IV Estimation of Panels with Factor Residuals

Donald Robertson[a] and Vasilis Sarafidis[b*]
[a]Cambridge University, [b]Monash University

This paper proposes a new instrumental variables approach for consistent and asymptotically efficient estimation of panel data models with weakly exogenous or endogenous regressors and residuals generated by a multi-factor error structure. In this case, the standard dynamic panel estimators fail to provide consistent estimates of the parameters. The novelty of our approach is that we introduce new parameters to represent the unobserved covariances between the instruments and the factor component of the residual; these parameters are estimable when N is large. Some important estimation and identification issues are studied in detail. The finite sample performance of the proposed estimators is investigated using simulated data. The results show that the method produces reliable estimates of the parameters over several parametrisations.

KEYWORDS: Generalised Method of Moments, Dynamic Panel Data, Factor Residuals.

JEL Classification: C23, C26.

1 Introduction

This paper develops a new approach based on instrumental variables for consistent and asymptotically efficient estimation of panel data models with errors generated

*Correspondence to: Department of Econometrics and Business Statistics, Monash University, Caulfield East, 3141, Australia. Tel.: +61 99032179. E-mail address: vasilis.sarafidis@monash.edu (V. Sarafidis).

by a multi-factor structure. The factor structure is an attractive framework as it permits general forms of unobserved heterogeneity that may otherwise contaminate estimation and statistical inference. Factor residuals can be motivated in several ways, depending on the application in mind. In macroeconometric panels, the factors may be thought of as economy-wide shocks that affect all individuals, albeit with different intensities; essentially, this allows cross sections to inhabit a common environment, to which they may respond differently. In microeconometric panels, the factor structure may capture different sources of unobserved individual-specific heterogeneity, the impact of which varies intertemporally in an arbitrary way. For instance, in studies of production functions, the factor loadings may capture distinct components of firm-specific technical efficiency, which varies through time. In models of earnings determination, the factor loadings may reflect an individual's set of unobserved skills, while the factors represent the industry-wide price of these skills, which is not necessarily constant over time (see also the detailed discussions in Ahn, Lee and Schmidt, 2013, and Bai, 2009). Systematic changes in tastes is another plausible example. In some circumstances such variables could be measured and directly included in the model, but often the details of measurement might be difficult, contentious and, in any case, outside the focus of the analysis.[1] In such cases it is inviting to allow the model residual to be composed of one or more unspecified factors, themselves to be estimated.

Panel data models with factor residuals and N, T both large have been proposed by Pesaran (2006), Bai (2009), Moon and Weidner (2013), Sarafidis and Yamagata (2013), among others. In the present paper we focus on the case where N is large and T fixed. In addition, unlike the models above, the set of regressors is allowed to include general weakly exogenous or endogenous variables, due to (say) errors of measurement, omitted variables and/or simultaneity. As a result, our method possesses an appealing generality.

When unobserved heterogeneity is subject to an error components structure, a popular method to estimate models with weakly exogenous, or endogenous regressors is the Generalised Method of Moments (GMM), analysed in the dynamic panel data context by Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1998) and others. However as shown

[1] For example, how does one measure monetary shocks? Does one look at interest rates or monetary aggregates? Which monetary aggregates? How does one handle financial innovation?

by Sarafidis and Robertson (2009), these procedures fail to provide consistent estimates of the parameters when the errors are generated by a multi-factor structure because the moment conditions they utilise are invalidated. Panel data models with a single factor structure and a small number of time series observations have been studied by Holtz-Eakin, Newey and Rosen (1988), Ahn, Lee and Schmidt (2001) and Nauges and Thomas (2003). All these studies utilise some form of quasi-differencing that eliminates the single factor component from the residuals. More recently, in a seminal paper Ahn, Lee and Schmidt (2013) develop a GMM estimator that allows for multiple factors using multi-quasi-differencing. Other recent contributions include the GMM estimator proposed by Sarafidis, Yamagata and Robertson (2009), which makes uses of strictly exogenous regressors as instruments for the endogenous variables and the conditional maximum likelihood approach proposed by Bai (2013), which is based on Chamberlain's projection method. The former requires that the covariates used to form the instrument set, if they are subject to common factors as well, have factor loadings that are uncorrelated with those in the disturbance. The latter requires in general strict exogeneity of the covariates.

In this paper we develop a new instrumental variables approach; instead of eliminating the factors using some form of quasi-differencing, our methodology introduces parameters that represent the unobserved covariances between the instruments and the factor component of the residual. The proposed estimator is shown to be more efficient than the existing quasi-differencing type GMM estimators and attains the semi-parametric efficiency bound discussed by Newey (1990). Furthermore, the estimator has the traditional attraction of method of moments estimators in that it exploits only the orthogonality conditions implied by the structure of the model, which in fact may be the implication of an underlying economic theory, and avoids imposing distributional assumptions about the idiosyncratic error term, or assumptions about the stochastic process that generates the regressors.

The remainder of this paper is organised as follows. The following section describes the model and its assumptions and provides the basic intuition of our methodology. Sections 3 and 4 analyse identification and estimation of the model. Section 5 reports results on the finite sample performance of the proposed estimator, while a final section concludes. Proofs are provided in the Appendix.

2 Stochastic Framework

We consider the following model:

$$\mathbf{x}_{it}'\boldsymbol{\phi} = \boldsymbol{\lambda}_i'\mathbf{f}_t + \varepsilon_{it}, \quad i = 1, ..., N, \ t = 1, ..., T, \tag{2.1}$$

where $\mathbf{x}_{it} = (y_{it}, x_{1it}, x_{2it}, \ldots, x_{Kit})'$ is a $(K+1) \times 1$ vector containing the (endogenous and exogenous) observed variables and $\boldsymbol{\phi} = (1, -\boldsymbol{\beta}')'$, where $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters. $\boldsymbol{\lambda}_i$ is a stochastic $n \times 1$ vector of factor loadings and $\mathbf{f}_t$ is an $n \times 1$ vector of factors which are treated as time-specific parameters; ε_{it} is a purely idiosyncratic disturbance.

The model (2.1) can be stacked over t to take the form

$$\mathbf{X}_i\boldsymbol{\phi} = (\mathbf{I}_T \otimes \boldsymbol{\lambda}_i')\mathbf{f} + \boldsymbol{\varepsilon}_i, \tag{2.2}$$

where $\mathbf{X}_i = [\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}]'$, $\mathbf{f} = \text{vec}(\mathbf{F}')$, $\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_T]'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$. The following assumptions are made.

Assumption 1. Existence of instruments. We assume potential instruments are given by a vector $\mathbf{w}_i$ of dimension d; these instruments may correspond to the variables of the model or be extraneous variables. In each period t, $c_t > 0$ instruments are available, expressed in vector form as follows:

$$\mathbf{w}_{it} = \mathbf{S}_t\mathbf{w}_i, \tag{2.3}$$

for which the condition $E(\mathbf{w}_{it}\varepsilon_{it}) = \mathbf{0}$ holds true.

Here $\mathbf{S}_t$ is the selector matrix of 0's and 1's that picks out from all potential instruments in $\mathbf{w}_i$, those that are valid at date t. The matrix $\mathbf{S}_t$ has dimension $c_t \times d$ where c_t is the number of orthogonality conditions associated with ε_{it}. The total number of moment conditions is $c = \sum_{t=1}^{T} c_t$. The instruments that are available in each time period depend on the structure of the model. For instance, in a model with a single explanatory variable, x_{it}, $\mathbf{w}_i$ could consist of all values of this variable, from $t = 1$ to $t = T$, i.e $\mathbf{w}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})'$. Then, if the variable is strictly exogenous with respect to ε_{it}, $\mathbf{S}_t$ could be the identity matrix $\mathbf{I}_T$ at each t. If the variable is only weakly exogenous then the selector matrix for each t would pick out values dated t and earlier.

Assumption 2. $\boldsymbol{x}_{it}$, $\mathbf{w}_i$, $\boldsymbol{\lambda}_i$ and ε_{it} are independently and identically distributed (i.i.d.) across i, with finite moments up to fourth order; furthermore, $E(\mathbf{w}_i\boldsymbol{\lambda}_i') = \mathbf{G}$

is a full column rank matrix and $\boldsymbol{\Sigma}_{\lambda} = E(\boldsymbol{\lambda}_i \boldsymbol{\lambda}_i')$ has rank $n_0 < \varpi(c)$, where n_0 denotes the true number of factors and $\varpi(c)$ is an upper bound that depends on the number of moment conditions available.

Assumption 2 is very similar to assumptions BA.1, BA.3 and BA.4 in Ahn, Lee and Schmidt (2013).[2] The i.i.d. assumption can be relaxed at the expense of considerable notational complexity. For example, ε_{it} could be heterogeneously distributed across both i and t. Conditional moments of $\boldsymbol{\lambda}_i$ could also be dependent on i. We avoid such generalisations only to simplify notation. Serial correlation in ε_{it} can also be allowed for by modifying $\mathbf{S}_t$, depending on the structure of the model. The remaining conditions assert that (i) every entry in $\boldsymbol{\lambda}_i$ is correlated with some or all variables in $\mathbf{w}_i$; (ii) the factors are non-degenerate; and (iii) the true number of factors cannot be too large to be able to identify the model. The full rank assumption on $\mathbf{G}$ is in the spirit of the fixed effects model and implies that n_0 in this paper is defined as the number of factors correlated with some entries in $\mathbf{w}_i$. All other factors can be absorbed into the error term, since the variance-covariance matrix of $\boldsymbol{\varepsilon}_i$ is left unrestricted. The rank assumption on $\boldsymbol{\Sigma}_{\lambda}$ implies that for the case where all regressors are strictly (treated as weakly) exogenous and there are no extraneous instruments available, $n_0 < T$ ($n_0 < (T+1)/2$). Essentially, this ensures that the degree of freedom of the model takes a non-negative value.

Let the matrix of instruments, $\mathbf{Z}_i$, be defined as

$$\mathbf{Z}_i' = \begin{bmatrix} \mathbf{w}_{i1} & 0 & .. & 0 \\ 0 & \mathbf{w}_{i2} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & .. & \mathbf{w}_{iT} \end{bmatrix}, \tag{2.4}$$

such that

$$E(\mathbf{Z}_i' \boldsymbol{\varepsilon}_i) = \mathbf{0}, \tag{2.5}$$

[2] Notice that the two approaches are fundamentally different, however. The methodology proposed by Ahn, Lee and Schmidt involves some form of quasi-differencing, which eliminates the incidental parameters, $\boldsymbol{\lambda}_i$, from the error term. On the other hand, as it will be shown shortly, our approach requires that certain functions of the incidental parameters are spanned by a finite set of parameters, which can then be estimated consistently. Essentially, different sensitivities to the factors (i.e. differences in the factor loadings) can be generated by different values of the variance of the cross sectional distribution of $\boldsymbol{\lambda}_i$."

where $\mathbf{Z}_i'$ is $c \times T$. In view of (2.3), the matrix of instruments can be written as

$$\mathbf{Z}_i' = \mathbf{S}(\mathbf{I}_T \otimes \mathbf{w}_i), \tag{2.6}$$

where

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & 0 & .. & 0 \\ 0 & \mathbf{S}_2 & .. & 0 \\ : & : & : & : \\ 0 & 0 & .. & \mathbf{S}_T \end{bmatrix}. \tag{2.7}$$

The matrix $\mathbf{S}$ has dimension $c \times Td$. The vector of orthogonality conditions we use to estimate the model parameters is

$$E[\mathbf{Z}_i'\mathbf{X}_i\boldsymbol{\phi} - \mathbf{Z}_i'(\mathbf{I}_T \otimes \boldsymbol{\lambda}_i')\mathbf{f} - \mathbf{Z}_i'\boldsymbol{\varepsilon}_i] = \mathbf{0}, \tag{2.8}$$

which, by use of (2.5) and (2.6), can be written as follows:

$$\mathbf{M}\boldsymbol{\phi} - \mathbf{S}(\mathbf{I}_T \otimes \mathbf{G})\mathbf{f} = \mathbf{0}, \tag{2.9}$$

where $\mathbf{M} = E(\mathbf{Z}_i'\mathbf{X}_i)$ and $\mathbf{G}$ is defined in Assumption 2. Matrices $\mathbf{M}$ and $\mathbf{G}$ have dimensions $c \times (K+1)$ and $d \times n$, respectively. Alternative forms of the second term in (2.9) are

$$\mathbf{S}(\mathbf{I}_T \otimes \mathbf{G})\mathbf{f} = \mathbf{S}\text{vec}(\mathbf{G}\mathbf{F}') = \mathbf{S}(\mathbf{F} \otimes \mathbf{I}_d)\mathbf{g}, \tag{2.10}$$

where $\mathbf{g} = \text{vec}(\mathbf{G})$. A compact expression of the orthogonality conditions is thus

$$\mathbf{M}\boldsymbol{\phi} - \mathbf{S}\text{vec}(\mathbf{G}\mathbf{F}') = \mathbf{0}. \tag{2.11}$$

When the instruments consist of current and all lagged values: the "canonical case" As an example, consider the case where all instruments available can be naturally arranged in a $T \times p$ matrix $\mathbf{V}_i$ of T observations on p variables (so that $\mathbf{w}_i = \text{vec}(\mathbf{V}_i)$), and ε_{it} is orthogonal to the block of potential instruments from $s = 1$ to $s = t$, i.e. the orthogonality conditions are given by

$$E(\mathbf{z}_{is}\varepsilon_{it}) = \mathbf{0},\ t = 1, ..., T,\ s = 1, ..., t, \tag{2.12}$$

where $\mathbf{z}_{is}'$ is the s^{th} row of $\mathbf{V}_i$. This can be viewed as a "canonical case" in the sense that there exists a collection of contemporaneous instruments and their lagged values; for instance, it arises when all variables in the model are (treated as) weakly exogenous, such as in the AR(1) dynamic panel data model with factor residuals, in which case $p = 1$. Define $\mathbf{M}_{st} = E(\mathbf{z}_{is}\mathbf{x}'_{it})$ and $\mathbf{G}_s = E(\mathbf{z}_{is}\boldsymbol{\lambda}'_i)$, which have dimensions $p \times (K+1)$ and $p \times n$, respectively. The orthogonality conditions are given by

$$\mathbf{M}_{st}\boldsymbol{\phi} - \mathbf{G}_s\mathbf{f}_t = \mathbf{0}, \ \ t = 1, ..., T, \ \ s = 1, ..., t. \tag{2.13}$$

These moments can be stacked as

$$\begin{bmatrix} \mathbf{M}_{11}\boldsymbol{\phi} \\ \mathbf{M}_{12}\boldsymbol{\phi} \\ \mathbf{M}_{22}\boldsymbol{\phi} \\ \vdots \\ \mathbf{M}_{1T}\boldsymbol{\phi} \\ \mathbf{M}_{2T}\boldsymbol{\phi} \\ \vdots \\ \mathbf{M}_{TT}\boldsymbol{\phi} \end{bmatrix} - \begin{bmatrix} \mathbf{G}_1\mathbf{f}_1 \\ \mathbf{G}_1\mathbf{f}_2 \\ \mathbf{G}_2\mathbf{f}_2 \\ \vdots \\ \mathbf{G}_1\mathbf{f}_T \\ \mathbf{G}_2\mathbf{f}_T \\ \vdots \\ \mathbf{G}_T\mathbf{f}_T \end{bmatrix} = \mathbf{0}. \tag{2.14}$$

More succinctly, we have

$$\mathbf{M}\boldsymbol{\phi} - \text{vech}(\mathbf{GF}') = \mathbf{0}, \tag{2.15}$$

where $\mathbf{M}$ is the stacked $\mathbf{M}_{st}$ terms and the vech operator is understood to act on $p \times 1$ submatrices. Let $\widetilde{\mathbf{S}}_T$ be the selector matrix of 0's and 1's that turns vec into vech, acting on $T \times T$ matrices. Then

$$\mathbf{M}\boldsymbol{\phi} - \text{vech}(\mathbf{GF}') = \mathbf{M}\boldsymbol{\phi} - (\widetilde{\mathbf{S}}_T \otimes \mathbf{I}_p)\text{vec}(\mathbf{GF}') = \mathbf{0}, \tag{2.16}$$

which is of the form of (2.11), with the selector matrix $\mathbf{S}$ given by $\mathbf{S} = \widetilde{\mathbf{S}}_T \otimes \mathbf{I}_p$.

3 The unrestricted estimator FIVU

Define the following moment function:

$$\boldsymbol{\psi}(\mathbf{Z}'_i\mathbf{X}_i; \boldsymbol{\theta}) = \mathbf{Z}'_i\mathbf{X}_i\boldsymbol{\phi}(\boldsymbol{\beta}) - \mathbf{S}\text{vec}(\mathbf{GF}'), \tag{3.1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{g}', \mathbf{f}')' \in \Omega$, and Ω is the full parameter space. Then by construction $E(\boldsymbol{\psi}(\mathbf{Z}'_i\mathbf{X}_i; \boldsymbol{\theta})) = \mathbf{0}$ at the true value $\boldsymbol{\theta}^0 = (\boldsymbol{\beta}^{0\prime}, \boldsymbol{g}^{0\prime}, \mathbf{f}^{0\prime})'$. However, as it stands $\boldsymbol{g}^0$ and $\mathbf{f}^0$ are not separately identified because

$$\mathbf{M}\boldsymbol{\phi}(\boldsymbol{\beta}) - \mathbf{S}\text{vec}(\mathbf{G}\mathbf{F}') = \mathbf{M}\boldsymbol{\phi}(\boldsymbol{\beta}) - \mathbf{S}\text{vec}(\mathbf{G}\mathbf{U}\mathbf{U}^{-1}\mathbf{F}'), \tag{3.2}$$

for any $n_0 \times n_0$ invertible matrix $\mathbf{U}$. When all regressors/instruments are strictly exogenous this particular indeterminancy can be eliminated by normalising an $n_0 \times n_0$ submatrix of $\mathbf{F}'$ to be some fixed invertible matrix, which is standard practice in factor models (see e.g. Bai and Ng, 2008). However, the aforementioned normalisation will not be sufficient for full identification in more general circumstances, as it will be made clear shortly. Thus, further normalisations may be required, which will vary depending upon the specification of the model and the number of factors. Let $\boldsymbol{\theta}_r = (\boldsymbol{\beta}', \mathbf{g}_r', \mathbf{f}_r')'$ denote the vector of the remaining free parameters corresponding to a particular set of normalisations on $\boldsymbol{\theta}$. Notice that the dimension of $\boldsymbol{\theta}_r$ is strictly smaller than that of $\boldsymbol{\theta}$ - in particular, $dim(\boldsymbol{\theta}) - dim(\boldsymbol{\theta}_r) \geq n_0^2$. In what follows we provide sufficient conditions for identification of the parameter vector $\boldsymbol{\theta}_r$ and then illustrate with examples.

Assumption 3. The true value of $\boldsymbol{\theta}_r$, denoted as $\boldsymbol{\theta}_0$, belongs to the interior of $\Theta_r \subseteq \Omega$ where Θ_r is obtainable by normalisations on the $\mathbf{G}, \mathbf{F}$ components of the vectors in Ω, together with some possible further restrictions excluding a closed set. Let $\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i; \boldsymbol{\theta}_r) := \boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i; \boldsymbol{\theta}_r, \boldsymbol{\theta} \setminus \boldsymbol{\theta}_r)$, where $\boldsymbol{\theta} \setminus \boldsymbol{\theta}_r$ is the part of $\boldsymbol{\theta}$ not in $\boldsymbol{\theta}_r$. We assume $\boldsymbol{\theta}_0$ is identified on Θ_r in the sense that $E(\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i; \boldsymbol{\theta}_r)) = \mathbf{0}$ for $\boldsymbol{\theta}_r \in \Theta_r$ implies $\boldsymbol{\theta}_r = \boldsymbol{\theta}_0$.

Let

$$\boldsymbol{\Gamma} = E\left(\left. \frac{\partial \boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i; \boldsymbol{\theta}_r)}{\partial \boldsymbol{\theta}_r'} \right|_{\boldsymbol{\theta}_r = \boldsymbol{\theta}_0} \right), \tag{3.3}$$

and

$$\boldsymbol{\Delta} = E\left(\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i; \boldsymbol{\theta}_0)\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i; \boldsymbol{\theta}_0)'\right). \tag{3.4}$$

Assumption 4. $\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$ exist and are full rank.

To see what these assumptions entail, consider the following model with a single regressor:

$$y_{it} = \beta x_{it} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{it}. \tag{3.5}$$

In the absence of extraneous instruments, the moment conditions can be obtained by multiplying (3.5) by x_{is} and taking expectations, which yields

$$E(x_{is}y_{it}) = \beta E(x_{is}x_{it}) + E(x_{is}\boldsymbol{\lambda}_i')\boldsymbol{f}_t + E(x_{is}\varepsilon_{it}),\ t = 1, \ldots, T.$$

Assuming that the instruments are valid, this can be written as

$$m_{st}^{xy} = \beta m_{st}^{xx} + \boldsymbol{g}_s'\boldsymbol{f}_t,\ t = 1, \ldots, T, \tag{3.6}$$

where $m_{st}^{xy} = E(x_{is}y_{it})$ and so on. Consider initially the case of a single factor i.e. $n_0 = 1$. The full parameter vector $\boldsymbol{\theta} \in \Omega$ is not identified at $\boldsymbol{\theta}_0$ so normalisations are required. Suppose that the regressor is strictly exogenous, i.e. $E(x_{is}\varepsilon_{it}) = 0$ for $s = 1, ..., T$. This implies that one can obtain T^2 estimating equations from (3.6). A normalisation that identifies this model is simply a rescaling of $\mathbf{g}$ and $\mathbf{f}$, obtained by setting one entry in $\mathbf{f}$ equal to 1, that is $f_\tau = 1$, for some $\tau \in \{1, ..., T\}$. The full rank assumption for $\boldsymbol{\Gamma}$ implies that at least one of the g's is nonzero. Thus, one may take $\Theta_r = \{(\boldsymbol{\theta}_r', f_\tau)';\ g_{\tau'} \neq 0,\ f_\tau = 1\}$ for some $\tau, \tau' \in \{1, ..., T\}$.

If the regressor is weakly exogenous, e.g. $x_{it} = y_{it-1}$, then given that $E(x_{is}\varepsilon_{it}) = 0$ for $s \leq t$ only, one can obtain $T(T+1)/2$ estimating equations from (3.6). In this case two columns of the matrix $\boldsymbol{\Gamma}$ consist of zeros except for a single entry that equals either g_1 or f_T. Thus, the full rank assumption for $\boldsymbol{\Gamma}$ implies that $g_1 \neq 0$ and $f_T \neq 0$, which is a stronger condition than in the strictly exogenous case. Hence here one may take $\Theta_r = \{(\boldsymbol{\theta}_r', f_T)';\ g_1 \neq 0,\ f_T = 1\}$. Notice that violations of the condition $g_1 \neq 0$ and $f_T \neq 0$ are testable. For example, $f_T = 0$ could be tested by imposing such restriction in estimation (which would involve dropping g_T from the objective function), normalising (say) $f_{T-1} = 1$ and examining the resulting overidentifying restrictions test statistic.[3]

Now consider the case where $n_0 = 2$ and x_{it} is weakly exogenous. One could impose the normalisation that the last n_0 columns of $\mathbf{F}'$ be some fixed invertible matrix. However, it turns out that further normalisations are required. To see this, notice that $\mathbf{G}_T$ in (2.14), which is of dimension 1×2 in this case, enters only in the last estimating equation but involves two unknown parameters, $g_T^{(1)}$ and $g_T^{(2)}$. These parameters can be identified up to a linear combination. Similarly $\mathbf{f}_1$, a 2×1 vector, enters only in the first estimating equation but involves two unknown parameters, $f_1^{(1)}$ and $f_1^{(2)}$. Thus for the first row of $\mathbf{F}$, $f_1^{(1)}$ and $f_1^{(2)}$ can be identified by the available moments up to a linear combination. Notice that the additional

[3] We would like to thank an anonymous referee for pointing this out.

normalisations required to identify the parameters do not affect the objective function and as a result they do not constitute overidentifying restrictions, i.e. they can be imposed without loss of generality.

Letting

$$\mathbf{F}^{+}=\begin{pmatrix} f_{T-1}^{(1)} & f_{T-1}^{(2)} \\ f_{T}^{(1)} & f_{T}^{(2)} \end{pmatrix}, \tag{3.7}$$

the full rank assumption for $\mathbf{\Gamma}$ implies here that $g_1^{(1)} \neq 0,\ g_1^{(2)} \neq 0,\ \det\left(\mathbf{F}^{+}\right) \neq 0$. As an example, normalising $\mathbf{F}^{+}$ to be the identity matrix, one may take $\Theta_r = \{(\boldsymbol{\theta}_r{}', (\boldsymbol{\theta} \setminus \boldsymbol{\theta}_r)')'; g_1^{(1)} \neq 0,\ g_1^{(2)} \neq 0,\ f_{T-1}^{(1)} = f_T^{(2)} = 1, f_T^{(1)} = f_{T-1}^{(2)} = 0\}$, where

$$\boldsymbol{\theta}_r = (\beta, g_1^{(1)}, \ldots, g_{T-1}^{(1)}, g_1^{(2)}, \ldots, g_T^{(2)}, f_2^{(1)}, \ldots, f_{T-2}^{(1)}, f_2^{(2)}, \ldots, f_{T-2}^{(2)}, f_1^{(1)}/f_1^{(2)})'. \tag{3.8}$$

Notice that this model with two factors is identified for $T \geq 5$. In particular, for $T = 5$ there are 15 moment conditions and whilst the full parameter vector $\boldsymbol{\theta}$ has dimension $dim(\boldsymbol{\theta}) = 1 + 4T = 21$, $dim(\boldsymbol{\theta}_r) = 15$; hence, the model is exactly identified. In other words, the condition $n_0 < (T+1)/2$ discussed below Assumption 2 is satisfied for $T \geq 5$.

The positive definiteness assumption for $\mathbf{\Delta}$ itself implies that $\boldsymbol{\theta}_0$ is locally identified. The above set of assumptions is sufficient to make an appeal to standard GMM theory in order to derive the asymptotic properties of FIVU. In our context the result is given in the following proposition:

Proposition 1. Distributional result for FIVU. *Let Θ_c be a compact subset of Θ_r containing $\boldsymbol{\theta}_0$ in its interior and let*

$$\widehat{\boldsymbol{\theta}}_r = \arg\min_{\boldsymbol{\theta}_r \in \Theta_c} \boldsymbol{\psi}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}_r)' \mathbf{C}_N \boldsymbol{\psi}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}_r), \tag{3.9}$$

where $\widehat{\mathbf{M}}_N = \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{X}_i / N$ and $\mathbf{C}_N$ is a given positive definite matrix. Then $\widehat{\boldsymbol{\theta}}_r$ converges in probability to $\boldsymbol{\theta}_0$ and

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, (\mathbf{\Gamma}'\mathbf{C}_N\mathbf{\Gamma})^{-1}(\mathbf{\Gamma}'\mathbf{C}_N\mathbf{\Delta}\mathbf{C}_N\mathbf{\Gamma})(\mathbf{\Gamma}'\mathbf{C}_N\mathbf{\Gamma})^{-1}). \tag{3.10}$$

Proof. This is straightforward enough; see e.g. Newey and McFadden (1994) for further details.[4] □

[4] It is easy to see that our assumptions imply the assumptions employed by Newey-McFadden, except perhaps for their assumption of dominance, i.e. the norm of the moment function is dominated by a function of $\widehat{\mathbf{M}}_N$ of finite expectation. In fact this follows easily in our case from compactness and the existence of second moments.

If $\mathbf{C}_N$ is chosen as $\boldsymbol{\Delta}^{-1}$ the covariance matrix of the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_r$ is $(\boldsymbol{\Gamma}'\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma})^{-1}$, in which case the estimator has certain optimality properties (Hansen, 1982). These distributional results hold as well if the unobserved $\boldsymbol{\Delta}$ is replaced by a consistent estimate, $\widehat{\boldsymbol{\Delta}}_N$.

Proposition 2. TESTING THE OVERIDENTIFYING RESTRICTIONS. *Under Assumptions 1-4 and for $n = n_0$, the minimised optimal GMM criterion scaled by N is asymptotically chi-square distributed with $\kappa = c - rank(\boldsymbol{\Gamma}) = c - dim(\boldsymbol{\theta}_r)$ degrees of freedom, such that*

$$J_N\left(\widehat{\dot{\boldsymbol{\theta}}}_r\right) = N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\dot{\boldsymbol{\theta}}}_r)'\widehat{\boldsymbol{\Delta}}_N^{-1}\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\dot{\boldsymbol{\theta}}}_r) \xrightarrow{d} \chi^2_\kappa, \tag{3.11}$$

where $\widehat{\dot{\boldsymbol{\theta}}}_r$ is the optimal GMM estimator.

Proof. cf. Proposition 1. □

Appendix II establishes a general identification scheme for FIVU under a multi-factor structure. In many circumstances, the full parameter vector is not the object of interest and one is interested only in estimating $\boldsymbol{\beta}$. In this case we show below that it is not essential to impose normalisations on the factors in FIVU estimation as the value of $\boldsymbol{\beta}$ obtained by unnormalised estimation (over Ω) will coincide with the normalised estimate (over Θ_c, which is defined in the distributional result for FIVU) under one further assumption.

ASSUMPTION 5. There exists an open set Θ, where $\Omega \supseteq \Theta \supseteq \Theta_r$ with Θ dense in Ω such that for all $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{g}', \mathbf{f}')' \in \Theta$

$$\mathbf{S}\text{vec}(\mathbf{GF}') = \mathbf{S}\text{vec}((\mathbf{GF}')_r), \tag{3.12}$$

for some $(\boldsymbol{\beta}', \mathbf{g}_r', \mathbf{f}_r')' \in \Theta_r$.[5] Assume as well that $\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta}_r)'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta}_r)$, $\boldsymbol{\theta}_r \in \Theta_r$, is bounded away from zero outside some given compact set.[6]

[5]Notice that $\mathbf{g}$ and $\mathbf{f}$ may be of larger dimension than $\mathbf{g}_r$, $\mathbf{f}_r$, which are obtained after imposing normalisations and can be used to compute $(\mathbf{GF}')_r$ together with the normalised entries of $\mathbf{g}$ and $\mathbf{f}$. $(\mathbf{GF}')_r$ is of the same dimension as $(\mathbf{GF}')$ of course.

[6]In the examples corresponding to equation (3.6), $\Theta_r \subset \Theta = \{\boldsymbol{\theta} = (\beta, g_1, \ldots, g_T, f_1, \ldots, f_T);\ g_{\tau'} \neq 0, f_\tau \neq 0\} \subset \Omega$ for the case of a strictly exogenous regressor, $\Theta_r \subset \Theta = \{\boldsymbol{\theta} = (\beta, g_1, \ldots, g_T, f_1, \ldots, f_T);\ g_1 \neq 0, f_T \neq 0\} \subset \Omega$ for the case of a weakly exogenous regressor and one factor, while $\Theta_r \subset \Theta = \{\boldsymbol{\theta} = (\beta, \boldsymbol{g}', \boldsymbol{f}')';\ g_1^1 \neq 0,\ g_1^2 \neq 0,\ det\,(\mathbf{F}^+) \neq 0\} \in \Omega$ for the case of a weakly exogenous regressor and two factors.

Theorem 3. Equivalence of normalised and unnormalised estimation. *Under Assumptions 1-5* $\widehat{\boldsymbol{\beta}}(\Omega) \to \widehat{\boldsymbol{\beta}}(\Theta_c)$ *in probability. Define* $\boldsymbol{\nu} = (\mathbf{g}', \mathbf{f}')'$ *and* $\boldsymbol{\nu}_r$ *as the subvector of free parameters in* $\boldsymbol{\nu}$*. If, moreover, at the true values of* $\boldsymbol{\nu}$ *and* $\boldsymbol{\nu}_r$

$$Span \frac{\partial \mathbf{S}vec(\mathbf{GF}')}{\partial \boldsymbol{\nu}'} = Span \frac{\partial \mathbf{S}vec((\mathbf{GF}')_r)}{\partial \boldsymbol{\nu}_r'}, \tag{3.13}$$

then the covariance matrix of $\widehat{\boldsymbol{\beta}}(\Omega)$ *obtained using the generalised inverse of* $(\partial \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right)/\partial \boldsymbol{\theta}')'\mathbf{C}_N \partial \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right)/\partial \boldsymbol{\theta}'$ *coincides with the covariance matrix of* $\widehat{\boldsymbol{\beta}}(\Theta_r)$ *inferred from the inverse of* $\left(\partial \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}_r\right)/\partial \boldsymbol{\theta}_r'\right)'\mathbf{C}_N \partial \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}_r\right)/\partial \boldsymbol{\theta}_r'$.

Proof. See Appendix I. □

Equation (3.13) ensures that the submatrix of the covariance matrix of $\hat{\boldsymbol{\theta}}$ corresponding to the parameters of interest has not been altered by the normalisations imposed on $\mathbf{G}$ and $\mathbf{F}$. Thus, under the conditions of Theorem 3 the distribution of $\hat{\boldsymbol{\beta}}$ obtained from estimation subject to a set of normalisations on $\mathbf{G}$ and $\mathbf{F}$ coincides with that obtained from optimisation without imposing these normalisations. If the normalisations constitute a set of identifying restrictions, Proposition 1 tells us that the distribution of $\hat{\boldsymbol{\beta}}$ over the restricted parameter space (and hence in this case the distribution also of $\hat{\boldsymbol{\beta}}$ without normalisations) is that given by Newey and McFadden (1994).

Essentially, Assumption 5 and the spanning condition (3.13) ensure that if *genuine* restrictions are required to identify the model (e.g. for the weakly exogenous case with $f_T = 0$, one would impose such restriction and drop g_T from the objective function), then the same restrictions are imposed on the unnormalised model.

In Appendix II we demonstrate that Assumptions 1-5 and condition (3.13) are satisfied under the identification scheme proposed for the AR(1) one-factor model, so that FIVU can be implemented for this model without normalisations.

Estimation for FIVU

The FIVU estimator is straightforward to obtain. Let $\mathbf{B}_N$ be a square-root matrix of $\mathbf{C}_N$. Then the objective function has the form

$$Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}) = \left\| \mathbf{B}_N \boldsymbol{\psi}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}) \right\|^2 = \left\| \mathbf{B}_N [\widehat{\mathbf{M}}_N \boldsymbol{\phi}(\boldsymbol{\beta}) - \mathbf{S}\text{vec}(\mathbf{GF}')] \right\|^2. \tag{3.14}$$

Since $\boldsymbol{\phi}$ is a linear function of $\boldsymbol{\beta}$, if either $\mathbf{G}$ or $\mathbf{F}$ is held fixed, the expression $\mathbf{B}_N[\widehat{\mathbf{M}}_N\boldsymbol{\beta}(\boldsymbol{\phi}) - \mathbf{S}\text{vec}(\mathbf{G}\mathbf{F}')]$ is a linear function of the remaining parameters and the conditional minimum of (3.14) may be found by a one pass least squares procedure. One may then seek a joint minimum by iteration over $\mathbf{G}$ and $\mathbf{F}$. This appears to work well in practice. In Appendix III we obtain first and second derivatives for the RHS in (3.14), so Gauss-Newton procedures are also available.

Equation (2.11) takes a particularly simple form when $f_t \equiv 1$ for all t, as in the one way error components model. In particular, one has

$$\mathbf{S}\text{vec}(\mathbf{G}\mathbf{F}') = \mathbf{S}(\boldsymbol{\iota}_T \otimes \mathbf{I}_d)\mathbf{g}. \tag{3.15}$$

Therefore one obtains based on (3.14)

$$\mathbf{B}_N\mathbf{M}\boldsymbol{\phi}(\boldsymbol{\beta}) - \mathbf{B}_N\mathbf{S}(\boldsymbol{\iota}_T \otimes \mathbf{I}_d)\boldsymbol{g} = \mathbf{0}, \tag{3.16}$$

which can be interpreted as a classical regression when $\mathbf{M}$ is replaced by its sample counterpart. Since $\boldsymbol{\phi}$ is a linear function of $\boldsymbol{\beta}$, FIVU may be obtained by a one pass least squares estimate of (3.16).

Quasi-differencing

An alternative approach to FIVU is obtained by quasi-differencing, which removes the factor component in (2.11). This is achieved by constructing a matrix $\mathbf{D} = \mathbf{D}(\mathbf{F})$ such that $\mathbf{D}(\mathbf{F})\mathbf{S}\text{vec}(\mathbf{G}\mathbf{F}') = \mathbf{0}$. The orthogonality conditions then become

$$\mathbf{D}(\mathbf{F})\mathbf{M}\boldsymbol{\phi}(\boldsymbol{\beta}) = \mathbf{0}. \tag{3.17}$$

Quasi-differencing is the method employed by Holtz-Eakin, Newey and Rosen (1988), Ahn, Lee and Schmidt (2001), Nauges and Thomas (2003) for the one factor case, and Ahn, Lee and Schmidt (2013) for the multi-factor case. In general, this approach eliminates the factor component from the error at the same cost in moment conditions. As shown in Appendix I, such transformations of moment conditions produce estimators of the same asymptotic efficiency as working with the untransformed moment conditions. This result is summarised in the following theorem:

Theorem 4. ASYMPTOTIC EQUIVALENCE RESULT. *Under Assumptions 1-4 FIVU in model (2.1) is asymptotically equivalent to a Generalised Method of Moments estimator based on quasi-differencing and upon constructing* $\mathbf{D}(\mathbf{F})$.

Proof. See Appendix I. □

To see this intuitively, consider without loss of generality a static model with a single strictly exogenous instrument, z_{it}, and a single factor structure.

$$y_{it} = \beta x_{it} + \lambda_i f_t + \varepsilon_{it,} \qquad t = 1, ..., T, \tag{3.18}$$

such that $E(z_{is}\varepsilon_{it}) = 0$ for $s = 1, ..., T$. The quasi-differencing procedure proposed by Holtz-Eakin, Newey and Rosen (1988) and adopted by Nauges and Thomas (2003) transforms the model as

$$\begin{aligned} y_{it} - r_t y_{it-1} &= \beta(x_{it} - r_t x_{it-1}) + \lambda_i(f_t - r_t f_{t-1}) + (\varepsilon_{it} - r_t \varepsilon_{it-1}) \\ &= \beta(x_{it} - r_t x_{it-1}) + (\varepsilon_{it} - r_t \varepsilon_{it-1}),\ t = 2, \ldots T, \end{aligned} \tag{3.19}$$

where $r_t = f_t/f_{t-1}$. Thus there exist $T(T-1)$ moment conditions and $1 + T - 1$ parameters to estimate (β and $T - 1$ r's). The procedure proposed by Ahn, Lee and Schmidt (2013) involves transforming the model as follows:

$$\begin{aligned} y_{it} - \widetilde{f}_t y_{iT} &= \beta(x_{it} - \widetilde{f}_t x_{iT}) + \lambda_i(f_t - \widetilde{f}_t f_T) + (\varepsilon_{it} - \widetilde{f}_t \varepsilon_{iT}) \\ &= \beta(x_{it} - \widetilde{f}_t x_{iT}) + (\varepsilon_{it} - \widetilde{f}_t \varepsilon_{iT}), \end{aligned} \tag{3.20}$$

where $\widetilde{f}_t$ is the normalised value of f_t such that $\widetilde{f}_t$=f_t/f_T with $\widetilde{f}_T = 1$. Again, this provides $T(T-1)$ moment conditions and requires estimating $1 + T - 1$ parameters. FIVU does not quasi-difference the model and as such it will use $T^2 = T(T-1)+T$ moment conditions at the expense of introducing T extra parameters (the g's) . The net difference between the number of moment conditions and parameters across all these methods is the same. Hence the resulting estimators are asymptoticallly equivalent.

Remark. Notice that r_t exists only if $f_t \neq 0$ for all t. In practice, this implies that the quasi-differencing procedure outlined in equation (3.19) might face computational problems if some of the factor values are close to zero. Similarly, Kruiniger (2008, pg. 16) points out that the normalisation $\widetilde{f}_t$=f_t/f_T can be restrictive as it requires that f_T be sufficiently far from zero. This issue is also discussed in an extensive simulation study by Juodis and Sarafidis (2014). For the case where the instruments are strictly exogenous as in the example above, Ahn, Lee and Schmidt (2013, Appendix A) analyse a continuous-updating type GMM estimator that requires conditional homoskedasticity in ε_{it} for asymptotic efficiency. Their estimator solves an eigenvalue problem yielding an estimate of the whole (unnormalised) quasi-differenced transformation matrix, $\widehat{\mathbf{D}}$; as a result, the estimator is invariant to the normalisation of the factors. Similarly, our FIVU approach can be implemented without imposing normalisations on the factors, and in addition it does not require conditional homoskedasticity in ε_{it}. We do need to assume the model is identified to invoke general GMM results; however, as Theorem 3 makes clear, any feasible identification scheme will suffice for this purpose.

Notice that it is also possible to construct a matrix $\mathbf{D} = \mathbf{D}(\mathbf{G})$ to eliminate the g terms. To see how this can be achieved, assume a single factor and consider the column vector $\mathbf{S}\text{vec}(\mathbf{gf}')$, consisting of scalar terms of the form $g_s f_t$. Consider the following operations on $\mathbf{S}\text{vec}(\mathbf{gf}')$:

1. Transform $\mathbf{S}\text{vec}(\mathbf{gf}')$ so that all coefficients of terms in the scalar g_1 are unity.

2. Choose one of the g_1 terms and use it to difference away the rest.

3. Eliminate the (single) remaining term in g_1.

One now repeats these operations for the remaining g's. The key point is that all these operations can be accomplished by left multiplication on $\mathbf{S}\text{vec}(\mathbf{gf}')$ by matrices of the form $\mathbf{D}(\mathbf{G})$. Where there is more than one factor, $\text{vec}(\mathbf{GF}')$ consists of sums of terms of the form $\text{vec}(\mathbf{gf}')$. Since the above operations preserve the structure of these terms, the operations may be applied sequentially to the later terms to eliminate them in their turn. Similarly as before, this approach eliminates dn_0 parameters (the g's) at the same cost in moment conditions and so there is no asymptotic efficiency gain/loss over the aforementioned methods.

Remark. The extension of our approach to unbalanced panels is a trivial exercise. This is not necessarily the case for procedures that involve some form of quasi-differencing. To see this, notice that the transformation in equation (3.20) that removes the factor component from the error is feasible only if the last period ($t = T$) is available for all individuals. Otherwise, the individuals for which the last observation is missing need to be dropped out altogether, or they could be used in a separate sub-sample that involves normalizing based on an earlier time period. Both alternatives are likely to result in a substantial loss in efficiency. This issue is discussed in Juodis and Sarafidis (2014).

Estimation of the number of factors

The true number of factors, n_0, is typically unknown in empirical applications. This quantity can be determined using a model information criterion, as in Proposition 3 of Ahn, Lee and Schmidt (2013). For example, the Schwarz Criterion (BIC) is of the form

$$S_N(n) = N \times Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n) - ln(N) \times h(n), \tag{3.21}$$

where $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n)$ is the value of the objective function evaluated at $\widehat{\boldsymbol{\theta}}_r$ using n factors, while $h(n) = \varrho \times \kappa(n) = O(1)$, a strictly increasing function of n with $0 < \varrho < \infty$ and $\kappa(n) = c - dim(\widehat{\boldsymbol{\theta}}_r)$. Observe that (i) $lim_{N\to\infty} N^{-1} ln(N) = 0$ and (ii) $lim_{N\to\infty} ln(N) = \infty$; the first condition prevents underfitting while the second condition prevents estimating too many factors asymptotically. Using the same line of arguments as in page 6 of Ahn, Lee and Schmidt, it is straightforward to show that (3.21) is consistent under our assumptions, i.e. $\hat{n} \xrightarrow{p} n_0$ as N $\to \infty$. In particular, consider initially the case where $\hat{n} > n_0$. We have

$$\begin{aligned} & Pr[S_N(n_0) - S_N(\hat{n}) > 0] \\ = & Pr\left[N\left(Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n_0) - Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|\hat{n})\right) - ln(N) \times (h(\hat{n}) - h(n_0)) > 0\right] \\ \leq & Pr\left[N \times Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n_0) - ln(N) \times (h(\hat{n}) - h(n_0)) > 0\right] \\ & \to 0, \text{ as } N \to \infty, \end{aligned} \tag{3.22}$$

since $N \times Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n_0)$ is $O_p(1)$, $h(\hat{n}) - h(n_0) > 0$ and $ln(N) \to \infty$.

For the case where $\hat{n} < n_0$ we have

$$\begin{aligned} & Pr[(S_N(n_0) - S_N(\hat{n})) > 0] \\ = & Pr[(Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n_0) - Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|\hat{n})) - N^{-1} ln(N) \times (h(\hat{n}) - h(n_0)) > 0] \\ & \to 0, \text{ as } N \to \infty, \end{aligned} \tag{3.23}$$

because both $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n_0) \xrightarrow{p} 0$ and $N^{-1} ln(N) \to 0$, while $N^{-1} Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|\hat{n}) \xrightarrow{p} \varpi > 0$ due to the identification assumption.

Notice that it is not necessary to use the optimal estimator, $\dot{\widehat{\boldsymbol{\theta}}}_r$, to derive this result because when a sub-optimal weighting matrix is used, then $N \times Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n_0)$ is asymptotically a weighted average of independent chi-square random variables.

The BIC criterion remains consistent when one uses the FIVU objective function without normalisations, i.e. based on $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}|n)$, as in (3.14). This is stated explicitly in the following theorem.

Theorem 5. Consistency of BIC without normalisations. *Under Assumptions 1-5, the BIC criterion in (3.21) with* $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}_r|n)$ *replaced by* $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \widehat{\boldsymbol{\theta}}|n)$

is consistent, i.e. $\hat{n} \xrightarrow{p} n_0$ *as* $N \to \infty$.

Proof. See Appendix I. □

We will investigate the finite sample performance of BIC defined in Theorem 5 in the Monte Carlo section of the paper.

4 Parameter restrictions: the FIVR estimator

When elements of $\mathbf{x}_{it}$ occur as instruments, model (2.1) implies restrictions on $\mathbf{G}$, the imposition of which will lead to greater efficiency. These restrictions require the following assumption:

ASSUMPTION 6. $E(\boldsymbol{\lambda}_i \varepsilon_{it}) = \mathbf{0}$, for all i and t.

The extra restrictions can be obtained by pre-multiplying (2.1) by $\boldsymbol{\lambda}_i$ and taking expectations, which yields (at $n = n_0$)

$$E(\boldsymbol{\lambda}_i \mathbf{x}'_{it})\boldsymbol{\phi} = \boldsymbol{\Sigma}_\lambda \mathbf{f}_t, \quad t = 1, ..., T. \tag{4.1}$$

The key point is that, when the instrument set includes elements of $\mathbf{x}_{it}$, the entries in $E(\boldsymbol{\lambda}_i \mathbf{x}'_{it})$ include terms in various of the g's so that the LHS of (4.1) is a linear function of the ensemble vector $\mathbf{g}$.

The restrictions take the matrix form

$$\mathbf{H}(\boldsymbol{\phi})\mathbf{P}_{d,n}\mathbf{g} = (\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\Lambda)\mathbf{f} + \mathbf{U}\boldsymbol{\delta}, \tag{4.2}$$

where $\mathbf{H}(\boldsymbol{\phi})$ is an $nT \times nd$ matrix that depends on the structure of the model, $\mathbf{P}_{m,n}$ is the permutation matrix such that $\mathbf{P}_{m,n}\text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}')$ for $m \times n$ matrices $\mathbf{A}$, $\mathbf{U}$ is a matrix of elementary column vectors and $\boldsymbol{\delta}$ is a vector of free parameters.

The FIVR estimator (restricted FIV estimator) chooses $\boldsymbol{\theta}$ to minimise the FIVU objective function subject to (4.2). Thus, FIVR will in general have fewer parameters to estimate than FIVU and as such it will be more efficient. The specific form of (4.2) is illustrated in several examples below.

Example 1. One lagged dependent variable and a single factor The model is

$$y_{it} = \beta y_{it-1} + \lambda_i f_t + \varepsilon_{it}. \tag{4.3}$$

Here $\mathbf{x}_{it} = (y_{it}, y_{it-1})'$, $\boldsymbol{\phi} = (1, -\beta)'$, $z_{it} = y_{it-1}$, $g_s = E(y_{is-1}\lambda_i)$. The linear restrictions in (4.1) take the form

$$g_{s+1} = \beta g_s + \sigma_\lambda^2 f_s, \tag{4.4}$$

where $\sigma_\lambda^2 = E(\lambda_i^2)$, which can be written in matrix form as

$$\begin{bmatrix} -\beta & 1 & 0 & .. & 0 \\ 0 & -\beta & & : & 0 \\ : & : & : & 1 & : \\ 0 & 0 & .. & -\beta & 1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ : \\ g_{T+1} \end{bmatrix} = \sigma_\lambda^2 \mathbf{f}. \tag{4.5}$$

Notice the appearance of the "out-of-sample" term g_{T+1}, which we regard as a constant to be estimated.[7] One can section this matrix equation into the form

$$\begin{bmatrix} \mathbf{H}(\boldsymbol{\phi}) & \mathbf{e}_T \end{bmatrix} \begin{bmatrix} \mathbf{g} \\ g_{T+1} \end{bmatrix} = \sigma_\lambda^2 \mathbf{f}, \tag{4.6}$$

where $\mathbf{g} = (g_1, ..., g_N)'$, $\mathbf{e}_T$ is the $T \times 1$ dimensional column vector with 1 in the T^{th} position and

$$\mathbf{H}(\boldsymbol{\phi}) = \begin{bmatrix} -\beta & 1 & 0 & .. \\ 0 & -\beta & & : \\ : & : & : & 1 \\ 0 & 0 & .. & -\beta \end{bmatrix}. \tag{4.7}$$

Thus, the restriction has the form

$$\mathbf{H}(\boldsymbol{\phi})\mathbf{g} = \sigma_\lambda^2 \mathbf{f} + \delta \mathbf{e}_T \qquad (\delta \in \mathbb{R}). \tag{4.8}$$

Remark. In the case of the one way error components model where $f_t = 1$ for $t = 1, ..., T$, the set of linear restrictions in (4.4) becomes

$$g_{s+1} = \beta g_s + \sigma_\lambda^2. \tag{4.9}$$

In this case, FIVR utilises the same set of orthogonality conditions as FIVU, $T(T+1)/2$ in total, but estimates only three parameters, namely β, g_1 and σ^2. Therefore, FIVR makes efficient use of second moment information and intuitively we should expect that it is asymptotically equivalent to the GMM estimator pro-

[7] Strictly speaking, the value of g_{T+1} is *defined* by the restriction it appears in (4.4). We adopt this convention so as to have a neat formula for the full vector f.

posed by Ahn and Schmidt (1995). Under stationary initial conditions there is an extra restriction in that $g_1 = \sigma_\lambda^2/(1-\beta)$. In this case the number of estimable parameters decreases by one and a version of FIVR that uses this extra restriction is asymptotically equivalent to the system GMM estimator proposed by Arellano and Bover (1995) and Blundell and Bond (1998).

Example 2. One lagged dependent variable and two factors. In this case $\mathbf{g}_s = E(y_{is-1}\boldsymbol{\lambda}_i')$ is a 1×2 row vector and the restrictions have the form $\mathbf{g}_{s+1}' = \beta\mathbf{g}_s' + \boldsymbol{\Sigma}_\lambda\mathbf{f}_s$. The matrix of restrictions is as in Example 1 except that $\mathbf{g}$ is replaced by $\text{vec}(\mathbf{G}')$ and $\boldsymbol{\delta} \in \mathbb{R}^2$. Therefore, we have

$$(\mathbf{H}(\boldsymbol{\phi}) \otimes \mathbf{I}_2)\mathbf{P}_{T,2}\mathbf{g} = (\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\lambda)\mathbf{f} + \mathbf{U}\boldsymbol{\delta}, \tag{4.10}$$

where $\mathbf{g}$ is a $2T \times 1$ vector and $\mathbf{U} = [\mathbf{e}_{2T-1}, \mathbf{e}_{2T}]$, a $2T \times 2$ matrix.

Example 3. One lagged dependent variable, one weakly exogenous covariate and one factor. The model is

$$y_{it} = \beta_1 y_{it-1} + \beta_2 x_{it} + \lambda_i f_t + \varepsilon_{it}. \tag{4.11}$$

In this case the instrument vector is $\mathbf{z}_{it} = (y_{it-1}, x_{it})'$. Note the g's are two-dimensional:

$$\mathbf{g}_s = (\ g_s^1,\ \ g_s^2\ )' = E[(\ y_{is-1}\lambda_i,\ \ x_{is}\lambda_i\ )']. \tag{4.12}$$

The restrictions are given by $\mathrm{g}_{s+1}^1 = \beta_1\mathrm{g}_s^1 + \beta_2\mathrm{g}_s^2 + \sigma_\lambda^2 f_s$. In matrix form we have

$$\begin{bmatrix} -\beta_1 & -\beta_2 & 1 & 0 & 0 & .. & 0 \\ 0 & 0 & -\beta_1 & -\beta_2 & 1 & .. & 0 \\ : & : & : & : & : & : & : \\ 0 & 0 & 0 & .. & -\beta_1 & -\beta_2 & 1 \end{bmatrix} \begin{bmatrix} g_1^1 \\ g_1^2 \\ : \\ g_T^1 \\ g_T^2 \\ g_{T+1}^1 \end{bmatrix} = \sigma_\lambda^2\mathbf{f}, \tag{4.13}$$

which can be written more generally as

$$\mathbf{H}(\boldsymbol{\phi})\mathbf{g} = \sigma_\lambda^2\mathbf{f} + \delta\mathbf{e}_{T,} \qquad \delta \in \mathbb{R}. \tag{4.14}$$

where

$$
\mathbf{H}(\boldsymbol{\phi}) = \begin{bmatrix} -\beta_1 & -\beta_2 & 1 & 0 & 0 & .. \\ 0 & 0 & -\beta_1 & -\beta^2 & 1 & .. \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & .. & -\beta_1 & -\beta^2 \end{bmatrix}, \tag{4.15}
$$

a $T \times 2T$ matrix.

Example 4. Two lagged dependent variables and one factor. The model is

$$
y_{it} = \beta_1 y_{it-1} + \beta_2 y_{it-2} + \lambda_i f_t + \varepsilon_{it}. \tag{4.16}
$$

In this case $\mathbf{w}_i = (y_{i0}, \ldots, y_{iT-1})'$, $z_{it} = y_{it-1}$ and the matrix of restrictions takes the form

$$
\begin{bmatrix} -\beta_2 & -\beta_1 & 1 & 0 & \cdots & 0 \\ 0 & -\beta_2 & -\beta_1 & 1 & \ldots & \vdots \\ 0 & 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & -\beta_2 & -\beta_1 & 1 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \\ g_T \\ g_{T+1} \end{bmatrix} = \sigma_\lambda^2 \mathbf{f}. \tag{4.17}
$$

This is partitioned conformably into

$$
\begin{bmatrix} -\beta_2 \mathbf{e}_1 & \mathbf{H}(\boldsymbol{\phi}) & \mathbf{e}_T \end{bmatrix} \begin{bmatrix} g_0 \\ \mathbf{g} \\ g_{T+1} \end{bmatrix} = \sigma_\lambda^2 \mathbf{f}, \tag{4.18}
$$

where

$$
\mathbf{H}(\boldsymbol{\phi}) = \begin{bmatrix} -\beta_1 & 1 & 0 & \cdots \\ -\beta_2 & -\beta_1 & 1 & \ldots \\ 0 & 0 & \ddots & \ddots \\ \vdots & \vdots & -\beta_2 & -\beta_1 \end{bmatrix}, \tag{4.19}
$$

with solution

$$
\mathbf{H}(\boldsymbol{\phi})\mathbf{g} = \sigma_\lambda^2 \mathbf{f} + \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_T \end{bmatrix} \boldsymbol{\delta} \qquad (\boldsymbol{\delta} \in \mathbb{R}^2). \tag{4.20}
$$

Identification and Estimation for FIVR One does not need to develop a separate theory of identification for FIVR; this can be inferred from the FIVU results.

If Assumptions 1-5 hold, and given the equivalence of normalised and unnormalised estimation, then the FIVU estimator may be obtained by minimising the criterion function over the whole parameter space. FIVR minimises the criterion over a closed neighbourhood of the parameter space and this implies straightforwardly that the FIVR estimates of the parameters of interest likewise have probability limit $\boldsymbol{\beta}^0$. Since FIVR is obtained by expressing some of the nuisance parameters in terms of the remaining parameters, its covariance matrix may be obtained from the FIVU matrix by application of the appropriate Jacobian (calculated in Appendix III). Of course, FIVR will be feasible in cases where FIVU is not, since FIVR estimates fewer parameters.

The standard method of solving a minimisation problem subject to an exact constraint is to use the constraint to solve out for some of the choice variables and substitute into the minimand. In particular, since $\mathbf{H}(\boldsymbol{\phi}) = \sum_{k=1}^{K+1} \mathbf{K}_k \boldsymbol{\phi}_k = \mathbf{K}(\boldsymbol{\phi} \otimes \mathbf{I}_{nd})$, where $\mathbf{K} = [\mathbf{K}_1, ..., \mathbf{K}_{K+1}]$ with $\mathbf{K}_k$ being a fixed $nT \times nd$ matrix that depends on the structure of the model, we have

$$\mathbf{H}(\boldsymbol{\phi})\mathbf{P}_{d,n}\mathbf{g} = \mathbf{K}(\mathbf{I}_{K+1} \otimes \mathbf{P}_{d,n}\mathbf{g})\boldsymbol{\phi}, \tag{4.21}$$

which is a linear function of $\boldsymbol{\phi}$ given $\mathbf{g}$. Furthermore, one can write

$$(\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\lambda)\mathbf{f} = (\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\lambda)(\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\lambda^{-1})(\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\lambda)\mathbf{f} = (\mathbf{I}_T \otimes \mathbf{I}_n)\tilde{\mathbf{f}}, \tag{4.22}$$

where $\tilde{\mathbf{f}} = (\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\lambda)\mathbf{f}$. Thus, the restrictions in (4.2) can be expressed as

$$\mathbf{K}(\mathbf{I}_{K+1} \otimes \mathbf{P}_{d,n}\mathbf{g})\boldsymbol{\phi} = \tilde{\mathbf{f}} + \mathbf{U}\boldsymbol{\delta}. \tag{4.23}$$

Solving in terms of $\tilde{\mathbf{f}}$ yields

$$\tilde{\mathbf{f}} = [\mathbf{K}(\mathbf{I}_{K+1} \otimes \mathbf{P}_{d,n}\mathbf{g})\boldsymbol{\phi} - \mathbf{U}\boldsymbol{\delta}]. \tag{4.24}$$

Hence, one can minimise (3.14) over $(\boldsymbol{\phi}(\boldsymbol{\beta}), \mathbf{g}, \boldsymbol{\delta})$, having substituted for $\tilde{\mathbf{f}}$ from (4.24). If there exist restrictions on some columns of $\mathbf{F}$, e.g. one wishes to *impose* that the model includes a fixed effect such that one column of $\mathbf{F}$ is set equal to unity, (4.23) still holds and can be used to eliminate $\tilde{\mathbf{f}}$ in the objective function at the cost of re-introducing some parameters in the variance-covariance matrix of the factor loadings corresponding to the restricted factors.

The FIVR estimator effects a more parsimonious parameterisation of the nuisance parameters, which leads to more efficient estimation of the parameters of interest. Thus FIVR is asymptotically more efficient than FIVU and since FIVU is itself asymptotically equivalent to quasi-differencing methods, FIVR is more efficient than these as well. This is summarised in the following theorem:

Theorem 6. DISTRIBUTION RESULT FOR FIVR. *Under Assumptions 1-6 and model (2.1) with $n = n_0$ FIVR is asymptotically more efficient than FIVU. Furthermore, it is the efficient estimator in the class of estimators that make use of second moment information.*

Proof. See Appendix I. □

5 Finite Sample Performance

In this section we investigate the performance of FIVU and FIVR in finite samples. Our focus is on the signal-to-noise ratio of the model, the proportion of the variance of the total error component that is due to the factor component and the degree of persistence in the model.

Design

The data generating process is given by[8]

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + u_{it};\ u_{it} = \boldsymbol{\lambda}_i' \mathbf{f}_t + \varepsilon_{it} = \sum_{j=1}^{n} \lambda_i^j f_t^j + \varepsilon_{it}, \tag{5.1}$$

for $i = 1, ..., N$, $t = -\varpi, ..., 0, 1, ..., T$, while

$$x_{it} = \rho x_{it-1} + \boldsymbol{\gamma}_i' \mathbf{f}_t + \upsilon_{it} = \rho x_{it-1} + \sum_{j=1}^{n} \gamma_i^j f_t^j + \upsilon_{it};\ \upsilon_{it} = \nu_{it} + \varphi \varepsilon_{it-1}, \tag{5.2}$$

where $\varepsilon_{it} \sim i.i.d.N\left(0, c_1 \sigma_{\varepsilon_i}^2\right)$, with $\sigma_{\varepsilon_i}^2 \sim i.i.d.U[0,2]$, $\nu_{it} \sim i.i.d.N\left(0, \sigma_\nu^2\right)$, $\lambda_i^j \sim i.i.d.N\left(0, c_2 \sigma_{\lambda_i}^2\right)$ with $\sigma_{\lambda_i}^2 \sim i.i.d.U[0,2]$ and $f_t^j \sim i.i.d.N\left(0,1\right)$ for all j, such that $E\left(c_1 \sigma_{\varepsilon_i}^2\right) = c_1 > 0$ and $E\left(c_2 \sigma_{\lambda_i}^2\right) = c_2 > 0$. Thus, our design allows for substantial cross-sectional heteroskedasticity in the idiosyncratic error and the factor loadings.

The zero mean assumption of the factor variates and the idiosyncratic error component is not restrictive since in practice one can remove the non zero mean for a multi-factor structure by adding individual- and time-specific effects. In

[8]In an earlier version of our paper we investigate the performance of our estimators based on a pure AR(1) panel model. That version is available on line at http://mpra.ub.uni-muenchen.de/26166/ .

particular, one can always reparameterise the error term $u_{it} = \boldsymbol{\lambda}_i'\mathbf{f}_t + \varepsilon_{it} = \eta_i + \tau_t + (\boldsymbol{\lambda}_i - \bar{\boldsymbol{\lambda}})'(\mathbf{f}_t - \bar{\mathbf{f}}) + \varepsilon_{it}$, where $\eta_i = \boldsymbol{\lambda}_i'\bar{\mathbf{f}}$ and $\tau_t = \bar{\boldsymbol{\lambda}}\mathbf{f}_t$. Similarly, adding a global intercept will remove the non zero mean of ε_{it}.

The factor loadings of the x and y processes are correlated such that

$$\gamma_i^j = \varrho_{\gamma\lambda}\lambda_i^j + \left(1 - \varrho_{\gamma\lambda}^2\right)^{1/2}\varpi_i^j,\ \ \varpi_i^j \sim i.i.d.N\left(0, c_2\sigma_{\lambda_i}^2\right)\ \ \forall j. \tag{5.3}$$

Since x_{it} can be expressed recursively as

$$x_{it} = \boldsymbol{\gamma}_i' \sum_{\tau=0}^{\infty} \rho^\tau \mathbf{f}_{t-\tau} + \sum_{\tau=0}^{\infty} \rho^\tau \upsilon_{it-\tau}, \tag{5.4}$$

we have

$$\begin{aligned} y_{it} &= \beta \sum_{s=0}^{\infty} \alpha^s x_{it-s} + \boldsymbol{\lambda}_i' \sum_{s=0}^{\infty} \alpha^s \mathbf{f}_{t-s} + \sum_{s=0}^{\infty} \alpha^s \varepsilon_{it-s} \\ &= \beta\boldsymbol{\gamma}_i' \sum_{s=0}^{\infty} \alpha^s \sum_{\tau=0}^{\infty} \rho^\tau \mathbf{f}_{t-s-\tau} + \beta \sum_{s=0}^{\infty} \alpha^s \sum_{\tau=0}^{\infty} \rho^\tau \upsilon_{it-s-\tau} + \boldsymbol{\lambda}_i' \sum_{s=0}^{\infty} \alpha^s \mathbf{f}_{t-s} + \sum_{s=0}^{\infty} \alpha^s \varepsilon_{it-s}. \end{aligned}$$

As described in Kiviet (1995) and Bun and Kiviet (2006), the variances of ν_{it} and $\boldsymbol{\lambda}_i$ are major determinants of the relative strength of the signal-to-noise ratio and the error components, respectively. Noticing that on average

$$\text{var}(\upsilon_{it}) = \sigma_\upsilon^2 = \sigma_\nu^2 + \varphi^2 c_1, \tag{5.5}$$

the average variance of the signal of the model, conditionally on $\boldsymbol{\lambda}_i'\mathbf{f}_t$ and $\boldsymbol{\gamma}_i'\mathbf{f}_t$, is given by

$$\begin{aligned}
\sigma_s^2 &= \mathrm{var}(y_{it}|\boldsymbol{\lambda}_i'\mathbf{f}_t, \boldsymbol{\gamma}_i'\mathbf{f}_t) - \mathrm{var}(\varepsilon_{it}) \\
&= \mathrm{var}(\beta \sum_{s=0}^{\infty} \alpha^s \sum_{\tau=0}^{\infty} \rho^\tau \upsilon_{it-s-\tau}) + \sum_{s=0}^{\infty} \alpha^s \varepsilon_{it-s} \\
&+ 2\mathrm{cov}(\beta \sum_{s=0}^{\infty} \alpha^s \sum_{\tau=0}^{\infty} \rho^\tau \upsilon_{it-s-\tau}, \sum_{s=0}^{\infty} \alpha^s \varepsilon_{it-s}) - \mathrm{var}(\varepsilon_{it}) \\
&= \frac{\beta^2}{(1-\alpha^2)(1-\rho^2)}\sigma_\nu^2 + \frac{\beta^2\varphi^2}{(1-\alpha^2)(1-\rho^2)}c_1 + \frac{1}{(1-\alpha^2)}c_1 \\
&+ \frac{2\beta\alpha\varphi}{(1-\alpha\rho)(1-\alpha^2)}c_1 - c_1 \\
&= \frac{\beta^2}{(1-\alpha^2)(1-\rho^2)}\sigma_\nu^2 + \frac{\beta^2\varphi^2 + (1-\alpha\rho)(1-\rho^2) + 2\beta\alpha\varphi(1-\rho^2)}{(1-\alpha^2)(1-\rho^2)(1-\alpha\rho)} - c_1.
\end{aligned} \tag{5.6}$$

The signal-to-noise ratio is defined as

$$SNR \equiv \frac{\sigma_s^2}{c_1}. \tag{5.7}$$

We normalise $c_1 = 1$, which implies that SNR depends on the value of σ_ν^2 only, as far as the variance parameters are concerned. Hence, we set σ_ν^2 such that SNR is controlled across experiments. In particular, solving for σ_ν^2 yields

$$\sigma_\nu^2 = \left(SNR + 1 - \frac{\beta^2\varphi^2 + (1-\alpha\rho)(1-\rho^2) + 2\beta\alpha\varphi(1-\rho^2)}{(1-\alpha^2)(1-\rho^2)(1-\alpha\rho)}\right)\frac{(1-\alpha^2)(1-\rho^2)}{\beta^2}. \tag{5.8}$$

Recalling that $E\left(c_2\sigma_{\lambda_i}^2\right) = c_2$, the value of c_2 is determined according to the average proportion of the variance of the total error, u_{it}, that is due to the factor component, $\boldsymbol{\lambda}_i'\mathbf{f}_t$. It is easy to show that this ratio equals

$$F_\lambda = nc_2(c_2+1)^{-1}.$$

Thus, for example, $F_\lambda = 1/4$ means that 25% of the variance of the total error is due to the unobserved factors; thus, the factor component has relatively small influence in this case. Solving for c_2 yields

$$c_2 = \frac{nF_\lambda}{1-F_\lambda}.$$

We specify $T = 10$, $\varrho_{\gamma\lambda} = 0.5$, $\varphi = 0.5$, $N \in \{150, 450\}$, $\rho \in \{0.5, 0.95\}$, $\alpha \in \{0.2, 0.8\}$, $F_\lambda \in \{1/4, 3/4\}$, $SNR \in \{3, 9\}$, $n_0 = 1, 2$, giving rise to 64 different experiments. $\rho = 0.95$ allows us to examine the case where the covariate is close to a unit root process. $\alpha = 0.8$ implies that the y process is highly persistent and receives relatively small influence from x. The SNR values are based on previous literature (e.g. Bun and Kiviet, 2006). To reduce the computational burden, for $n_0 = 1$ we fit models with $n = 0, 1, 2$ factors and for $n_0 = 2$ we fit models with $n = 1, 2, 3$ factors. The number of factors is estimated based on the value of n that corresponds to the minimum value of the model information criterion described in Section 3. Following Ahn, Lee and Schmidt (2013, pg. 8) we set $\varrho = 0.75/T^{0.3}$. 2,000 replications are performed.

Results

The results are reported in Tables 1-4. We distinguish between one step and two step GMM estimators; $FIVU_j$ ($FIVR_j$) refers to the j step FIVU (FIVR) estimator, $j = 1, 2$. One step estimators make use of the identity matrix as a weighting matrix. Two step estimators make use of the optimal weighting matrix, computed using estimates of the parameters obtained from the first stage. The moment conditions utilised are of the form $E(y_{is}\varepsilon_{it}) = 0$ for $1 \leq s < t$, $t = 2, ..., 10$, and $E(x_{is}\varepsilon_{it}) = 0$ for $1 \leq s \leq t$, $t = 2, ..., 10$. For FIVU minima are found by the iterative least squares procedure described in Section 3. For FIVR we use a constrained nonlinear optimisation algorithm based on Matlab's fmincon function. Convergence is deemed to have occurred when the modulus of the gradient vector is less than 10^{-5}. Starting values for the factors, $\mathbf{f}$, in FIVU are obtained based on the n largest principal components of the residual $\tilde{u}_{it} = y_{it} - \tilde{\alpha}y_{it-1} - \tilde{\beta}x_{it}$, where $\tilde{\alpha}$, $\tilde{\beta}$ correspond either to the OLS estimates, or to nine sets of uniform random variables on $[0, 1]$. The preferred $\mathbf{f}$ initialisation corresponds to the value that minimises the objective function. Starting values for FIVR are obtained from FIVU. Notice that normalisations on the factor parameters are not imposed.

For comparison, we examine the performance of two popular estimators in dynamic panels, the first-differenced GMM estimator proposed by Arellano and Bond (1991), hereafter DIF, and the system GMM estimator (see e.g. Blundell and Bond, 1998), hereafter SYS. Although these estimators are not consistent under a multi-factor error structure, it is useful to examine their performance under this situation given that the two way error components model can be viewed as a special case of a common factor structure.[9] DIF_a and SYS_a make use of the three most recent available instruments for both y and x with respect to the equations in first differences, while DIF_b and SYS_b make use of all available instruments with respect to the equations in first differences. The SYS estimators use, in addition, Δy_{it-1} as an instrument for y_{it-1} in the model in levels, $t = 3, ..., T$. Thus, DIF_a, DIF_b, SYS_a and SYS_b utilise 37, 72, 58, and 88 moment conditions respectively, quantities that are well below the size of N. In all cases, estimators make use of the optimal weighting matrix.

The results are reported using the following format: average, (standard deviation), [RMSE], {size} of the z-statistic for the structural parameters of the model and |size| of the overidentifying restrictions test statistic, reported only for two step estimators as it is invalid otherwise. Nominal size is set equal to 5%.[10] For FIVU and FIVR the statistics are computed based on the quantities corresponding to the optimal number of factors in each replication. π denotes the proportion of times the correct number of factors has been selected using the two step FIVU. Similar results are obtained for FIVR and therefore we do not report these here.

It is clear that FIVU and FIVR perform well under all circumstances. Naturally, their performance improves when the signal-to-noise ratio increases. The same holds as F_λ increases, for $\alpha = .5$, especially when x is highly persistent. Bias for two step FIVU and FIVR is negligible in all experiments. FIVR has lower standard deviation than FIVU and therefore it performs better in terms of RMSE, often by a substantial margin. The difference in the performance of the two estimators with regards to RMSE appears to become larger with higher values of ρ and α, especially when the factor component has a relatively small contribution in the variance of the total error (i.e. $F_\lambda = 1/4$). For example, for $SNR = 9$ the ratio of the standard deviation of the estimated autoregressive parameter for $FIVR_2$ over the standard deviation of $FIVU_2$ is roughly about 73% when $\alpha = .5$ and $\rho = .5$ and decreases to around 62% for $\alpha = .8$ and $\rho = .95$. Gains in terms of dispersion and RMSE obtained using FIVR appear to be slightly smaller for β compared to α. As expected two step estimators outperform their one step counterparts, especially when x is highly persistent. All estimators perform well in terms of the empirical size of the z-statistic for the structural parameters of the model. The overidentifying restrictions test statistic is valid only for the optimal (two step) GMM estimators and in this case there are only small size distortions. As

[9] We do not examine the quasi-differenced GMM estimator of Ahn, Lee and Schmidt (2013) here, as this is asymptotically equivalent to FIVU.

[10] For DIF and SYS, since the moment conditions are invalid under a factor structure, the entries in | . | reflect power, as opposed to size.

indicated by π, the frequency of selecting the correct number of factors is above, or close to 90%.

The performance of DIF and SYS is generally poor and highly sensitive to the design. As expected, bias is smaller when $F_\lambda = 1/4$ relative to $F_\lambda = 3/4$. Even in the former case however, bias can be very large, especially when the signal-to-noise ratio is not large, $\rho = 0.95$ and/or $\alpha = 0.8$. Naturally, there also appears to be large size distortions for the z-statistic, especially when bias is large, in which case the null hypothesis is rarely not rejected. The power of the overidentifying restrictions test statistic can be small, particularly for SYS_b when $F_\lambda = 1/4$. Practically this means is that provided the number of moment conditions used is large enough, it is not unlikely that one would fail to reject the validity of the model based on SYS_b, even if the model is not specified correctly.

Similar conclusions apply for the two factor model in that FIVU and FIVR perform well in all experiments. Compared to the one factor case, the dispersion of FIVU increases slightly, while FIVR appears to remain largely unaffected. The performance of the estimators improves for $N = 450$ and, as expected, their standard deviation decreases roughly at the rate of $N^{1/2}$. To save space we do not report these results.

6 Concluding Remarks

The Generalised Method of Moments is a popular approach for estimating dynamic panel data models with large N and T fixed. This approach has the appealing feature that it can handle endogeneity, it requires relatively weak assumptions about the initial conditions of the data generating process. Furthermore, it avoids full specification of the serial correlation and heteroskedasticity properties of the idiosyncratic error, or indeed any other distributional assumptions. On the other hand, under a multi-factor error structure standard dynamic panel estimators can be inconsistent as the moment conditions they utilise are invalidated. In this paper we develop a new GMM type approach for consistent and asymptotically efficient estimation of panel data models with factor residuals. One novelty of our approach is that we introduce new parameters to represent the unobserved covariances between the instruments and the factor component of the residual. We develop estimators that are asymptotically efficient and appear to behave well in small samples under a wide range of parameterisations.

Acknowledgements

An earlier version of this paper circulated as Robertson, Sarafidis and Symons. We are particularly grateful to Jim Symons for his enormous contribution to this work. We would also like to thank two anonymous referees and the Editor for providing us with several constructive comments and suggestions, as well as seminar participants at the ESWC 2010 and SETA 2011. We have benefited greatly from discussions with Arturas Juodis, Mervyn Silvapulle and Neville Weber. All remaining errors are ours.

References

[1] Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. Journal of Econometrics 68, 5-28.

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

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

[4] Arellano, M., 2003, Panel Data Econometrics. Oxford University Press, Oxford.

[5] Arellano, M., Bond, S.R., 1991. Some specification tests for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277-298.

[6] Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-component models. Journal of Econometrics 68, 29-51.

[7] Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

[8] Bai, J., 2013. Likelihood approach to dynamic panel models with interactive effects. Mimeo.

[9] Bai, J., Ng, S. 2008. Large Dimensional Factor Analysis. Foundations and Trends in Econometrics 3:2.

[10] Blundell, R., Bond, S.R., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115-143.

[11] Bun, M., Kiviet, J., 2006. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. Journal of Econometrics 132, 409-444.

[12] Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. Journal of Econometrics 34, 305-334.

[13] Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

[14] Holtz-Eakin D., Newey, W., Rosen, H., 1988. Estimating vector autoregressions with panel data. Econometrica 56, 1371-1395.

[15] Juodis A., Sarafidis, V., 2014. Fixed T dynamic panel data estimators with multi-factor errors. MPRA paper 57659, University Library of Munich, Germany. http://mpra.ub.uni-muenchen.de/57659/1/MPRA_paper_57659.pdf

[16] Kiviet, J., 1995. On bias, inconsistency and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68, 53-78.

[17] Kruiniger, H., 2008. Not So Fixed Effects: Correlated structural breaks in panel data. Mimeo, Queen Mary, University of London. http://www.iza.org/conference_files/pada2009/kruiniger_h5168.pdf

[18] Moon, R., Weidner, M., 2013. Dynamic linear panel regression models with interactive fixed effects. Mimeo.

[19] Nauges, C., Thomas, A., 2003. Consistent estimation of dynamic panel data models with time-varying individual effects. Annales d'Economie et de Statistique 70, 54-75.

[20] Newey, W.K., 1990. Semiparametric efficiency bound. Journal of Applied Econometrics 5, 99-136.

[21] Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. Handbook of Econometrics, Vol 4.

[22] Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967-1012.

[23] Sarafidis, V., Robertson, D., 2009. On the impact of cross-sectional dependence in short dynamic panel estimation. The Econometrics Journal 12, 62-81.

[24] Sarafidis, V., Yamagata, T., 2013. Instrumental variable estimation of dynamic linear panel models with defactored regressors under cross-sectional dependence. Mimeo, Monash University.

[25] Sarafidis, V., Yamagata, T., Robertson, D., 2009. A test of error cross section dependence for a linear dynamic panel model with regressors. Journal of Econometrics 148, 149-161.

[26] Sarafidis, V., Wansbeek, T., 2012. Cross-sectional dependence in panel data analysis. Econometric Reviews 31, 483-531.

Appendix I: Proofs of Theorems

Theorem 3

Proof. Assumption 5 guarantees that $\widehat{\boldsymbol{\beta}}(\Theta) = \widehat{\boldsymbol{\beta}}(\Theta_r)$. According to the boundedness assumption, we may choose Θ_c such that the objective function is bounded away from zero outside of this set. Since the minimised value over this set converges to $\boldsymbol{\theta}_0$ in probability, it follows that, for N sufficiently large, $\widehat{\boldsymbol{\beta}}(\Theta_c) = \widehat{\boldsymbol{\beta}}(\Theta_r)$ with arbitrarily high probability. The result that $\widehat{\boldsymbol{\beta}}(\Omega) \to \widehat{\boldsymbol{\beta}}(\Theta_c)$ now follows from the density of Θ in Ω.[11] The result for the covariance matrices follows from the following observation. Let $\mathbf{X}$ and $\mathbf{Y}$ be matrices with the same number of rows. Then the submatrix in the north west corner of the inverse or generalised inverse of $\begin{bmatrix} \mathbf{X} & \mathbf{Y} \end{bmatrix}'\begin{bmatrix} \mathbf{X} & \mathbf{Y} \end{bmatrix}$, which is of dimension that of $\mathbf{X}'\mathbf{X}$, is $(\mathbf{X}'\mathbf{M_Y}\mathbf{X})^{-1}$, where $\mathbf{M_Y}$ is the projection that removes $\mathbf{Y}$, i.e. $\mathbf{M_Y} = \mathbf{I} - \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'$. This follows from the partitioned inverse formula. Thus the covariance matrix of the parameters of interest is obtained by removing from $\mathbf{\Gamma}$ the linear space spanned by the columns corresponding to the nuisance variables; two sets of nuisance variables generating the same span will yield the same covariance matrix. □

[11] "Dense subset" means that one can find something in the subset arbitrarily close to any element in the superset. For example the set of invertible square matrices is dense in the set of all square matrices, because one can find an invertible matrix arbitrarily close to a given singular matrix. In our context, certain arguments concerning identification will not go through if certain submatrices of $\mathbf{F}$ and $\mathbf{G}$ are singular. For example in the AR(1), one factor case, we require $g_1 \neq 0$. Density allows us to assume away $g_1 = 0$ and thus obtain identification.

Theorem 4

Proof. Let $\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\theta}_r)$ be a c-dimensional moment function and consider the optimal GMM estimator of the true value of $\boldsymbol{\theta}_r$ based on $\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\theta}_r)$. This has asymptotic variance

$$\operatorname{var}(\widehat{\boldsymbol{\theta}}_r) = (\boldsymbol{\Gamma}'\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma})^{-1}, \tag{6.1}$$

where

$$\boldsymbol{\Gamma} = E\left[\frac{\partial\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\theta}_r)}{\partial\boldsymbol{\theta}_r'}\right]; \qquad \boldsymbol{\Delta} = E(\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\theta}_r)\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\theta}_r)'), \tag{6.2}$$

both evaluated at the true value $\boldsymbol{\theta}_0$. Assume $\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$ have full rank and let $\boldsymbol{\theta}_r = (\boldsymbol{\varphi}_r', \boldsymbol{\xi}_r')'$ be a decomposition of the parameter space into two subsets; $\boldsymbol{\varphi}_r$ is a vector that includes the parameters of interest, $\boldsymbol{\beta}$, together with some possible nuisance parameters, while the vector $\boldsymbol{\xi}_r$ contains the remaining nuisance parameters. Let $\boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \end{bmatrix}$, where

$$\mathbf{Q} = E\left[\frac{\partial\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\theta}_r)}{\partial\boldsymbol{\varphi}_r'}\right]; \qquad \mathbf{R} = E\left[\frac{\partial\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\theta}_r)}{\partial\boldsymbol{\xi}_r'}\right]. \tag{6.3}$$

Since $\boldsymbol{\Gamma}$ is of full rank, so too are $\mathbf{Q}$ and $\mathbf{R}$. Assume that, for some $\ell \times c$ matrix $\mathbf{D}(\boldsymbol{\varphi}_r)$ of full rank ($\ell \leq c$)

$$\mathbf{D}(\boldsymbol{\varphi}_r)\boldsymbol{\psi}\left(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\varphi}_r,\boldsymbol{\xi}_r\right) = \bar{\boldsymbol{\psi}}\left(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\varphi}_r\right), \qquad \text{for all } \boldsymbol{\varphi}_r, \boldsymbol{\xi}_r, \tag{6.4}$$

i.e. $\mathbf{D}(.)$ represents a set of transformations that eliminates the nuisance parameters $\boldsymbol{\xi}_r$ at the cost of some loss of moment conditions. Then $\bar{\boldsymbol{\psi}}\left(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\varphi}_r\right)$ is a moment function and inference about $\boldsymbol{\varphi}_r$ may be based on it. Denote the resulting estimator by $\bar{\boldsymbol{\varphi}}_r$, which has the asymptotic variance matrix

$$\operatorname{var}(\bar{\boldsymbol{\varphi}}_r) = (\bar{\boldsymbol{\Gamma}}'\bar{\boldsymbol{\Delta}}^{-1}\bar{\boldsymbol{\Gamma}})^{-1}, \tag{6.5}$$

where $\bar{\boldsymbol{\Gamma}} = E\left[\partial\bar{\boldsymbol{\psi}}\left(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\varphi}_r\right)/\partial\boldsymbol{\varphi_r}'\right]$ and $\bar{\boldsymbol{\Delta}} = E\left[\bar{\boldsymbol{\psi}}\left(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\varphi}_r\right)\bar{\boldsymbol{\psi}}\left(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\varphi}_r\right)'\right]$, both evaluated at the true value of $\boldsymbol{\varphi}_r$. Differentiating (6.4) with respect to $\boldsymbol{\varphi}_r$ and using the fact that $E\left(\boldsymbol{\psi}\left(\mathbf{Z}_i'\mathbf{X}_i;\boldsymbol{\theta}_0\right)\right) = \mathbf{0}$ one has

$$\mathbf{D}(\boldsymbol{\varphi}_r)\mathbf{Q} = \bar{\boldsymbol{\Gamma}}. \tag{6.6}$$

Differentiating (6.4) with respect to $\boldsymbol{\xi}_r$ one has

$$\mathbf{D}(\varphi_r)\mathbf{R} = \mathbf{0}, \tag{6.7}$$

where, in both cases, $\mathbf{D}(\varphi_r)$ is evaluated at the true value of $\boldsymbol{\varphi}_r$. One has as well that

$$\bar{\boldsymbol{\Delta}} = \mathbf{D}(\varphi_r)\boldsymbol{\Delta}\mathbf{D}(\varphi_r)'. \tag{6.8}$$

The asymptotic covariance matrix of $\bar{\boldsymbol{\varphi}}_r$ is now

$$\mathrm{var}(\bar{\varphi}_r) = [\mathbf{Q}'\mathbf{D}(\varphi_r)'(\mathbf{D}(\varphi_r)\boldsymbol{\Delta}\mathbf{D}(\varphi_r)')^{-1}\mathbf{D}(\varphi_r)\mathbf{Q}]^{-1}. \tag{6.9}$$

Make the transformations $\mathbf{D}_{\boldsymbol{\Delta}} = \mathbf{D}(\varphi_r)\boldsymbol{\Delta}^{1/2}$, where the dependence of $\mathbf{D}_{\boldsymbol{\Delta}}$ on $\boldsymbol{\varphi}_r$ is dropped for notational simplicity, $\boldsymbol{\Gamma}_{\boldsymbol{\Delta}} = \boldsymbol{\Delta}^{-1/2}\boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{Q}_{\boldsymbol{\Delta}} & \mathbf{R}_{\boldsymbol{\Delta}} \end{bmatrix}$. Then, using results for partitioned inverses, one finds

$$\mathrm{var}(\hat{\boldsymbol{\varphi}}_r) = (\mathbf{Q}'_{\boldsymbol{\Delta}}(\mathbf{I}_c - \mathbf{P}_{\mathbf{R}_\Delta})\mathbf{Q}_{\boldsymbol{\Delta}})^{-1}, \tag{6.10}$$

where $\mathbf{P}_{\mathbf{R}_{\boldsymbol{\Delta}}} = \mathbf{R}_{\boldsymbol{\Delta}}(\mathbf{R}'_{\boldsymbol{\Delta}}\mathbf{R}_{\boldsymbol{\Delta}})^{-1}\mathbf{R}'_{\boldsymbol{\Delta}}$. One also has

$$\mathrm{var}(\bar{\boldsymbol{\varphi}}_r) = (\mathbf{Q}'_{\boldsymbol{\Delta}}\mathbf{P}_{\mathbf{D}_{\boldsymbol{\Delta}}}\mathbf{Q}_{\boldsymbol{\Delta}})^{-1}, \tag{6.11}$$

where $\mathbf{P}_{\mathbf{D}_{\boldsymbol{\Delta}}} = \mathbf{D}'_{\boldsymbol{\Delta}}(\mathbf{D}_{\boldsymbol{\Delta}}\mathbf{D}'_{\boldsymbol{\Delta}})^{-1}\mathbf{D}_{\boldsymbol{\Delta}}$. Then $\mathrm{var}(\bar{\boldsymbol{\varphi}}_r) > \mathrm{var}(\hat{\boldsymbol{\varphi}}_r)$ (as positive matrices) if and only if

$$\mathbf{Q}'_{\boldsymbol{\Delta}}(\mathbf{I}_c - \mathbf{P}_{\mathbf{R}_{\boldsymbol{\Delta}}} - \mathbf{P}_{\mathbf{D}_{\boldsymbol{\Delta}}})\mathbf{Q}_{\boldsymbol{\Delta}} > \mathbf{0}. \tag{6.12}$$

Now (6.7) implies that the matrices inside the brackets are orthogonal projections so the sandwich matrix is a projection of rank $c - \ell - \dim(\mathbf{R})$. There are thus no losses in efficiency from eliminating the $\boldsymbol{\nu}_r$ parameters in this way if $\dim(\mathbf{R}) = c - \ell$, i.e. the number of eliminated parameters is equal to the number of lost moment conditions. □

Remark. In the case where $n = 1$ and $f_t \equiv 1$ for all t the moment conditions are linear of the form

$$\mathbf{m} + \mathbf{Q}\boldsymbol{\beta} + \mathbf{R}\boldsymbol{\xi} = \mathbf{0}, \tag{6.13}$$

where $\mathbf{m}$, $\mathbf{Q}$ and $\mathbf{R}$ consist of observable moments. The parameters $\boldsymbol{\xi}$ are here the g's from the development in the text, while $\boldsymbol{\beta} = \boldsymbol{\varphi}_r$ in the discussion above. The first-differenced GMM estimator proposed by Arellano and Bond (1991) introduces a differencing matrix of full rank to eliminate $\mathbf{R}$:

$$\mathbf{Dm} + \mathbf{DQ}\boldsymbol{\beta} = \mathbf{0}. \tag{6.14}$$

Both forms give rise to GMM estimates of the parameters of interest $\boldsymbol{\beta}$ by a one pass regression, given estimates of the error variance-covariance matrix. Let $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ be such estimates for (6.13) and (6.14) respectively. Call these estimates *compatible* if $\boldsymbol{\Omega}_2 = \mathbf{D}\boldsymbol{\Omega}_1\mathbf{D}'$. One might form compatible estimates by first developing an estimate of the covariance matrix for (6.13) and then adjusting it appropriately for (6.14). The following is true:

Proposition. *GMM estimates based on (6.13) and (6.14) are arithmetically equal if they employ compatible estimates of the error variance-covariance matrix.*

To prove this one shows

$$\mathbf{Q}'\boldsymbol{\Omega}^{-1/2}(\mathbf{I}-\mathbf{P})_{\boldsymbol{\Omega}^{-1/2}\mathbf{R}}\boldsymbol{\Omega}^{-1/2}\mathbf{Q} = \mathbf{QD}'(\mathbf{D}\boldsymbol{\Omega}\mathbf{D}')^{-1}\mathbf{DQ}, \tag{6.15}$$

for any conformable full rank symmetric $\boldsymbol{\Omega}$. This is will be so if $(\mathbf{I}-\mathbf{P})_{\boldsymbol{\Omega}^{-1/2}\mathbf{R}} = \mathbf{P}_{\boldsymbol{\Omega}^{1/2}\mathbf{D}}$. It is easy to see that $\mathbf{P}_{\boldsymbol{\Omega}^{-1/2}\mathbf{R}}\mathbf{P}_{\boldsymbol{\Omega}^{1/2}\mathbf{D}} = \mathbf{0}$, so that the projections are orthogonal. Consideration of ranks now delivers the result.

In our context, this result shows the first-differenced GMM of the error components model is precisely the FIVU estimator, given compatible covariance matrix estimates. In practice, first-differenced GMM estimates and FIVU estimates need not be the same as first step estimates of the structural parameters may differ when the two equations are considered in isolation. In this case, equality is only asymptotic.

Theorem 5.

Proof. Assume we have found the minimiser of the objective function $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}|n)$, $\widehat{\boldsymbol{\theta}}$. Clearly this minimiser is not unique as $\widehat{\mathbf{G}}\widehat{\mathbf{F}}' = \widehat{\mathbf{G}}\mathbf{U}\mathbf{U}^{-1}\widehat{\mathbf{F}}' = \widehat{\widehat{\mathbf{G}}}\widehat{\widehat{\mathbf{F}}}'$. Assumption 5 implies that for this given $\widehat{\boldsymbol{\theta}}$ we can find $(\widetilde{\mathbf{g}}_r, \widetilde{\mathbf{f}}_r)$ such that

$$\mathbf{S}\text{vec}\left(\widehat{\mathbf{G}}\widehat{\mathbf{F}}'\right) = \mathbf{S}\text{vec}\left(\widetilde{\mathbf{GF}'}\right)_r, \tag{6.16}$$

where $\left(\widetilde{\mathbf{GF}'}\right)_r$ is a function of $(\widetilde{\mathbf{g}}_r, \widetilde{\mathbf{f}}_r)$ such that some identifying normalisations are satisfied. By definition the parameter vector $\widetilde{\boldsymbol{\theta}}_r = \left(\widehat{\boldsymbol{\beta}}', \widetilde{\mathbf{g}}'_r, \widetilde{\mathbf{f}}'_r\right)'$ belongs to the restricted set Θ_r. Now using the definition of $\widehat{\boldsymbol{\theta}}_r$ as an *argmin*, we have

$$\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}}_r)'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}}_r) \leq \boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta}_r)'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta}_r)\ \forall \boldsymbol{\theta}_r \in \Theta_r. \tag{6.17}$$

In particular, it is valid for $\widetilde{\boldsymbol{\theta}}_r \in \Theta_r$. As a result, we have

$$\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}}_r)'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}}_r) \leq \boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widetilde{\boldsymbol{\theta}}_r)'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widetilde{\boldsymbol{\theta}}_r) = Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}}|n). \tag{6.18}$$

The inverse holds true as well. That is, assume we have found the minimiser of the objective function $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta}_r|n)$, $\widehat{\boldsymbol{\theta}}_r$. Assumption 5 implies that for this given $\widehat{\boldsymbol{\theta}}_r$ there exists $(\widetilde{\mathbf{g}},\widetilde{\mathbf{f}})$ such that

$$\mathbf{S}\text{vec}\left(\widehat{\mathbf{GF}'}\right)_r = \mathbf{S}\text{vec}\left(\widetilde{\mathbf{G}}\widetilde{\mathbf{F}}'\right), \tag{6.19}$$

where $\left(\widetilde{\mathbf{G}}\widetilde{\mathbf{F}}'\right)$ is the matrix product of $(\widetilde{\mathbf{g}},\widetilde{\mathbf{f}})$, the normalised estimated parameters under some normalisation scheme. By definition, the parameter vector $\widetilde{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\beta}}_r{}',\widetilde{\mathbf{g}}',\widetilde{\mathbf{f}}'\right)'$ belongs to Θ. Given the definition of $\widehat{\boldsymbol{\theta}}$ as an *argmin*, we have

$$\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}})'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}}) \leq \boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta})'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta})\ \forall \boldsymbol{\theta} \in \Theta. \tag{6.20}$$

Since this inequality holds for $\widetilde{\boldsymbol{\theta}} \in \Theta$ we have

$$\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}})'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}}) \leq \boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widetilde{\boldsymbol{\theta}})'\mathbf{C}_N\boldsymbol{\psi}(\widehat{\mathbf{M}}_N;\widetilde{\boldsymbol{\theta}}) = Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N;\widehat{\boldsymbol{\theta}}_r|n). \tag{6.21}$$

Thus, the minimised value of $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta}_r|n)$ and the minimised value of $Q_{\mathbf{B}}(\widehat{\mathbf{M}}_N;\boldsymbol{\theta}|n)$ are equal. It follows that the BIC that makes use of the latter is consistent as $N \to \infty$. □

Theorem 6.

Proof. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}',\boldsymbol{\nu}')'$, $\boldsymbol{\nu} = \boldsymbol{\nu}(\boldsymbol{\beta},\boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is a vector of nuisance parameters which has lower dimension than $\boldsymbol{\nu}$. We assume $\boldsymbol{\nu}(.)$ is linear in $\boldsymbol{\tau}$, i.e. $\boldsymbol{\nu}(\boldsymbol{\beta},\boldsymbol{\tau}) = \boldsymbol{V}(\boldsymbol{\beta})\boldsymbol{\tau}$, though the argument to be presented would go through under the assumption of sufficient differentiability at the true value. We consider the estimator $\bar{\boldsymbol{\beta}}$ based on the moment conditions in terms of $\boldsymbol{\beta},\boldsymbol{\tau}$. One has $\overline{\boldsymbol{\Gamma}} = \left[\begin{array}{cc} \mathbf{Q}+\mathbf{RJ} & \mathbf{RV} \end{array}\right]$ where now $\mathbf{Q} = \partial\boldsymbol{\psi}(\boldsymbol{\beta},\boldsymbol{\nu})/\partial\boldsymbol{\beta}'$, $\mathbf{R} = \partial\boldsymbol{\psi}(\boldsymbol{\beta},\boldsymbol{\nu})/\partial\boldsymbol{\nu}'$

$\mathbf{J} = \partial\boldsymbol{\nu}(\boldsymbol{\beta},\boldsymbol{\tau})/\partial\boldsymbol{\beta}'$and $\mathbf{V} = \mathbf{V}(\boldsymbol{\beta}) = \partial\boldsymbol{\nu}(\boldsymbol{\beta},\boldsymbol{\tau})/\partial\boldsymbol{\tau}'$ so, as in (6.10)

$$\mathrm{var}(\overline{\boldsymbol{\beta}}) = [(\mathbf{Q}+\mathbf{RJ})'_{\boldsymbol{\Delta}}(\mathbf{I}_c - \mathbf{P}_{(\mathbf{RV})_{\boldsymbol{\Delta}}})(\mathbf{Q}+\mathbf{RJ})_{\boldsymbol{\Delta}})]^{-1}. \tag{6.22}$$

Since $(\mathbf{I}_c - \mathbf{P}_{\mathbf{R}_{\boldsymbol{\Delta}}})((\mathbf{Q}+\mathbf{RJ})_{\boldsymbol{\Delta}}) = (\mathbf{I}_c - \mathbf{P}_{\mathbf{R}_{\boldsymbol{\Delta}}})\mathbf{Q}$ and $\mathbf{P}_{\mathbf{R}_{\boldsymbol{\Delta}}} > \mathbf{P}_{(\mathbf{RV})_{\boldsymbol{\Delta}}}$, one sees from (6.10) that

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) \geq \mathrm{var}(\bar{\boldsymbol{\beta}}) \tag{6.23}$$

where $\hat{\boldsymbol{\beta}}$ is the estimates based on the moment conditions in terms of $\boldsymbol{\beta}, \boldsymbol{\nu}$ with equality if and only if $(\mathbf{P}_{\mathbf{R}_{\boldsymbol{\Delta}}} - \mathbf{P}_{(\mathbf{RV})_{\boldsymbol{\Delta}}})(\mathbf{Q}+\mathbf{RJ})_{\boldsymbol{\Delta}} = \mathbf{0}$. Since in general there is no particular reason for this equality to hold, it follows that a more parsimonious parameterisation of the nuisance parameters will typically deliver a more efficient estimator of the parameters of interest.[12] □

It is also straightforward to prove that FIVR is efficient in the class of estimators that make use of second moment information, based on an argument similar to that provided by Ahn and Schmidt (1995, section 4). Therefore this proof is omitted. In summary, FIVR reaches the semi-parametric efficiency bound discussed by Newey (1990) using standard results of Chamberlain (1987). Thus, FIVR is asymptotically efficient relative to a QML estimator, but the estimators are equally efficient under normality.

Appendix II: Identification for FIVU

Here we show how an identification scheme for the FIVU model could be developed. Note that the implementation of FIVU discussed in the text does not require the imposition of an identification scheme and optimises freely over the whole parameter space, but the distribution of this proposed FIVU estimator can only be obtained if there does exist some scheme that would identify $\boldsymbol{\theta}_0$.

We focus on the canonical case, where the set of instruments consists of current and lagged values of the variables. Extension to the general case is straightforward. The moment conditions are of the form given in (2.15), i.e. $\mathbf{M}\boldsymbol{\phi} - \mathrm{vech}(\mathbf{GF}') = \mathbf{0}$. The problem is to impose restrictions on $\mathrm{vech}(\mathbf{GF}')$ so that the values of $\mathbf{G}$ and $\mathbf{F}$ can be uniquely inferred from knowledge of $\mathrm{vech}(\mathbf{GF}')$, at the same time ensuring that the original $\mathrm{vech}(\mathbf{GF}')$ can be obtained from the restricted $\mathbf{G}$ and $\mathbf{F}$. Consider

[12] The condition will hold if $\mathbf{J} = \mathbf{0}$ and $\mathbf{Q}'_{\boldsymbol{\Delta}}\mathbf{R}_{\boldsymbol{\Delta}} = \mathbf{0}$. This will be so when the reparameterisation can be accomplished independently of $\boldsymbol{\beta}$ and the GMM estimates of the parameters of interest are independent of the estimates of the nuisance parameters.

the upper triangular elements of the product $\mathbf{GF}'$:

$$\begin{bmatrix} \mathbf{G}_1\mathbf{f}_1 & \mathbf{G}_1\mathbf{f}_2 & \dots & \mathbf{G}_1\mathbf{f}_T \\ & \mathbf{G}_2\mathbf{f}_2 & \dots & \mathbf{G}_2\mathbf{f}_T \\ & & \ddots & \vdots \\ & & & \mathbf{G}_T\mathbf{f}_T \end{bmatrix}. \tag{6.24}$$

One can impose the restriction that (say) the last n_0 columns of $\mathbf{F}'$ be $\mathbf{I}_{n_0}$. We assume $n_0 < (T+1)/2$, so that an $n_0 \times n_0$ block of terms exists above the main diagonal in (6.24). If this is done, all $\mathbf{G}_s$, for s =1,... $T - n_0 + 1$, may be inferred from the values of the terms in (6.24). When $s > T - n_0 + 1$ this is no longer so, as such terms as $\mathbf{G}_{T-n_0+2}\mathbf{f}_{T-n_0+1}$ drop out from the objective function. In this case one can impose the restrictions that the first $s - T + n_0 - 1$ columns of $\mathbf{G}_s$ are zero. This enables the unique inference of all the $\mathbf{G}_s$ in (6.24) i.e. the full $\mathbf{G}$ matrix. Consider now the problem of inferring $\mathbf{f}_t$ when $t \leq T - n_0$. The matrix

$$\widetilde{\mathbf{G}}_t\mathbf{f}_t = \begin{bmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_t \end{bmatrix} \mathbf{f}_t$$

is observable. The number of rows of $\widetilde{\mathbf{G}}_t$ is pt. When $pt \geq n_0$ we impose the restriction that the null space of $\widetilde{\mathbf{G}}_t$ be zero, the full rank assumption on $\widetilde{\mathbf{G}}_t$. When $pt < n_0$ (which need not occur), we set the last $n_0 - pt$ entries of $\widetilde{\mathbf{G}}_t$ equal to unity and impose the condition that the appropriately truncated submatrix of $\widetilde{\mathbf{G}}_t$ be of full rank. This establishes the identification of $\mathbf{G}$ and $\mathbf{F}$. The scheme has the following characteristics:

1. The last n_0 columns of $\mathbf{F}'$ form $\mathbf{I}_n$.

2. There are additional zero/one restrictions on $\mathbf{G}$ and $\mathbf{F}$.

3. There is a collection of full rank conditions on submatrices of $\mathbf{G}$.

Let Θ_r be the collection of parameters such that 1-3 hold and Θ be the collection such that both 3 holds and the matrix formed from the last n_0 columns of $\mathbf{F}'$ is of full rank. The following facts are straightforward to show:

PROPERTIES OF THE IDENTIFICATION SCHEME.

Assume $n_0 < (T+1)/2$.

1. *With* $\boldsymbol{\phi}$ *held fixed, any* $\boldsymbol{\theta} \in \Theta_r$ *is identified from the moment conditions.*

2. *For any* $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\psi}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}) = \boldsymbol{\psi}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}_r)$ *for some* $\boldsymbol{\theta}_r \in \Theta_r$. Θ *is dense in the unrestricted parameter set* Ω.

3. $E\left(\partial\boldsymbol{\psi}(\mathbf{Z}_i'\mathbf{X}_i; \boldsymbol{\theta})/\partial\boldsymbol{\nu}_r'\right)$ *is of full rank where* $\boldsymbol{\nu}_r$ *is the vector of free parameters in restricted* $\mathbf{G}, \mathbf{F}$.

4. *For any* $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\psi}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}) = \boldsymbol{\psi}(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}_r)$ *for some* $\boldsymbol{\theta}_r \in \Theta_r$.

5. *The spanning condition* (3.13) *holds.*

These results establish all of Assumption 5 in the canonical case except the boundedness condition for $\boldsymbol{\theta} \in \Theta_r$. To see this, assume $\boldsymbol{\beta}$ is restricted to a compact set. Then

$$\|\mathbf{B}_N(\mathbf{M}\boldsymbol{\phi}(\boldsymbol{\beta}) - \text{vech}(\mathbf{G}\mathbf{F}')\| \geq \left| \|\mathbf{G}\| \left\|\mathbf{B}_N \text{vech}(\bar{\mathbf{G}}\mathbf{F}')\right\| - \|\mathbf{B}_N\mathbf{M}\boldsymbol{\phi}(\boldsymbol{\beta})\| \right|,$$

where $\|\mathbf{G}\|$ is the Hilbert-Schmidt norm of $\mathbf{G}$ and $\left\|\bar{\mathbf{G}}\right\|=1$, where $\bar{\mathbf{G}} = \mathbf{G}/\left\|\bar{\mathbf{G}}\right\|$. The second term can be made arbitrarily large by choice of $\|\mathbf{G}\|$ provided $\left\|\mathbf{B}_N \text{vech}(\bar{\mathbf{G}}\mathbf{F}')\right\|$ can be bounded away from zero. Now $\left\|\mathbf{B}_N \text{vech}(\bar{\mathbf{G}}\mathbf{F}')\right\| \geq b \left\|\text{vech}(\bar{\mathbf{G}}\mathbf{F}')\right\|$ where b is the smallest eigenvalue of $\mathbf{B}_N$.[13] The identification restrictions on $\mathbf{G}$ are such that each element of the matrix either appears as a separate term in $\text{vech}(\bar{\mathbf{G}}\mathbf{F}')$ or is zero. This implies $\left\|\text{vech}(\bar{\mathbf{G}}\mathbf{F}')\right\| \geq \left\|\bar{\mathbf{G}}\right\| = 1$, thus delivering the result.

These conditions suffice to identify the factors; it remains to consider identification for the full vector $\boldsymbol{\theta}$. We shall give a condition for the *one factor* case. We examine when $\boldsymbol{\Gamma} = E(\partial\boldsymbol{\psi}\left(\mathbf{Z}_i'\mathbf{X}_i; \boldsymbol{\theta}_r\right)/\partial\boldsymbol{\theta}_r')$ is of full rank. Local identification will follow from the full rank of $\boldsymbol{\Gamma}$. Write the moment condition (2.14) in terms of upper-triangular matrices

$$\begin{bmatrix} \mathbf{M}_{11}\boldsymbol{\phi} & \mathbf{M}_{12}\boldsymbol{\phi} & \dots & \mathbf{M}_{1T}\boldsymbol{\phi} \\ & \mathbf{M}_{22}\boldsymbol{\phi} & \dots & \mathbf{M}_{2T}\boldsymbol{\phi} \\ & & \ddots & \vdots \\ & & & \mathbf{M}_{TT}\boldsymbol{\phi} \end{bmatrix} - \begin{bmatrix} \mathbf{g}_1 f_1 & \mathbf{g}_1 f_2 & \dots & \mathbf{g}_1 f_T \\ & \mathbf{g}_2 f_2 & \dots & \mathbf{g}_2 f_T \\ & & \ddots & \vdots \\ & & & \mathbf{g}_T f_T \end{bmatrix} = \mathbf{0}. \qquad (6.25)$$

The identification restriction is here that $f_T = 1$ and $\mathbf{g}_1 \neq 0$, the latter being the full rank condition on submatrices of $\mathbf{G}$. If this is so, and given that the full rank

[13]This argument is facilitated by the assumption that $\mathbf{B}_N$ is the symmetric square root of the weight matrix $\mathbf{C}_N$ rather than the Choleski matrix.

of $\partial\boldsymbol{\psi}/\partial\boldsymbol{\nu}'_r$ is established, $\mathbf{\Gamma}$ can fail to have full rank only if

$$\text{vech}(\mathbf{M}^{\dagger}(I_T \otimes \boldsymbol{\beta}^*)) = \frac{\partial\text{vech}(\mathbf{gf}')}{\partial\mathbf{g}'}\mathbf{g}^* + \frac{\partial\text{vech}(\mathbf{gf}')}{\partial\mathbf{f}'}\mathbf{f}^* \tag{6.26}$$

for some non-zero $(\boldsymbol{\beta}^{*\prime}, \mathbf{g}^{*\prime}, \mathbf{f}^{*\prime})'$, where $\mathbf{M}^{\dagger}$ is the $Tp \times KT$ matrix comprised of the $p \times (K+1)$ matrices $\mathbf{M}_{st}$ with their first columns removed. In this expression $f_T^* = 0$ since the identification procedure has removed the last column of $\partial\boldsymbol{\psi}/\partial\mathbf{f}'$. Making use of (2.10), this can be written as

$$\text{vech}(\mathbf{M}^{\dagger}(\mathbf{I}_T \otimes \boldsymbol{\beta}^*)) = \text{vech}(\boldsymbol{g}^*\mathbf{f}') + \text{vech}(\mathbf{gf}^{*\prime}), \tag{6.27}$$

such that the term on the left hand side is $T^2p \times 1$. One can give a condition under which this relationship cannot hold, and thus $\mathbf{\Gamma}$ calculated for the unrestricted elements of $\boldsymbol{\theta}$ must be of full rank. Assume $T \geq 3$. For the 2×2 submatrix $\mathbf{M}$ of terms from the north east of $\mathbf{M}^{\dagger}$ one finds

$$\mathbf{M}(\mathbf{I}_2 \otimes \boldsymbol{\beta}^*) = \mathbf{g}^*\mathbf{f}' + \mathbf{gf}^{*\prime}, \tag{6.28}$$

where the terms on the right now each consist of two elements of the original vectors on the right of (6.27), dated $1, 2$ for both $\mathbf{g}$ vectors and $T-1, T$ for the $\mathbf{f}$ vectors. Exploiting the conditions $f_T = 1$, $f_T^* = 0$, one can show that $(\mathbf{M}^{(1)} - f_{T-1}\mathbf{M}^{(2)})\boldsymbol{\beta}^* = f_{T-1}^*\mathbf{g}$ where $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(2)}$ are the first and second blocks of K columns of $\mathbf{M}$, respectively. Thus $\mathbf{\Gamma}$ being not of full rank implies that the subvector $\mathbf{g} \in Span(\mathbf{M}^{(1)} - f_{T-1}\mathbf{M}^{(2)})$ i.e the $2p \times 1$ vector $\mathbf{g}$ is a linear combination of the K columns of $\mathbf{M}^{(1)} - f_{T-1}\mathbf{M}^{(2)}$. Thus:

Identification in the canonical case with one factor *Assume $T \geq 3$. Then $\mathbf{\Gamma}$ has full rank in the case of one factor if $\mathbf{g}_1 \neq 0$, $f_T = 1$ and*

$$\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} \notin Span(\mathbf{M}^{(1)} - f_{T-1}\mathbf{M}^{(2)}) \tag{6.29}$$

at the true values of the parameters.

As a specific example of the canonical case, consider a single lagged dependent variable with this (and its lags) as the instrument and assume $0 < |\beta| < 1$. The model is

$$y_{it} = \beta y_{it-1} + \lambda_i f_t + \varepsilon_{it}. \tag{6.30}$$

If one assumes that the observed data are generated by a process beginning in the distant past, this can be solved as

$$\begin{aligned} \mathrm{y}_{it} &= \lambda_i (I - \beta L)^{-1} f_t + (I - \phi L)^{-1} \varepsilon_{it} && (6.31) \\ &= \lambda_i f_t^c + \eta_{it}, && (6.32) \end{aligned}$$

where the $f_t^c = (I - \beta L)^{-1} f_t$ are redefined factors and η_{it} is a stationary AR(1) (if the ε_{it} are homoskedastic). If we assume λ_i and ε_{it} are independent, it follows that

$$m_{st}^{\dagger} = E(y_{is-1} y_{it}) = \sigma_\lambda^2 f_t^c f_{s-1}^c + \sigma_\eta^2 \beta^{|t-s+1|}, \;\; s = 1, \ldots, t; \; t = 1, \ldots, T. \quad (6.33)$$

One has as well that

$$g_s = E(\lambda_i y_{is-1}) = \sigma_\lambda^2 f_{s-1}^c. \quad (6.34)$$

Using these formulae, one can show $\mathbf{\Gamma}$ has full rank unless

$$\begin{bmatrix} f_0^c \\ f_1^c \end{bmatrix} \propto \begin{bmatrix} \beta \\ 1 \end{bmatrix}. \quad (6.35)$$

If this condition is false the structural parameter of the AR(1) model is identified.

There is a somewhat more complicated version of (6.29) for the multi-factor case. If this condition is satisfied then Assumptions 1-5 can be taken to hold (save for $\mathbf{\Delta}$ being full rank) and hence the distributional result; since the spanning condition has been demonstrated, the equivalence of restricted and unrestricted estimation may be invoked in the canonical case. One caveat is that the condition (6.29) is not in terms of primitive parameters (i.e. those giving a complete description of the stochastic process generating the data) so it is possible in principle that the condition is in fact vacuous. We have shown this is not the case for the AR(1).

Appendix III: Derivatives

We shall derive the gradient function and the Hessian for a number of FIV models. The notation will be as follows. If $\mathbf{A}(\boldsymbol{\theta})$ is a (column) vector-valued function of $\boldsymbol{\theta}$ then $D_{\boldsymbol{\theta}}\mathbf{A}(\boldsymbol{\theta}) = \partial\mathbf{A}/\partial\boldsymbol{\theta}'$. If $\mathbf{A}$ is a matrix then $\mathrm{D}_{\boldsymbol{\theta}}\mathbf{A}(\boldsymbol{\theta}) = \partial\mathrm{vec}(\mathbf{A})/\partial\boldsymbol{\theta}'$. The chain rule takes the form $\mathrm{D}_{\boldsymbol{\theta}}(\mathbf{A}(\mathbf{B}(\boldsymbol{\theta}))) = \mathrm{D}_{\mathrm{vec}\mathbf{B}}(\mathbf{A}(\mathbf{B}))\mathrm{D}_{\boldsymbol{\theta}}\mathbf{B}$. The product rule is

$$\mathrm{D}_{\boldsymbol{\theta}}(\mathbf{A}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})) = (\mathbf{B}' \otimes \mathbf{I}_m)\mathrm{D}_{\boldsymbol{\theta}}\mathbf{A} + (\mathbf{I}_{K+1} \otimes \mathbf{A})\mathrm{D}_{\boldsymbol{\theta}}\mathbf{B}, \quad (6.36)$$

where $\mathbf{A}$ is $m \times p$ and $\mathbf{B}$ is $p \times (K+1)$. The gradient vector is defined as $\nabla_{\boldsymbol{\theta}} \mathbf{A} = (\mathrm{D}_{\boldsymbol{\theta}} \mathbf{A})'$.

FIVU gradient vector

In this case the minimand is

$$\mathrm{Q}_{\mathbf{B}} = \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right)' \mathbf{B}_N' \mathbf{B}_N \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right), \tag{6.37}$$

where

$$\boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right) = \widehat{\mathbf{M}}_N \boldsymbol{\phi}(\boldsymbol{\beta}) - \mathbf{S}\mathrm{vec}(\mathbf{G}\mathbf{F}'). \tag{6.38}$$

This is optimised with respect to $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{f}', \mathbf{g}')'$. One has

$$\mathrm{D}_{\boldsymbol{\theta}} Q_{\mathbf{B}} = 2\boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right)' \mathbf{B}_N' \mathbf{B}_N D_{\boldsymbol{\theta}} \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right), \tag{6.39}$$

and, using (2.10),

$$\mathrm{D}_{\boldsymbol{\theta}} \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right) = \left[\begin{array}{ccc} (\widehat{\mathbf{M}}_N \mathrm{D}_{\boldsymbol{\theta}} \boldsymbol{\phi}(\boldsymbol{\beta}) & -\mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) & -\mathbf{S}(\mathbf{F} \otimes \mathbf{I}_d) \end{array}\right]. \tag{6.40}$$

The gradient vector is then calculated as

$$\nabla Q_{\mathbf{B}} = 2\left(\mathrm{D}_{\boldsymbol{\theta}} \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right)\right)' \mathbf{B}_N' \mathbf{B}_N \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right). \tag{6.41}$$

FIVR gradient vector

As a general principle, the derivatives of the restricted models can be obtained from the FIVU derivatives by use of appropriate Jacobian matrices. Assume the restrictions effect a reparameterisation $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\vartheta})$ and let $\mathbf{J}_{\boldsymbol{\vartheta}}(\boldsymbol{\theta}) = \mathrm{D}_{\boldsymbol{\vartheta}} \boldsymbol{\theta}$ be the Jacobian. Then

$$(\nabla_R Q_{\mathbf{B}}(\boldsymbol{\vartheta}))' = \partial Q_{\mathbf{B}} / \partial \boldsymbol{\vartheta}' = \partial Q_{\mathbf{B}} / \partial \boldsymbol{\theta}' \mathbf{J}_{\boldsymbol{\vartheta}}(\boldsymbol{\theta}) = (\nabla_U Q_{\mathbf{B}})' \mathbf{J}_{\boldsymbol{\vartheta}}(\boldsymbol{\theta}). \tag{6.42}$$

The FIVR minimisation is in terms of the $\boldsymbol{\vartheta}$ vector consisting of $\boldsymbol{\beta}, \mathbf{g}, \boldsymbol{\delta}$ where $\mathbf{f} = \mathbf{H}\mathbf{P}_{d,n}\mathbf{g} - \mathbf{U}\boldsymbol{\delta}$. The Jacobian matrix is given by

$$\mathbf{J} = \left[\begin{array}{ccc} \mathbf{I}_K & \mathbf{0}_{K \times nd} & \mathbf{0}_{K \times \varsigma} \\ \mathbf{K}(I_{K+1} \otimes \mathbf{P}_{d,n}\mathbf{g})\mathrm{D}_{\boldsymbol{\phi}}\boldsymbol{\beta} & \mathbf{H}(\boldsymbol{\beta})\mathbf{P}_{d,n} & -\mathbf{U} \\ \mathbf{0}_{nd \times K} & \mathbf{I}_{nd} & \mathbf{0}_{nd \times \varsigma} \end{array}\right], \tag{6.43}$$

where ς denotes the number of columns in $\mathbf{U}$, defined in (4.2).

Second derivatives

Write $Q_{\mathbf{B}} = \mathbf{u}'\mathbf{u}$ where $\mathbf{u} = \mathbf{B}_N \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right)$. For any parameter vector $\boldsymbol{\theta}$ one has

$$\nabla Q_{\mathbf{B}} = 2\frac{\partial \mathbf{u}'}{\partial \boldsymbol{\theta}}\mathbf{u}, \tag{6.44}$$

so

$$\begin{aligned} \mathrm{D}^2_{\boldsymbol{\theta}} Q_{\mathbf{B}} &= \mathrm{D}_{\boldsymbol{\theta}} \nabla Q_{\mathbf{B}} & (6.45)\\ &= 2\mathrm{D}_{\boldsymbol{\theta}}[\frac{\partial \mathbf{u}'}{\partial \boldsymbol{\theta}}\mathbf{u}] & (6.46)\\ &= 2[(\mathbf{u}' \otimes \mathbf{I}_{\dim \boldsymbol{\theta}})\mathrm{D}_{\boldsymbol{\theta}}(\frac{\partial \mathbf{u}'}{\partial \boldsymbol{\theta}}) + (\mathrm{D}_{\boldsymbol{\theta}}\mathbf{u})'(\mathrm{D}_{\boldsymbol{\theta}}\mathbf{u}). & (6.47) \end{aligned}$$

Denote the first term within the brackets $\boldsymbol{\Upsilon}(\boldsymbol{\theta})$. One can show that

$$\boldsymbol{\Upsilon} = \sum_{j=1}^{\dim \mathbf{u}} u_j \left(\mathrm{D}^2_{\boldsymbol{\theta}} Q_{\mathbf{B}}\right) u_j, \tag{6.48}$$

where u_j denotes the j^{th} element of $\mathbf{u}$. For both FIVU and FIVR the $\mathbf{u}$ vector is linear in the stochastic term $\widehat{\mathbf{M}}_N \boldsymbol{\phi}$ so the second derivatives are nonstochastic functions of $\boldsymbol{\theta}$. Since the $\mathbf{u}$ vector is zero in expectation at the true value of the parameter vector, in GMM-type models we have

$$E\left(\mathrm{D}^2_{\boldsymbol{\theta}} Q_{\mathbf{B}}\right) = E\left[(\mathrm{D}_{\boldsymbol{\theta}}\mathbf{u})'(\mathrm{D}_{\boldsymbol{\theta}}\mathbf{u})\right], \tag{6.49}$$

which suggests that the non-negative matrix $(\mathrm{D}_{\boldsymbol{\theta}}\mathbf{u})'(\mathrm{D}_{\boldsymbol{\theta}}\mathbf{u})$ may give a good approximation to the Hessian close to convergence.

FIVU second derivatives in the canonical case.

For the FIVU residual vector $\boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right)$, write $\boldsymbol{\psi}^* = \mathbf{B}'_N \mathbf{B}_N \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right)$ and section it into $p \times 1$ submatrices so that $\boldsymbol{\psi}^* = (\boldsymbol{\psi}^{*\prime}_1, ..., \boldsymbol{\psi}^{*\prime}_{T(T+1)/2})'$. Create a $T \times T$ upper semi-triangular matrix $\boldsymbol{\Upsilon}^*$, with dimensions $pT \times T$, from these submatrices so that $\mathrm{vech}(\boldsymbol{\Upsilon}^*) = \boldsymbol{\psi}^{*}$. Then one can show that

$$\boldsymbol{\Upsilon}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{0}_{K\times K} & \mathbf{0}_{K\times nT} & \mathbf{0}_{K\times npT} \\ \mathbf{0}_{nT\times K} & \mathbf{0}_{nT\times nT} & \mathbf{I}_n \otimes \boldsymbol{\Upsilon}^{*\prime} \\ \mathbf{0}_{npT} & \mathbf{I}_n \otimes \boldsymbol{\Upsilon}^{*} & \mathbf{0}_{npT\times npT} \end{bmatrix}. \tag{6.50}$$

The Hessian for FIVU is thus

$$\mathrm{D}^2_{\boldsymbol{\theta}} Q_{\mathbf{B}} = \boldsymbol{\Upsilon} + (\mathrm{D}_{\boldsymbol{\theta}}\mathbf{u})'(\mathrm{D}_{\boldsymbol{\theta}}\mathbf{u}). \tag{6.51}$$

It is easy to see that the eigenvalues of $\boldsymbol{\Upsilon}$ are $\pm\sqrt{\mu_j}$, $j = 1, ..., nT$ (plus zero), where the μ_j are the eigenvalues of $\boldsymbol{\Upsilon}^{*\prime}\boldsymbol{\Upsilon}^*$. Thus the positivity of the Hessian is not assured in (6.51). In fact, observe that the second term is independent of $\boldsymbol{\beta}$ (see (6.40)), whereas the first term is not. If one imagines a scale increase in $\boldsymbol{\beta}$ then eventually the first term will grow as the square of the expansion factor and the resulting Hessian will have saddlepoints. This shows that an original bad approximation to $\boldsymbol{\beta}$ may lead to problems with algorithms based on the unmodified Hessian.

Concentrations.

For FIVU one has

$$\mathbf{u} = \mathbf{B}_N \boldsymbol{\psi}\left(\widehat{\mathbf{M}}_N; \boldsymbol{\theta}\right) = \mathbf{B}_N(\widehat{\mathbf{M}}_N \boldsymbol{\phi} - \mathbf{S}\mathrm{vec}(\mathbf{G}\boldsymbol{F}')). \tag{6.52}$$

By use of (2.10) one has

$$\mathbf{u} = \mathbf{B}_N \begin{bmatrix} \widehat{\mathbf{M}}_N & -\mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi} \\ \mathbf{f} \end{bmatrix} = \mathbf{B}_N \begin{bmatrix} \widehat{\mathbf{M}}_N & -\mathbf{S}(\mathbf{F} \otimes \mathbf{I}_d) \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi} \\ \mathbf{g} \end{bmatrix}. \tag{6.53}$$

These relationships imply that, given $\mathbf{F}$ one can minimise the criterion function by a one pass linear regression, and similarly for $\mathbf{G}$. Iterating these procedures will produce a declining sequence of values of the criterion which usually in practice converges to a local minimum. As a general rule in FIVU estimation we use these concentrations as they are much swifter than line-search methods based on the Hessian. No such concentrations are available for FIVR as, after substituting out for $\mathbf{f}$, the resulting residual vector $\mathbf{u}$ is quadratic in $\mathbf{g}$, so there we are forced to rely on Hessian methods.

Table 1: *Monte Carlo results,* $\boldsymbol{\rho} = \mathbf{0.5}$

SNR	F_λ	$FIVU_1$	$FIVU_2$	$FIVR_1$	$FIVR_2$	π	DIF_a	DIF_b	SYS_a	SYS_b
					$\boldsymbol{\alpha} = \mathbf{0.5}$					
3	1/4	.498 (.031) [.031] {.068}	.499 (.025) [.025] {.081} |.033|	.501 (.025) [.025] {.061}	.500 (.021) [.021] {.076} |.035|	.891	.403 (.085) [.129] {.441} |.481|	.442 (.062) [.084] {.431} |.231|	.455 (.059) [.074] {.287} |.342|	.471 (.044) [.053] {.346} |.080|
3	3/4	.499 (.029) [.029] {.059}	.498 (.026) [.026] {.077} |.038|	.500 (.023) [.023] {.063}	.499 (.019) [.019] {.074} |.041|	.942	.365 (.135) [.191] {.772} |.992|	.288 (.144) [.256] {.907} |.823|	.360 (.126) [.188] {.805} |.975|	.367 (.122) [.180] {.834} |.479|
9	1/4	.499 (.019) [.019] {.065}	.500 (.017) [.017] {.077} |.045|	.500 (.014) [.014] {.058}	.500 (.013) [.013] {.074} |.042|	.932	.494 (.020) [.021] {.105} |.453|	.472 (.038) [.048] {.300} |.223|	.489 (.026) [.029] {.165} |.366|	.488 (.024) [.027] {.201} |.069|
9	3/4	.500 (.018) [.018] {.056}	.500 (.017) [.017] {.078} |.051|	.500 (.012) [.012] {.064}	.500 (.011) [.011] {.081} |.046|	.967	.424 (.100) [.126] {.734} |.999|	.302 (.133) [.239] {.906} |.669|	.367 (.104) [.160] {.822} |.945|	.378 (.104) [.160] {.332} |.040|
					$\boldsymbol{\beta} = \mathbf{0.5}$					
3	1/4	.497 (.027) [.027] {.031}	.498 (.025) [.025] {.073}	.501 (.022) [.022] {.043}	.498 (.020) [.020] {.078}		.592 (.076) [.012] {.390}	.574 (.061) [.096] {.388}	.570 (.068) [.098] {.315}	.562 (.057) [.084] {.369}
3	3/4	.502 (.024) [.024] {.058}	.501 (.025) [.025] {.069}	.503 (.021) [.021] {.055}	.502 (.017) [.017] {.074}		.662 (.079) [.180] {.847}	.875 (.107) [.390] {.986}	.899 (.126) [.381] {.970}	.848 (.118) [.368] {.975}
9	1/4	.501 (.014) [.014] {.042}	.500 (.014) [.014] {.064}	.500 (.012) [.012] {.058}	.499 (.011) [.011] {.069}		.503 (.012) [.012] {.057}	.527 (.032) [.042] {.242}	.518 (.030) [.035] {.136}	.520 (.029) [.035] {.204}
9	3/4	.501 (.013) [.013 {.063}	.499 (.014) [.014] {.066}	.500 (.010) [.010] {.055}	.499 (.010) [.010] {.068}		.555 (.045) [.071] {.605}	.778 (.106) [.297] {.973}	762 (.125) [.290] {.935}	.749 (.114) [.274] {.955}

$N = 150$; $T = 10$; $n_0 = 1$. Results are reported in terms of average point estimates (standard deviation) [RMSE] {size} and |size of J statistic| computed at the 5% level. π denotes the proportion of times the true number of factors is selected for FIVU_2.

Table 2: *Monte Carlo results,* $\boldsymbol{\rho} = \mathbf{0.95}$

SNR	F_λ	$FIVU_1$	$FIVU_2$	$FIVR_1$	$FIVR_2$	π	DIF_a	DIF_b	SYS_a	SYS_b
					$\boldsymbol{\alpha} = \mathbf{0.5}$					
		.493	.496	.497	.498		.401	.394	.420	.428
		(.055)	(.053)	(.047)	(.043)		(.067)	(.071)	(.066)	(.061)
3	1/4	[.055]	[.053]	[.047]	[.043]	.883	[.119]	[.127]	[.104]	[.094]
		{.059}	{.071}	{.063}	{.079}		{.498}	{.616}	{.451}	{.498}
			|.032|		|.034|		|.689|	|.339|	|.495|	|.130|
		.502	.503	.498	.499		.189	.202	.286	.294
		(.053)	(.051)	(.050)	(.040)		(.143)	(.155)	(.157)	(.155)
3	3/4	[.053]	[.051]	[.050]	[.040]	.951	[.343]	[.336]	[.265]	[.258]
		{.045}	{.069}	{.055}	{.064}		{.922}	{.931}	{.846}	{.858}
			|.039|		|.035|		|.994|	|.948|	|.997|	|.997|
		.497	.499	.498	.499		.189	.202	.286	.294
		(.044)	(.043)	(.041)	(.030)		(.143)	(.155)	(.157)	(.155)
9	1/4	[.044]	[.043]	[.041]	[.030]	.920	[.344]	[.336]	[.265]	[.258]
		{.059}	{.073}	{.047}	{.063}		{.922}	{.948}	{.846}	{.858}
			|.041|		|.048|		|1.00|	|.931|	|.997|	|.755|
		.499	.501	.499	.499		.250	.214	.329	.335
		(.045)	(.038)	(.044)	(.031)		(.119)	(.155)	(.132)	(.135)
9	3/4	[.045]	[.038]	[.044]	[.031]	.969	[.276]	[.325]	[.217]	[.214]
		{.057}	{.075}	{.048}	{.066}		{.917}	{.934}	{.794}	{.803}
			|.045|		|.054|		|1.00|	|.938|	|.847|	|.046|
					$\boldsymbol{\beta} = \mathbf{0.5}$					
		.504	.502	.503	.500		.815	.773	.633	.627
3	1/4	(.052)	(.050)	(.049)	(.042)		(.173)	(.141)	(.082)	(.075)
		[.052]	[.0508]	[.049]	[.042]		[.359]	[.307]	[.156]	[.148]
		{.043}	{.078}	{.055}	{.075}		{.671}	{.717}	{.556}	{.601}
		.498	.501	.503	.501		.889	.876	.776	.776
3	3/4	(.047)	(.043)	(.043)	(.039)		(.135)	(.130)	(.145)	(.144)
		[.047]	[.043]	[.043]	[.039]		[.412]	[.398]	[.312]	[.311]
		{.055}	{.068}	{.046}	{.075}		{.968}	{.979}	{.884}	{.905}
		.502	.501	.501	.500		.889	.876	.776	.776
9	1/4	(.037)	(.030)	(.033)	(.026)		(.135)	(.130)	(.145)	(.144)
		[.037]	[.030]	[.033]	[.026]		[.412]	[.398]	[.312]	[.311]
		{.054}	{.073}	{.049}	{.072}		{.968}	{.979}	{.884}	{.905}
		.501	.500	.501	.500		.862	.867	.745	.738
9	3/4	(.036)	(.030)	(.029)	(.024)		(.138)	(.131)	(.139)	(.133)
		[.036]	[.030]	[.029]	[.024]		[.388]	[.390]	[.282]	[.273]
		{.061}	{.068}	{.053}	{.072}		{.957}	{.984}	{.892}	{.897}

$N = 150$; $T = 10$; $n_0 = 1$. Results are reported in terms of average point estimates (standard deviation) [RMSE] {size} and |size of J statistic| computed at the 5% level. π denotes the proportion of times the true number of factors is selected for $FIVU_2$.

Table 3: *Monte Carlo results,* $\boldsymbol{\rho} = \mathbf{0.5}$

SNR	F_λ	$FIVU_1$	$FIVU_2$	$FIVR_1$	$FIVR_2$	π	DIF_a	DIF_b	SYS_a	SYS_b
					$\boldsymbol{\alpha} = \mathbf{0.8}$					
3	1/4	.797 (.039) [.039] {.064}	.799 (.030) [.030] {.078} |.031|	.801 (.022) [.022] {.061}	.801 (.024) [.024] {.074} |.034|	.917	.713 (.110) [.140] {.377} |.644|	.673 (.104) [.164] {.599} |.285|	.756 (.047) [.052] {.224} |.421|	.780 (.042) [.046] {.268} |.095|
3	3/4	.798 (.045) [.045] {.074}	.799 (.040) [.040] {.083} |.041|	.801 (.030) [.030] {.066}	.802 (.025) [.025] {.082} |.042|	.952	.508 (.211) [.360] {.910} |.999|	.491 (.203) [.370] {.950} |.925|	.673 (.143) [.191] {.808} |.991|	.685 (.135) [.178] {.831} |.677|
9	1/4	.800 (.020) [.020] {.059}	.802 (.016) [.016] {.074} |.059|	.798 (.014) [.014] {.063}	.799 (.012) [.012] {.070} |.056|	.937	.777 (.042) [.048] {.163} |.500|	.743 (.060) [.083] {.388} |.270|	.796 (.017) [.018] {.115} |.384|	.796 (.017) [.017] {.134} |.082|
9	3/4	.801 (.023) [.023] {.078}	.801 (.020) [.020] {.088} |.072|	.800 (.011) [.011] {.066}	.799 (.011) [.011] {.079} |.067|	.963	.588 (.201) [.292] {.856} |.999|	.526 (.205) [.342] {.930} |.780|	.722 (.102) [.124] {.806} |.961|	.728 (.097) [.121] {.840} |.467|
					$\boldsymbol{\beta} = \mathbf{0.2}$					
3	1/4	.197 (.022) [.022] {.051}	.198 (.021) [.021] {.067}	.202 (.019) [.019] {.055}	.201 (.018) [.018] {.064}		.216 (.044) [.047] {.146}	.272 (.062) [.095] {.334}	.265 (.060) [.088] {.292}	.267 (.053) [.086] {.384}
3	3/4	.198 (.022) [.022] {.060}	.198 (.020) [.020] {.081}	.201 (.019) [.019] {.049}	.201 (.017) [.017] {.073}		.407 (.108) [.828] {.234}	.540 (.097) [.353] {.989}	.539 (.109) [.356] {.985}	.530 (.101) [.345] {.989}
9	1/4	.202 (.009) [.009] {.046}	.201 (.008) [.008] {.061}	.201 (.007) [.007] {.057}	.200 (.007) [.007] {.060}		.197 (.011) [.012] {.081}	.206 (.014) [.015] {.094}	.204 (.013) [.014] {.080}	.206 (.012) [.013] {.111}
9	3/4	.198 (.009) [.009] {.095}	.199 (.010) [.010] {.059}	.199 (.006) [.006] {.064}	.200 (.006) [.006] {.070}		.190 (.057) [.058] {.574}	.306 (.058) [.121] {.875}	.310 (.067) [.129] {.885}	.308 (.062) [.125] {.907}

$N = 150$; $T = 10$; $n_0 = 1$. Results are reported in terms of average point estimates (standard deviation) [RMSE] {size} and |size of J statistic| computed at the 5% level. π denotes the proportion of times the true number of factors is selected for FIVU_2.

Table 4: *Monte Carlo results,* $\boldsymbol{\rho = 0.95}$

SNR	F_λ	$FIVU_1$	$FIVU_2$	$FIVR_1$	$FIVR_2$	π	DIF_a	DIF_b	SYS_a	SYS_b
					$\boldsymbol{\alpha = 0.8}$					
		.796	.797	.798	.799		.650	.633	.751	.762
		(.046)	(.041)	(.035)	(.032)		(.095)	(.092)	(.052)	(.044)
3	1/4	[.046]	[.041]	[.035]	[.032]	.897	[.177]	[.190]	[.071]	[.058]
		{.056}	{.085}	{.063}	{.072}		{.563}	{.732}	{.306}	{.329}
			|.034|		|.036|		|.705|	|.351|	|.512|	|.143|
		.797	.799	.798	.799		.439	.458	.644	.656
		(.045)	(.042)	(.046)	(.037)		(.183)	(.190)	(.159)	(.148)
3	3/4	[.045]	[.042]	[.046]	[.037]	.957	[.405]	[.392]	[.223]	[.207]
		{.070}	{.082}	{.065}	{.073}		{.953}	{.950}	{.825}	{.839}
			|.039|		|.042|		|.991|	|.969|	|.995|	|.778|
		.798	.799	.800	.800		.757	.731	.781	.782
		(.035)	(.028)	(.026)	(.018)		(.054)	(.060)	(.035)	(.034)
9	1/4	[.035]	[.028]	[.026]	[.018]	.955	[.069]	[.092]	[.039]	[.039]
		{.054}	{.072}	{.046}	{.057}		{.180}	{.388}	{.234}	{.234}
			|.046|		|.065|		|.594|	|.276|	|.104|	|.104|
		.800	.801	.800	.800		.490	.472	.637	.650
		(.043)	(.027)	(.031)	(.017)		(.145)	(.174)	(.169)	(.158)
9	3/4	[.043]	[.027]	[.031]	[.017]	.976	[.342]	[.371]	[.235]	[.218]
		{.079}	{.083}	{.061}	{.069}		{.933}	{.949}	{.855}	{.868}
			|.056|		|.053|		|.999|	|.931|	|.987|	|.679|
					$\boldsymbol{\beta = 0.2}$					
		.206	.202	.201	.200		.554	.542	.440	.441
3	1/4	(.045)	(.043)	(.041)	(.037)		(.137)	(.129)	(.106)	(.100)
		[.045]	[.043]	[.041]	[.037]		[.379]	[.365]	[.262]	[.261]
		{.056}	{.068}	{.051}	{.059}		{.839}	{.874}	{.775}	{.855}
		.201	.201	.201	.201		.574	.569	.486	.497
3	3/4	(.029)	(.026)	(.027)	(.023)		(.125)	(.117)	(.125)	(.123)
		[.029]	[.026]	[.027]	[.023]		[.395]	[.387]	[.312]	[.321]
		{.058}	{.074}	{.052}	{.064}		{.989}	{.990}	{.944}	{.965}
		.202	.201	.200	.199		.241	.241	.215	.215
9	1/4	(.017)	(.016)	(.015)	(.012)		(.126)	(.084)	(.021)	(.020)
		[.017]	[.016]	[.015]	[.012]		[.136]	[.093]	[.026]	[.025]
		{.054}	{.062}	{.062}	{.068}		{.133}	{.195}	{.126}	{.171}
		.204	.202	.201	.200		.552	.515	.343	.336
9	3/4	(.020)	(.016)	(.017)	(.012)		(.182)	(.135)	(.098)	(.090)
		[.020]	[.016]	[.017]	[.012]		[.396]	[.343]	[.174]	[.163]
		{.065}	{.072}	{.049}	{.058}		{.888}	{.963}	{.853}	{.869}

$N = 150$; $T = 10$; $n_0 = 1$. Results are reported in terms of average point estimates (standard deviation) [RMSE] {size} and |size of J statistic| computed at the 5% level. π denotes the proportion of times the true number of factors is selected for FIVU_2.