windmeijer (2005) type correction is not trivial here because the proposed expression applies to linear estimators only. Truncating the moment conditions for $T = 8$ seems to have a negligible effect on the size properties of the one-step estimator but does improve size for the two-step estimator quite substantially. This result seems to apply for all overidentified GMM estimators actually. The J statistic exhibits small size distortions upwards.

Tables A.5 -A.8 report results for FIVU and FIVR based on either the full or the truncated sets of moment conditions, proposed by Robertson and Sarafidis (2013). Similarly to Ahn et al. (2013), both estimators have very small median bias in all circumstances. Furthermore, they perform well in terms of qStd. Especially the two-step versions have small dispersion regardless of the design. Naturally, the dispersion decreases further with high values of T because the degree of overidentification of the model increases. As expected, RMSE appears to go down roughly at the rate of $\sqrt{N}$. FIVR dominates FIVU, which is not surprising given that the former imposes overidentifying restrictions arising from the structure of the model and thus it estimates a smaller number of parameters. The size of one-step FIVU and FIVR estimators is close to its nominal value in all circumstances. On the other hand, the two-step versions appear to be size distorted when T is large, although the distortion decreases when only a subset of the moment conditions is used. Thus, one may conclude that using the full set of moment conditions and relying on inferences based on first-step estimates is a sensible strategy. From the empirical point of view this is appealing because it simplifies matters regarding how many instruments to be used $-$ an important question that often arises in two-way error components models estimated using linear GMM estimators. Finally, the size of the J statistic is often slightly distorted when N is small, but improves rapidly as N increases.

The projection GMM estimator proposed by Hayakawa (2012) has small bias and performs well in general in terms of qStd unless α is close to unity, in which case outliers seem to occur relatively more frequently. One could suspect that this design is the worst case scenario for the estimator because only $y_{i,0}$ is included in the set of instruments, while lagged values of $x_{i,t}$ are only weakly correlated with $y_{i,t-1}$. Inferences based on the first-step estimator are reasonably accurate, certainly more so compared to the two-step version, although the latter improves for the truncated set of moment conditions. The J statistic seems to be size-distorted downwards but it slowly improves for larger values of N.

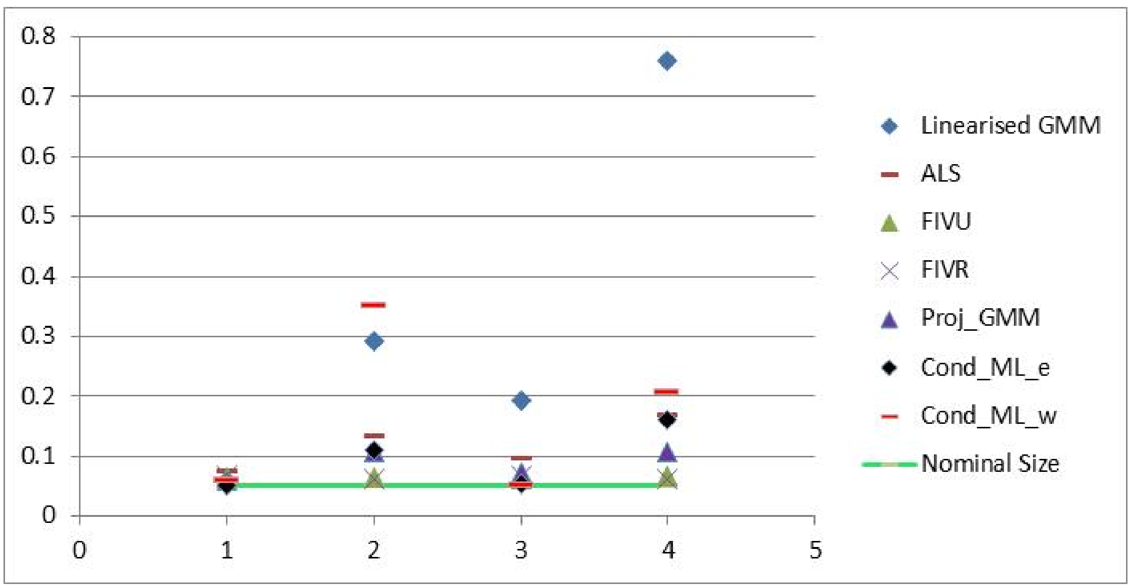
Finally, Table A.11 reports results for the conditional maximum likelihood estimator proposed by Bai (2013b). The left panel corresponds to the estimator that treats x as strictly exogenous with respect to the idiosyncratic error, while the panel on the right-hand side corresponds to the estimator that is consistent under weak exogeneity of a first-order form[12], which is satisfied in our design, assuming that $\rho = 0$. Interestingly, the former appears to exhibit negligible median bias in

[11] Notice that imposing a different normalization, e.g. $f_{T-1} = 1$ would result in losing T moment conditions, as explained in the main text.

[12] That is, when x follows an AR(1) process.

all cases, even when both δ and ρ take non-zero values. The dispersion of the estimator is small as well, unless $T = 4$ and $\delta = 0.3$. Likewise the size of the estimator is distorted upwards when $\delta = 0.3$ and gets worse with higher values of N, which is natural given that the estimator is not consistent in this case. However, for cases where this estimator is consistent ($\delta = 0$ and $\rho = 0$), it may serve as a benchmark because it has negligible bias and excellent size. This can be expected given the asymptotic optimality of this estimator. The conclusion is pretty much invariant to different values of N, T or ρ. The second estimator, in designs with $\rho = 0.6$ where it is not consistent, tends to have substantial bias for both α and β. On the other hand, when it is supposed to be consistent ($\delta = 0.3, \rho = 0.0$) it is more size distorted than the first estimator that is inconsistent. This is a somewhat puzzling finding.

The following picture provides a snapshot illustration of our discussion regarding the size properties of the estimators. The numbers $1, ..., 4$ on the horizontal axis correspond to the designs where ($\delta = 0.0; \rho = 0.0$) and ($\delta = 0.3; \rho = 0.6$) respectively when $\alpha = 0.4$, followed by the same values of δ, ρ for $\alpha = 0.8$.



5. Conclusion

In this paper we have provided a synopsis for a growing group of fixed T dynamic panel data estimators with a multi-factor error structure. All currently available estimators have been presented using a unified notational approach. Both their theoretical properties as well as possible limitations are discussed. We have considered a model with a lag dependent variable and additional regressors, possibly weakly exogenous or endogenous. We found that the number of identifiable parameters for the GMM estimators can be smaller than what can be found in the literature. This result is of major importance for practitioners when performing model selection based on overidentifying test statistics. Theoretical discussions in this paper were complemented by a finite sample study based on Monte Carlo simulation. We designed our Monte Carlo exercise to shed some light on the relative merits of the various estimation approaches. It was found that the likelihood estimator of

Bai (2013b), when consistent, can serve as a benchmark in that it has negligible bias and good size control, irrespective of the sample size. Under such circumstances, the FIVR estimator proposed by Robertson and Sarafidis (2013) performs closely as well. However, FIVR is more robust to violations from strict exogeneity, as well as from no conditional correlation between the factor loadings. The latter applies to other GMM estimators as well, at least provided that the cross-sectional dimension is large enough.

This paper assumes that the time-series dimension is fixed. A natural question to ask is whether GMM estimators in models where the number of parameters grows with T suffer from an incidental parameters problem. Based on the large T proof in Bai (2013b), where it is shown that the presence of factors does not result in an incidental parameters problem for the conditional maximum likelihood estimator as far as the structural parameters are concerned, one may suspect that a similar result is also valid for the GMM estimators. We leave a proof of this assertion for future.

References

Abadir, K. M. and J. R. Magnus (2002): "Notation in Econometrics: A Proposal for a Standard," *Econometrics Journal*, 5, 76–90.

Abrevaya, J. (2013): "The Projection Approach for Unbalanced Panel Data," *The Econometrics Journal*, 16, 161–178.

Ahn, S. C., Y. H. Lee, and P. Schmidt (2001): "GMM estimation of linear panel data models with time-varying individual effects," *Journal of Econometrics*, 101, 219–255.

——— (2013): "Panel data models with multiple time-varying individual effects," *Journal of Econometrics*, 174, 1–14.

Alvarez, J. and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71(4), 1121–1159.

Anderson, T. W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Bai, J. (2013a): "Fixed-Effects Dynamic Panel Models, a Factor Analytical Method," *Econometrica*, 81, 285–314.

——— (2013b): "Likelihood approach to dynamic panel models with interactive effects," Working Paper.

Bond, S. and F. Windmeijer (2002): "Projection Estimators for Autoregressive Panel Data Models," *The Econometrics Journal*, 5, 457–479.

Bun, M. J. G. and J. F. Kiviet (2006): "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, 132, 409–444.

Chamberlain, G. (1982): "Multivariate regression models for panel data," *Journal of Econometrics*, 18, 5–46.

Doornik, J. (2009): *An Object-Oriented Matrix Language Ox 6*, London: Timberlake Consultants Press.

Hayakawa, K. (2012): "GMM Estimation of Short Dynamic Panel Data Model with Interactive Fixed Effects," *Journal of the Japan Statistical Society*, 42, 109–123.

Holtz-Eakin, D., W. K. Newey, and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–1395.

Juodis, A. (2014): "Linear Pseudo Panel Data Models with Multi Factor Error Structure," Working Paper.

Kruiniger, H. (2013): "Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions," *Journal of Econometrics*, 173, 175–188.

Magnus, J. R. and H. Neudecker (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons.

Mundlak, Y. (1978): "On The Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85.

Nauges, C. and A. Thomas (2003): "Consistent estimation of dynamic panel data models with time-varying individual effects," *Annales d'Economie et de Statistique*, 70, 54–75.

Norkutė, M. (2014): "A Monte Carlo study of a factor analytical method for fixed-effects dynamic panel models," *Economics Letters*, 123, 348–351.

Robertson, D. and V. Sarafidis (2013): "IV Estimation of Panels with Factor Residuals," Working Paper.

Robertson, D., V. Sarafidis, and J. Westerlund (2014): "GMM Unit Root Inference in Generally Trending and Cross-Correlated Dynamic Panels," Working Paper.

Sarafidis, V. and D. Robertson (2009): "On the Impact of Error Cross-Sectional Dependence in Short Dynamic Panel Estimation," *Econometrics Journal*, 12, 62–81.

Sarafidis, V. and T. Wansbeek (2012): "Cross-Sectional Dependence in Panel Data Analysis," *Econometric Reviews*, 31, 483–531.

Sarafidis, V., T. Yamagata, and D. Robertson (2009): "A test of cross section dependence for a linear dynamic panel model with regressors," *Journal of Econometrics*, 148, 149–161.

Windmeijer, F. (2005): "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics*, 126, 25–51.

Appendices

Appendix A. Implementation

Appendix A.1. Starting values for non-linear estimators

This appendix discusses the choice of starting values used for the non-linear optimization algorithms.

Ahn et al. (2013). Under conditional homoskedasticity in $\varepsilon_{i,t}$, this estimator can be implemented through an iterative procedure. Iterations start given some set of initial values for the structural parameters, α, β. For this purpose, we use both the one- and two-step linearized GMM estimator as proposed by Hayakawa (2012), as well as the OLS estimator. The two-step estimator is implemented in exactly the same way except that the set of initial values for the structural parameters includes the one-step estimator. Once final estimates of $\hat{\alpha}, \hat{\beta}$ and $\hat{\boldsymbol{F}}$ are obtained, these are used as initial values in the non-linear optimization algorithm, which optimises all parameters at once. This is implemented in order to make sure that we indeed find the global minimum of the objective function.

FIVU. Similarly to the previous estimator, FIVU can also be implemented in steps. Iterations start given a set of starting values for the factors $\boldsymbol{F}$. This set is obtained using the linearized GMM estimator, estimates of the principal components extracted from OLS residuals, and one set of uniform random variables on $[-1; 1]$. Unlike for Ahn et al. (2013), joint non-linear optimization is not used as a final step in order to save computational time.

FIVR. For this estimator the main source of starting values is obtained from FIVU with the starting value of $\boldsymbol{g}_T$ implied in terms of other parameters. Other starting values include those based on the OLS estimator and the one- and two-step linearized GMM estimator. In this case starting values for the nuisance parameters $\boldsymbol{G}$ are simply drawn from uniform $[-1; 1]$.

Projection GMM. This estimator is implemented in exactly the same way as Ahn et al. (2013), i.e. firstly an iterative procedure is used, followed by a non-linear one. Starting values for the factors are obtained using the principal components extracted from OLS residuals, the estimate of $\boldsymbol{f}$ obtained from the linearized GMM estimator, and two sets of uniform random variables on $[-1; 1]$. In order to uniquely identify all parameters up to rotation, we impose $f_T = 1$ in estimation. We suspect that in principle, similarly to FIVU, one can estimate the model without normalizations and perform a degrees of freedom correction at the end. We leave this question open for future research.

Projection MLE. Starting values for the structural parameters are obtained using the linearized GMM estimator, OLS, and two sets of uniform random variables on $[-1; 1]$. The remaining parameters (including $\log(\sigma^2)$) are drawn as uniform random variables on $[0; 1]$. In the preliminary study we also tried $[-1; 1]$, however the results were identical. Alternatively, one could also use the principal component estimates of $\boldsymbol{F}$ obtained from OLS residuals, as suggested by Bai (2013b).

Subset GMM estimators. For $T = 8$ when both the subset and full-set GMM estimators are available, we estimate the subset estimators first using the algorithms as described above and then use the subset estimator as starting values for the estimators that make use of the full set of moment conditions.

Appendix A.2. Specifics

Appendix A.2.1. Ahn, Lee, and Schmidt (2013)

To describe the procedure assume for simplicity that there no $x's$, such that the only available moment conditions are

$$\boldsymbol{m}_l = \frac{1}{N}\,\mathrm{vech}\left(\boldsymbol{J}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1}\right)'\boldsymbol{Y}_{-1}\boldsymbol{J}' + \boldsymbol{F}^*\boldsymbol{J}_L\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1}\right)'\boldsymbol{Y}_{-1}\boldsymbol{J}'\right).$$

The objective function for this estimator is simply given by

$$f(\alpha, \mathrm{vec}\,(\boldsymbol{F}^*)) = \boldsymbol{m}_l'\boldsymbol{W}_N\boldsymbol{m}_l.$$

For any given value of α, the moment conditions are linear $\mathrm{vec}\,(\boldsymbol{F}^*)$. That is,

$$\boldsymbol{m}_l = \mathrm{vech}(\boldsymbol{Z}) + \boldsymbol{B}_{(T-L)}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{T-L})\,\mathrm{vec}\,(\boldsymbol{F}^*) = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}.$$

Here $\boldsymbol{Z}$ and $\boldsymbol{Q}$ are given by

$$\begin{aligned}
\boldsymbol{Z} &= \frac{1}{N}\boldsymbol{J}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1}\right)'\boldsymbol{Y}_{-1}\boldsymbol{J}';\\
\boldsymbol{Q} &= \frac{1}{N}\boldsymbol{J}_L\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1}\right)'\boldsymbol{Y}_{-1}\boldsymbol{J}';\\
\boldsymbol{y} &= \mathrm{vech}(\boldsymbol{Z});\\
\boldsymbol{X} &= \boldsymbol{B}_{(T-L)}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{T-L});\\
\boldsymbol{\beta} &= -\,\mathrm{vec}\,(\boldsymbol{F}^*).
\end{aligned}$$

Hence the usual formula for the OLS estimator implies that

$$-\,\mathrm{vec}\,(\boldsymbol{F}^*) = \boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{W}_N\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}_N\boldsymbol{y}.$$

If, on the other hand, $\boldsymbol{F}^*$ is known then α is obtained in exactly the same way with $\boldsymbol{\beta} = \alpha$, while

$$\begin{aligned}
\boldsymbol{y} &= \frac{1}{N}\,\mathrm{vech}(\boldsymbol{D}(\boldsymbol{\Phi}^*)\boldsymbol{Y}'\boldsymbol{Y}_{-1}\boldsymbol{J}');\\
\boldsymbol{X} &= \frac{1}{N}\,\mathrm{vech}(\boldsymbol{D}(\boldsymbol{\Phi}^*)\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\boldsymbol{J}').
\end{aligned}$$

Appendix A.2.2. Restricted estimator of Robertson and Sarafidis (2013)

The moment conditions are given by

$$\begin{aligned}
\boldsymbol{m}_l &= \mathrm{vech}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{Y}_{-1} - \boldsymbol{F}\boldsymbol{G}'\right);\\
\boldsymbol{m}_k &= \mathrm{vech}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{X}_k - \boldsymbol{F}\boldsymbol{G}_k'\right) \quad \forall k.
\end{aligned}$$

$\boldsymbol{F}$ obeys the following restriction:

$$\boldsymbol{F} = (\boldsymbol{L}_T' - \alpha \boldsymbol{I}_T)\, \boldsymbol{G} + \boldsymbol{e}_T \boldsymbol{g}_T' - \sum_{k=1}^{k} \beta_k \boldsymbol{G}_k.$$

The differential of $\operatorname{vec} \boldsymbol{F}$ is simply given by

$$\begin{aligned} \mathrm{dvec}\, \boldsymbol{F} = &- \operatorname{vec}(\boldsymbol{G})\, \mathrm{d}\alpha + (\boldsymbol{I}_L \otimes (\boldsymbol{L}_T' - \alpha \boldsymbol{I}_T))\, \mathrm{dvec}\, \boldsymbol{G} \\ &- \sum_{k=1}^{K} \operatorname{vec}(\boldsymbol{G}_k)\, \mathrm{d}\beta_k - (\boldsymbol{I}_L \otimes \boldsymbol{I}_T) \sum_{k=1}^{K} \beta_k\, \mathrm{dvec}\, \boldsymbol{G}_k \\ &+ (\boldsymbol{I}_L \otimes \boldsymbol{e}_T)\, \mathrm{d}\boldsymbol{g}_T. \end{aligned}$$

By the chain rule for differentials we have

$$\begin{aligned} \mathrm{d}\boldsymbol{m}_l = &-\frac{1}{N} \operatorname{vech}\left(\boldsymbol{Y}_{-1}' \boldsymbol{Y}_{-1}\right) \mathrm{d}\alpha - \sum_{k=1}^{K} \frac{1}{N} \operatorname{vech}(\boldsymbol{X}_k' \boldsymbol{Y}_{-1})\, \mathrm{d}\beta_k \\ &- \boldsymbol{B}_T \left(\mathbf{K}_{T,T}(\boldsymbol{F} \otimes \boldsymbol{I}_T)\, \mathrm{d}(\operatorname{vec} \boldsymbol{G}) + (\boldsymbol{G} \otimes \boldsymbol{I}_T)\, \mathrm{d}(\operatorname{vec} \boldsymbol{F})\right). \end{aligned}$$

The result for $\mathrm{d}\boldsymbol{m}_k$ follows analogously.

Appendix A.2.3. Bai (2013b)

Some specific results for this estimator can be written as follows:

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_\tau + \boldsymbol{F}\boldsymbol{F}'; \\ \boldsymbol{\Sigma}_\tau &= \sigma^2 \boldsymbol{I}_T; \\ \boldsymbol{v}_i &= \boldsymbol{y}_i - \boldsymbol{W}_i \boldsymbol{\gamma} - \boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{z}_i. \end{aligned}$$

The corresponding differentials are

$$\begin{aligned} \mathrm{d}\boldsymbol{\Sigma} &= \boldsymbol{I}_T\, \mathrm{d}\sigma^2 + \boldsymbol{F}(\mathrm{d}\boldsymbol{F})' + (\mathrm{d}\boldsymbol{F})\boldsymbol{F}'; \\ \mathrm{d}^2\boldsymbol{\Sigma} &= 2(\mathrm{d}\boldsymbol{F}\, \mathrm{d}\boldsymbol{F}'); \\ \mathrm{d}\boldsymbol{v}_i &= -\boldsymbol{W}_i(\mathrm{d}\boldsymbol{\gamma}) - \mathrm{d}(\boldsymbol{F})\boldsymbol{\Phi}\boldsymbol{z}_i - \boldsymbol{F}\, \mathrm{d}(\boldsymbol{\Phi})\boldsymbol{z}_i; \\ \mathrm{d}^2\boldsymbol{v}_i &= -2(\mathrm{d}(\boldsymbol{F})\, \mathrm{d}(\boldsymbol{\Phi})\boldsymbol{z}_i). \end{aligned}$$

Denoting as $\boldsymbol{V}(\boldsymbol{\theta})$ the following $[N \times T]$ matrix (with the i'th row being simply $\boldsymbol{v}_i'$)

$$\boldsymbol{V}(\boldsymbol{\theta}) = \frac{1}{N}\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right),$$

then the score vector, using matrix notation rather than sums, is simply given by

$$\nabla(\boldsymbol{\theta}) = \begin{pmatrix} \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{Y}_{-1}) \\ \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{X}_1) \\ \vdots \\ \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{X}_K) \\ -0.5 \operatorname{tr}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) \\ -\operatorname{vec}((\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1})\, \boldsymbol{F}) + \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{Z}\boldsymbol{\Phi}') \\ \operatorname{vec}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{Z}) \end{pmatrix}.$$

Appendix A.2.4. Hessians of likelihood based-estimators

Observe that the general structure of the likelihood function is given by

$$-\frac{2}{N}\ell(\theta) = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \operatorname{tr}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{S}(\boldsymbol{\theta})\right).$$

Using the rules for differentials (see e.g. Magnus and Neudecker (2007)) the first differential of the two components is given by

$$\begin{aligned}
\mathrm{d}\log|\boldsymbol{\Sigma}| &= \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\right);\\
\mathrm{d}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) &= -\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{S})\right),
\end{aligned}$$

where for simplicity the dependence on $\boldsymbol{\theta}$ has been dropped. By the chain rule for differentials it follows similarly that the second differential for the log-determinant is of the following form:

$$\mathrm{d}^2\log|\boldsymbol{\Sigma}| = \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{\Sigma})\right) - \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\right),$$

while the trace component is given by

$$\begin{aligned}
\mathrm{d}^2\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) &= 2\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\\
&\quad - 2\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{S})\right)\\
&\quad - \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\\
&\quad + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{S})\right).
\end{aligned}$$

We can combine both terms such that

$$\begin{aligned}
-\frac{2}{N}\,\mathrm{d}^2\ell(\theta) &= \operatorname{tr}\left(\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)\mathrm{d}^2\boldsymbol{\Sigma}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{S})\right)\\
&\quad + \operatorname{tr}\left(\left(2\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\right)(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\right)\\
&\quad - 2\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{S})\right).
\end{aligned}$$

Note that, evaluated at any consistent estimate of $\hat{\boldsymbol{\theta}}$, we have

$$\begin{aligned}
\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1} &= o_p(1);\\
2\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Sigma}^{-1} + o_p(1).
\end{aligned}$$

Hence from the asymptotic point of view this is equivalent to considering the following consistent estimate of the Hessian:

$$-\frac{2}{N}\,\mathrm{d}^2\ell(\theta) = \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{S})\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\right) - 2\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{S})\right).$$

In our Monte Carlo study we will make use of these facts and ignore the $o_p(1)$ terms. Now let us consider the differentials of $\boldsymbol{S}$ in more detail. We have

$$\begin{aligned}
\mathrm{d}\boldsymbol{S} &= \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{v}_i\,\mathrm{d}(\boldsymbol{v}_i)' + \mathrm{d}(\boldsymbol{v}_i)\boldsymbol{v}_i'\right);\\
\mathrm{d}^2\boldsymbol{S} &= \frac{1}{N}\sum_{i=1}^{N}\left(2\,\mathrm{d}(\boldsymbol{v}_i)\,\mathrm{d}(\boldsymbol{v}_i)' + \mathrm{d}^2(\boldsymbol{v}_i)\boldsymbol{v}_i' + \boldsymbol{v}_i\,\mathrm{d}^2(\boldsymbol{v}_i)'\right).
\end{aligned}$$

Note that if evaluated at any consistent estimator of $\hat{\boldsymbol{\theta}}$

$$\frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{d}^2(\boldsymbol{v}_i)\boldsymbol{v}_i' + \boldsymbol{v}_i\,\mathrm{d}^2(\boldsymbol{v}_i)'\right) = o_p(1).$$

However, in our Monte Carlo study we retain the corresponding terms in the formula of the estimate for the Hessian matrix. Furthermore, note that

$$\operatorname{vec}\mathrm{d}\boldsymbol{S} = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{v}_i \otimes \boldsymbol{I}_T + \boldsymbol{I}_T \otimes \boldsymbol{v}_i\right)\mathrm{d}(\boldsymbol{v}_i).$$

Appendix A.3. Tables

Table A.1: OLS estimator and System GMM estimator by Sarafidis, Yamagata, and Robertson (2009)

Designs					OLS								Sub-System								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	.022	.048	.135	.609	-.008	.025	.069	.247	-.002	.029	.089	.060	-.002	.021	.065	.060	.041
200	4	.4	.0	.3	.005	.051	.146	.438	-.048	.062	.146	.485	-.080	.094	.228	.351	.037	.069	.204	.310	.707
200	4	.4	.6	.0	-.035	.051	.139	.633	.088	.088	.092	.851	-.035	.056	.152	.405	.086	.087	.130	.638	.720
200	4	.4	.6	.3	-.170	.170	.162	.921	.141	.141	.162	.817	-.320	.320	.237	.907	.289	.289	.320	.878	.866
200	4	.8	.0	.0	-.048	.050	.097	.662	.009	.013	.035	.139	-.038	.091	.299	.105	-.012	.032	.108	.082	.044
200	4	.8	.0	.3	-.066	.066	.102	.647	-.031	.045	.114	.412	-.301	.305	.684	.649	-.007	.096	.299	.413	.823
200	4	.8	.6	.0	-.083	.083	.102	.835	.064	.064	.059	.893	-.113	.147	.397	.488	.029	.054	.158	.360	.587
200	4	.8	.6	.3	-.181	.181	.137	.964	.109	.109	.131	.799	-.445	.445	.403	.907	.246	.246	.319	.808	.818
200	8	.4	.0	.0	.037	.045	.110	.691	-.018	.024	.061	.355	-.003	.014	.044	.148	-.001	.012	.036	.135	.029
200	8	.4	.0	.3	.032	.051	.129	.519	-.060	.065	.118	.567	-.122	.122	.160	.772	.090	.095	.165	.653	.901
200	8	.4	.6	.0	-.013	.041	.116	.667	.077	.077	.067	.934	-.045	.047	.089	.669	.087	.087	.079	.933	.768
200	8	.4	.6	.3	-.149	.149	.122	.971	.154	.154	.126	.952	-.362	.362	.148	1	.393	.393	.207	.999	.988
200	8	.8	.0	.0	-.016	.031	.084	.641	.003	.010	.029	.103	-.033	.041	.115	.248	-.007	.013	.039	.179	.039
200	8	.8	.0	.3	-.023	.036	.101	.444	-.059	.063	.111	.564	-.404	.404	.465	.960	.095	.139	.396	.692	.990
200	8	.8	.6	.0	-.045	.046	.082	.760	.062	.062	.040	.980	-.097	.099	.204	.766	.038	.040	.073	.653	.680
200	8	.8	.6	.3	-.177	.177	.108	.999	.165	.165	.135	.952	-.570	.570	.211	1	.513	.513	.299	1	.972
800	4	.4	.0	.0	.031	.079	.221	.846	-.012	.031	.089	.437	-.001	.018	.056	.053	.000	.016	.049	.048	.053
800	4	.4	.0	.3	.004	.054	.152	.714	-.057	.069	.155	.719	-.075	.088	.193	.565	.050	.071	.190	.544	.949
800	4	.4	.6	.0	-.064	.085	.202	.867	.217	.217	.166	.987	-.122	.127	.177	.857	.267	.267	.171	.957	.986
800	4	.4	.6	.3	-.181	.181	.168	.970	.154	.154	.182	.928	-.364	.364	.212	.960	.366	.366	.325	.968	.985
800	4	.8	.0	.0	-.069	.071	.137	.858	.005	.013	.038	.086	-.014	.045	.148	.069	-.002	.017	.057	.053	.048
800	4	.8	.0	.3	-.061	.061	.106	.805	-.048	.057	.136	.703	-.295	.305	.630	.807	.015	.104	.323	.645	.978
800	4	.8	.6	.0	-.110	.110	.134	.929	.208	.208	.135	.989	-.209	.220	.359	.878	.232	.233	.183	.933	.979
800	4	.8	.6	.3	-.199	.199	.148	.993	.136	.136	.171	.919	-.515	.515	.305	.965	.399	.399	.331	.967	.971
800	8	.4	.0	.0	.063	.074	.162	.876	-.029	.034	.086	.549	-.001	.010	.030	.074	.000	.010	.030	.059	.042
800	8	.4	.0	.3	.035	.051	.124	.740	-.067	.069	.106	.791	-.104	.104	.123	.871	.081	.082	.117	.773	1
800	8	.4	.6	.0	-.036	.057	.148	.841	.205	.205	.118	1	-.129	.129	.086	.974	.236	.236	.088	1	1
800	8	.4	.6	.3	-.158	.158	.116	.998	.168	.168	.125	.992	-.362	.362	.111	1	.403	.403	.163	1	1
800	8	.8	.0	.0	-.023	.040	.111	.863	.002	.010	.032	.083	-.006	.019	.061	.096	.000	.009	.026	.057	.046
800	8	.8	.0	.3	-.023	.034	.092	.704	-.057	.058	.091	.769	-.365	.365	.453	.974	.069	.109	.319	.772	1
800	8	.8	.6	.0	-.068	.068	.095	.925	.209	.209	.081	1	-.200	.200	.182	.982	.210	.210	.084	.999	1
800	8	.8	.6	.3	-.169	.169	.095	1	.157	.157	.118	.993	-.530	.530	.180	1	.462	.462	.240	.998	1

Table A.2: Linear Estimator of Hayakawa (2012) with strict exogeneity assumption

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	-.004	.030	.097	.060	-.005	.030	.099	.057	-.003	.025	.077	.120	-.008	.024	.076	.111	.125
200	4	.4	.0	.3	-.059	.065	.160	.214	-.160	.168	.247	.504	-.032	.051	.142	.306	-.189	.190	.215	.812	.239
200	4	.4	.6	.0	-.012	.031	.109	.079	.000	.029	.103	.068	-.007	.026	.085	.147	-.005	.025	.080	.128	.133
200	4	.4	.6	.3	-.085	.086	.181	.291	-.160	.174	.262	.503	-.059	.065	.150	.404	-.195	.196	.228	.831	.216
200	4	.8	.0	.0	-.060	.077	.216	.193	-.010	.025	.084	.085	-.060	.074	.209	.281	-.014	.023	.077	.194	.179
200	4	.8	.0	.3	-.322	.322	.301	.768	-.125	.127	.134	.643	-.348	.348	.320	.930	-.157	.157	.096	.905	.098
200	4	.8	.6	.0	-.075	.090	.242	.236	-.008	.025	.084	.095	-.072	.088	.243	.345	-.017	.026	.076	.207	.193
200	4	.8	.6	.3	-.347	.347	.305	.761	-.126	.130	.134	.627	-.380	.380	.334	.938	-.157	.157	.089	.905	.082
200	8	.4	.0	.0	-.003	.022	.075	.094	.000	.023	.078	.092	.000	.015	.048	.339	-.003	.015	.047	.333	.108
200	8	.4	.0	.3	-.064	.070	.168	.279	-.015	.063	.219	.157	-.029	.039	.102	.525	-.058	.068	.142	.695	.642
200	8	.4	.6	.0	-.012	.024	.092	.117	.010	.022	.091	.113	-.006	.017	.056	.372	.004	.015	.051	.331	.114
200	8	.4	.6	.3	-.080	.080	.200	.374	-.007	.063	.267	.186	-.042	.044	.117	.584	-.057	.073	.164	.707	.583
200	8	.8	.0	.0	-.024	.029	.092	.165	-.003	.015	.051	.080	-.020	.024	.071	.433	-.005	.011	.036	.311	.118
200	8	.8	.0	.3	-.201	.201	.179	.820	-.048	.074	.215	.317	-.193	.193	.149	.991	-.086	.095	.126	.852	.600
200	8	.8	.6	.0	-.029	.033	.106	.216	.004	.015	.063	.111	-.025	.027	.079	.476	-.002	.011	.038	.319	.104
200	8	.8	.6	.3	-.208	.208	.185	.884	-.048	.077	.252	.340	-.200	.200	.137	.996	-.089	.097	.137	.869	.508
800	4	.4	.0	.0	-.005	.028	.102	.081	-.007	.032	.117	.078	-.002	.023	.074	.143	-.006	.023	.076	.117	.149
800	4	.4	.0	.3	-.066	.069	.122	.478	-.192	.194	.227	.726	-.037	.055	.128	.603	-.215	.215	.178	.979	.818
800	4	.4	.6	.0	-.008	.028	.108	.093	-.003	.033	.114	.087	-.004	.023	.083	.160	-.005	.024	.084	.142	.160
800	4	.4	.6	.3	-.078	.078	.125	.549	-.200	.203	.194	.773	-.054	.057	.118	.605	-.229	.229	.175	.980	.732
800	4	.8	.0	.0	-.082	.098	.302	.255	-.020	.035	.123	.144	-.073	.087	.292	.339	-.021	.031	.122	.266	.203
800	4	.8	.0	.3	-.389	.389	.307	.892	-.148	.149	.121	.806	-.436	.436	.321	.981	-.178	.178	.067	.995	.549
800	4	.8	.6	.0	-.106	.118	.316	.307	-.022	.037	.120	.156	-.099	.112	.341	.422	-.028	.036	.118	.312	.233
800	4	.8	.6	.3	-.409	.409	.311	.887	-.151	.152	.107	.824	-.458	.458	.308	.985	-.182	.182	.051	.991	.436
800	8	.4	.0	.0	-.003	.019	.079	.088	-.002	.024	.099	.112	.000	.011	.035	.208	-.004	.012	.039	.199	.167
800	8	.4	.0	.3	-.066	.069	.117	.515	-.019	.052	.157	.290	-.013	.025	.066	.528	-.085	.087	.089	.915	1
800	8	.4	.6	.0	-.007	.020	.077	.106	.002	.022	.092	.113	-.003	.012	.036	.209	-.002	.012	.037	.173	.163
800	8	.4	.6	.3	-.072	.073	.117	.585	-.027	.053	.166	.314	-.019	.024	.057	.511	-.094	.096	.083	.952	1
800	8	.8	.0	.0	-.027	.029	.107	.242	-.004	.019	.071	.091	-.023	.024	.075	.415	-.008	.012	.040	.250	.193
800	8	.8	.0	.3	-.185	.185	.141	.884	-.057	.067	.125	.531	-.182	.182	.140	.984	-.103	.104	.089	.974	1
800	8	.8	.6	.0	-.031	.033	.112	.275	-.003	.019	.071	.091	-.025	.026	.079	.459	-.008	.013	.039	.259	.192
800	8	.8	.6	.3	-.192	.192	.136	.926	-.062	.073	.133	.572	-.188	.188	.124	.993	-.109	.110	.076	.977	1

Table A.3: GMM estimator of Ahn, Lee, and Schmidt (2013)

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	.001	.028	.087	.075	-.002	.026	.085	.056	-.001	.022	.067	.137	.000	.021	.065	.102	.097
200	4	.4	.0	.3	-.001	.055	.200	.109	-.005	.057	.199	.111	-.007	.038	.134	.148	.000	.041	.137	.158	.085
200	4	.4	.6	.0	-.005	.029	.097	.094	.004	.025	.083	.063	-.004	.023	.074	.150	.002	.020	.063	.091	.094
200	4	.4	.6	.3	-.020	.048	.211	.134	.013	.049	.217	.117	-.013	.037	.134	.141	.005	.037	.127	.138	.081
200	4	.8	.0	.0	-.004	.029	.107	.096	-.001	.016	.056	.058	-.005	.022	.083	.146	.000	.013	.045	.099	.102
200	4	.8	.0	.3	-.014	.043	.424	.182	-.004	.038	.292	.166	-.013	.034	.373	.197	-.003	.029	.270	.198	.122
200	4	.8	.6	.0	-.007	.032	.117	.110	.003	.016	.053	.067	-.007	.022	.086	.142	.002	.013	.044	.092	.106
200	4	.8	.6	.3	-.016	.039	.323	.168	.006	.034	.125	.098	-.013	.032	.273	.193	.001	.027	.103	.151	.107
200	8	.4	.0	.0	-.001	.022	.077	.109	.000	.022	.081	.100	-.001	.015	.049	.315	.000	.014	.045	.257	.106
200	8	.4	.0	.3	.008	.054	.205	.133	-.011	.057	.219	.128	.001	.029	.105	.341	-.002	.029	.102	.332	.078
200	8	.4	.6	.0	-.006	.024	.092	.142	.004	.020	.076	.100	-.004	.017	.058	.356	.002	.013	.043	.239	.085
200	8	.4	.6	.3	-.014	.046	.235	.144	.010	.047	.246	.141	-.007	.027	.116	.323	.006	.027	.110	.296	.091
200	8	.8	.0	.0	-.005	.021	.072	.104	.001	.013	.044	.063	-.002	.015	.050	.288	.001	.009	.028	.197	.095
200	8	.8	.0	.3	-.005	.035	.133	.099	.003	.037	.133	.096	-.004	.022	.079	.280	.002	.023	.076	.263	.074
200	8	.8	.6	.0	-.006	.021	.080	.113	.002	.012	.045	.076	-.003	.015	.054	.295	.001	.008	.027	.195	.093
200	8	.8	.6	.3	-.010	.033	.134	.118	.010	.036	.146	.113	-.005	.021	.075	.264	.006	.023	.076	.241	.075
800	4	.4	.0	.0	-.002	.025	.085	.090	.002	.029	.105	.092	-.001	.018	.057	.123	.001	.021	.068	.120	.096
800	4	.4	.0	.3	-.002	.033	.124	.106	-.001	.033	.126	.119	-.003	.021	.070	.122	.000	.022	.072	.124	.105
800	4	.4	.6	.0	-.005	.024	.086	.102	.005	.025	.097	.086	-.003	.019	.060	.136	.002	.019	.064	.091	.096
800	4	.4	.6	.3	-.008	.028	.115	.111	.005	.027	.121	.111	-.005	.019	.063	.110	.002	.019	.066	.109	.100
800	4	.8	.0	.0	-.004	.020	.076	.096	.000	.018	.059	.078	-.004	.017	.058	.136	.000	.015	.048	.093	.088
800	4	.8	.0	.3	-.005	.022	.094	.127	-.002	.021	.079	.124	-.004	.017	.067	.132	-.001	.016	.059	.130	.111
800	4	.8	.6	.0	-.006	.019	.073	.101	.001	.019	.063	.064	-.005	.016	.065	.143	.000	.016	.052	.085	.090
800	4	.8	.6	.3	-.006	.021	.089	.127	.002	.021	.074	.098	-.005	.017	.070	.138	.000	.017	.054	.115	.106
800	8	.4	.0	.0	.001	.022	.083	.136	-.001	.027	.111	.123	-.001	.010	.035	.220	.000	.013	.041	.186	.141
800	8	.4	.0	.3	.003	.029	.115	.109	-.004	.030	.118	.119	-.001	.012	.040	.176	.001	.012	.040	.173	.123
800	8	.4	.6	.0	-.004	.019	.079	.143	.003	.021	.088	.113	-.001	.012	.038	.237	.001	.012	.037	.154	.139
800	8	.4	.6	.3	-.005	.023	.117	.133	.004	.024	.120	.126	-.002	.012	.039	.170	.002	.012	.038	.150	.114
800	8	.8	.0	.0	-.002	.013	.045	.083	.000	.015	.051	.076	-.001	.008	.027	.175	.000	.009	.027	.125	.110
800	8	.8	.0	.3	-.002	.017	.063	.083	.001	.017	.063	.083	-.001	.009	.029	.137	.001	.010	.030	.134	.097
800	8	.8	.6	.0	-.003	.013	.046	.087	.000	.015	.052	.083	-.001	.008	.030	.183	.000	.009	.027	.115	.116
800	8	.8	.6	.3	-.003	.015	.056	.093	.002	.016	.063	.088	-.001	.008	.027	.117	.001	.009	.028	.108	.095

Table A.4: Subset GMM estimator of Ahn, Lee, and Schmidt (2013)

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	8	.4	.0	.0	.000	.022	.072	.102	-.001	.021	.074	.094	-.001	.014	.046	.262	-.001	.013	.041	.193	.128
200	8	.4	.0	.3	.008	.050	.185	.125	-.012	.052	.188	.125	.001	.025	.090	.263	-.002	.026	.088	.258	.087
200	8	.4	.6	.0	-.006	.023	.085	.134	.004	.019	.067	.090	-.003	.016	.053	.299	.002	.013	.041	.189	.098
200	8	.4	.6	.3	-.012	.044	.205	.131	.009	.042	.208	.124	-.006	.025	.094	.256	.005	.024	.089	.225	.087
200	8	.8	.0	.0	-.004	.021	.071	.094	.000	.013	.041	.060	-.002	.014	.047	.261	.000	.008	.027	.168	.116
200	8	.8	.0	.3	-.005	.034	.127	.092	.003	.035	.126	.090	-.004	.021	.073	.235	.002	.022	.070	.214	.088
200	8	.8	.6	.0	-.006	.022	.079	.115	.002	.012	.042	.072	-.003	.015	.054	.273	.001	.008	.026	.152	.097
200	8	.8	.6	.3	-.010	.032	.121	.109	.009	.034	.132	.101	-.006	.020	.071	.213	.006	.022	.071	.190	.088
800	8	.4	.0	.0	.001	.020	.079	.119	-.002	.026	.101	.116	.000	.011	.034	.189	.000	.012	.039	.166	.135
800	8	.4	.0	.3	.004	.027	.109	.105	-.004	.029	.111	.113	-.001	.012	.039	.166	.001	.012	.039	.152	.129
800	8	.4	.6	.0	-.004	.018	.076	.127	.002	.020	.081	.115	-.001	.012	.037	.208	.000	.011	.036	.130	.131
800	8	.4	.6	.3	-.004	.021	.099	.124	.004	.021	.100	.123	-.002	.012	.038	.151	.001	.012	.037	.130	.110
800	8	.8	.0	.0	-.002	.013	.046	.084	.000	.014	.051	.077	-.001	.009	.028	.162	.000	.009	.028	.121	.103
800	8	.8	.0	.3	-.003	.017	.060	.082	.001	.017	.061	.082	-.001	.009	.029	.132	.001	.009	.030	.131	.101
800	8	.8	.6	.0	-.003	.013	.047	.092	-.001	.014	.050	.078	-.001	.009	.030	.170	.000	.009	.027	.105	.108
800	8	.8	.6	.3	-.003	.014	.053	.089	.001	.015	.058	.082	-.001	.008	.027	.120	.001	.009	.028	.102	.094

Table A.5: FIVU estimator of Robertson and Sarafidis (2013)

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	.001	.023	.068	.064	-.002	.022	.065	.048	.000	.021	.061	.073	-.001	.021	.060	.061	.031
200	4	.4	.0	.3	.008	.045	.132	.072	-.004	.043	.136	.068	-.003	.036	.111	.085	.001	.038	.113	.085	.031
200	4	.4	.6	.0	.000	.023	.069	.063	.001	.020	.060	.041	.000	.022	.064	.079	.000	.019	.057	.064	.029
200	4	.4	.6	.3	-.008	.036	.107	.064	.006	.036	.116	.064	-.006	.033	.100	.068	.003	.034	.102	.079	.031
200	4	.8	.0	.0	.000	.024	.075	.063	.000	.014	.042	.053	-.001	.020	.061	.070	.001	.012	.040	.069	.035
200	4	.8	.0	.3	-.003	.030	.099	.060	.003	.026	.088	.065	-.003	.028	.089	.076	.002	.024	.079	.080	.038
200	4	.8	.6	.0	-.002	.025	.079	.063	.001	.013	.041	.043	-.002	.020	.066	.071	.002	.012	.038	.066	.033
200	4	.8	.6	.3	-.006	.029	.093	.068	.004	.026	.082	.069	-.004	.028	.089	.084	.002	.025	.079	.085	.035
200	8	.4	.0	.0	.002	.014	.042	.072	-.002	.013	.041	.071	.001	.012	.036	.182	.000	.011	.034	.160	.032
200	8	.4	.0	.3	.012	.034	.097	.080	-.014	.034	.099	.085	.004	.021	.063	.173	-.004	.022	.065	.180	.035
200	8	.4	.6	.0	.000	.014	.042	.065	.000	.012	.035	.061	.000	.013	.037	.179	.000	.011	.033	.135	.032
200	8	.4	.6	.3	-.004	.025	.080	.056	.003	.026	.079	.054	-.002	.020	.060	.174	.002	.020	.061	.158	.034
200	8	.8	.0	.0	-.001	.013	.038	.053	.000	.008	.025	.050	.000	.011	.034	.168	.000	.007	.023	.143	.037
200	8	.8	.0	.3	-.001	.022	.066	.051	.001	.023	.068	.048	-.001	.018	.054	.163	.001	.018	.057	.155	.036
200	8	.8	.6	.0	-.001	.014	.039	.051	.000	.008	.023	.055	.000	.012	.035	.164	.001	.007	.022	.140	.037
200	8	.8	.6	.3	-.004	.020	.060	.048	.005	.023	.066	.048	-.003	.018	.053	.156	.002	.019	.057	.153	.030
800	4	.4	.0	.0	.000	.020	.061	.060	.000	.022	.073	.066	.000	.017	.051	.069	-.001	.020	.060	.069	.052
800	4	.4	.0	.3	.002	.024	.078	.072	-.001	.024	.081	.068	-.001	.020	.059	.059	.000	.020	.061	.063	.055
800	4	.4	.6	.0	-.002	.019	.055	.068	.002	.019	.058	.056	-.001	.017	.053	.074	.002	.018	.057	.066	.050
800	4	.4	.6	.3	-.004	.021	.063	.064	.002	.020	.067	.059	-.002	.018	.054	.060	.001	.018	.055	.065	.046
800	4	.8	.0	.0	-.002	.016	.053	.058	.000	.015	.047	.050	-.001	.016	.048	.067	.000	.013	.042	.056	.050
800	4	.8	.0	.3	-.002	.017	.055	.056	.001	.017	.053	.053	-.002	.015	.048	.058	.001	.015	.047	.052	.051
800	4	.8	.6	.0	-.004	.015	.051	.071	.000	.016	.049	.058	-.003	.014	.047	.077	.001	.015	.046	.059	.048
800	4	.8	.6	.3	-.004	.016	.052	.069	.002	.016	.050	.059	-.002	.015	.047	.066	.000	.015	.046	.058	.049
800	8	.4	.0	.0	.002	.013	.038	.056	-.003	.017	.050	.066	.000	.008	.025	.079	.000	.010	.031	.081	.050
800	8	.4	.0	.3	.005	.018	.055	.063	-.007	.019	.055	.064	.000	.010	.030	.080	.000	.010	.031	.083	.047
800	8	.4	.6	.0	-.001	.011	.031	.054	.000	.012	.035	.055	-.001	.009	.026	.078	.001	.010	.030	.080	.055
800	8	.4	.6	.3	-.001	.013	.039	.054	.000	.013	.038	.052	-.001	.010	.029	.078	.001	.010	.030	.077	.051
800	8	.8	.0	.0	-.001	.008	.026	.049	.000	.010	.030	.059	.000	.007	.021	.077	.000	.008	.024	.080	.050
800	8	.8	.0	.3	.000	.011	.034	.050	.001	.011	.034	.056	.000	.008	.024	.065	.000	.008	.025	.079	.052
800	8	.8	.6	.0	-.001	.008	.025	.050	-.001	.009	.028	.057	-.001	.007	.021	.084	.000	.008	.024	.079	.051
800	8	.8	.6	.3	-.001	.009	.029	.056	.000	.010	.031	.053	.000	.007	.024	.076	.000	.008	.025	.073	.059

Table A.6: Subset FIVU estimator of Robertson and Sarafidis (2013)

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	8	.4	.0	.0	.002	.013	.042	.076	-.003	.013	.041	.069	.000	.012	.035	.126	-.001	.011	.035	.116	.029
200	8	.4	.0	.3	.011	.032	.094	.088	-.012	.033	.094	.087	.001	.021	.064	.125	-.002	.021	.065	.119	.030
200	8	.4	.6	.0	.000	.014	.042	.068	.000	.012	.037	.049	.000	.013	.037	.127	-.001	.011	.034	.104	.032
200	8	.4	.6	.3	-.005	.025	.075	.057	.005	.024	.074	.059	-.002	.020	.060	.121	.002	.020	.060	.109	.030
200	8	.8	.0	.0	.000	.014	.042	.066	.000	.008	.025	.057	-.001	.012	.037	.136	.000	.008	.023	.115	.031
200	8	.8	.0	.3	-.002	.023	.068	.057	.001	.023	.068	.047	-.003	.018	.057	.125	.002	.019	.056	.116	.035
200	8	.8	.6	.0	-.002	.014	.044	.069	.001	.008	.024	.058	-.001	.012	.038	.134	.000	.008	.023	.101	.028
200	8	.8	.6	.3	-.005	.020	.061	.052	.005	.022	.067	.044	-.004	.018	.054	.122	.003	.019	.058	.103	.039
800	8	.4	.0	.0	.002	.013	.038	.059	-.003	.016	.047	.060	.000	.009	.026	.072	.000	.011	.033	.063	.044
800	8	.4	.0	.3	.004	.017	.051	.069	-.005	.017	.051	.072	.000	.010	.032	.076	.000	.011	.033	.074	.045
800	8	.4	.6	.0	.000	.011	.032	.060	.000	.012	.035	.058	.000	.009	.028	.077	.000	.010	.032	.071	.048
800	8	.4	.6	.3	-.001	.012	.038	.055	.001	.012	.038	.069	-.001	.010	.030	.079	.000	.010	.031	.071	.044
800	8	.8	.0	.0	.000	.010	.029	.059	.000	.010	.031	.055	.000	.008	.024	.068	.000	.008	.025	.072	.041
800	8	.8	.0	.3	-.001	.011	.034	.059	.001	.011	.032	.056	-.001	.008	.026	.068	.001	.008	.026	.068	.047
800	8	.8	.6	.0	-.001	.009	.029	.059	.000	.009	.029	.061	.000	.008	.025	.072	.000	.008	.025	.073	.046
800	8	.8	.6	.3	-.002	.010	.030	.049	.001	.010	.031	.056	.000	.008	.025	.078	.000	.009	.025	.067	.050

Table A.7: FIVR estimator of Robertson and Sarafidis (2013)

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	.001	.019	.058	.068	-.002	.020	.060	.058	.000	.016	.047	.081	-.001	.018	.052	.081	.035
200	4	.4	.0	.3	.008	.037	.113	.081	-.006	.038	.122	.071	-.002	.027	.083	.081	-.001	.030	.090	.080	.033
200	4	.4	.6	.0	.000	.019	.057	.061	.000	.019	.055	.046	.000	.016	.048	.081	.000	.017	.051	.073	.031
200	4	.4	.6	.3	-.002	.031	.095	.062	.003	.034	.106	.065	-.001	.026	.079	.068	.000	.029	.088	.077	.032
200	4	.8	.0	.0	.001	.017	.055	.066	.000	.012	.038	.063	.000	.014	.044	.072	.000	.011	.035	.085	.035
200	4	.8	.0	.3	.000	.023	.073	.061	.002	.024	.076	.057	.000	.021	.061	.067	.000	.022	.067	.082	.039
200	4	.8	.6	.0	-.001	.018	.054	.059	.000	.012	.037	.060	.000	.014	.044	.068	.000	.011	.035	.086	.038
200	4	.8	.6	.3	-.001	.023	.071	.062	.002	.024	.076	.066	.000	.021	.062	.071	.000	.022	.072	.084	.041
200	8	.4	.0	.0	.001	.012	.037	.069	-.002	.013	.039	.068	.001	.011	.031	.181	-.001	.011	.033	.172	.043
200	8	.4	.0	.3	.015	.034	.095	.086	-.017	.036	.099	.087	.005	.020	.057	.214	-.006	.021	.061	.215	.043
200	8	.4	.6	.0	.000	.012	.036	.067	-.001	.011	.033	.062	.001	.011	.032	.189	.000	.011	.032	.163	.040
200	8	.4	.6	.3	-.002	.025	.077	.054	.001	.027	.080	.051	-.001	.018	.055	.197	.001	.020	.060	.186	.038
200	8	.8	.0	.0	.000	.011	.032	.054	.000	.008	.023	.051	.001	.009	.028	.179	.000	.007	.022	.155	.037
200	8	.8	.0	.3	.000	.019	.057	.047	.000	.022	.066	.045	.001	.015	.046	.183	.000	.018	.054	.174	.037
200	8	.8	.6	.0	.000	.011	.031	.054	.000	.007	.022	.051	.001	.009	.028	.181	.000	.007	.022	.159	.036
200	8	.8	.6	.3	-.003	.018	.055	.051	.004	.022	.066	.046	-.001	.016	.047	.176	.002	.018	.056	.177	.038
800	4	.4	.0	.0	-.001	.015	.045	.059	.000	.019	.061	.063	-.001	.012	.036	.066	.001	.016	.049	.066	.051
800	4	.4	.0	.3	.000	.021	.064	.068	-.001	.022	.070	.066	-.001	.015	.044	.068	.000	.016	.049	.060	.048
800	4	.4	.6	.0	-.001	.013	.041	.059	.002	.017	.052	.051	-.001	.012	.037	.062	.001	.015	.048	.056	.051
800	4	.4	.6	.3	-.002	.017	.051	.062	.002	.018	.058	.059	-.001	.014	.043	.059	.001	.016	.050	.058	.048
800	4	.8	.0	.0	-.001	.011	.034	.061	.000	.014	.043	.056	.000	.010	.030	.075	.000	.012	.038	.060	.045
800	4	.8	.0	.3	.000	.014	.042	.051	.000	.015	.046	.052	-.001	.011	.035	.062	.000	.013	.040	.059	.047
800	4	.8	.6	.0	-.001	.011	.033	.069	.000	.015	.044	.056	.000	.010	.029	.075	.000	.014	.042	.062	.042
800	4	.8	.6	.3	-.001	.013	.041	.064	.002	.015	.048	.056	.000	.011	.035	.057	.000	.014	.042	.059	.044
800	8	.4	.0	.0	.001	.011	.033	.050	-.002	.015	.047	.064	.000	.007	.020	.093	.000	.010	.028	.082	.054
800	8	.4	.0	.3	.005	.017	.053	.070	-.006	.018	.056	.073	.000	.008	.026	.082	.000	.010	.028	.082	.054
800	8	.4	.6	.0	.000	.009	.026	.051	-.001	.011	.033	.054	.000	.007	.021	.079	.000	.009	.028	.077	.053
800	8	.4	.6	.3	-.001	.012	.037	.052	.000	.013	.038	.056	.000	.009	.026	.078	.000	.009	.029	.079	.053
800	8	.8	.0	.0	.000	.006	.019	.053	.000	.010	.028	.061	.000	.005	.016	.082	.000	.007	.023	.080	.053
800	8	.8	.0	.3	.000	.010	.029	.055	.000	.011	.032	.054	.000	.007	.020	.078	.000	.008	.023	.081	.053
800	8	.8	.6	.0	.000	.006	.018	.051	.000	.009	.028	.057	.000	.005	.016	.078	.000	.008	.024	.080	.050
800	8	.8	.6	.3	.000	.009	.027	.055	.000	.010	.031	.055	.000	.007	.021	.079	.000	.008	.024	.079	.049

Table A.8: Subset FIVR estimator of Robertson and Sarafidis (2013)

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	8	.4	.0	.0	.002	.012	.038	.072	-.002	.012	.039	.071	.000	.010	.031	.131	-.001	.010	.032	.124	.035
200	8	.4	.0	.3	.012	.032	.092	.093	-.012	.034	.097	.089	.002	.018	.056	.142	-.003	.019	.059	.140	.038
200	8	.4	.6	.0	.001	.012	.037	.067	.000	.011	.034	.054	.000	.011	.032	.139	-.001	.010	.032	.124	.033
200	8	.4	.6	.3	-.002	.023	.070	.063	.003	.025	.074	.060	-.001	.018	.053	.133	.000	.020	.057	.122	.035
200	8	.8	.0	.0	.000	.011	.032	.066	.000	.008	.023	.061	.000	.010	.029	.142	.000	.008	.022	.116	.035
200	8	.8	.0	.3	-.001	.019	.058	.053	.001	.022	.065	.052	.000	.015	.046	.136	.001	.018	.052	.126	.039
200	8	.8	.6	.0	.000	.011	.032	.064	.000	.007	.023	.054	.000	.009	.029	.143	.000	.007	.022	.118	.038
200	8	.8	.6	.3	-.003	.018	.054	.056	.004	.022	.065	.049	-.001	.015	.047	.131	.001	.018	.055	.121	.042
800	8	.4	.0	.0	.001	.010	.032	.061	-.002	.014	.043	.064	.000	.007	.022	.077	.000	.010	.029	.074	.053
800	8	.4	.0	.3	.004	.015	.048	.071	-.005	.017	.051	.073	.000	.009	.027	.073	.000	.010	.029	.076	.048
800	8	.4	.6	.0	.000	.009	.026	.052	.000	.011	.033	.055	.000	.007	.023	.077	.000	.010	.029	.073	.055
800	8	.4	.6	.3	.000	.011	.035	.053	.000	.012	.036	.058	.000	.009	.026	.066	.000	.010	.029	.074	.049
800	8	.8	.0	.0	.000	.006	.020	.055	.000	.009	.028	.064	.000	.006	.017	.074	.000	.007	.023	.069	.044
800	8	.8	.0	.3	.000	.009	.028	.061	.000	.010	.030	.060	.000	.007	.020	.070	.000	.008	.023	.070	.052
800	8	.8	.6	.0	.000	.006	.019	.056	.000	.009	.028	.057	.000	.006	.017	.073	.000	.008	.024	.076	.046
800	8	.8	.6	.3	.000	.009	.026	.059	.000	.010	.030	.059	.000	.007	.021	.072	.000	.008	.024	.070	.050

Table A.9: Projection GMM estimator of Hayakawa (2012) with weak exogeneity

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	.000	.025	.076	.058	-.001	.023	.075	.053	-.002	.023	.072	.087	.002	.023	.072	.074	.020
200	4	.4	.0	.3	.003	.056	.181	.078	-.003	.054	.172	.083	-.011	.055	.171	.113	.007	.050	.166	.113	.026
200	4	.4	.6	.0	-.001	.026	.081	.077	.003	.021	.070	.062	-.003	.026	.081	.106	.005	.022	.073	.097	.028
200	4	.4	.6	.3	-.016	.055	.191	.106	.015	.052	.191	.097	-.019	.056	.206	.153	.016	.050	.199	.141	.021
200	4	.8	.0	.0	-.001	.033	.107	.073	.001	.016	.050	.047	-.001	.031	.101	.092	.002	.015	.046	.062	.020
200	4	.8	.0	.3	-.009	.050	.179	.088	.002	.034	.116	.079	-.013	.052	.179	.132	.004	.035	.117	.111	.033
200	4	.8	.6	.0	-.003	.032	.104	.069	.003	.015	.050	.053	-.005	.032	.108	.108	.004	.015	.051	.088	.021
200	4	.8	.6	.3	-.013	.056	.212	.106	.009	.041	.142	.084	-.018	.059	.253	.167	.010	.041	.150	.122	.025
200	8	.4	.0	.0	.001	.015	.046	.075	-.001	.014	.046	.075	.000	.013	.039	.143	.000	.012	.038	.131	.018
200	8	.4	.0	.3	.015	.045	.134	.089	-.015	.044	.135	.099	.002	.031	.093	.147	-.002	.031	.092	.145	.021
200	8	.4	.6	.0	.000	.014	.044	.063	.001	.012	.037	.056	.000	.014	.042	.144	.001	.012	.035	.118	.028
200	8	.4	.6	.3	-.008	.038	.120	.066	.008	.038	.118	.051	-.006	.031	.089	.136	.006	.030	.089	.128	.029
200	8	.8	.0	.0	-.001	.016	.050	.059	.001	.009	.028	.068	-.001	.016	.046	.140	.001	.008	.026	.129	.021
200	8	.8	.0	.3	-.001	.033	.104	.052	.002	.030	.094	.056	-.004	.031	.090	.128	.004	.027	.081	.123	.019
200	8	.8	.6	.0	-.001	.015	.046	.045	.000	.009	.025	.058	-.002	.015	.047	.136	.001	.008	.025	.118	.026
200	8	.8	.6	.3	-.010	.041	.125	.059	.007	.038	.121	.059	-.009	.035	.106	.135	.008	.033	.100	.138	.026
800	4	.4	.0	.0	-.001	.026	.074	.065	.000	.026	.082	.079	-.001	.021	.064	.072	.001	.023	.069	.073	.035
800	4	.4	.0	.3	.000	.032	.105	.072	-.001	.031	.102	.075	-.003	.028	.090	.079	.003	.027	.088	.074	.037
800	4	.4	.6	.0	-.004	.024	.079	.086	.004	.022	.072	.069	-.004	.024	.077	.103	.004	.022	.074	.095	.038
800	4	.4	.6	.3	-.006	.031	.107	.081	.006	.030	.108	.072	-.005	.027	.089	.078	.004	.025	.089	.075	.042
800	4	.8	.0	.0	-.006	.037	.113	.066	.001	.019	.057	.054	-.007	.036	.108	.098	.002	.017	.052	.059	.036
800	4	.8	.0	.3	-.003	.028	.094	.067	.001	.021	.067	.060	-.004	.025	.088	.074	.001	.020	.063	.066	.044
800	4	.8	.6	.0	-.007	.028	.105	.081	.004	.021	.069	.071	-.008	.028	.105	.096	.003	.020	.067	.084	.045
800	4	.8	.6	.3	-.007	.031	.115	.077	.005	.026	.091	.064	-.007	.030	.106	.085	.005	.024	.083	.072	.043
800	8	.4	.0	.0	.003	.018	.053	.105	-.003	.022	.065	.118	.000	.011	.032	.082	.000	.012	.036	.077	.030
800	8	.4	.0	.3	.008	.025	.073	.072	-.007	.025	.072	.074	.000	.016	.048	.076	.000	.016	.047	.073	.027
800	8	.4	.6	.0	-.001	.014	.041	.060	.000	.013	.040	.055	-.001	.012	.037	.084	.001	.012	.035	.079	.042
800	8	.4	.6	.3	-.003	.022	.068	.054	.003	.022	.068	.056	-.001	.017	.047	.067	.001	.016	.047	.073	.035
800	8	.8	.0	.0	-.001	.019	.056	.068	.000	.013	.039	.079	-.002	.015	.046	.087	.001	.010	.029	.075	.031
800	8	.8	.0	.3	.000	.018	.055	.057	.000	.016	.048	.057	-.001	.014	.042	.080	.001	.012	.036	.079	.037
800	8	.8	.6	.0	.000	.015	.048	.046	.001	.011	.033	.055	-.001	.014	.046	.083	.001	.010	.030	.073	.040
800	8	.8	.6	.3	-.002	.020	.062	.053	.001	.019	.058	.055	-.001	.015	.047	.071	.000	.015	.043	.069	.041

Table A.10: Subset Projection GMM estimator of Hayakawa (2012) with weak exogeneity

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	8	.4	.0	.0	.000	.014	.045	.074	.000	.014	.044	.064	.000	.013	.041	.123	-.001	.013	.039	.108	.013
200	8	.4	.0	.3	.009	.043	.133	.083	-.009	.043	.134	.088	.000	.033	.099	.120	.000	.032	.098	.128	.026
200	8	.4	.6	.0	.000	.014	.045	.071	.001	.012	.037	.054	.000	.015	.046	.140	.001	.012	.037	.111	.029
200	8	.4	.6	.3	-.009	.038	.118	.076	.009	.038	.118	.067	-.006	.031	.093	.110	.005	.030	.091	.109	.035
200	8	.8	.0	.0	-.001	.017	.055	.071	.001	.009	.029	.068	-.002	.017	.052	.138	.000	.009	.027	.115	.020
200	8	.8	.0	.3	-.004	.038	.117	.068	.004	.033	.105	.063	-.006	.033	.102	.116	.004	.029	.087	.109	.031
200	8	.8	.6	.0	-.002	.016	.050	.059	.000	.009	.026	.058	-.002	.017	.052	.122	.001	.009	.026	.104	.025
200	8	.8	.6	.3	-.011	.044	.139	.079	.009	.042	.130	.070	-.007	.037	.116	.131	.007	.035	.108	.120	.032
800	8	.4	.0	.0	.003	.017	.049	.083	-.003	.020	.057	.097	-.001	.011	.034	.072	.000	.013	.038	.072	.033
800	8	.4	.0	.3	.006	.024	.072	.067	-.005	.024	.072	.070	.000	.018	.051	.072	.001	.018	.051	.072	.034
800	8	.4	.6	.0	-.002	.014	.042	.053	.001	.014	.040	.051	-.001	.013	.039	.075	.001	.012	.036	.070	.044
800	8	.4	.6	.3	-.002	.022	.066	.058	.003	.022	.067	.054	-.001	.017	.050	.072	.000	.017	.048	.072	.044
800	8	.8	.0	.0	.000	.020	.064	.068	.000	.013	.039	.070	-.002	.017	.052	.092	.001	.010	.030	.076	.035
800	8	.8	.0	.3	-.001	.019	.057	.058	.000	.015	.048	.061	-.001	.015	.047	.076	.001	.013	.039	.069	.036
800	8	.8	.6	.0	-.001	.016	.052	.054	.001	.011	.034	.055	-.002	.015	.052	.079	.001	.011	.031	.065	.044
800	8	.8	.6	.3	-.003	.021	.065	.059	.002	.019	.061	.062	-.001	.017	.051	.068	.000	.016	.046	.068	.042

Table A.11: Conditional likelihood estimator of Bai (2013b)

Designs					Strict								Weak							
					α				β				α				β			
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size
200	4	.4	.0	.0	.001	.013	.040	.052	-.001	.013	.038	.050	-.001	.013	.039	.059	.002	.013	.036	.066
200	4	.4	.0	.3	.003	.027	.081	.150	-.015	.031	.103	.207	-.001	.025	.074	.127	.000	.027	.078	.161
200	4	.4	.6	.0	.000	.014	.040	.053	.000	.013	.038	.052	-.011	.017	.043	.129	.024	.024	.039	.302
200	4	.4	.6	.3	.000	.025	.074	.109	-.006	.029	.090	.167	-.040	.042	.081	.350	.050	.051	.078	.445
200	4	.8	.0	.0	.000	.013	.040	.054	.000	.009	.026	.059	.000	.013	.039	.052	-.002	.009	.025	.069
200	4	.8	.0	.3	-.005	.026	.225	.234	-.016	.030	.134	.313	.000	.019	.058	.093	.000	.020	.060	.142
200	4	.8	.6	.0	.000	.013	.039	.048	.000	.009	.027	.059	-.005	.014	.041	.066	.011	.012	.026	.166
200	4	.8	.6	.3	-.003	.022	.075	.162	-.002	.026	.082	.194	-.025	.028	.060	.205	.035	.035	.055	.347
200	8	.4	.0	.0	.000	.008	.024	.056	.000	.008	.024	.051	-.001	.008	.024	.053	.001	.008	.024	.064
200	8	.4	.0	.3	.005	.015	.045	.086	-.005	.016	.047	.096	.001	.016	.049	.120	-.003	.018	.053	.144
200	8	.4	.6	.0	.000	.009	.025	.059	.000	.008	.024	.057	-.006	.010	.025	.088	.012	.013	.025	.190
200	8	.4	.6	.3	.003	.015	.044	.076	-.003	.016	.047	.090	-.023	.025	.054	.290	.027	.028	.057	.328
200	8	.8	.0	.0	.000	.008	.024	.053	.000	.006	.017	.062	.000	.008	.024	.050	-.001	.006	.018	.064
200	8	.8	.0	.3	-.008	.015	.044	.131	.007	.018	.054	.155	.000	.015	.047	.148	-.001	.019	.057	.179
200	8	.8	.6	.0	.000	.008	.024	.052	.000	.006	.017	.059	-.003	.008	.024	.065	.006	.007	.017	.122
200	8	.8	.6	.3	-.009	.015	.042	.128	.011	.019	.052	.150	-.021	.022	.050	.256	.027	.029	.057	.332
800	4	.4	.0	.0	.000	.010	.031	.060	.001	.012	.035	.051	-.003	.011	.033	.095	.004	.015	.043	.172
800	4	.4	.0	.3	.002	.022	.072	.339	-.014	.028	.116	.438	.001	.020	.061	.301	-.002	.025	.078	.415
800	4	.4	.6	.0	.000	.010	.031	.057	.001	.012	.035	.052	-.025	.026	.051	.449	.064	.064	.076	.798
800	4	.4	.6	.3	-.002	.021	.063	.297	-.003	.028	.099	.409	-.044	.044	.073	.642	.056	.056	.074	.741
800	4	.8	.0	.0	-.001	.009	.027	.057	.000	.010	.030	.049	.000	.009	.027	.058	-.006	.012	.038	.182
800	4	.8	.0	.3	-.008	.024	.250	.448	-.019	.035	.170	.578	-.002	.016	.049	.263	.001	.022	.067	.411
800	4	.8	.6	.0	.000	.009	.027	.055	.000	.010	.032	.052	-.007	.011	.030	.134	.040	.040	.045	.722
800	4	.8	.6	.3	-.008	.022	.077	.388	.005	.031	.110	.516	-.034	.034	.049	.616	.049	.049	.055	.779
800	8	.4	.0	.0	.000	.006	.018	.058	.000	.008	.023	.056	-.001	.007	.020	.081	.002	.010	.029	.154
800	8	.4	.0	.3	.005	.011	.030	.211	-.006	.012	.035	.241	.002	.013	.039	.314	-.004	.016	.048	.385
800	8	.4	.6	.0	-.001	.006	.019	.054	.000	.008	.023	.050	-.014	.015	.025	.403	.035	.035	.044	.708
800	8	.4	.6	.3	.003	.010	.029	.175	-.003	.012	.035	.224	-.026	.026	.047	.586	.030	.031	.054	.629
800	8	.8	.0	.0	.000	.005	.015	.050	.000	.007	.019	.058	.000	.005	.015	.052	-.004	.008	.024	.163
800	8	.8	.0	.3	-.006	.009	.024	.212	.007	.012	.035	.301	.000	.010	.031	.270	-.001	.014	.041	.369
800	8	.8	.6	.0	.000	.005	.015	.046	.000	.007	.020	.057	-.004	.006	.016	.118	.021	.021	.031	.553
800	8	.8	.6	.3	-.007	.010	.023	.214	.010	.013	.033	.323	-.020	.020	.032	.568	.026	.027	.040	.655