Panel Vector Autoregression in R with the Package panelvar

Michael Sigmund[1], Robert Ferstl[2]

Abstract

In this paper we extend two general methods of moment (GMM) estimators to panel vector autoregression models (PVAR) with p lags of endogenous variables, predetermined and strictly exogenous variables. We first extend the first difference GMM estimator to this extended PVAR model. Second, we do the same for the system GMM estimator. We implement these estimators in the R package panelvar. In addition to the GMM estimators, we contribute to the empirical literature by implementing common specification tests (Hansen overidentification test, lag selection criterion and stability test of the PVAR polynomial) and classical structural analysis for PVAR models such as orthogonal and generalized impulse response functions, bootstrapped confidence intervals for impulse response analysis and forecast error variance decompositions. Finally, we implement the first difference and the forward orthogonal transformation to remove the fixed effects.

Keywords: Panel vector autoregression model, generalized method of moments, first difference and system GMM, **R**
JEL: Classification Numbers G20, G30

1. Introduction

Over the past decades important advances have been made in the study of dynamic panel data models with fixed effects for the typical setting that cross-sectional dimension (N) is large and the time dimension (T) is short. Classical OLS-based regression methods cannot be applied because of the Nickell bias (Nickell, 1981) that does not disappear asymptotically if $N \to \infty$ and T is fixed. One solution to this problem is to apply generalized method of moments estimators popularized by Hansen (1982) in economics. Important contributions are Anderson and Hsiao (1982), Holtz-Eakin et al. (1988), Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998).

Email addresses: `michael.sigmund@oenb.at` (Michael Sigmund), `robert.ferstl@wiwi.uni-regensburg.de` (Robert Ferstl)

[1]Oesterreichische Nationalbank (OeNB), Otto-Wagner-Platz 3, A-1090 Vienna, Austria.

[2]Department of Finance, University of Regensburg, 93053 Regensburg, Germany

Within the well established GMM estimators it is useful to distinguish between the first difference GMM estimator (Holtz-Eakin et al., 1988; Arellano and Bond, 1991) that uses lags of the endogenous variable(s) as instruments and the system GMM estimator (Blundell and Bond, 1998) that uses additional moment conditions based on information contained in the "levels".

First difference and system GMM estimators for single equation dynamic panel data models have been implemented in STATA: **xtabond2** (Roodman, 2009b) and some of the features are also available in the R package: **plm** (Croissant and Millo, 2008).

However, with the exception of Holtz-Eakin et al. (1988) and later Binder et al. (2005) the theoretical literature has primarily focused on single equation dynamic panel data models, whereas there are many applications that requires a simultaneous treatment of the decision problems of households, firm, banks or other economic agents.

Ever since the seminal paper of Sims (1980) on macroeconomic reality, vector autoregressive models (VAR) are considered as a starting point in economics to study models with more than one endogenous variable. These models have been extensively studied in the time series literature (see Lütkepohl, 2007; Pfaff, 2008). Again, the standard ordinary least square equation-by-equation estimation procedure for VAR models does not provide unbiased estimates for PVAR models.

The popularity of PVAR model in empirical economics (and other social sciences) is documented by over 1000 citations of Love and Zicchino (2006). They provide an unofficial STATA code that has been extended recently by Abrigo and Love (2016). Abrigo and Love (2016) use the first generation GMM estimator suggested by Anderson and Hsiao (1982) to deal with the Nickell bias. Our code implements the direct extension of the Anderson and Hsiao (1982), the first difference GMM estimator (Holtz-Eakin et al., 1988; Arellano and Bond, 1991) and the more complex system GMM by Blundell and Bond (1998) for PVAR models.

In the **panelvar** package we basically extend all features of **xtabond2** to a system of dynamic panel models. We extend the first difference GMM and system GMM estimator as laid out in Binder et al. (2005) for PVAR models with predetermined and strictly exogenous variables.

In addition to the GMM estimators we also provide structural analysis functions for PVAR models that are well established for (time-series) VAR models. These functions include orthogonal impulse response functions (see Lütkepohl, 2007), generalized impulse response functions (see Pesaran and Shin, 1998) and forecast error variance decomposition. For the impulse response function we also provide a GMM specific bootstrap method for estimating confidence intervals.

Furthermore we extend the Hansen overidentification test (Hansen, 1982; Roodman, 2009b), the model selection procedure of Andrews and Lu (2001) and the Windmeijer corrected standard errors (Windmeijer, 2005) from (single equation) dynamic panel models to PVAR models.

The paper is organized as follows: Section 2 describes the methodology behind the extended PVAR model. First, we define the extended PVAR model. Next, we derive the first difference GMM estimator for this model. In the next step we define the system GMM moment conditions and corresponding system GMM estimator. Next, the Windmeijer correction (Windmeijer, 2005) is extended to the standard errors of a PVAR model. The next subsections introduce the orthogonal impulse response function, the generalized impulse response function and confidence bands for impulse response analysis for PVAR models. The final subsection of Section 2 defines the Hansen overidentification test and the Andrews-Lu model selection procedure. In Section 3 we show how to use the **panelvar** package for dynamic panel models and for PVAR models. We also compare our code with existing software.

2. Methodology

In this section we introduce the extended PVAR model by adding predetermined and strictly exogenous variables and allowing p lags of endogenous variables. Next, we define the first difference moment conditions (Holtz-Eakin et al., 1988; Arellano and Bond, 1991), formalize the ideas to reduce the number of moment conditions by linear transformations of the instrument matrix and define the one- and two-step GMM estimator. In the following two subsections the system moment conditions are defined (Blundell and Bond, 1998) and the extended GMM estimator is presented. Extending the Windmeijer correction (Windmeijer, 2005) for the two-step GMM estimator of the variance to our PVAR model is next. After these GMM related topics, we introduce the orthogonal and generalized impulse response analysis to PVAR models. Finally, we extend the Hansen overidentification test (Hansen, 1982) and the model selection procedure of Andrews and Lu (2001) to PVAR models.

2.1. The Extended PVAR model

The first vector autoregressive panel model (PVAR) was introduced by Holtz-Eakin et al. (1988). We extend their model to allow for p lags of m endogenous variables, k predetermined variables and n strictly exogenous variables. Therefore, we consider the following stationary PVAR with fixed effects.[3]

[3]A random effects specification in a dynamic panel context is possible but requires strong assumption on the individual effects. Empirical applications mostly use a fixed-effects specification. We do not consider a random-effects implementation at this stage. See Binder et al. (2005) for more details.

$$\mathbf{y}_{i,t} = \mu_i + \sum_{l=1}^{p} \mathbf{A}_l \mathbf{y}_{i,t-l} + \mathbf{B}\mathbf{x}_{i,t} + \mathbf{C}\mathbf{s}_{i,t} + \epsilon_{i,t} \tag{1}$$

$\mathbf{I}_m$ denotes an $m \times m$ identity matrix. Let $\mathbf{y}_{i,t} \in \mathbb{R}^m$ be an $m \times 1$ vector of endogenous variables for the ith cross-sectional unit at time t. Let $\mathbf{y}_{i,t-l} \in \mathbb{R}^m$ be an $m \times 1$ vector of lagged endogenous variables. Let $\mathbf{x}_{i,t} \in \mathbb{R}^k$ be an $k \times 1$ vector of predetermined variables that are potentially correlated with past errors. Let $\mathbf{s}_{i,t} \in \mathbb{R}^n$ be an $n \times 1$ vector of strictly exogenous variables that neither depend on ϵ_t nor on ϵ_{t-s} for $s = 1, ..., T$. Moreover, the disturbances $\epsilon_{i,t}$ are independently[4] and identically distributed (i.i.d.) for all i and t with $\mathbb{E}[\boldsymbol{\epsilon}_{\mathbf{i,t}}] = 0$ and $Var[\epsilon_{i,t}] = \Sigma_\epsilon$. Σ_ϵ is a positive semidefinite matrix.

We assume that all unit roots of $\mathbf{A}$ in Eq. (1) fall inside the unit circle to assure covariance stationarity. The cross section i and the time section t are defined as follows: $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$. In this specification we assume parameter homogeneity for $\mathbf{A}_l$ $(m \times m)$, $\mathbf{B}$ $(m \times k)$ and $\mathbf{C}$ $(m \times n)$ for all i.

A PVAR model is hence a combination of a single equation dynamic panel model (DPM) and a vector autoregressive model (VAR).

2.2. *First difference moment conditions*

Before we set up the first difference estimator moment conditions, we apply the first difference or the forward orthogonal transformation[5] to Eq. (1)

$$\Delta^* \mathbf{y}_{i,t} = \sum_{l=1}^{p} \mathbf{A}_l \Delta^* \mathbf{y}_{i,t-l} + \mathbf{B}\Delta^* \mathbf{x}_{i,t} + \mathbf{C}\Delta^* \mathbf{s}_{i,t} + \Delta\epsilon_{i,t} \tag{2}$$

Δ^* either refers to the first difference or the forward orthogonal transformation. The first difference transformation exists for $t \in \{p+2, \cdots, T\}$ and the forward orthogonal transformation exists for $t \in \{p+1, \cdots, T-1\}$. We denote the set of indexes t, for which the transformation exists by $\mathbb{T}_{\Delta^*}$.

Binder et al. (2005) extend the equation-by-equation estimator of Holtz-Eakin et al. (1988) for a PVAR model with only endogenous variables that are lagged by one period. We further extend Binder et al. (2005) by adding more lags of the endogenous variables,

[4]Under certain circumstances with N and T sufficiently large, it is possible to relax this cross-sectional independence assumption. However, exploring the issue of cross-sectional dependence in the context of PVAR model is beyond the scope of this paper, but a very interesting future research topic.

[5]$y^{\perp}_{i,t+1} = c_{i,t}(y_{i,t} - 1/T_{i,t} \sum_{s>t} y_{i,s})$. Where $c_{i,t} = \sqrt{T_{i,t}/(T_{i,t}+1)}$. This transformation is suggested by Arellano and Bover (1995) to minimize data losses due to data gaps.

predetermined and strictly exogenous variables. Moreover, we follow Binder et al. (2005) by setting up the GMM conditions for each individual i.

First, we express the moment conditions for the lagged endogenous, the predetermined and the strictly exogenous variables for an individual i.[6]

Definition 2.1. *(First difference GMM moment conditions):*

$$\begin{aligned}
&\mathbb{E}[\Delta^{*}\epsilon_{i,t}\mathbf{y}_{i,j}^{\intercal}] = \mathbf{0} \quad j \in \{1, \cdots, T-2\} \textit{ and } t \in \mathbb{T}_{\Delta^{*}},\\
&\mathbb{E}[\Delta^{*}\epsilon_{i,t}\mathbf{x}_{i,j}^{\intercal}] = \mathbf{0} \quad j \in \{1, \cdots, T-1\} \textit{ and } t \in \mathbb{T}_{\Delta^{*}},\\
&\mathbb{E}[\Delta^{*}\epsilon_{i,t}\Delta^{*}\mathbf{s}_{i,t}^{\intercal}] = \mathbf{0} \quad t \in \mathbb{T}_{\Delta^{*}}
\end{aligned} \tag{3}$$

The dimension of these matrices are the following: $\Delta^{*}\epsilon_{i,t}$ is $m \times 1$, $\mathbf{y}_{i,j}$ is $m \times 1$, $\mathbf{x}_{i,j}$ is $k \times 1$ and $\Delta^{*}\mathbf{s}_{i,t}$ is $n \times 1$.

For later derivations it is useful to define $q_{i,t}$ by

$$\mathbf{q}_{i,t}^{\intercal} := (\mathbf{y}_{i,t-p-1}^{\intercal}, \mathbf{y}_{i,t-p-2}^{\intercal}, ..., \mathbf{y}_{i,1}^{\intercal}, \mathbf{x}_{i,t-1}^{\intercal}, \mathbf{x}_{i,t-2}^{\intercal}, ..., \mathbf{x}_{i,1}^{\intercal}, \Delta^{*}\mathbf{s}_{i,t}^{\intercal}) \quad t \in \{p+2, \cdots, T\}$$

and stacking over t results in:

$$\mathbf{Q}_i := \begin{pmatrix} \mathbf{q}_{i,p+2}^{\intercal} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{q}_{i,p+3}^{\intercal} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{q}_{i,T}^{\intercal} \end{pmatrix} \tag{4}$$

Next, we redefine Eq. (2) by stacking over t:

$$\Delta^{*}\mathbf{Y}_i = \sum_{l=1}^{p} \Delta^{*}\mathbf{Y}_{i,l}A_l^{\intercal} + \Delta^{*}\mathbf{X}_i B^{\intercal} + \Delta^{*}\mathbf{S}_i C^{\intercal} + \Delta^{*}\mathbf{E}_i \tag{5}$$

$\Delta^{*}\mathbf{Y}_i$, $\Delta^{*}\mathbf{Y}_{i,l}$ and $\Delta^{*}\mathbf{E}_i$ are an $(T-1-p) \times m$ matrices. A, B and C have the same dimension as in Eq (1). $\Delta^{*}\mathbf{X}_i$ is $(T-1-p) \times k$. $\Delta^{*}\mathbf{S}_i$ is $(T-1-p) \times n$.

Based on Eq. (5), we finally set up the stacked moment conditions for each i.

$$\mathbb{E}[\mathbf{Q}_i^{\intercal}(\Delta^{*}\mathbf{E}_i)] = 0 \tag{6}$$

With the stacked moment conditions, we are able to provide the next definition.

[6] $\Delta^{*}\epsilon_{i,t}$ follows from rewriting Eq. (2).

Definition 2.2. *(First difference GMM moment function):*

$$\begin{aligned} \hat{\mathbf{g}}(\mathbf{\Phi}) &= \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \hat{\mathbf{g}}_{\mathbf{i}}(\mathbf{\Phi}) \\ \hat{\mathbf{g}}_{\mathbf{i}}(\mathbf{\Phi}) &= (\mathbf{Q}_i \otimes \mathbf{I}_{m\times m})\,(vec(\Delta^{*}\mathbf{E}_i)) \end{aligned} \tag{7}$$

The first difference moment function states that the sample average $\hat{\mathbf{g}}_{\mathbf{i}}(\mathbf{\Phi})$ *is close to its true parameter* $\mathbf{\Phi_0}$ *in Eq.* (6)*, given that* $\Delta^{*}\mathbf{E}_i = \Delta^{*}\mathbf{W}_i - \Delta^{*}\mathbf{W}_{minus,i}\mathbf{\Phi}_0$*, where* $\mathbf{W}_i = \mathbf{Y}_i$ *and* $\Delta^{*}\mathbf{W}_{minus} = (\Delta^{*}\mathbf{Y}_{i,-1}, \Delta^{*}\mathbf{X}_i, \Delta^{*}\mathbf{S}_i)$.

2.3. A note on the number of first difference moment conditions

Looking at $\mathbf{Q}_i$ in Eq. (4) shows that the dimension is very large. In particular, the dimension of $\mathbf{Q}_i$ depends on the number of lags of the endogenous variables (p), the number of endogenous variables (m), the number of predetermined variables (k) and the number of strictly exogenous variables (n).

Proposition 2.1. *(Dimension of* $\mathbf{Q}_i$*):*
The dimension consists of three parts which can be attributed to

1. *the lagged endogenous variables,*
2. *the predetermined variables and*
3. *the strictly exogenous variables.*

For each of the lagged endogenous variables and the predetermined variables the number of moment conditions follow the sum of two parts: The first part is an arithmetic progressions $S_1 = \frac{n_1}{2} * (a_1 + (n_1 - 1) * d)$ *where one has to identify the value of the first element* a_1*, the number of elements* n_1 *and the difference between two consecutive elements* $d = 1$*. The second part consists of the sum of two elements* a_2 *and* n_2*:* $S_2 = a_2 * n_2$ *and apply if the number of moment conditions are restricted before a certain lag* (L^{endo}_{min} *and* L^{pre}_{min}) *or after a certain lag* (L^{endo}_{max} *and* L^{pre}_{max}).

$$\begin{aligned} \dim(\mathbf{Q}_i) &= (T - 1 - p) \times S \\ S &= S_m + S_k + S_n \end{aligned} \tag{8}$$

where S_m*,* S_k *and* S_n *are defined as follows:*

$$
\begin{aligned}
&S_m = m * (S_m(1) + S_m(2)) \\
&S_m(1) = n_1/2 * (2 * a_1 + (n_1 - 1) * d_1) \\
&a_1 = p - \min\left(L_{min}^{endo} - 2, p - 1\right) \\
&n_1 = \max\left(\min(L_{max}^{endo}, T - 2) - \max(L_{min}^{endo} - 2, p - 1), 1\right) \\
&d_1 = 1 \\
&S_m(2) = a_2 * n_2 \\
&a_2 = a_1 + n_1 - 1 \\
&n_2 = T - n_1 - \left(\min(p, 2) + \max\left(L_{min}^{endo} - p, p - 1\right)\right)
\end{aligned} \tag{9}
$$

$$
\begin{aligned}
S_k &= k * (S_k(1) + S_k(2)) \\
S_k(1) &= o_1/2 * (2 * b_1 + (o_1 - 1) * e_1) \\
b_1 &= p + 1 - \max\left(L_{min}^{endo} - 1, p - 1\right) \\
o_1 &= \max(\min(T - 2, L_{max}^{pre}) - \max(L_{min}^{pre} - 1, p - 1), \\
&\quad \min(T - p, L_{max}^{pre}) + 1 - b_1, 1) \\
e_1 &= 1 \\
S_k(2) &= o_2 * b_2 \\
o_2 &= T - o_1 - \left(\min(p, 1) + max(L_{min}^{pre} - p, p)\right) \\
b_2 &= b_1 + (o_1 - 1)
\end{aligned} \tag{10}
$$

$$
S_n = n \tag{11}
$$

Proof. Given Definition 2.1, it is clear that every added time period t increases the number of available instruments for the endogenous and predetermined variables in the later periods by one. The starting value a_1 depends on the number of lags and L_{min}^{endo} and is usually one. Identifying the starting value for L_{min}^{pre} is similar. These observations proof that $S_m(1)$ and $S_k(1)$ follow a arithmetic progression. If L_{max}^{endo} and/or L_{max}^{pre} are binding (being smaller than T), then the constant parts $S_m(2)$ and $S_k(2)$ are added.

□

We illustrate the number of moment condition based on Eq. (8) with a simple example in Appendix B. As the dimension of $\mathbf{Q}_i$ signals, there are many moment conditions to identify a possibly small number of parameters. There are a number of reasons why the reduction of the number of moment conditions could be of major importance in many cases.

First, some authors detect the problem of instrument proliferation (Roodman, 2009a). Instrument proliferation is intrinsic in GMM estimation of dynamic panel models when all the lags of the endogenous explanatory variables (predetermined variables) are exploited, as the number of moment conditions increases with T and with the dimension of the vectors of endogenous regressors and predetermined variables. Newey and Smith (2004) showed that this problem is due to higher order properties of GMM estimators that are related to the choice of weighting matrix of the two-step estimator (see Section 2.4 for more details).

Second, a related problem is concerned with asymptotic theory. Alvarez and Arellano (2003) also showed that it is possible to get consistency of GMM estimators in panel with large T ($N \to \infty$ and $T \to \infty$) under the following condition: $log(T)^2/N \to 0$ provided that the optimal weighting matrix is used. In practical, for large T, if we can fix the number of moment conditions q after $T > c$, the first difference GMM remains consistent. We can also rely on an alternative proof of Koenker and Machado (1999) that the first difference GMM estimator remains consistent and asymptotically normally distributed in the case of $T \to \infty$. Koenker and Machado (1999) state the following condition: $q^3/N \to 0$.

Third, reducing the number of moment conditions is necessary for very large panel data sets to make an estimation computationally feasible.

As a consequence, the single equation dynamic panel literature offers two possibilities to reduce this number. The first idea is to reduce the number of moment conditions by fixing a maximal lag length L_{max} after which no further instruments are used, if available. It is less popular but statistically possible not to start with the first possible lag of the instruments L_{min} but with a deeper lag. Both parts of the first idea can be implemented. Only the L_{max} part of the first idea is formulated by Mehrhoff (2009) for single equation dynamic panel models. We extend his idea by adding the L_{min} and the extension to PVAR models.

Due to different L_{min} assumptions for the lagged endogenous ($L^{endo}_{min} = 2$) and the predetermined variables ($L^{pre}_{min} = 1$) we implemented two transformation matrices in our code but only explain the transformation matrix for the "endogenous" block of the instrument matrix $\mathbf{Q}_i$. The transformation matrix for the "predetermined" block of the instrument matrix is derived in a similar fashion.

In principle, we apply a linear transformation on the matrix $\mathbf{Q}^{endo}_i$ to reduce the number of rows and columns (that have non-zero elements). For each $\mathbf{q}^{endo}_{i,t}$ in Eq. (4) there is a corresponding transformation matrix $\mathbf{f}^L_{i,j}$ which are identity matrices of growing dimension depending on the number of periods T, the lags p, L_{min} and L_{max}. The columns of the identity matrices are cut off in the following way:

$$\mathbf{f}^L_{i,j} = \left\{ \begin{array}{lll} (1,..,j) \times (1,..,j) & \text{if} & j < L^{endo}_{max}, L^{endo}_{min} = 2 \\ (1,..,j) \times (1,..,L^{endo}_{max}) & \text{if} & j \geq L^{endo}_{max}, L^{endo}_{min} = 2 \\ (1,...,j) \times (L^{endo}_{min},..,j) & \text{if} & j < L^{endo}_{max}, L^{endo}_{min} > 2 \\ (1,...,j) \times (L^{endo}_{min},..,L^{endo}_{max}) & \text{if} & j \geq L^{endo}_{max}, L^{endo}_{min} > 2 \end{array} \right\}$$

The $\mathbf{f}^L_{i,j}$ are defined for each $j \in \{p, ..., T-2\}$, where p is the lag order of the PVAR model. The $\mathbf{f}^L_{i,j}$ are block-diagonalized to complete the full transformation matrix $\mathbf{F}^L_i$:

$$\mathbf{F}_i^L = \begin{pmatrix} \mathbf{f}_{i,j}^L & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{f}_{i,j+1}^L & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{f}_{i,T-2}^L \end{pmatrix} \otimes \mathbf{I}_{m\times m}$$

The second idea to reduce the number of moment conditions is called collapsing of instruments. This idea has also been used by Roodman (2009a) in the context of single equation dynamic panel models.

Definition 2.3. *(Collapsed first difference GMM moment conditions):*

$$\begin{array}{l} \mathbb{E}[\sum_{j=1}^{T-2} \Delta^* \epsilon_{i,t} \mathbf{y}_{i,j}^\intercal] = \mathbf{0} \quad t \in \mathbb{T}_{\Delta^*}, \\ \mathbb{E}[\sum_{j=1}^{T-1} \Delta^* \epsilon_{i,t} \mathbf{x}_{i,j}^\intercal] = \mathbf{0} \quad t \in \mathbb{T}_{\Delta^*}, \\ \mathbb{E}[\Delta^* \epsilon_{i,t} \Delta^* \mathbf{s}_{i,t}^\intercal] = \mathbf{0} \quad t \in \mathbb{T}_{\Delta^*} \end{array} \tag{12}$$

The collapsed version of Eq. (4) reduced Q_i to:

$$\mathbf{Q}_i^{collapse} := \begin{pmatrix} \mathbf{q}_{i,p+2}^\intercal & \mathbf{q}_{i,p+3}^\intercal & \cdots & \mathbf{q}_{i,T}^\intercal \\ \mathbf{0} & & & \\ \vdots & \mathbf{0} & \ddots & \\ \mathbf{0} & \vdots & & \\ \mathbf{0} & \mathbf{0} & \cdots & \end{pmatrix} \tag{13}$$

$\mathbf{Q}_i^{collapse}$ reduced to a $(T-2) \times (T-2)$ matrix.

Following Mehrhoff (2009), the transformation matrix for collapsing the instrument for the lagged endogenous variables is made up of identity matrices of increasing dimension stacked one upon the other with blocks of zero matrices to the right.[7]

$$\mathbf{F}_i^C = \begin{pmatrix} \mathbf{f}_{i,j}^L & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{f}_{i,j+1}^L & & \mathbf{0} & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{f}_{i,T-2}^L & & & \end{pmatrix} \otimes \mathbf{I}_{m\times m} \tag{14}$$

Not surprisingly, both ideas to reduce the number of moment conditions can be applied at the same time as $\mathbf{F}_i^L$ and $\mathbf{F}_i^C$ have the same "block" elements.

[7] Again, the matrix for collapsing the predetermined variables is derived in similar fashion.

2.4. GMM estimator with first difference moment conditions

Based on the moment conditions in Eq. (6) and the derived set of instruments $\mathbf{Q}_i$, we can now formulate the following value function:

$$\begin{aligned}\Pi(\mathbf{\Phi}) \quad = \quad & \left(\sum_{i=1}^{N} \mathbf{Z}_i^\intercal \text{vec}(\Delta^*\mathbf{Y}_i - \begin{bmatrix}\Delta^*\mathbf{Y}_{i,-1} & \Delta^*\mathbf{X}_i & \Delta^*\mathbf{S}_i\end{bmatrix} \mathbf{\Phi})\right)^\intercal \mathbf{\Lambda}_Z^{-1} \\ & \left(\sum_{i=1}^{N} \mathbf{Z}_i^\intercal \text{vec}(\Delta^*\mathbf{Y}_i - \begin{bmatrix}\Delta^*\mathbf{Y}_{i,-1} & \Delta\mathbf{X}_i & \Delta^*\mathbf{S}_i\end{bmatrix} \mathbf{\Phi})\right)\end{aligned} \tag{15}$$

Where $\mathbf{\Phi}$ is defined as $\begin{bmatrix}\mathbf{A} & \mathbf{B} & \mathbf{C}\end{bmatrix}$ which is a $m \times (p+k+n)$ matrix. Λ_Z is the GMM weighting matrix. We also define $\Delta^*\mathbf{W}_{minus} := \left(\Delta^*\mathbf{Y}_{i,-1} \quad \Delta^*\mathbf{X}_i \quad \Delta^*\mathbf{S}_i\right)$ and $\Delta^*\mathbf{W}_i = \Delta^*\mathbf{Y}_i$.

In order to show that there exists a consistent estimator $\hat{\Phi}$ that minimizes Eq. (15) we adapt Theorem 2.7 by Newey and McFadden (1994) for our set up:[8]

Theorem 2.1. *(Consistency of* $\mathbf{\Phi}$*):*
If there is a function $\Pi_0(\mathbf{\Phi})$ *such that (i)* $\Pi_0(\mathbf{\Phi})$ *is uniquely minimized at* $\mathbf{\Phi}_0$*; (ii)* $\mathbf{\Phi}_0$ *is an element of the interior of a convex set* Θ *and* $\Pi_N(\mathbf{\Phi})$ *is convex, (iii)* $\Pi_N(\mathbf{\Phi}) \xrightarrow{p} \Pi_0(\mathbf{\Phi})$ *for all* $\mathbf{\Phi} \in \Theta$ *and (iv)* $\mathbf{\Lambda}_Z$ *is positive definite, then* $\hat{\mathbf{\Phi}}_N$ *exists with probability approaching one and* $\hat{\mathbf{\Phi}}_n \xrightarrow{p} \mathbf{\Phi}_0$.

Proof. Based on the proof of Theorem 2.7 in Newey and McFadden (1994) we need to show that (i), (ii), (iii) and (iv) hold for the value function in Eq. (15).

(i): We apply Lemma 2.3 (GMM identification) from Newey and McFadden (1994). A necessary order condition for GMM identification is that there be at least as many moment condition functions as parameters. Based on Eq. (9), even if we set $L_{max}^{endo} = L_{min}^{endo} = 2$, $L_{max}^{pre} = L_{min}^{pre} = 1$, p = 1 and T = 3 (minimal number of observations to apply the first difference GMM moment conditions), it follows that the number of rows in $vec(\Phi)$ is less or equal to the number of columns in $\mathbf{Q}_i$: We have to show that ncol$(\mathbf{Q}_i \otimes \mathbf{I}_{m\times m}) \geq p*m^2 + k*m + n$, where ncol refers to the number of columns and $p*m^2+k*m+n$ defines the number of parameters in a PVAR model. We look at the endogenous, predetermined and exogenous instruments separately. Applying Eq. (9) yields:

[8] In particular, Newey and McFadden (1994) multiply Eq. (15) by -1. Moreover, Theorem 10.8 of Rockafellar (1970) states that pointwise convergence of *convex* (instead of concave) functions on a dense subset of an open set implies uniform convergence on any compact subset of the open set.

$$
\begin{aligned}
&S_m = m * 1\\
&S_k = k * 2\\
&S_n = n\\
&S\ = m + 2k + n\\
&\mathbf{Q}_i = 1 \times (m + 2k + n)\\
&\mathrm{ncol}(\mathbf{Q}_i \otimes \mathbf{I}_{m\times m}) = m * (m + 2k + n) \geq m^2 + k * m + n
\end{aligned}
\tag{16}
$$

If we added more lags ($p > 1$), we would also need $T > 3$ for the first difference GMM estimator to be applicable. Thus, identification would be still possible. Adding the system moment conditions (Blundell and Bond, 1998) would also increase $\mathrm{ncol}(Q_i \otimes I_{m\times m})$ and therefore did not alter the identification.

(ii): Usually most textbooks and articles assume that $\mathbf{\Phi_0}$ is an element of the interior of $\mathbf{\Theta}$ to make many statistical inference techniques work. As we make the assumptions on the PVAR process in Section 1, especially on the stationarity, $\mathbf{\Phi_0}$ is in the interior of the parameter space if the model is stationary. As Eq. (7) implies the moment functions are linear in parameters and $\mathbf{\Lambda}_Z$ is positive definite (see iv), the value function as defined in Eq. (15) is convex.

(iii): According to Theorem 10.8 by Rockafellar (1970), point-wise convergence of convex functions on a dense subset of an open sets implies uniform convergence on any compact subset of the open set. It then follows that $\Pi_N(\mathbf{\Phi}) \xrightarrow{p} \Pi_0(\mathbf{\Phi})$ for all $\mathbf{\Phi} \in \Theta$ (uniform convergence).

(iv): The positive definiteness of $\mathbf{\Lambda}_Z$ follows from Eq. (17) and for the two-step and m-step estimators from Eq. (22). It makes sure that every observation gets a positive weight in the value function.

Following the proof of Theorem 2.7 in Newey and McFadden (1994), we construct a closed sphere Ψ around $\mathbf{\Phi}_0$ with radius 2δ. We use the fact that every convex function defined on a finite-dimensional separated topological linear space is continuous on the interior of its effective domain (see Barbu and Precupanu (2012) Proposition 2.17), so that $\Pi_0(\mathbf{\Phi})$ is continuous in Ψ. It then follows that $\tilde{\mathbf{\Phi}}_n$ converges uniformly on any compact subset of Θ, especially on Ψ. Then suppose $\tilde{\mathbf{\Phi}}_n$ is within δ distance of $\mathbf{\Phi}_0$, so that $\Pi(\tilde{\mathbf{\Phi}}_n) \leq \min_\Psi \Pi_n(\mathbf{\Phi})$. So for any $\mathbf{\Phi}$ outside Ψ, there is a linear convex combination $\lambda\tilde{\mathbf{\Phi}}_n + (1+\lambda)\mathbf{\Phi}$ that lies in Ψ, so that $\Pi(\tilde{\mathbf{\Phi}}_n) \leq \Pi(\lambda\tilde{\mathbf{\Phi}}_n + (1-\lambda)\mathbf{\Phi})$. By convexity of $\Pi(\cdot)$, we have $\Pi(\lambda\tilde{\mathbf{\Phi}}_n + (1-\lambda)\Phi) \leq \lambda\Pi(\tilde{\mathbf{\Phi}}_n) + (1-\lambda)\Pi(\mathbf{\Phi})$. Combining these last two inequalities yields $(1-\lambda)\Pi(\tilde{\mathbf{\Phi}}_n) \leq (1-\lambda)\Pi(\mathbf{\Phi})$ which implies that $\tilde{\mathbf{\Phi}}_n$ is a minimand over Θ and $\tilde{\mathbf{\Phi}}_n = \hat{\mathbf{\Phi}}_n$.

□

We follow the standard GMM literature (Newey and McFadden, 1994; Hansen, 2012) which proposes a one-step and a two-step estimation procedure that differ in how $\mathbf{\Lambda}_Z$ is defined. Since the two-step estimation builds on the residuals of the one-step estimation, we start with the one-step (or initial) estimate $\mathbf{\Phi}_{IE}$.

For the one-step estimation we follow Binder et al. (2005) and define $\mathbf{\Lambda}_Z$ in the following way:

$$\mathbf{\Lambda}_Z = \left(\sum_{i=1}^{N} \mathbf{Q}_i^\intercal \mathbf{D}\mathbf{D}^\intercal \mathbf{Q}_i\right) \otimes \mathbf{I}_{m\times m} \tag{17}$$

Let $\mathbf{V}_i$ be the matrix with untransformed time series data of cross section i. Then there exist a $(T-1)\times T$ linear transformation matrix $\mathbf{D}$ such that $\mathbf{D}\mathbf{V}_i = \Delta^*\mathbf{V}_i$.

If the first difference transformation is used to remove the fixed effect, then $\mathbf{D}$ has the following structure $((T-1-p)\times(T-p))$:

$$\mathbf{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \tag{18}$$

If the forward orthogonal transformation is applied to remove the fixed effect, then $\mathbf{D}$ has the following structure $((T-1-p)\times(T))$:

$$\mathbf{D} = \begin{pmatrix} \sqrt{\frac{T-1}{T}} & -\sqrt{\frac{1}{T(T-1)}} & -\sqrt{\frac{1}{T(T-1)}} & \cdots & -\sqrt{\frac{1}{T(T-1)}} \\ 0 & \sqrt{\frac{T-2}{T-1}} & -\sqrt{\frac{1}{(T-1)(T-2)}} & \cdots & -\sqrt{\frac{1}{(T-1)(T-2)}} \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\frac{1}{2}} & -\sqrt{\frac{T-1}{T}}\frac{1}{T-(T-1)} \end{pmatrix}. \tag{19}$$

Using Eq. (17) as the initial weighting matrix gives the one-step solution for $\hat{\mathbf{\Phi}}_{IE}$:

$$\mathrm{vec}(\hat{\mathbf{\Phi}}_{IE}) = (\mathbf{S}_{ZX}^\intercal \mathbf{\Lambda}_Z^{-1} \mathbf{S}_{ZX})^{-1} \mathbf{S}_{ZX}^\intercal \mathbf{\Lambda}_Z^{-1} \mathrm{vec}(\mathbf{S}_{Zy}) \tag{20}$$

In the two-step estimation the choice of the optimal weighting matrix Λ_Z requires the residuals of the one-step estimation $(\Delta^*\hat{\mathbf{E}}_i = \Delta^*\mathbf{Y}_i - \Delta^*\mathbf{W}_{minus,i}\hat{\mathbf{\Phi}}_{IE})$. The feasible efficient general methods of moment estimator (FEGMM) reads as follows:

$$\mathrm{vec}(\hat{\mathbf{\Phi}}_{FEGMM}) = (\mathbf{S}_{ZX}^\intercal \mathbf{\Lambda}_{Z_{\hat{e}}}^{-1} \mathbf{S}_{ZX})^{-1} \mathbf{S}_{ZX}^\intercal \mathbf{\Lambda}_{Z_{\hat{e}}}^{-1} \mathrm{vec}(\mathbf{S}_{Zy}) \tag{21}$$

where

$$
\begin{aligned}
\Delta^*\mathbf{W}_i &= \left(\Delta^*\mathbf{Y}_i\right), \\
\Delta^*\mathbf{W}_{minus} &= \begin{pmatrix}\Delta^*\mathbf{Y}_{i,-1} & \Delta^*\mathbf{X}_i & \Delta^*\mathbf{S}_i\end{pmatrix}, \\
\mathbf{S}_{QX} &= \sum_{i=1}^{N} \mathbf{Q}_i^\intercal \Delta^*\mathbf{W}_{i,minus}, \\
\mathbf{S}_{Qy} &= \sum_{i}^{N} \mathbf{Q}_i^\intercal \Delta^*\mathbf{W}_i, \\
\mathbf{Z}_i &= \mathbf{Q}_i \otimes \mathbf{I}_{m\times m}, \\
\mathbf{S}_{ZX} &= \mathbf{S}_{QX} \otimes \mathbf{I}_{m\times m}, \\
\mathbf{S}_{Zy} &= \mathrm{vec}(\mathbf{S}_{Qy}^\intercal), \\
\mathbf{\Lambda}_Q &= \sum_{i=1}^{N} \mathbf{Q}_i^\intercal \mathbf{D}\mathbf{D}^\intercal \mathbf{Q}_i, \\
\mathbf{\Lambda}_Z &= \mathbf{\Lambda}_Q \otimes \mathbf{I}_{m\times m}, \\
\hat{\mathbf{e}}_i &= \mathrm{vec}(\hat{\mathbf{E}}_i), \\
\hat{\mathbf{E}}_i &= \Delta\mathbf{W}_i - \Delta\mathbf{W}_{minus,i}\hat{\mathbf{\Phi}}_{IE}, \\
\mathbf{\Lambda}_{\mathbf{Z}_{\hat{e}}} &= \sum_{i=1}^{N} \mathbf{Z}_i^\intercal \mathbf{\Gamma}_{\hat{e}} \mathbf{Z}_i, \\
\mathbf{\Gamma}_{\hat{e}} &= \sum_{i=1}^{N} \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^\intercal.
\end{aligned}
\tag{22}
$$

In our code, we also implement the iterated efficient GMM, where we minimize $\mathbf{\Phi}^m$ (the m-step estimator) based on Eq. (15) with the weighting matrix of the (m-1)-step ($\mathbf{\Lambda}_{\mathbf{Z}_{\hat{e}}}$). In the PVAR context it is an open question, if the gains in efficiency outweight the numerical problems of inverting $\mathbf{\Lambda}_{\mathbf{Z}_{\hat{e}}}$.

2.5. System moment conditions

Additional moment conditions can be constructed when imposing the following assumption (see Blundell and Bond (1998) for the case $m = 1$) on the structure of the process.

Definition 2.4. *(System GMM moment conditions):*

$$
\begin{aligned}
\mathbb{E}[(\epsilon_{i,t} + \mu_i)(\mathbf{y}_{i,t-1} - \mathbf{y}_{i,t-2})^\intercal] &= \mathbf{0} \quad t \in \{3, 4, \cdots, T\} \\
\mathbb{E}[(\epsilon_{i,t} + \mu_i)(\mathbf{x}_{i,t} - \mathbf{x}_{i,t-1})^\intercal] &= \mathbf{0} \quad t \in \{2, 3, \cdots, T\} \\
\mathbb{E}[(\epsilon_{i,t} + \mu_i)\mathbf{s}_{i,t}^\intercal] &= \mathbf{0} \quad t \in \{2, 3, \cdots, T\}
\end{aligned} \tag{23}
$$

So, this assumption is valid if changes in $y_{i,t}$ are not systematically related to μ_i. Following Blundell and Bond (1998), this assumption is clearly satisfied in a stationary PVAR model (all eigenvalues of the PVAR polynomial are less than 1).

Blundell and Bond (1998) also argue that the system GMM estimator performs better than the first difference GMM estimator because the additional instruments remain good predictors for the endogenous variables in this model even when the series are very persistent.

In matrix notation, the new set of moment conditions P_i for the case $p = 1$ reads as follows ("Δ" denotes the first difference operator):

$$
\mathbf{P}_i := \begin{pmatrix}
\mathbf{0} & \Delta\mathbf{y}_{i,2} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \Delta\mathbf{y}_{i,3} & & \mathbf{0} \\
\vdots & & & \ddots & \\
\mathbf{0} & \mathbf{0} & \cdots & & \Delta\mathbf{y}_{i,T-1} \\
\Delta\mathbf{x}_{i,2} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \Delta\mathbf{x}_{i,3} & \mathbf{0} & & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \Delta\mathbf{x}_{i,4} & & \mathbf{0} \\
\vdots & & & \ddots & \\
\mathbf{0} & \mathbf{0} & \cdots & & \Delta\mathbf{x}_{i,T} \\
\mathbf{s}_{i,2} & \mathbf{s}_{i,3} & \cdots & & \mathbf{s}_{i,T}
\end{pmatrix}
$$

Next we define a new matrix for the instruments:[9]

$$
\mathbf{Q}_i^* := \begin{pmatrix} \mathbf{Q}_i & 0 \\ 0 & \mathbf{P}_i \end{pmatrix}
$$

2.6. System GMM estimator

The extended GMM estimator is derived in the same way as Eq. (20) and reads as follows (IEE stands for initial estimator extended):

$$
\mathrm{vec}(\hat{\mathbf{\Phi}}_{IEE}) = (\mathbf{S}_{\mathbf{Z}^*\mathbf{X}}^\intercal \Lambda_{Z^*}^{-1} \mathbf{S}_{\mathbf{Z}^*\mathbf{X}})^{-1} \mathbf{S}_{\mathbf{Z}^*\mathbf{X}}^\intercal \Lambda^{-1} \mathrm{vec}(\mathbf{S}_{\mathbf{Z}^*\mathbf{y}}) \tag{24}
$$

[9] In the case where $p > 1$ the matrices P_i have to be adapted by deleting columns and rows such that Eq. (23) holds.

Where

$$\Delta^*\mathbf{W}_i^* = \begin{pmatrix} \Delta^*\mathbf{Y}_i \\ \mathbf{Y}_i \end{pmatrix},$$

$$\Delta^*\mathbf{W}_{minus}^* = \begin{pmatrix} \Delta^*\mathbf{Y}_{i,-1} & \Delta^*\mathbf{X}_i & \Delta^*\mathbf{S}_i & \mathbf{0} \\ \mathbf{Y}_{i,-1} & \mathbf{X_i} & \mathbf{S}_i & \mathbf{1} \end{pmatrix},$$

$$\mathbf{S_{Q^*X}} = \sum_{i=1}^{N} \mathbf{Q}_i^{*\intercal}\Delta^*\mathbf{W}_{minus}^*,$$

$$\mathbf{S_{Z^*X}} = \mathbf{S_{Q^*X}} \otimes \mathbf{I}_{m\times m},$$

$$\mathbf{S_{Q^*Y}} = \sum_{i=1}^{N} \mathbf{Q}_i^*\Delta^*\mathbf{W}_i^*,$$

$$\mathbf{S_{Z^*Y}} = \mathbf{S_{Q^*Y}} \otimes \mathbf{I}_{m\times m},$$

$$\mathbf{\Lambda_{Q^*}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Q}_i^*\mathbf{D}^*(\mathbf{D}^*)^\intercal(\mathbf{Q}_i^*)^\intercal,$$

D^* is defined as follows $(((T-1-p)+(T-p))\times(T-p))$ in the case of the first difference transformation:[10]

$$\mathbf{D}^* = \begin{pmatrix} \mathbf{D} \\ \mathbf{I}_{(T-p)\times(T-p)} \end{pmatrix} \tag{25}$$

Where $\mathbf{D}$ is defined in Eq. (18) for the first difference transformation and in Eq. (19) for the forward orthogonal transformation.

For the two-step estimation we use

$$\mathbf{\Lambda}_{Z^*_{\hat{e}}} = \sum_{i=1}^{N} \mathbf{Z}_i^*\mathbf{\Gamma}_{\hat{e}}(\mathbf{Z}_i^*)^\intercal$$

as the weighting matrix and arrive at the feasible and efficient two-step GMM estimator:

$$\text{vec}(\hat{\mathbf{\Phi}}_{EFEGMM}) = (\mathbf{S}^\intercal_{Z^*X}\mathbf{\Lambda}^{-1}_{Z^*_{\hat{e}}}\mathbf{S}_{Z^*X})^{-1}\mathbf{S}^\intercal_{Z^*X}\mathbf{\Lambda}^{-1}_{Z^*_{\hat{e}}}\text{vec}(\mathbf{S}_{Z^*y}) \tag{26}$$

[10]In the case of the forward orthogonal transformation D^* is defined as $(((T-p)+(T-p))\times(T))$ matrix.

2.7. *Estimating the asymptotic covariance matrix of the estimator*

In this section, we define the estimates of the asymptotic covariance matrix of the GMM estimators.

The covariance matrix for the one-step GMM estimator defined in Eq. (20) and Eq. (24) can be defined in a straightforward way. Ruud (2000) shows that the one-step GMM estimators are asymptotically normally distributed.

$$\begin{aligned}\hat{\mathbf{V}}\mathbf{ar}_{\hat{\mathbf{\Phi}}_{IE}} := &\left(\mathbf{S}_{ZX}^{\intercal}\mathbf{\Lambda}_{Z}^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\mathbf{\Lambda}_{Z}^{-1}\mathbf{\Lambda}_{Z_{\hat{e}}}\\ &\times\mathbf{\Lambda}_{Z}^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{S}_{ZX}^{\intercal}\mathbf{\Lambda}_{Z}^{-1}\mathbf{S}_{ZX}\right)^{-1},\end{aligned}$$

with $\mathbf{\Lambda}_Z = \mathbf{\Lambda}_Q \otimes I_m$ and $\mathbf{\Lambda}_{Z_{\hat{e}}}$ as defined in Eq. (22).

Since the choice of the weighting matrix is in general not optimal in the one-step estimation, we cannot expect that the one-step estimator is asymptotically efficient. For the two-step GMM estimators in Eq. (21), the asymptotic variance is defined as follows:

$$\hat{\mathbf{V}}\mathbf{ar}_{\hat{\mathbf{\Phi}}_{FEGMM}} := \frac{1}{N}\left(\mathbf{S}_{ZX}^{\intercal}\mathbf{\Lambda}_{Z_{\hat{e}}}^{-1}\mathbf{S}_{ZX}\right)^{-1}$$

However, Windmeijer (2005) shows that this estimator does not perform well in finite samples on simulated data of a dynamic panel process ($m = 1$). Windmeijer (2005) therefore suggested a small sample correction for a dynamic panel model (see Roodman, 2009b), which we extend to a PVAR model (see Appendix A for a derivation).

$$\begin{aligned}\hat{\mathbf{V}}\mathbf{ar}_{\hat{\mathbf{\Phi}}_{FEGMM}}^{Wc} = &\left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\hat{\mathbf{\Phi}}_{IE})\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\\ &+ D_{\hat{\mathbf{\Phi}}_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\hat{\mathbf{\Phi}}_{IE})}\left(\mathbf{S}_{ZX}\left(\mathbf{\Lambda}_{Z}\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z}\right)^{-1}\\ &+ \left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\hat{\mathbf{\Phi}}_{IE})\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}D_{\hat{\mathbf{\Phi}}_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\hat{\mathbf{\Phi}}_{IE})}^{\intercal}\\ &+ D_{\hat{\mathbf{\Phi}}_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\hat{\mathbf{\Phi}}_{IE})}\hat{\mathbf{V}}\mathbf{ar}_{\hat{\mathbf{\Phi}}_{IE}}D_{\hat{\mathbf{\Phi}}_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\hat{\mathbf{\Phi}}_{IE})}^{\intercal}\end{aligned} \tag{27}$$

2.8. *Orthogonal impulse response analysis*

In this section responses to orthogonal impulses are calculated. This section closely follows Lütkepohl (2007) on theory and Pfaff (2008) on the implementation.

Considering Eq. (1), the impulse response analysis in a vector autoregression context is concerned with the response of one (endogenous) variables to an impulse in another (endogenous) variable. This idea can be formalized by first deriving the so-called PVMA-X representation (panel vector moving average representation with exogenous variables) of a PVAR-X(1) process:[11]

$$\mathbf{y}_{i,t} = \nu_i + \left(\sum_{j=0}^{\infty} \mathbf{A}^{j-1}[\mathbf{B}\ \mathbf{C}]\right)\begin{bmatrix}\mathbf{x}_{i,t-j}\\ \mathbf{s}_{i,t-j}\end{bmatrix} + \left(\sum_{j=0}^{\infty} \mathbf{A}^{j}\right)\left[\epsilon_{i,t-j}\right] \tag{28}$$

With $\nu_i = (\mathbf{I}_m - \mathbf{A})^{-1}\,\mu_i^*$.[12]

It is also important to note that in the impulse response analysis we treat predetermined and strictly exogenous variables in the same way.

Based on the PVMA-X representation the impulse response function can be stated as follows:

$$\mathrm{IRF}(k,r) = \frac{\partial \mathbf{y}_{i,t+k}}{\partial(\epsilon_{i,t})_r} = \mathbf{A}^k\mathbf{e}_r$$

where k is the number of periods after the shock to the r-th component of $\epsilon_{i,t}$ with $\mathbf{e}_r$ being a $m \times 1$ vector with a 1 in the r-th column and 0 otherwise.

Let $\mathbf{\Sigma}_\epsilon$ be the covariance matrix of ϵ_t. Usually the off diagonal elements of $\mathbf{\Sigma}_\epsilon$ are different from 0, so shocks across the m equations are not independent of each other. Therefore, the parameters of the PVAR model have to be adjusted such that the responses to "independent" shocks are transferred through the PVAR system accordingly.

Since we assume that $\mathbf{\Sigma}_\epsilon$ is a symmetric positive definite matrix, there exist a unique Cholesky decomposition such that $\mathbf{\Sigma}_\epsilon = \mathbf{P}\mathbf{P}^\intercal$, where $\mathbf{P}$ is a lower triangular matrix. Defining $\mathbf{\Theta}_k = \mathbf{A}^k\mathbf{P}$ and $\mathbf{u}_{i,t} = \mathbf{P}^{-1}\epsilon_{i,t}$ we obtain the orthogonal impulse response function:

$$\mathrm{OIRF}(k,r) = \frac{\partial \mathbf{y}_{i,t+k}}{\partial(\mathbf{u}_{i,t})_r} = \mathbf{\Theta}_k\mathbf{e}_r \tag{29}$$

As stated in Lütkepohl (2007) and many others, although the Cholesky-decomposition is unique it depends on the ordering of variables which has been criticized in the literature. An alternative to the OIRF that meets some of the critics is presented in the next section.

Following Lütkepohl (2007) a closely related tool for interpreting PVAR models is available, namely the forecast error variance decomposition. It determines how much of the forecast error variance of each of the variables can be explained by exogenous shocks to the other variables.

We start the forecast error variance decomposition by defining the h-step forecast error in the MA representation:

[11] As a PVAR(p) process can be expressed as a PVAR(1) process (see Lütkepohl (2007)) we need not derive a more general expression.

[12] $\mu_i^* = \left(\mathbf{I}_m - \sum_{l=1}^{p}\mathbf{A}_l\right)\mu_i$.

$$y_{i,t+h} - y_{i,t} = \sum_{k=0}^{h-1} \mathbf{\Theta}_k u_{i,t-k}$$

Let $\theta_{k,m,n}$ be the $m-n^{th}$ component of $\mathbf{\Theta}_k$. Then it is possible to define the contribution of innovations in variable n to the forecast error variance or MSE of the h-step forecast of variable m.

$$y_{i,m,t+h} - y_{i,m,t} = \sum_{k=0}^{h-1} (e_m^\intercal \mathbf{\Theta}_k e_n)^2 \tag{30}$$

If we divide Eq. (30) by the mean squared error of the h-step forecast of $y_{i,m,t+h}$, then the forecast error variance of variable $y_{i,m}$ yields:

$$\omega^o_{m,n,h} = \sum_{k=0}^{h-1} (e_m^\intercal \mathbf{\Theta}_k e_n)^2 / \left(\sum_{k=0}^{h-1} \sum_{m=1}^{m} \theta^2_{k,m,n} \right) \tag{31}$$

2.9. Generalized impulse response analysis

In Pesaran and Shin (1998) an alternative approach to the orthogonal impulse response analysis in Eq. (29) is suggested.

Instead of shocking all the elements of $\epsilon_{i,t}$ Pesaran and Shin (1998) choose to shock only one element, say its r-th element and integrate out the effects of other shocks using the historically observed distribution of the errors. In this case we have

$$\text{GIRF}(k, r, \mathbf{\Sigma}_\epsilon) = \mathbb{E}\left[\mathbf{y}_{i,t+k} | \epsilon_{i,t,r} = \delta_r, \mathbf{\Sigma}_\epsilon\right] - \mathbb{E}\left[\mathbf{y}_{i,t+k} | \mathbf{\Sigma}_\epsilon\right]$$

By setting $\delta_r = \sqrt{\mathbf{\Sigma}_{\epsilon,r,r}}$ we obtain the generalized impulse response function by

$$\text{GIRF}(k, r, \mathbf{\Sigma}_\epsilon) = \mathbf{A}^k \mathbf{\Sigma}_\epsilon \left(\sigma_{r,r}\right)^{-1/2}$$

where $\sigma_{r,r}$ is the r-th diagonal element of $\mathbf{\Sigma}_\epsilon$.

Lin (2006) states that when $\mathbf{\Sigma}_\epsilon$ is diagonal OIRF and GIRF are the same. Moreover GIRF is unaffected by the ordering of variables. The GIRF of the effect of an unit shock to the r-th equation is the same as that of an orthogonal impulse response but different for other shocks. Hence, the GIRF can easily computed by using OIRF with each variable as the leading one.

In analogy to Eq. (31), the forecast error variance decomposition for the generalized impulse response function is defined as follows:

$$\omega^{g}_{m,n,h} = \left(\sigma^{-1}_{r,r} \sum_{k=0}^{h-1} \left(e_{m}{}^{\top} \mathbf{A}^{k} \mathbf{\Sigma}_{\epsilon} e_{n} \right)^{2} \right) / \left(\sum_{k=0}^{h-1} \sum_{m=1}^{m} \theta^{2}_{k,m,n} \right), \tag{32}$$

2.10. Confidence bands for impulse response analysis

In this section we do not enter the theoretical discussion which methods are best for estimating confidence bands for our (orthogonal and generalized) impulse response functions but focus on the ideas presented in Lütkepohl (2007). He states that if the distributions of a VAR model under consideration is unknown, so-called bootstrap or resampling methods may be applied to investigate the distributions of functions of stochastic processes or multiple time series. It is important to observe that our problem does not only involve a standard VAR model but a PVAR model with GMM estimators.

For panel datasets, there are in general three resampling schemes, temporal resampling, cross-sectional resampling and a combined resampling. In contrast to the earlier literature Kapetanios (2008) suggests to use cross-sectional resampling, where subsets of the data with the same panel individual panel identifier are drawn completely with replacement. Whereas Kapetanios (2008) shows that this bootstrapping procedure works well with many cross-sectional units for general panel models (not only dynamic panel models), Yan (2012) additionally presents Monte Carlo simulations that show the superiority of this procedure in combination with the first difference GMM estimator.

Kapetanios (2008) defines the bootstrapping procedure that we implement as follows:

Definition 2.5. *(cross-sectional resampling):*
For a $T \times N$ matrix of random variables Θ, cross-sectional resampling is defined as the operation of constructing a $T \times N^{*}$ matrix Θ^{*} where the columns of Θ^{*} are a random sample with replacement of blocks of the columns of Θ and N^{*} is not necessarily equal to N.

Based on Definition (2.5) our bootstrapping procedure can be defined as follows:

1. Let $P(i) = 1/N$ be the uniformly distributed probability of drawing i from the set $i = 1, ..., N$. Let Y_i, X_i, S_i be such a draw from the full panel data set. Repeat this draw N^{*}-times with replacement.
2. Depending on the selected argument, estimate $\mathbf{\Phi}$ (one-step/two-step, system or first difference GMM) for the drawn dataset.
3. Depending on what impulse response function was selected, calculate the orthogonal or generalized impulse response function.

2.11. *Specification tests*

2.11.1. *Hansen overidentification test for the validity of instrument subsets*

Following Hansen (1982) a critical assumption for the validity of the GMM estimator is that the instruments are exogenous. If the model is overidentified, the Hansen overidentification test (Hansen, 1982) can be applied. The test can be applied to one-step ($\mathbf{\Lambda_Z}$) and two-step ($\mathbf{\Lambda}_{\mathbf{Z}_{\hat{e}}}$) GMM estimations:

$$N\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{\hat{E}_i}\right)^{\top}\mathbf{\Lambda}_{\mathbf{Z}_{\hat{e}}}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{\hat{E}_i}\right)\overset{a}{\sim}\tilde{\chi}^2_{L-K} \tag{33}$$

$L = ncol(Q^*)$ represents the number of instruments. $K = m * (p + k + n)$ counts the number of parameters in the model (see Section 2.4). Newey (1985) proofs that Eq. (33) is the asymptotically chi-squared distributed.

2.11.2. *Andrews-Lu model selection procedure*

In Andrews and Lu (2001) a class of model and moment selection criteria (MMSC) are introduced that are analogous to the well-known model section criteria for choosing between competing models.

In the original notation of Andrews and Lu (2001) the basic MMSC reads as follows:

$$MMSC_n(b,c) = J_n(b,c) - h(|c| - |b|)\kappa_n \tag{34}$$

$J_n(b,c)$ is the Hansen overidentification test statistics from Eq. (33). b stands for the number of parameters, c is the number of moment conditions and n is the number of total observations. We implemented three different versions of MMSC function:

$$\begin{aligned}
MMSC_{BIC,n}(b,c) &= J_n(b,c) - (|c| - |b|)\cdot ln(n)\\
MMSC_{AIC,n}(b,c) &= J_n(b,c) - (|c| - |b|)\cdot 2\\
MMSC_{HQIC,n}(b,c) &= J_n(b,c) - Q\cdot(|c| - |b|)\cdot ln(ln(n))
\end{aligned} \tag{35}$$

Andrews and Lu (2001) recommend the MMSC-BIC (Bayesian information criterion) or the MMSC-HQIC (Hannan-Quinn information criterion). The MMSC-AIC (Akaike information criterion) does not fulfill their consistency criterion as it has positive probability even asymptotically of selecting too few over-identifying restrictions.

2.11.3. *Stability of the PVAR model*

The standard stability condition of the panel VAR coefficients is based on the modulus of each eigenvalue of the estimated model. Lütkepohl (2007) and Hamilton (1994) both show that a VAR model is stable if all moduli of the companion matrix are strictly less than one.

Stability implies that the panel VAR is invertible and has an infinite-order vector moving-average representation.

3. Applying the panelvar package

In this section, we demonstrate how to use the **panelvar** package. We start with a dynamic panel model for the ”EmplUK” (”abdata”) datacompare the results with STATA’s **xtabond2**. For the same specification, we apply first difference and system GMM estimations with the first difference and the forward orthogonal transformation. We compare these results also for two-step estimator (including robust standard errors calculated with the Windmeijer correction).[13]

Second, we apply the package **panelvar** to examples with more than one endogenous variable to the ”Dahlberg” data set (Dahlberg and Johansson, 2000). Third, we perform a simulation exercise in which we compare our code with the existing STATA codes of Abrigo and Love (2016) who claim to use the Anderson and Hsiao (1982) GMM estimator.[14]

3.1. Dynamic panel estimation with panelvar

In the following section, a well-known data sets from the literature is used. Our code can be easily installed in R by typing: install.packages(“panelvar”). The ”EmplUK” (or ”abdata”) data set[15] was used in the dynamic panel literature by Arellano and Bond (1991), Blundell and Bond (1998) and Roodman (2009b). This data set describes employment, wages, capital and output of 140 firms in the United Kingdom from 1976 to 1984. To be in line with the literature, we choose the specification of table 4b in Arellano and Bond (1991) and apply the first difference GMM estimator:

$$\begin{aligned} log\_emp_{i,t} &= log\_emp_{i,t-1} + log\_emp_{i,t-2} + log\_wage_{i,t} + log\_wage_{i,t-1} + \\ &\quad log\_output_{i,t} + log\_output_{i,t-1} + log\_capital_{i,t} + \textstyle\sum_{t=2}^{T} year_t \end{aligned} \tag{36}$$

Employment is explained by past values of employment (two lags), current and first lag of wages and output and current value of capital. We estimate the model specified in Eq. (36) with the following STATA code:

[13] The R package is available on CRAN: `https://cran.r-project.org/web/packages/panelvar/`.

[14] In Abrigo and Love (2016) it is not clear which instruments are used. The possibilities are $y_{t,i2}$ or $\Delta y_{t,i2}$. Based on our code, we could rule out $y_{t,i2}$ as instruments, as this would correspond to setting $max\_instr\_dependent\_vars = 2$.

[15] In STATA, you load the data by executing the following: ”webuse abdata”.

```
clear all
webuse abdata
#delimit;
xtabond2 n L1.n L2.n w L1.w k ys L1.ys i.year, gmmstyle(n, lag(2 99))
ivstyle(w L1.w k ys L1.ys i.year) nolevel twostep robust;
#delimit cr

Dynamic panel-data estimation, two-step difference GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       611
Time variable : year                            Number of groups   =       140
Number of instruments = 38                      Obs per group: min =         4
Wald chi2(13) =    822.28                                      avg =      4.36
Prob > chi2   =     0.000                                      max =         6
------------------------------------------------------------------------------
             |              Corrected
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .4741506   .1853986     2.56   0.011     .1107761    .8375251
         L2. |  -.0529677   .0517491    -1.02   0.306    -.1543941    .0484587
             |
           w |
         --. |  -.5132049   .1455655    -3.53   0.000    -.7985081   -.2279017
         L1. |     .22464   .1419498     1.58   0.114    -.0535765    .5028564
             |
           k |   .2927232   .0626271     4.67   0.000     .1699764      .41547
             |
          ys |
         --. |    .609776   .1562628     3.90   0.000     .3035065    .9160455
         L1. |  -.4463736   .2173026    -2.05   0.040    -.8722788   -.0204684
             |
        year |
       1979  |    .010509   .0099019     1.06   0.289    -.0088983    .0299163
       1980  |   .0246513   .0157698     1.56   0.118     -.006257    .0555596
       1981  |  -.0158017   .0267314    -0.59   0.554    -.0681943    .0365908
       1982  |  -.0374419   .0299934    -1.25   0.212    -.0962278    .0213441
       1983  |  -.0392887   .0346649    -1.13   0.257    -.1072307    .0286534
       1984  |  -.0495093   .0348579    -1.42   0.156    -.1178296    .0188109
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    D.(w L.w k ys L.ys 1976b.year 1977.year 1978.year 1979.year 1980.year
    1981.year 1982.year 1983.year 1984.year)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(2/8).n
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -1.54  Pr > z =  0.124
Arellano-Bond test for AR(2) in first differences: z =  -0.28  Pr > z =  0.780
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(25)   =  75.46  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(25)   =  30.11  Prob > chi2 =  0.220
  (Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:
  iv(w L.w k ys L.ys 1976b.year 1977.year 1978.year 1979.year 1980.year 1981.year 1982.year 1983.year 1984.year)
    Hansen test excluding group:     chi2(14)   =  13.33  Prob > chi2 =  0.501
```

```
Difference (null H = exogenous): chi2(11)   =  16.78  Prob > chi2 =  0.115
```

The same example can be estimated with the **pvargmm** function in the package **panelvar**.

```
library("panelvar")
data("abdata")
Arellano_Bond_1991_table4b <-pvargmm(
                dependent_vars = c("n"),
                lags = 2,
                exog_vars = c("w", "wL1", "k", "ys", "ysL1",
                              "yr1979", "yr1980", "yr1981", "yr1982",
                              "yr1983", "yr1984"),
                transformation = "fd",
                data = abdata,
                panel_identifier = c("id", "year"),
                steps = c("twostep"),
                system_instruments = FALSE,
                max_instr_dependent_vars = 99,
                min_instr_dependent_vars = 2L,
                collapse = FALSE)

summary(Arellano_Bond_1991_table4b)

---------------------------------------------------
Dynamic Panel VAR estimation, two-step GMM
---------------------------------------------------
Transformation: First-differences
Group variable: id
Time variable: year
Number of observations = 751
Number of groups = 140
Obs per group: min = 5
               avg = 5.364286
               max = 7
Number of instruments = 38

===================
        n
-------------------
lag1_n   0.4742 *
        (0.1854)
lag2_n  -0.0530
        (0.0517)
w       -0.5132 ***
        (0.1456)
wL1      0.2246
        (0.1419)
k        0.2927 ***
        (0.0626)
ys       0.6098 ***
        (0.1563)
ysL1    -0.4464 *
        (0.2173)
yr1979   0.0105
        (0.0099)
yr1980   0.0247
```

```
        (0.0158)
yr1981  -0.0158
        (0.0267)
yr1982  -0.0374
        (0.0300)
yr1983  -0.0393
        (0.0347)
yr1984  -0.0495
        (0.0349)
===================
*** p < 0.001, ** p < 0.01, * p < 0.05

---------------------------------------------------
Instruments for first differences equation
 Standard
  FD.(w wL1 k ys ysL1 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984)
 GMM-type
  Dependent vars: L(2, 7)
  Collapse =  FALSE
---------------------------------------------------

Hansen test of overid. restrictions: chi2(25) = 30.11 Prob > chi2 = 0.22
(Robust, but weakened by many instruments.)
```

Next, we look at two important options for any GMM estimation: (1) *collapse* = *TRUE*, as defined by Eq. (13), and (2) *transformation* = *"fod"*, the forward orthogonal transformation. We begin with the code for the estimation in R.

```
library("panelvar")
data("abdata")
Arellano_Bond_1991_table4b_coll_fod <-pvargmm(dependent_vars = c("n"),
                                     lags = 2,
                                     exog_vars = c("w", "wL1", "k", "ys", "ysL1",
                                                   "yr1979", "yr1980", "yr1981", "yr1982", "yr1983",
                                                   "yr1984"),
                                     transformation = "fod",
                                     data = abdata,
                                     panel_identifier = c("id", "year"),
                                     steps = c("twostep"),
                                     system_instruments = FALSE,
                                     max_instr_dependent_vars = 99,
                                     max_instr_predet_vars = 99,
                                     min_instr_dependent_vars = 2L,
                                     min_instr_predet_vars = 1L,
                                     collapse = TRUE
)

summary(Arellano_Bond_1991_table4b_coll_fod)

---------------------------------------------------
Dynamic Panel VAR estimation, two-step GMM
---------------------------------------------------
Transformation: Forward orthogonal deviations
Group variable: id
Time variable: year
```

```
Number of observations = 751
Number of groups = 140
Obs per group: min = 5
               avg = 5.364286
               max = 7
Number of instruments = 18

==================
        n
------------------
lag1_n   1.3783 **
        (0.4523)
lag2_n  -0.2526 **
        (0.0955)
w       -0.5626 **
        (0.2036)
wL1      0.5399
        (0.4064)
k        0.0966
        (0.1482)
ys       0.5777 *
        (0.2454)
ysL1    -0.8983 *
        (0.4463)
yr1979   0.0134
        (0.0133)
yr1980   0.0130
        (0.0202)
yr1981  -0.0403
        (0.0262)
yr1982  -0.0358
        (0.0238)
yr1983  -0.0149
        (0.0304)
yr1984  -0.0249
        (0.0260)
==================
*** p < 0.001, ** p < 0.01, * p < 0.05

---------------------------------------------------
Instruments for orthogonal deviations equation
 Standard
  FOD.(w wL1 k ys ysL1 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984)
 GMM-type
  Dependent vars: L(2, 7)
  Collapse =  TRUE
---------------------------------------------------

Hansen test of overid. restrictions: chi2(5) = 7.79 Prob > chi2 = 0.168
(Robust, but weakened by many instruments.)
```

The corresponding STATA: xtabond2 reads as follows:

```
clear all
webuse abdata
#delimit;
xtabond2 n L1.n L2.n w L1.w k ys L1.ys i.year, gmmstyle(n, lag(2 99) collapse) ivstyle(w L1.w k ys L1.ys i.year) nolevel twos
```

```
#delimit cr

Dynamic panel-data estimation, two-step difference GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       611
Time variable : year                            Number of groups   =       140
Number of instruments = 18                      Obs per group: min =         4
Wald chi2(13) =   1146.57                                      avg =      4.36
Prob > chi2   =     0.000                                      max =         6
------------------------------------------------------------------------------
             |              Corrected
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |    1.37832   .4523282     3.05   0.002     .4917729    2.264867
         L2. |  -.2526192    .095528    -2.64   0.008    -.4398507   -.0653877
             |
           w |
         --. |  -.5626032   .2036126    -2.76   0.006    -.9616766   -.1635298
         L1. |   .5399246   .4064035     1.33   0.184    -.2566115    1.336461
             |
           k |   .0965644   .1481514     0.65   0.515    -.1938071    .3869358
             |
          ys |
         --. |   .5776872   .2454191     2.35   0.019     .0966746      1.0587
         L1. |  -.8982714   .4463081    -2.01   0.044    -1.773019   -.0235236
             |
        year |
       1979  |   .0133686   .0132508     1.01   0.313    -.0126024    .0393397
       1980  |   .0129923   .0201691     0.64   0.519    -.0265385    .0525231
       1981  |  -.0402913   .0262247    -1.54   0.124    -.0916907    .0111081
       1982  |  -.0358365   .0238203    -1.50   0.132    -.0825235    .0108504
       1983  |  -.0149097   .0303587    -0.49   0.623    -.0744117    .0445922
       1984  |  -.0248517   .0260187    -0.96   0.340    -.0758474    .0261441
------------------------------------------------------------------------------
Instruments for orthogonal deviations equation
  Standard
    FOD.(w L.w k ys L.ys 1976b.year 1977.year 1978.year 1979.year 1980.year
    1981.year 1982.year 1983.year 1984.year)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(2/8).n collapsed
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -2.23  Pr > z =  0.026
Arellano-Bond test for AR(2) in first differences: z =  -0.07  Pr > z =  0.947
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(5)    =   9.04  Prob > chi2 =  0.108
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(5)    =   7.79  Prob > chi2 =  0.168
  (Robust, but weakened by many instruments.)
```

Second, we apply our code to reproduce a slightly different version of table 4 column (2) and (4) in Blundell and Bond (1998) with the **system GMM estimator**.[16]

[16]Our code is not designed for instrumenting each predetermined variables with a different number of

$$
\begin{aligned}
log\_emp_{i,t} \quad = \quad & log\_emp_{i,t-1} + log\_wage_{i,t} + log\_wage_{i,t-1} + \\
& log\_capital_{i,t} + log\_capital_{i,t-1} + \textstyle\sum_{t=2}^{T} year_t
\end{aligned}
\tag{37}
$$

The model in Eq. (37) can be estimated with the following STATA code:

```
clear all
webuse abdata
#delimit;
xtabond2 n L1.n w L1.w k L1.k i.year, gmmstyle(L1.n w L1.w k L1.k k, lag(1 99))
ivstyle(i.year) level robust;
#delimit cr

Dynamic panel-data estimation, one-step system GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       891
Time variable : year                            Number of groups   =       140
Number of instruments = 133                     Obs per group: min =         6
Wald chi2(12) =  12423.26                                      avg =      6.36
Prob > chi2   =     0.000                                      max =         8
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .9288048   .0257337    36.09   0.000     .8783678    .9792418
             |
           w |
         --. |  -.5293467   .1654901    -3.20   0.001    -.8537013   -.2049921
         L1. |   .2882503   .1369533     2.10   0.035     .0198268    .5566737
             |
           k |
         --. |   .3810652   .0625376     6.09   0.000     .2584937    .5036367
         L1. |  -.3219537   .0631328    -5.10   0.000    -.4456917   -.1982157
             |
        year |
       1978  |  -.0052381    .018024    -0.29   0.771    -.0405645    .0300882
       1979  |   .0022239   .0206624     0.11   0.914    -.0382737    .0427215
       1980  |  -.0180138   .0216991    -0.83   0.406    -.0605431    .0245156
       1981  |   -.054249   .0284007    -1.91   0.056    -.1099133    .0014153
       1982  |  -.0218739   .0288166    -0.76   0.448    -.0783533    .0346055
       1983  |  -.0054128   .0251686    -0.22   0.830    -.0547423    .0439166
       1984  |  -.0136591   .0292565    -0.47   0.641    -.0710007    .0436825
             |
       _cons |   .8458168    .225243     3.76   0.000     .4043487    1.287285
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    D.(1976b.year 1977.year 1978.year 1979.year 1980.year 1981.year 1982.year
    1983.year 1984.year)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(1/8).(L.n w L.w k L.k k)
```

---

lagged instruments. Furthermore, please note that the results in Blundell and Bond (1998) are misleading, as they do not use the correct weighting matrix, in particular Eq. (25).

```
Instruments for levels equation
  Standard
    1976b.year 1977.year 1978.year 1979.year 1980.year 1981.year 1982.year
    1983.year 1984.year
    _cons
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    D.(L.n w L.w k L.k k)
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -4.34  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -0.49  Pr > z =  0.625
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(120)  = 266.74  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(120)  = 129.23  Prob > chi2 =  0.266
  (Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:
  GMM instruments for levels
    Hansen test excluding group:     chi2(92)   = 105.58  Prob > chi2 =  0.158
    Difference (null H = exogenous): chi2(28)   =  23.65  Prob > chi2 =  0.700
  iv(1976b.year 1977.year 1978.year 1979.year 1980.year 1981.year 1982.year 1983.year 1984.year)
    Hansen test excluding group:     chi2(113)  = 124.33  Prob > chi2 =  0.219
    Difference (null H = exogenous): chi2(7)    =   4.90  Prob > chi2 =  0.672
```

We estimate the same model in R with the **pvargmm** function:

```
library(panelvar)
data("abdata")
Blundell_Bond_1998_table4 <-pvargmm(
          dependent_vars = c("n"),
          lags = 1,
          predet_vars = c("w", "wL1", "k", "kL1"),
          exog_vars = c("yr1978", "yr1979", "yr1980", "yr1981", "yr1982",
                        "yr1983", "yr1984"),
          transformation = "fd",
          data = abdata,
          panel_identifier = c("id", "year"),
          steps = c("onestep"),
          system_instruments = TRUE,
          max_instr_dependent_vars = 99,
          max_instr_predet_vars = 99,
          min_instr_dependent_vars = 2L,
          min_instr_predet_vars = 1L,
          collapse = FALSE
          )

summary(Blundell_Bond_1998_table4)
---------------------------------------------------
Dynamic Panel VAR estimation, one-step GMM
---------------------------------------------------
Transformation: fd
Group variable: id
Time variable: year
Number of observations = 891
```

```
Number of groups = 140
Obs per group: min = 6
               avg = 6.364286
               max = 8
Number of instruments = 215

===================
        n
-------------------
lag1_n   0.9288 ***
        (0.0257)
w       -0.5293 **
        (0.1655)
wL1      0.2883 *
        (0.1370)
k        0.3811 ***
        (0.0625)
kL1     -0.3220 ***
        (0.0631)
yr1978  -0.0052
        (0.0180)
yr1979   0.0022
        (0.0207)
yr1980  -0.0180
        (0.0217)
yr1981  -0.0542
        (0.0284)
yr1982  -0.0219
        (0.0288)
yr1983  -0.0054
        (0.0252)
yr1984  -0.0137
        (0.0293)
const    0.8458 ***
        (0.2252)
===================
*** p < 0.001, ** p < 0.01, * p < 0.05

---------------------------------------------------
Instruments for first differences equation
 Standard
  FD.(yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984)
 GMM-type
  Dependent vars: L(2, 8)
  Predet vars: L(1, 8)
  Collapse =  FALSE
---------------------------------------------------

Hansen test of overid. restrictions: chi2(202) = 2.67 Prob > chi2 = 1
(Robust, but weakened by many instruments.)
```

Thus, in contrast to other software, the result estimated with our code is identical to STATA's xtabond2 for dynamic panel models.

3.2. *Panel vector autoregression with panelvar*

We apply our package to the Dahlberg data set which was used by Dahlberg and Johansson (2000) and many other papers. The panel data set consists of 265 Swedish municipalities and covers 9 years (1979-1987). These variables include total expenditures (expenditures), total own-source revenues (revenues) and intergovernmental grants received by the municipality (grants). Following Dahlberg and Johansson (2000) grants from the central to the local government are of three kinds: support to municipalities with small tax capacity, grants toward the running of certain local government activities and grants toward certain investments.

```
library(panelvar)
data("Dahlberg")
ex1_dahlberg_data <-
  pvargmm(dependent_vars = c("expenditures", "revenues", "grants"),
          lags = 1,
          transformation = "fod",
          data = Dahlberg,
          panel_identifier=c("id", "year"),
          steps = c("twostep"),
          system_instruments = FALSE,
          max_instr_dependent_vars = 99,
          max_instr_predet_vars = 99,
          min_instr_dependent_vars = 2L,
          min_instr_predet_vars = 1L,
          collapse = FALSE
  )
summary(ex1_dahlberg_data)
```

```
---------------------------------------------------
Dynamic Panel VAR estimation, two-step GMM
---------------------------------------------------
Transformation: Forward orthogonal deviations
Group variable: id
Time variable: year
Number of observations = 2120
Number of groups = 265
Obs per group: min = 8
               avg = 8
               max = 8
Number of instruments = 252

=========================================================
                   expenditures  revenues     grants
---------------------------------------------------------
lag1_expenditures   0.2846 ***    0.2583 **    0.0167
                   (0.0664)      (0.0795)     (0.0172)
lag1_revenues      -0.0470        0.0588      -0.0405 **
                   (0.0637)      (0.0726)     (0.0151)
lag1_grants        -1.6746 ***   -2.2367 ***   0.3204 ***
                   (0.2818)      (0.2846)     (0.0521)
=========================================================
*** p < 0.001, ** p < 0.01, * p < 0.05
```

```
---------------------------------------------------
Instruments for  equation
 Standard

 GMM-type
  Dependent vars: L(2, 8)
  Collapse =  FALSE
---------------------------------------------------

Hansen test of overid. restrictions: chi2(243) = 263.01 Prob > chi2 = 0.18
(Robust, but weakened by many instruments.)= 263.01 Prob > chi2 = 0.18
(Robust, but weakened by many instruments.)
```

With the following sequence of commands, we show how to use the model selection procedure of Andrews and Lu (2001) to select the optimal lag length for our example.

```
Andrews_Lu_MMSC(ex1_dahlberg_data)

$MMSC_BIC
[1] -1610.877

$MMSC_AIC
[1] -234.9924

$MMSC_HQIC
[1] -792.3698

ex2_dahlberg_data <- pvargmm(dependent_vars = c("expenditures", "revenues", "grants"),
                                         lags = 2,
                                         transformation = "fod",
                                         data = Dahlberg,
                                         panel_identifier=c("id", "year"),
                                         steps = c("twostep"),
                                         system_instruments = FALSE,
                                         max_instr_dependent_vars = 99,
                                         max_instr_predet_vars = 99,
                                         min_instr_dependent_vars = 2L,
                                         min_instr_predet_vars = 1L,
                                         collapse = FALSE
)

> Andrews_Lu_MMSC(ex2_dahlberg_data)

$MMSC_BIC
[1] -1486.935

$MMSC_AIC
[1] -213.8917

$MMSC_HQIC
[1] -734.1071
```

All three included moment and model selection criteria (Andrews and Lu, 2001) select the model with one lag over the models with two, three and four lags.[17]

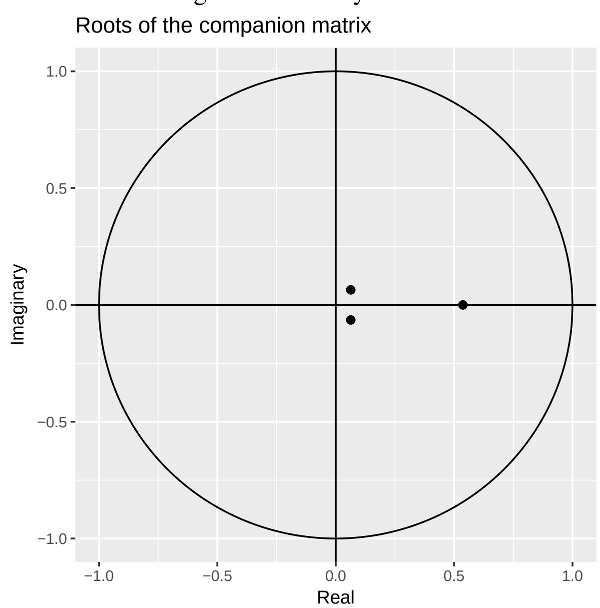
Next, we can test the stability of the autoregressive process:

```
stab_ex1_dahlberg_data <- stability(ex1_dahlberg_data)
print(stab_ex1_dahlberg_data)

Eigenvalue stability condition:

             Eigenvalue    Modulus
1 0.5370657+0.00000000i 0.53706574
2 0.0633651+0.06442426i 0.09036383
3 0.0633651-0.06442426i 0.09036383

All the eigenvalues lie inside the unit circle.
PVAR satisfies stability condition.
```

We generate a plot of the roots by executing the following line:

```
plot(stab_ex1_dahlberg_data)
```

Figure 1: Stability check



Then we can generate impulse response functions.

[17]Lag = 3: BIC = -1301.451, AIC = -180.5911, HQIC = -643.3513, Lag = 4: BIC = -1098.728, AIC = -175.047, HQIC = -561.2191.

```
ex1_dahlberg_data_oirf <-  oirf(ex1_dahlberg_data, n.ahead = 8)
ex1_dahlberg_data_girf <-  girf(ex1_dahlberg_data, n.ahead = 8, ma_approx_steps= 8)
```

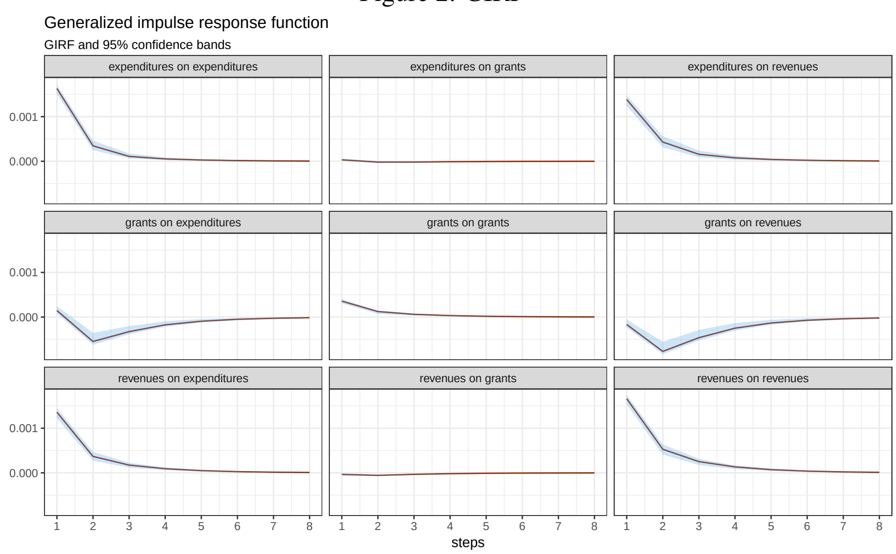
Next, we use the bootstrap function of calculate confidence intervals.

```
ex1_dahlberg_data_bs <-  bootstrap_irf(ex1_dahlberg_data, typeof_irf = c("GIRF"),
                                    n.ahead = 8,
                                    nof_Nstar_draws = 500,
                                    confidence.band = 0.95)
```

Then we plot the impulse response functions including the bootstrapped confidence intervals.

```
plot(ex1_dahlberg_data_girf, ex1_dahlberg_data_bs)
```

Figure 2: GIRF



3.3. Comparison of PVAR STATA Code with panelvar package

Comparing Abrigo and Love (2016) (**STATA: pvar**) with our code (**R: panelvar**) is more complex due to the fact that there is no widely accepted third code that we can use as a benchmark. We do not compare our code to Love and Zicchino (2006) as we think that Abrigo and Love (2016) is a major improvement, especially with respect to transformations (fd and fod) applied to remove the fixed effects.[18]

[18] Some parts of the documentation in Love and Zicchino (2006) are also simply wrong. E.g. it is recommended to apply the fod/fd transformation twice before starting the estimation.

We make two comparisons. First, we use the EmplUk (abdata) set with a simple example. Second we apply both codes to simulated data.

For the EmplUK (abdata), we set up a model with two endogenous variables, employment and wage.[19]

```
clear all
webuse abdata
pvar emp wage, instl(2/99) gmmstyle fod

Panel vector autoregresssion



GMM Estimation

Final GMM Criterion Q(b) =     .0912
Initial weight matrix: Identity
GMM weight matrix:     Robust
                                                    No. of obs      =       751
                                                    No. of panels   =       140
                                                    Ave. no. of T   =     5.364


------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
emp          |
         emp |
         L1. |   1.831062    .106205    17.24   0.000     1.622904     2.03922
             |
        wage |
         L1. |      .3553   .0526694     6.75   0.000     .2520699      .45853
-------------+----------------------------------------------------------------
wage         |
         emp |
         L1. |    .214861   .0737082     2.92   0.004     .0703955    .3593265
             |
        wage |
         L1. |   1.454015   .1737963     8.37   0.000     1.113381     1.79465
------------------------------------------------------------------------------
Instruments : l(2/99).(emp wage)

pvar emp wage, instl(2/99) gmmstyle fd

Panel vector autoregresssion



GMM Estimation

Final GMM Criterion Q(b) =     .0777
```

[19] This model is for illustration only, we do not claim that it is correctly specified or makes any economic sense.

```
Initial weight matrix: Identity
GMM weight matrix:     Robust
                                                   No. of obs      =       751
                                                   No. of panels   =       140
                                                   Ave. no. of T   =     5.364


------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
emp          |
         emp |
         L1. |   1.058351   .1031948    10.26   0.000     .8560933     1.26061
             |
        wage |
         L1. |   .0659102   .0245567     2.68   0.007       .01778    .1140405
-------------+----------------------------------------------------------------
wage         |
         emp |
         L1. |   -.382159    .098159    -3.89   0.000    -.5745471   -.1897709
             |
        wage |
         L1. |   .0646176   .0689922     0.94   0.349    -.0706047    .1998399
------------------------------------------------------------------------------
Instruments : l(2/99).(emp wage)
```

The results in table 1 should be quite similar. However, **STATA: PVAR** estimation with the fod transformation is quite different from the other columns. Given the balancedness of the data, this is unexpected.

Table 1: Comparison **STATA: PVAR** with **R: panelvar**

	STATA: pvar collapse = T	**STATA: pvar** collapse = T	**R: pvargmm** collapse = T	**R: pvargmm** collapse = T
	fd_emp	fod_emp	fd_emp	fod_emp
fd_lag1_emp	1.0584 ***	1.8311 ***	1.1024 ***	1.1021 ***
	(0.1032)	(0.1062)	(0.1428	(0.1416)
fd_lag1_wage	0.0659 ***	0.3553 ***	-0.0472	-0.0569
	(0.2455)	(0.0526)	(0.0848)	(0.0884)
	fd_wage	fod_wage	fd_wage	fod_wage
fd_lag1_emp	-0.3821 ***	0.2148 ***	-0.4038 ***	-0.4128 ***
	(0.0981)	(0.0737)	(0.1300)	(0.1301)
fd_lag1_wage	0.0646	1.4540 ***	0.1626	0.1765
	(0.0689)	(0.1737)	(0.0922)	(0.0911)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Source: OeNB and own calculations.

Second, we simulate data with $N = 500$ and $T = 10$. The true parameter of the equations are simulated as follows:

$$\begin{aligned} \text{y1\_endo}_{i,t} &= \alpha_{i,1} + 0.7 * \text{y1\_endo}_{i,t-1} - 0.1 * \text{y2\_endo}_{i,t-1} \\ \text{y2\_endo}_{i,t} &= \alpha_{i,2} - 0.4 * \text{y1\_endo}_{i,t-1} + 0.8 * \text{y2\_endo}_{i,t-1} \end{aligned} \tag{38}$$

The PVAR code for Abrigo and Love (2016) reads as follows:

```
*Plz adjust input path for the simulated data:
global inputfiles "C:\PVAR\Supplementary Files\Simulated Data"

cd "$inputfiles"
insheet using "PVAR-Simulated-data-2-endo.csv", double delimiter(";") case

* Example 1.A.:
pvar y1_endo y2_endo, instl(2/99) gmmstyle fd

Panel vector autoregresssion



GMM Estimation
```

```
Final GMM Criterion Q(b) =     .0309
Initial weight matrix: Identity
GMM weight matrix:     Robust
                                                   No. of obs      =      4000
                                                   No. of panels   =       500
                                                   Ave. no. of T   =     8.000


------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1_endo      |
     y1_endo |
         L1. |   .5336504   .1265049     4.22   0.000     .2857054    .7815954
             |
     y2_endo |
         L1. |  -.1604724   .0479966    -3.34   0.001    -.2545441   -.0664008
-------------+----------------------------------------------------------------
y2_endo      |
     y1_endo |
         L1. |  -.2474478     .15373    -1.61   0.107    -.5487532    .0538575
             |
     y2_endo |
         L1. |   .7897402   .0567985    13.90   0.000     .6784171    .9010632
------------------------------------------------------------------------------
Instruments : l(2/99).(y1_endo y2_endo)

* Example 1.B.:
pvar y1_endo y2_endo, instl(2/99) gmmstyle fod

Panel vector autoregresssion



GMM Estimation

Final GMM Criterion Q(b) =     .0201
Initial weight matrix: Identity
GMM weight matrix:     Robust
                                                   No. of obs      =      4000
                                                   No. of panels   =       500
                                                   Ave. no. of T   =     8.000


------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1_endo      |
     y1_endo |
         L1. |   .5694483   .0872861     6.52   0.000     .3983706     .740526
             |
     y2_endo |
         L1. |  -.1507859   .0352422    -4.28   0.000    -.2198593   -.0817125
-------------+----------------------------------------------------------------
y2_endo      |
     y1_endo |
         L1. |   -.358661   .0824727    -4.35   0.000    -.5203046   -.1970174
             |
```

```
     y2_endo |
         L1. |   .7983379   .0333429    23.94   0.000     .7329869    .8636888
------------------------------------------------------------------------------
Instruments : l(2/99).(y1_endo y2_endo)
```

The corresponding R code reads as follows:

```
library(panelvar)

# Load data: please adjust your input folder:
input_files <- "K:/PVAR/Supplementary Files/Simulated Data/"
file_name_simluated_data1 <- c("PVAR-Simulated-data-2-endo.csv")

simluated_data1 <- read.table(paste(input_files, file_name_simluated_data1, sep = ""),
                              header = TRUE, sep = ";", dec=".", na.strings = "NA", comment.char="")

ex1A_abigo_fd_nosystem <-  pvargmm(dependent_vars = c("y1_endo",        "y2_endo"),
                                        lags = 1,
                                        transformation = "fd",
                                        data = simulated_data1,
                                        panel_identifier = c("cat", "time"),
                                        steps = c("twostep"),
                                        system_instruments = FALSE,
                                        max_instr_dependent_vars = 99,
                                        max_instr_predet_vars = 99,
                                        min_instr_dependent_vars = 2L,
                                        min_instr_predet_vars = 1L,
                                        collapse = FALSE)

summary(ex1A_abigo_fd_nosystem)

---------------------------------------------------
Dynamic Panel VAR estimation, two-step GMM
---------------------------------------------------
Transformation: fd
Group variable: cat
Time variable: time
Number of observations = 4500
Number of groups = 500
Obs per group: min = 9
               avg = 9
               max = 9
Number of instruments = 144

======================================
              y1_endo      y2_endo
--------------------------------------
lag1_y1_endo   0.7220 ***  -0.3889 ***
              (0.0262)     (0.0290)
lag1_y2_endo  -0.0818 ***   0.7906 ***
              (0.0134)     (0.0142)
======================================
*** p < 0.001, ** p < 0.01, * p < 0.05
```

```
---------------------------------------------------
Instruments for first differences equation
 Standard

 GMM-type
  Dependent vars: L(2, 9)
  Collapse =  FALSE
---------------------------------------------------

Hansen test of overid. restrictions: chi2(140) = 265.69 Prob > chi2 = 0
(Robust, but weakened by many instruments.)

ex1B_abigo_fod_nosystem <-  pvargmm(dependent_vars = c("y1_endo", "y2_endo"),
                                  lags = 1,
                                  transformation = "fod",
                                  data = simulated_data1,
                                  panel_identifier = c("cat", "time"),
                                  steps = c("twostep"),
                                  system_instruments = FALSE,
                                  max_instr_dependent_vars = 99,
                                  max_instr_predet_vars = 99,
                                  min_instr_dependent_vars = 2L,
                                  min_instr_predet_vars = 1L,
                                  collapse = FALSE)

 summary(ex1B_abrigo_fod_nosystem)

---------------------------------------------------
Dynamic Panel VAR estimation, two-step GMM
---------------------------------------------------
Transformation: fod
Group variable: cat
Time variable: time
Number of observations = 4500
Number of groups = 500
Obs per group: min = 9
               avg = 9
               max = 9
Number of instruments = 144

======================================
              y1_endo      y2_endo
--------------------------------------
lag1_y1_endo   0.7220 ***  -0.3889 ***
              (0.0262)     (0.0290)
lag1_y2_endo  -0.0818 ***   0.7906 ***
              (0.0134)     (0.0142)
======================================
*** p < 0.001, ** p < 0.01, * p < 0.05

---------------------------------------------------
Instruments for orthogonal deviations equation
 Standard

 GMM-type
  Dependent vars: L(2, 9)
  Collapse =  FALSE
---------------------------------------------------
```

```
Hansen test of overid. restrictions: chi2(140) = 265.69 Prob > chi2 = 0
(Robust, but weakened by many instruments.)
```

The results of example 1.A. and 1.B. are compared in Table 2. The first columns repeats the true parameters. The second and third columns contain the estimations with STATA: **pvar** for the first difference and the forward orthogonal transformation. The last two columns show the **panelvar** estimation results. It is important to note that only our code passes the following consistency check (Roodman, 2009b; Arellano and Bover, 1995): For a balanced panel the forward orthogonal and the first difference transformation should yield the same estimation results, if the full set of available instruments is used (no *collapse*, no *max_instr_dependent_vars* or *min_instr_dependent_vars*).[20]

Table 2: Two Simulated endogenous variables with Abrigo STATA Code

	True values	STATA: pvar	STATA: pvar	**R:pvargmm**	**R:pvargmm**
	y1_endo	fd_y1_endo	fod_y1_endo	fd_y1_endo	fod_y1_endo
L.y1_endo	0.7	0.5337 ***	0.5695 ***	0.7220 ***	0.7220 ***
		(0.1265)	(0.0873)	(0.0262)	(0.0262)
L.y2_endo	-0.1	-0.1605 ***	-0.1508 ***	-0.0818 ***	-0.0818 ***
		(0.0480)	(0.0352)	(0.0134)	(0.0134)
	y2.endo	fd_y2_endo	fod_y2_endo	fd_y2_endo	fod_y2_endo
L.y1_endo	-0.4	-0.2474	-0.3587 ***	-0.3889 ***	-0.3889 ***
		(0.1537)	(0.0825)	(0.0290)	(0.0290)
L.y2_endo	0.8	0.7897 ***	0.7983 ***	0.7906 ***	0.7906 ***
		(0.0568)	(0.0333)	(0.0142)	(0.0142)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Source: OeNB and own calculations.

4. Conclusion

Implementing the most GMM estimators for panel vector autoregression models has been a long unsolved problem in the literature ever since the contribution of Holtz-Eakin et al. (1988). First, new and improved theory on GMM estimators where suggested by

[20]In fact, Arellano and Bover (1995) show that in balanced panels, any two transformations of full row rank will yield numerically identical estimators, holding the instrument set fixed.

Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998). In the next step, Roodman (2009b) was able to implement them in dynamic panel models in STATA.

Some early multi equation PVAR GMM estimators were suggested by Love and Zicchino (2006) and Abrigo and Love (2016). In Abrigo and Love (2016) the simple GMM estimator of Anderson and Hsiao (1982) is implemented. However, their contribution did not include the standard first difference and system GMM estimator. Their codes also have limitations with respect to the number of endogenous, exogenous and predetermined variables.

Our code can be considered as an extension of Roodman (2009b) to a PVAR model. The key step of our implementation is the setup of the instrument matrix. We are able to detect the formula behind its dimension. We do not use any routines from existing packages for the estimation and therefore the user can access each element in each matrix of the instrumental variable estimator.

From a theoretical point of view, there are still open questions with respect to improving the GMM estimations by adding additional moment conditions, by finding more efficient weighting matrices, etc. We therefore think that our code could serve as a platform for incorporating such new elements.

Finally, we hope that our code could be used in numerous empirical applications and replace existing codes for PVAR model that rely on less efficient estimators as shown in Section 3.3..

References

Abrigo, M. and Love, I. (2016). Estimation of Panel Vector Autoregression in Stata. *The Stata Journal*, 16(3):778–804.

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

Anderson, T. W. and Hsiao, C. (1982). Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics*, 18(1):47–82.

Andrews, D. and Lu, B. (2001). Consistent Model and Moment Selection Procedures for GMM Estimation with application to dynamic panel data models. *Journal of Econometrics*, 101(1):123–164.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297.

Arellano, M. and Bover, O. (1995). Another Look at the Instrumental Variable Estimation of Error-components Models. *Journal of Econometrics*, 68(1):29–51.

Barbu, V. and Precupanu, T. (2012). *Convexity and Optimization in Banach Spaces*. Springer Netherlands.

Binder, M., Hsiao, C., and Pesaran, M. H. (2005). Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration. *Econometric Theory*, 21(4):795–837.

Blundell, R. and Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, 87(1):115–143.

Croissant, Y. and Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, 27(2):1–43.

Dahlberg, M. and Johansson, E. (2000). Behavior of Local Governments Using GMM Bootstrapping Methods. *Journal of Applied Econometrics*, 15(4):401–416.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054.

Hansen, L. P. (2012). Proofs for Large Sample Properties of Generalized Method of Moments Estimators. Working paper, University of Chicago.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating Vector Autoregressions with Panel Data. *Econometrica*, 56(6):1371–1395.

Kapetanios, G. (2008). A Bootstrap Procedure for Panel Data Sets with Many Cross-Sectional Units. *The Econometrics Journal*, 11(2):377–395.

Koenker, R. and Machado, J. (1999). GMM Inference when the Number of Moment Conditions is Large. *Journal of Econometrics*, 93(2):327–344.

Lin, J. (2006). Teaching Notes on Impulse Response Function and Structural VAR. Technical report, Department of Economics, National Chengchi University.

Love, I. and Zicchino, L. (2006). Financial Development and Dynamic Investment Behavior: Evidence from Panel VAR. *The Quarterly Review of Economics and Finance*, 46(2):190–210.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag Berlin Heidelberg, 2 edition.

Mehrhoff, J. (2009). A Solution to the Problem of Too Many Instruments in Dynamic Panel Data GMM. Technical report, Deutsche Bundesbank, Research Centre.

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier, 1 edition.

Newey, W. and Smith, R. (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, 72(1):219–255.

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics*, 29(3):229–256.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417–1426.

Pesaran, H. and Shin, Y. (1998). Generalized Impulse Response Analysis in Linear Multivariate Models. *Economics Letters*, 58(1):17–29.

Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within r package vars. *Journal of Statistical Software*, 27(4).

Rockafellar, R. (1970). *Convex Analysis*. Princeton University Press.

Roodman, D. (2009a). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158.

Roodman, D. (2009b). How to do xtabond2: An Introduction to Difference and System GMM in Stata. *The Stata Journal*, 9(1):86–136.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory*. Oxford University Press.

Sims, C. (1980). Macroeconomic Reality. *Econometrica*, 48(1):1–48.

Windmeijer, F. (2005). A Finite Sample Correction for the Variance of Linear Efficient two-step GMM Estimators. *Journal of Econometrics*, 126(1):25–51.

Yan, M. (2012). Essay in Econometrics. Master's thesis, Universitaet Wien.

Appendix A. Windmeijer correction

Without loss of generality we focus on the two-step first difference GMM estimator in Eq. (21) to derive the Windmeijer correction for the two-step covariance matrix.

First, it is important to note that $\mathbf{\Lambda}_{\mathbf{Z}_{\hat{e}}}$ is a function of the one-step estimator $\mathbf{\Phi}_{IE}$, so $\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})$. Suppose $\mathbf{\Phi}_0$ is the true parameter and its weighting matrix is $\mathbf{\Lambda}_{\mathbf{Z}_{\hat{e}}}(\mathbf{\Phi}_0)$.

Using this notation we can write the finite sample bias of $\mathbf{\Phi}_{FEGMM}$ as follows:[21]

$$\begin{aligned}
\mathbf{\Phi}_{FEGMM} - \mathbf{\Phi}_0 &= \left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1} S_{Zy} - \mathbf{\Phi}_0 \\
&= \left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1} \\
&\quad \times \left(\mathbf{S}_{ZX}\mathbf{\Phi}_0 + \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i(\Delta\mathbf{e}_i)^{\intercal}\right) - \mathbf{\Phi}_0 \qquad \text{(A.1)} \\
&= \left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\mathbf{\Phi}_{IE}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\frac{1}{N}\left(\sum_{i=1}^{N} Z_i(\Delta\mathbf{e}_i)^{\intercal}\right)
\end{aligned}$$

Windmeijer (2005) suggests to use Taylor's formula to approximate $\mathbf{\Phi}_{FEGMM} - \mathbf{\Phi}_0$ at $\mathbf{\Phi}_0$.

$$\begin{aligned}
\mathbf{\Phi}_{FEGMM} - \mathbf{\Phi}_0 &= \left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}\mathbf{\Phi}_0\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\frac{1}{N}\left(\sum_{i=1}^{N} Z_i(\Delta\mathbf{e}_i)^{\intercal}\right) \\
&\quad + D_{\Phi_0,(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0))^{-1}}(\mathbf{\Phi}_{IE} - \Phi_0) + o(\|\mathbf{\Phi}_{IE} - \Phi_0\|)
\end{aligned}$$

Most importantly the j-th column of the matrix is defined by[22],

[21] See Roodman (2009a) p. 89 for a similar derivation of a simpler model.

[22] See Windmeijer (2005) for more details in the case of $m = 1$.

$$D_{\Phi_0,\Lambda(\Phi_0)}e_j = \left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}\mathbf{\Phi}_0\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}\frac{\partial\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi})}{\partial\beta_j}|_{\Phi=\Phi_0}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}\mathbf{S}_{ZX}$$
$$\times\left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}Z_i(\Delta\mathbf{e}_i)^{\intercal}\right)$$
$$-\left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}\frac{\partial\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)}{\partial\Phi_j}|_{\Phi=\Phi_0}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)\right)^{-1}$$
$$\times\left(\frac{1}{N}\sum_{i=1}^{N}Z_i(\Delta\mathbf{e}_i)^{\intercal}\right) \tag{A.2}$$

with $\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)$ defined in Eq. (22). Define

$$\frac{\partial\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi})}{\partial\Phi_j} = -\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}^{\intercal}\left(\mathbf{S}_{ZX}\Delta\mathbf{e}_i^{\intercal}+\Delta\mathbf{e}_i\mathbf{S}_{ZX}^{\intercal}\right)\mathbf{Z}_i.$$

Similar to the derivation of Eq. (A.1), $\mathbf{\Phi}_{IE}-\mathbf{\Phi}_0$ can be obtained by:

$$\Phi_{IE}-\Phi_0 = \left(\mathbf{S}_{ZX}^{\intercal}(\mathbf{\Lambda}_Z)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}(\mathbf{\Lambda}_Z)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}Z_i(\Delta\mathbf{e}_i)^{\intercal}\right)$$

Substituting $\mathbf{\Lambda}(\Phi_0)$ by $\mathbf{\Lambda}(\Phi_{IE})$ in Eq. (A.2) and $\mathbf{e}$ by $\hat{\mathbf{e}}$ (as defined in Eq. (22)) one gets

$$D_{\Phi_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\Phi_{IE})}e_j = -\left(\mathbf{S}_{ZX}^{\intercal}(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE}))^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\frac{\partial\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi})}{\partial\Phi_j}|_{\Phi=\Phi_{IE}}$$
$$\times\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}Z_i(\Delta\hat{\mathbf{e}}_i)^{\intercal}\right) \tag{A.3}$$

Again, substituting $D_{\Phi_0,\Lambda(\Phi_0)}e_j$ (Eq. (A.2)) by $D_{\Phi_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\Phi_{IE})}e_j$ (Eq. (A.3)) and replacing $\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_0)$ by $\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})$ it is possible to approximate $\mathbf{\Phi}_{FEGMM}-\mathbf{\Phi}_0$ by:

$$\begin{aligned}\mathbf{\Phi}_{FEGMM} - \mathbf{\Phi}_0 \approx &\left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\frac{1}{N}\left(\sum_{i=1}^{N} Z_i(\Delta\mathbf{e}_i)^{\intercal}\right)\\ &+ D_{\Phi_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\Phi_{IE})}\left(\mathbf{S}_{ZX}\left(\mathbf{\Lambda}_{Z}\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z}\right)^{-1}\\ &\times\left(\frac{1}{N}\sum_{i=1}^{N} Z_i(\Delta\hat{\mathbf{e}}_i)^{\intercal}\right) + o(\|\mathbf{\Phi}_{IE} - \Phi_0\|)\end{aligned}\tag{A.4}$$

Based on Eq. (A.4) we derive the Windmeijer corrected asymptotic variance of the two-step (FEGMM) estimator:

$$\begin{aligned}\hat{\mathbf{V}}\mathbf{ar}_{\Phi_{FEGMM}}^{Wc} = &\left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\\ &+ D_{\Phi_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\Phi_{IE})}\left(\mathbf{S}_{ZX}\left(\mathbf{\Lambda}_{Z}\right)^{-1}\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z}\right)^{-1}\\ &+ \left(\mathbf{S}_{ZX}^{\intercal}\left(\mathbf{\Lambda}_{Z_{\hat{e}}}(\mathbf{\Phi}_{IE})\right)^{-1}\mathbf{S}_{ZX}\right)^{-1} D_{\Phi_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\Phi_{IE})}^{\intercal}\\ &+ D_{\Phi_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\Phi_{IE})}\hat{\mathbf{V}}\mathbf{ar}_{\Phi_{IE}} D_{\Phi_{FEGMM},\mathbf{\Lambda}_{Z_{\hat{e}}}(\Phi_{IE})}^{\intercal}\end{aligned}$$

Appendix B. Number of Moment Conditions

To describe the dimension of $\mathbf{Q}_i$, we present a table with six examples. Suppose $T = 9$ and all data of individual i are available.

Table B.3: Number of first difference moment conditions for an individual

Time Periods	$m = p = 1$	$m = p = k = 1$	$m = 1, p = 2$	$m = 1, p = 2, k = 1$	$m = 1, p = 3$	$m = 1, p = 3, k = 2$ $L_{max}^{endo,pre} = 5$
$t = 1$	0	0	0	0	0	0
$t = 2$	0	0	0	0	0	0
$t = 3$	1	3	0	0	0	0
$t = 4$	2	5	2	5	0	0
$t = 5$	3	7	3	7	3	11
$t = 6$	4	9	4	9	4	14
$t = 7$	5	11	5	11	5	17
$t = 8$	6	13	6	13	6	17
$t = 9$	7	15	7	15	7	17
Total Number	28	63	27	60	25	76

This table applies Theorem 2.1 to 6 different settings. Each row counts the number of moment conditions for time period t. In the sixth setting we fix the maximal number of available lags to 5, instead of using all available lags.
First setting (column 2): The number of endogenous variables $m = 1$, lags (p) are set to 1. $L_{min}^{endo} = 2$ (the lag to start the first difference instrumenting), $L_{max}^{endo} = 9$ (which means that all available instruments are used for instrumenting). $T = 9$ number of time periods for individual i. Applying Theorem 2.1 implies $S = S_m$. $a_1 = 1, n_1 = 7 \Rightarrow S_m(1) = 28$. $a_2 = 7$ and $n_2 = 0$, so $S = 28$ and $\dim(\mathbf{Q}_i) = 7 \times 28$.
Second setting (column 3): The number of endogenous variables $m = 1$, lags (p) are set to 1. $L_{min}^{endo} = 2$ (the lag to start the first difference instrumenting), $L_{max}^{endo} = 9$ (which means that all available instruments are used for instrumenting). The number of predetermined variables $k = 1$. $L_{min}^{pre} = 1$ (the lag to start the first difference instrumenting), $L_{max}^{pre} = 9$ (which means that all available instruments are used for instrumenting). $T = 9$ number of time periods for individual i. Applying Theorem 2.1 implies $S = S_m + S_k$. $a_1 = 1, n_1 = 7 \Rightarrow S_m(1) = 28$. $a_2 = 7$ and $n_2 = 0$. $o_1 = 7, b_1 = 2, o_2 = 0$ and $b_2 = 9$. Therefore, $S = 28 + 35$ and $\dim(\mathbf{Q}_i) = 7 \times 63$.
Third setting: $m = 1$ and $p = 2$. As a consequence, there are now two lags of the endogenous variable. Again, $L_{min}^{endo} = 2$, and $L_{max}^{endo.pre} = 9$. $S = S_m$. $a_1 = 3, n_1 = 5 \Rightarrow S_m(1) = 27$. $\dim(\mathbf{Q}_i) = 27 \times 60$.
Fourth setting (column 4): $m = 1$, $p = 2$ and $k = 1$. As a consequence, there are now two lags of the endogenous variable and one predetermined variable. For simplicity, we keep: $L_{min}^{endo} = 2$, $L_{min}^{pre} = 1$ and $L_{max}^{endo,pre} = 9$. $S = S_m$. $a_1 = 3, n_1 = 5, a_2 = 7$ and $n_2 = 0$. $\Rightarrow S_m(1) = 27$. $o_1 = 6, b_1 = 3, o_2 = 0$ and $b_2 = 8$. $\Rightarrow S_k = 33$. Therefore, $S = 27$ and $\dim(\mathbf{Q}_i) = 7 \times 60$.
Fifth setting (column 5): $m = 1$, $p = 3$, $L_{min}^{endo} = 2$ and $L_{max}^{endo} = 9$ $S = S_m$. $a_1 = 3, n_1 = 5, a_2 = 7$ and $n_2 = 0$. $\Rightarrow S_m(1) = 25$ and $\dim(\mathbf{Q}_i) = 7 \times 25$.
Sixth setting (column 6): $m = 1$, $p = 3$, $L_{min}^{endo} = 2$ and $L_{max}^{endo} = 5$. $k = 2$ with $L_{min}^{pre} = 1$ and $L_{max}^{pre} = 5$. Be aware, that we set the maximal lag length for the first difference instruments to 5, which is a change to the first five settings. $S = S_m + S_k$. $a_1 = 3, n_1 = 3$, $a_2 = 5$ and $n_2 = 2$. $\Rightarrow S_m = 12 + 10 = 22$. For the predetermined variable we add the following: $o_1 = 3, b_1 = 4, o_2 = 2$ and $b_2 = 6$. $\Rightarrow S_k = 54 \Rightarrow S = 76$.

Appendix C. List of mathematical symbols

Variable N. (Paper)	Variable N. (Code)	Dimension	Description
$\mathbf{A}_l$	(Part of) `Phi`	$m \times m$	Coefficients of lagged dependent variables
$\mathbf{B}$	(Part of) `Phi`	$m \times k$	Coefficients of predetermined variables
$\mathbf{C}$	(Part of) `Phi`	$m \times n$	Coefficients of strictly exogenous variables
$\Delta \mathbf{Y}_i$	`delta_W[[i]]`	$(T-1-p) \times m$	First difference or forward orthogonal transformation of the dependent variables
$\Delta \mathbf{W}_{minus}$	`delta_W_minus[[i]]`	$(T-1-p) \times (m * p + k + n)$	First difference or forward orthogonal transformation of lagged dependent, predetermined and strictly exogenous variables
$\mathbf{Q}_i$	`Q[[i]]`	see Eq. (4)	Matrix of instruments
$\mathbf{\Lambda}_Q$	`sum_Lambda`	$\text{nrow}(\mathbf{Q}_i) \times \text{nrow}(\mathbf{Q}_i)$	Weighting matrix for the first step estimator
$\mathbf{D}$	`V`	$(T-1) \times T$	Transformation matrix
$\hat{\mathbf{\Phi}}_{IE}$	`Phi_first_step`	$m \times (m * p + k + n)$	Onestep estimator of $\mathbf{\Phi}$
$\mathbf{S}_{QX}$	`sum_S_ZX_a`	$\text{nrow}(\mathbf{Q}_i) \times (T-1-p)$	$\sum_{i=1}^{N} \mathbf{Q}_i^{\intercal} \Delta \mathbf{W}_{minus}$
$\mathbf{S}_{Qy}$	`sum_S_Zy_a`	$\text{nrow}(\mathbf{Q}_i) \times (T-1-p)$	$\sum_{i}^{N} \mathbf{Q}_i^{\intercal} \Delta \mathbf{W}_i$
$\hat{\mathbf{\Phi}}_{FEGMM}$	`Phi_second_step`	$m \times (m * p + k + n)$	Twostep estimator of $\mathbf{\Phi}$
$\mathbf{Z}_i$	`Z[[i]]`	$\mathbf{Q}_i \otimes \mathbf{I}_{m \times m}$	Kronecker product of the instrument matrix and the number of endogenous variables
$\hat{\mathbf{E}}_i$	`Delta_E_first_step[[i]]`	$(T-1-p) \times m$	Fitted values with onestep estimator
$\hat{\mathbf{e}}_i$	`e_hat`	$(T-1-p) * m \times 1$	Vectorization of $\hat{\mathbf{E}}_i$
$\mathbf{\Lambda}_{\mathbf{Z}_{\hat{e}}}$	`sum_D_e`	$\text{nrow}(\mathbf{Z}_i) \times \text{nrow}(\mathbf{Z}_i)$	Optimal weighting matrix for the second step estimator
$\hat{\mathbf{V}}\mathbf{ar}_{\hat{\mathbf{\Phi}}_{IE}}$	`var_first_step`	$m \times (m * p + k + n)$	Variance of the first step estimator coefficients
$\hat{\mathbf{V}}\mathbf{ar}^{Wc}_{\hat{\mathbf{\Phi}}_{FEGMM}}$	`var_second_step`	$m \times (m * p + k + n)$	Windmeijer corrected variance of the second step estimator coefficients