pdynmc – An **R**-package for estimating linear dynamic panel data models based on nonlinear moment conditions

Markus Fritsch
University of Passau

Andrew Adrian Yu Pua
Xiamen University

Joachim Schnurbus
University of Passau

Abstract

pdynmc is an R-package for IV- or GMM-estimation of linear dynamic panel data models that are based on nonlinear moment conditions as proposed by Ahn and Schmidt (1995). This paper provides a description of the variety of options regarding instrument type, covariate type, estimation methodology, and general configuration from the perspective of an applied statistician. All functionality is demonstrated for a publicly available unbalanced panel data set (Arellano and Bond 1991) and we relate to other software and packages.

Keywords: dynamic panel, generalized method of moments, instrument variable, nonlinear moment conditions, R.

1. Introduction

The linear dynamic panel data model allows to account for dynamics and unobserved individual-specific heterogeneity simultaneously. Due to the presence of unobserved individual-specific effects and lagged dependent variables, standard estimation techniques like pooled ordinary least squares (OLS) or the within estimation generally do not lead to consistent estimates (see, e.g., Hsiao 2014). A suitable alternative for obtaining parameter estimates in linear dynamic panel data models is deriving moment conditions (or population orthogonality conditions) from the model assumptions. The moment conditions may be linear (Anderson and Hsiao 1982; Holtz-Eakin, Newey, and Rosen 1988; Arellano and Bover 1995) or nonlinear (Ahn and Schmidt 1995) in parameters and determine the natural instruments available for estimation. Usually, the number of moment conditions exceeds the number of parameters and the moment conditions need to be aggregated appropriately. This can be achieved by the generalized method of moments (GMM), where (weighted) linear combinations of the moment conditions are employed to obtain parameter estimates.

Theoretical results and evidence from Monte Carlo simulations in the literature suggest that incorporating the nonlinear moment conditions proposed by Ahn and Schmidt (1995) may prove valuable for particular data generating processes (DGPs). One example is when the process exhibits high persistence and the linear moment conditions fail to identify the model parameters: The nonlinear moment conditions may still provide identification (Bun and Kleibergen 2014; Bun and Sarafidis 2015; Gorgens, Han, and Xue 2016). Further note that the nonlinear moment conditions only impose standard assumptions about the (unknown)

underlying DGP. Despite these results, however, and although the nonlinear moment conditions were proposed by Ahn and Schmidt more than 20 years ago, standard estimation routines are generally not available across statistical software. To the best of our knowledge, there is currently only the implementation provided by (Kripfganz 2019) for the commercial statistical software Stata (Stata Corporation 2011) that is explicitly designed to incorporate nonlinear moment conditions into GMM estimation.

Our package **pdynmc** provides an implementation of GMM estimation of linear dynamic panel data models based on different sets of moment conditions in the statistical open source software R (R Core Team 2019). The building blocks from which the sets of moment conditions available for GMM estimation can be constructed are the nonlinear (in parameters) Ahn and Schmidt (1995) moment conditions and the two different types of linear moment conditions proposed by Holtz-Eakin *et al.* (1988) and Arellano and Bover (1995). Our package allows to use various combinations of these moment conditions to obtain parameter estimates. In their standard form, the Holtz-Eakin *et al.* (1988), Arellano and Bover (1995), and Ahn and Schmidt (1995) moment conditions are derived from the lagged dependent variable. Additional moment conditions, which may arise from assumptions about the non-lagged dependent explanatory variables, can also be included in estimation. Since the moment conditions employed in GMM estimation of linear dynamic panel data models are derived from model assumptions, a basic understanding of these assumptions is vital for setting up a plausible estimation routine. The methodological part of this paper briefly reviews the assumptions implied when using particular moment conditions in estimation and provides further references for more detailed overviews.

The structure of the paper is as follows. Section 2 briefly sketches the linear dynamic panel data model, states the underlying assumptions frequently used in the literature, and describes the moment conditions arising from the model assumptions. Section 3 covers GMM estimation of linear dynamic panel data models and illustrates the minimization criterion, estimation in one, two, or multiple steps, and closed form solutions. Section 4 outlines the computation of standard errors, specification- and overidentifying restrictions testing, and the testing of general linear hypotheses. Related software and R-packages are summarized in Section 5. Section 6 illustrates the estimation of linear dynamic panel data models with **pdynmc** for the data set of Arellano and Bond (1991) on adjustments of employment of firms located in the United Kingdom. Section 7 concludes and sketches functionality we plan to add to future releases of the package.

2. Linear dynamic panel data model

2.1. Model and standard assumptions

For a given sample with a cross section dimension n and a time series dimension T, consider the two equations

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + u_{i,t}, \quad i = 1, \ldots, n; \ t = 2, \ldots, T, \tag{1}$$

$$u_{i,t} = \eta_i + \varepsilon_{i,t}, \tag{2}$$

where $y_{i,t}$ and $y_{i,t-1}$ denote the dependent variable and its lag, α is the lag parameter, and $x_{i,t}$ is a non-lagged dependent explanatory variable with corresponding slope coefficient β. The

second equation requires that the (unobserved) composite error term $u_{i,t}$ can be separated into an unobserved individual-specific effect η_i and an idiosyncratic remainder component $\varepsilon_{i,t}$.[1] The initial time period is denoted by $t = 1$.

Combining the Equations (1) and (2) yields the single equation form of the model

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \eta_i + \varepsilon_{i,t}, \quad i = 1, \dots, n; \; t = 2, \dots, T. \tag{3}$$

We impose the following set of standard assumptions (SA) from the literature (see Ahn and Schmidt 1995):

$$\text{The data are independently distributed across } i, \tag{4}$$
$$E(\eta_i) = 0, \quad i = 1, ..., n,$$
$$E(\varepsilon_{i,t}) = 0, \quad i = 1, ..., n; \; t = 2, ..., T,$$
$$E(\varepsilon_{i,t} \cdot \eta_i) = 0, \quad i = 1, ..., n; \; t = 2, ..., T,$$
$$E(\varepsilon_{i,t} \cdot \varepsilon_{i,s}) = 0, \quad i = 1, ..., n; \; t \neq s,$$
$$E(y_{i,1} \cdot \varepsilon_{i,t}) = 0, \quad i = 1, ..., n; \; t = 2, ..., T,$$
$$n \to \infty, \text{ while } T \text{ is fixed, such that } \frac{n}{T} \to 0.$$

The six assumptions in Equation (4) imply: First, the assumption that the data are independently distributed across individuals allows for dependence of the model components across time, but not across individuals. Second, the unobserved individual-specific effect and the idiosyncratic remainder component need to be zero in expectation (if this is not the case, a constant can be included in the model to ensure the property). Third, orthogonality of the $\varepsilon_{i,t}$ with the following model components is required: The unobserved individual-specific effects, the idiosyncratic remainder components of all other time periods, and the initial conditions of the $y_{i,t}$-process. Due to the zero mean assumption concerning the $\varepsilon_{i,t}$, uncorrelatedness follows from orthogonality of these model components. The last assumption requires that the cross section dimension is large, while the time series dimension is finite.

2.2. Moment conditions from standard assumptions

Usual approaches in applied statistics obtain (OLS) estimates of the model parameters of Equation (3) by: (i) ignoring the unobserved individual-specific effects, (ii) deducting the individual-specific mean over time from all left-hand- and right-hand side variables (also referred to as the within transformation) of the equation, or (iii) including one dummy per observation in the estimation (the least squares dummy variables – or LSDV – approach; the within estimation and LSDV yield identical slope coefficient estimates). However, due to the presence of the lagged dependent variable and the unobserved individual-specific effects, the techniques (i)-(iii) do not yield consistent estimates without imposing additional restrictions on the model (see, e.g., Hsiao 2014).

---

[1]We only include one lag of the dependent variable, one non-lagged dependent explanatory variable, and omit unobserved time-specific effects for simplicity of exposition and notational convenience. Extending the representation is straightforward. Unobserved time-specific effects can, e.g., be incorporated by including time dummies.

The unobserved individual-specific effects can be eliminated from Equation (3) by first differencing the equation. Utilizing the Δ-operator to indicate the first differencing gives

$$\Delta y_{i,t} = \alpha \Delta y_{i,t-1} + \beta \Delta x_{i,t} + \Delta \varepsilon_{i,t}, \quad i = 1, \ldots, n;\ t = 2, \ldots, T. \tag{5}$$

Due to the first differenced lagged dependent variables $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ and the first differenced error terms $\Delta \varepsilon_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$ not being orthogonal, estimating Equation (5) with OLS still leads to biased coefficient estimates. The standard assumptions stated in Equation (4) provide a remedy: The assumptions imply two sets of moment conditions, whose sample analogues can be used in estimation. Note that the following moment conditions refer to the population and that the expectation is taken over the cross section dimension.

Holtz-Eakin *et al.* (1988) (hereafter HNR) propose the linear (in parameters) moment conditions

$$E(y_{i,s} \cdot \Delta u_{i,t}) = 0, \qquad t = 3, \ldots, T;\ s = 1, \ldots, t - 2. \tag{6}$$

Depending on the time series dimension available for estimation, Equation (6) provides $0.5(T-1)(T-2)$ moment conditions. Equivalent moment conditions can be derived from the non-lagged dependent explanatory variables. Endogenous (x^{end}), predetermined (x^{pre}), and (strictly) exogenous (x^{ex}) variables provide the linear moment conditions (see the Equations (9.5)-(9.7) of Blundell, Bond, and Windmeijer 2001):

$$\begin{aligned}
E(x_{i,s} \cdot \Delta u_{i,t}) = 0, \qquad t &= 3, \ldots, T, \qquad \text{where} \\
s &= 1, \ldots, t-2, \qquad \text{for} \qquad x = x^{\text{end}}, \\
s &= 1, \ldots, t-1, \qquad \text{for} \qquad x = x^{\text{pre}}, \\
s &= 1, \ldots, T, \qquad \text{for} \qquad x = x^{\text{ex}}.
\end{aligned} \tag{7}$$

For endogenous non-lagged dependent explanatory variables, moment conditions analogous to Equation (6) result. When the non-lagged dependent explanatory variables are predetermined, one more moment condition per time period is available and for exogenous non-lagged dependent explanatory variables, all non-lagged dependent explanatory variables can be used as instruments for time periods $t = 3, \ldots, T$ – compared to the case for endogenous non-lagged dependent explanatory variables.

A further set of moment conditions implied by the SA in Equation (4) is elaborated on by Ahn and Schmidt (1995) (hereafter AS). The authors point out that the following $T - 3$ additional moment conditions can be used in estimation:

$$E(u_{i,t} \cdot \Delta u_{i,t-1}) = 0, \qquad t = 4, \ldots, T. \tag{8}$$

Rewriting the equation and expressing the moment conditions in terms of parameters and observable variables reveals that the AS moment conditions are nonlinear in parameters. Equations (6) and (8) are slightly adjusted versions of the AS-Equations (3) and (4).[2]

---

[2]The notation is adjusted to reflect the time periods of a data set. Hence, $t = 0$ of AS is changed to $t = 1$. Additionally note, that the AS moment conditions could be built on reference period T instead of t via $E(u_{i,T} \cdot \Delta u_{i,t}) = 0$, with $t = 3, \ldots, T - 1$ – as originally proposed in Ahn and Schmidt (1995). The HNR moment conditions could then also be expressed based on the reference period T by $E(y_{i,s} \cdot \Delta u_{i,T}) = 0$, with $s = 1, \ldots, T - 2$. We adjust the AS moment conditions here, for all moment conditions to be expressed based on the same reference period.

For a given panel data set, parameter estimates can be obtained by using the sample analogues of the moment conditions. This yields the m sample moment conditions $\overline{\mathbf{M}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{M}_i$. For the linear dynamic panel data model specified in Equation (3), consider the following moment conditions and the corresponding vector of individual moment condition contributions[3] $\mathbf{M}_i$ to be available for estimation:

$$
\begin{pmatrix}
E(y_{i,1} \cdot \Delta u_{i,3}) \\
E(y_{i,1} \cdot \Delta u_{i,4}) \\
E(y_{i,2} \cdot \Delta u_{i,4}) \\
E(y_{i,1} \cdot \Delta u_{i,5}) \\
\vdots \\
E(y_{i,3} \cdot \Delta u_{i,5}) \\
\vdots \\
E(y_{i,T-2} \cdot \Delta u_{i,T}) \\
\hline
E(x_{i,1} \cdot \Delta u_{i,3}) \\
E(x_{i,2} \cdot \Delta u_{i,3}) \\
E(x_{i,1} \cdot \Delta u_{i,4}) \\
\vdots \\
E(x_{i,3} \cdot \Delta u_{i,4}) \\
\vdots \\
E(x_{i,T-1} \cdot \Delta u_{i,T}) \\
\hline
E(u_{i,4} \cdot \Delta u_{i,3}) \\
\vdots \\
E(u_{i,T} \cdot \Delta u_{i,T-1})
\end{pmatrix}_{m \times 1}
=
\begin{pmatrix}
0 \\
0 \\
0 \\
\vdots \\
0 \\
0 \\
\vdots \\
0 \\
0 \\
\vdots \\
0
\end{pmatrix},
\qquad
\underbrace{\mathbf{M}_i}_{m \times 1} =
\begin{pmatrix}
y_{i,1} \cdot \widetilde{\Delta u}_{i,3} \\
y_{i,1} \cdot \widetilde{\Delta u}_{i,4} \\
y_{i,2} \cdot \widetilde{\Delta u}_{i,4} \\
y_{i,1} \cdot \widetilde{\Delta u}_{i,5} \\
\vdots \\
y_{i,3} \cdot \widetilde{\Delta u}_{i,5} \\
\vdots \\
y_{i,T-2} \cdot \widetilde{\Delta u}_{i,T} \\
\hline
x_{i,1} \cdot \widetilde{\Delta u}_{i,3} \\
x_{i,2} \cdot \widetilde{\Delta u}_{i,3} \\
x_{i,1} \cdot \widetilde{\Delta u}_{i,4} \\
\vdots \\
x_{i,3} \cdot \widetilde{\Delta u}_{i,4} \\
\vdots \\
x_{i,T-1} \cdot \widetilde{\Delta u}_{i,T} \\
\hline
\widetilde{u}_{i,4} \cdot \widetilde{\Delta u}_{i,3} \\
\vdots \\
\widetilde{u}_{i,T} \cdot \widetilde{\Delta u}_{i,T-1}
\end{pmatrix}.
$$

The dashed lines separate the different sets of moment conditions shown here: Two sets of HNR moment conditions (derived from the lagged dependent variable and one predetermined $x_{i,t}$) and the nonlinear moment conditions. Compared to the case illustrated, T moment conditions are available for each time period $t = 3, \ldots, T$ from the $x_{i,t}$-process if $x_{i,t}$ is exogenous – while when $x_{i,t}$ is endogenous, one moment condition per time period is lost and the moment conditions resulting from the $x_{i,t}$ are structured equivalently to the ones that arise from the $y_{i,t}$-process[4].

Further consider decomposing the individual moment condition contributions into $\boldsymbol{M}_i = \boldsymbol{Z}_i' \cdot \tilde{\boldsymbol{s}}_i$, where $\boldsymbol{Z}_i'$ denotes the transpose of a matrix that does not depend on parameter estimates, while the column vector $\tilde{\boldsymbol{s}}_i$ does. For the linear dynamic panel data model in Equation (3) with a predetermined $x_{i,t}$, we obtain:

---

[3]In the following, a tilde sign denotes the estimates during optimization, while a hat sign indicates the final optimization results (i.e., the coefficient estimates and the corresponding residuals).

[4]When T is used as reference period (and all moment conditions involve $u_{i,T}$), the number of HNR moment conditions reduces substantially and only one moment condition is available per time period. For the nonlinear moment conditions, the number of moment conditions remains unchanged.

$$
\boldsymbol{Z}'_i =
\begin{pmatrix}
y_{i,1} & 0 & \cdots & 0 & | & 0 & \cdots & & 0 \\
0 & y_{i,1} & & & | & & & & \\
& y_{i,2} & & & | & & & & \\
& 0 & & & | & & & & \\
\vdots & \vdots & & \vdots & | & \vdots & & & \vdots \\
& & & 0 & | & & & & \\
& & & y_{i,1} & | & & & & \\
& & & \vdots & | & & & & \\
0 & 0 & & y_{i,T-2} & | & 0 & \cdots & & 0 \\
\hline
x_{i,1} & 0 & \cdots & 0 & | & 0 & \cdots & & 0 \\
x_{i,2} & 0 & & & | & & & & \\
0 & x_{i,1} & & & | & & & & \\
& x_{i,2} & & & | & & & & \\
& x_{i,3} & & & | & & & & \\
& 0 & & & | & & & & \\
\vdots & \vdots & & \vdots & | & \vdots & & & \vdots \\
& & & 0 & | & & & & \\
& & & x_{i,1} & | & & & & \\
& & & \vdots & | & & & & \\
0 & 0 & & x_{i,T-1} & | & 0 & \cdots & & 0 \\
\hline
0 & & \cdots & 0 & | & 1 & 0 & \cdots & 0 \\
\vdots & & & \vdots & | & \vdots & \ddots & & \vdots \\
& & & & | & & & & 0 \\
0 & & \cdots & 0 & | & 0 & \cdots & 0 & 1
\end{pmatrix}
\underbrace{}_{m \times (2T-5)}
, \qquad
\tilde{\boldsymbol{s}}_i =
\begin{pmatrix}
\widetilde{\Delta u}_{i,3} \\
\widetilde{\Delta u}_{i,4} \\
\vdots \\
\widetilde{\Delta u}_{i,T} \\
\hline
\widetilde{u}_{i,4} \cdot \widetilde{\Delta u}_{i,3} \\
\widetilde{u}_{i,5} \cdot \widetilde{\Delta u}_{i,4} \\
\vdots \\
\widetilde{u}_{i,T} \cdot \widetilde{\Delta u}_{i,T-1}
\end{pmatrix}
\underbrace{}_{(2T-5)\times 1}
.
$$

Compared to the case shown, $x_{i,1}, \ldots, x_{i,T}$ can be used for each time period $t = 3, \ldots, T$ in the HNR-part of the matrix $\boldsymbol{Z}'_i$ if $x_{i,t}$ is strictly exogenous – while an equivalent structure to the $y_{i,t}$-part results for the $x_{i,t}$-part of the matrix if $x_{i,t}$ is endogenous. Changing the reference period from t to T reduces the HNR-part of the matrix to a column vector; the structure of the AS-part remains unchanged.

Stacking the $\boldsymbol{Z}'_i$ for all cross sectional observations horizontally yields the $m \times n(2T - 5)$ matrix $\boldsymbol{Z}' = (\boldsymbol{Z}'_1, \ldots, \boldsymbol{Z}'_n)$. Concatenating the column vectors $\tilde{\boldsymbol{s}}_i$ yields the $n(2T - 5)$ vector $\tilde{\boldsymbol{s}}$.

2.3. Moment conditions from extended assumptions

Under SA, the moment conditions stated in Equations (6), (7), and (8) can be employed to obtain coefficient estimates. Additional moment conditions can be derived from the assumption

$$
E(\Delta y_{i,t} \cdot \eta_i) = 0, \quad i = 1, \ldots, n. \tag{9}
$$

This expression requires that the dependent variable and the unobserved individual-specific effects are constantly correlated over time for each individual. Deviations from the assumption are required to be unsystematic over both, the cross section and the time series dimension (see Section 6.5 in Arellano 2003, which also provides an example). For the case of non-

lagged dependent explanatory variables, Blundell *et al.* (2001) state that if $\Delta y_{i,t}$ and η_i are correlated, it is still possible that $\Delta x_{i,t}$ and η_i are uncorrelated – while the reverse is unlikely to be the case (for a derivation confirming this statement see Fritsch 2019).

From the 'constant correlated effects'[5] assumption, the additional $T-2$ Arellano and Bover (1995) (hereafter AB) linear moment conditions can be derived:

$$E(\Delta y_{i,t-1} \cdot u_{i,t}) = 0, \quad t = 3, \ldots, T. \tag{10}$$

By rewriting these moment conditions, it can be shown that the AB moment conditions encompass the nonlinear AS moment conditions and render them redundant for estimation (for a derivation see Fritsch 2019).

Additional AB moment conditions can be derived from the non-lagged dependent explanatory variables. Depending on the nature of the $x_{i,t}$-process, the further AB moment conditions are available for estimation:

$$E(\Delta x_{i,v} \cdot u_{i,t}) = 0, \qquad \text{where}$$
$$v = 2, \ldots, t-1;\ t = 3, \ldots T, \quad \text{for} \quad x = x^{end},$$
$$v = t;\ t = 2, \ldots, T, \quad \text{for} \quad x = x^{ex} \quad \text{or} \quad x = x^{pre}.$$

From an endogenous $x_{i,t}$, $T-2$ moment conditions can be derived – while $T-1$ moment conditions are available for an exogenous or predetermined $x_{i,t}$.

When using the HNR and AB moment conditions to estimate the linear dynamic panel data model in Equation (3) with a predetermined explanatory variable, $\boldsymbol{M}_i$ is as follows:

---

[5]Bun and Sarafidis (2015) use this term and point out that this assumption is also referred to as 'effect stationarity' (Kiviet 2007a) or 'mean stationarity' (Arellano 2003) in the literature.

$$\boldsymbol{Z}_i' = \underbrace{\begin{pmatrix}
y_{i,1} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \begin{matrix} y_{i,1} \\ y_{i,2} \end{matrix} & & & & & & & & \\
\vdots & 0 & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
& & & 0 & & & & & & \\
& & & y_{i,1} & & & & & & \\
& & & \vdots & & & & & & \\
0 & \cdots & 0 & y_{i,T-2} & 0 & \cdots & 0 & 0 & \cdots & 0 \\
x_{i,1} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
x_{i,2} & 0 & & & & & & & & \\
0 & \begin{matrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{matrix} & & & & & & & & \\
\vdots & 0 & & \vdots & \vdots & & & \vdots & & \vdots \\
& & & 0 & & & & & & \\
& & & x_{i,1} & & & & & & \\
& & & \vdots & & & & & & \\
0 & \cdots & 0 & x_{i,T-1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & & \cdots & 0 & \Delta y_{i,2} & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & & & \vdots & 0 & \Delta y_{i,3} & & \vdots & \vdots & & \vdots \\
& & & & \vdots & & \ddots & 0 & & & \\
0 & & \cdots & 0 & 0 & \cdots & 0 & \Delta y_{i,T-1} & 0 & \cdots & 0 \\
0 & & \cdots & 0 & 0 & \cdots & 0 & \Delta x_{i,2} & 0 & \cdots & 0 \\
& & & & & & & 0 & \Delta x_{i,3} & & \vdots \\
\vdots & & & \vdots & \vdots & & \vdots & \vdots & & \ddots & 0 \\
0 & & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \Delta x_{i,T}
\end{pmatrix}}_{m \times (3T-5)},$$

$$\tilde{\boldsymbol{s}}_i' = \underbrace{\left(\widetilde{\Delta u}_{i,3}, \quad \widetilde{\Delta u}_{i,4}, \quad \cdots, \quad \widetilde{\Delta u}_{i,T} \mid \tilde{u}_{i,3}, \quad \tilde{u}_{i,4}, \quad \cdots, \quad \tilde{u}_{i,T}, \mid \tilde{u}_{i,2}, \quad \tilde{u}_{i,3}, \quad \cdots, \quad \tilde{u}_{i,T} \right)}_{1 \times (3T-5)}.$$

Compared to the case shown, no additional AB moment conditions arise if $x_{i,t}$ is strictly exogenous. When $x_{i,t}$ is endogenous, the structure of the $x_{i,t}$-part corresponding to the AB moment conditions of $\boldsymbol{Z}_i'$ reduces by the last element and is equivalent to the $y_{i,t}$-part corresponding to the AB moment conditions. The changes of the HNR-part of the matrix are as illustrated in Section 2.2. Changing the reference period from t to T yields the changes in the number of HNR moment conditions discussed earlier; the number of AB moment conditions remains the same, with the structure of the AB-part of the matrix $\boldsymbol{Z}_i'$ reducing to

a column vector[6].

3. GMM estimation

3.1. Minimization criterion

For a given loss function, the p parameters of the linear dynamic panel data model in Equation (3) can be estimated by employing the moment conditions derived from the model assumptions. According to the necessary (but not sufficient) order condition for identification (see, e.g., Hayashi 2000), at least as many moment conditions need to be available as there are model parameters for the parameters to be estimable[7]. Each moment condition is a function of the p model parameters. If the number of moment conditions and the number of model parameters coincide ($m = p$), the system of equations (or moment conditions) possesses a unique solution. Due to the number of moment conditions increasing with the time series dimension, the number of available moment conditions typically exceeds the number of model parameters with linear dynamic panel data models. Therefore, obtaining parameter estimates from the system of equations defined by the moment conditions requires an aggregation scheme such as the generalized method of moments (GMM). For a given sample, GMM estimation minimizes the aggregated squared distance of the moment conditions from zero and can be represented as

$$L_{\boldsymbol{W}}^2 = \overline{\boldsymbol{M}}' \cdot \boldsymbol{W} \cdot \overline{\boldsymbol{M}}. \tag{11}$$

The index of the $L_{\boldsymbol{W}}^2$-norm expresses that the norm depends on the weighting matrix $\boldsymbol{W}$ and the superscript indicates that the norm is a quadratic form. The $m \times m$ weighting matrix $\boldsymbol{W}$ guides the aggregation of the m moment conditions. Recall the notation developed in Section 2, where the moment conditions are decomposed into a vector that depends on the parameter estimates $\tilde{\boldsymbol{s}}$ and a matrix $\boldsymbol{Z}$ that does not. Plugging these two terms into Equation (11) gives:

$$L_{\boldsymbol{W}}^2 = \frac{1}{n^2} \cdot \tilde{\boldsymbol{s}}' \boldsymbol{Z} \cdot \boldsymbol{W} \cdot \boldsymbol{Z}' \tilde{\boldsymbol{s}}.$$

Minimizing the equation yields the GMM estimator $\hat{\boldsymbol{\theta}}$.

3.2. One-step, two-step, and multiple-step estimation

---

[6]When using the reference period T instead of t, the AB moment conditions can be built on

$$\begin{aligned} E(\Delta y_{i,v} \cdot u_{i,T}) = 0, \quad &\text{with} \quad t = 3, \ldots, T, \\ E(\Delta x_{i,v} \cdot u_{i,T}) = 0, \quad &\text{where} \\ & v = 2, \ldots, T-1, \quad \text{for} \quad x = x^{end}, \\ & v = 2, \ldots, T, \quad \text{for} \quad x = x^{ex} \quad \text{or} \quad x = x^{pre}. \end{aligned}$$

[7]A discussion of the assumptions required for identification, consistency, and asymptotic normality of the GMM estimator when estimating linear dynamic panel data models is provided in Fritsch (2019).

In practice, GMM estimation is frequently carried out in multiple steps. In order to start the estimation process, an initial estimate of the weighting matrix $\widehat{W}$ is required. Obviously, plugging in different weighting matrices into Equation (11) yields varying objective function values and different estimates for the model parameters. Different propositions for the first step weighting matrix – with varying asymptotic efficiency – exist in the literature (see Blundell *et al.* 2001) for the various types of moment conditions which can be employed in the estimation of the linear dynamic panel data model in Equation (3). Common examples involve identity or tridiagonal matrices. Generally, the proposed weighting matrices are based on the expected variances and covariances of the moment conditions and are derived from the underlying model assumptions. Assuming consistency and asymptotic normality of the GMM estimator, the optimal W is proportional (up to a multiplicative constant) to the inverse of the variance covariance matrix of the moment conditions (see, e.g., Arellano 2003). A possible estimate for the first step weighting matrix $\widehat{W}_1$ of the one-step GMM estimator (GMM1S) is

$$\widehat{W}_1 = \left(\frac{1}{n} \cdot Z'HZ \right)^{-1}. \tag{12}$$

The structure of the matrix H varies depending on the types of moment conditions employed in estimation. When only HNR moment conditions are used, Arellano and Bond (1991) propose to set the matrix to

$$H_{HNR} = \begin{pmatrix} 2 & -1 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & & \\ 0 & -1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & 0 \\ & & & & 2 & -1 \\ 0 & & \dots & 0 & -1 & 2 \end{pmatrix}.$$

The tridiagonal matrix H_{HNR} accounts for the serial correlation in the idiosyncratic remainder components introduced by first differencing Equation (3) to eliminate the unobserved individual-specific effects from the equation.

When using only the AB moment conditions in estimation, a choice for H often encountered in practice is the identity matrix with $T - 2$ diagonal elements

$$H_{AB} = \begin{pmatrix} 1 & 0 & \dots & & 0 \\ 0 & 1 & & & \\ \vdots & & \ddots & & \vdots \\ & & & & 0 \\ 0 & & \dots & 0 & 1 \end{pmatrix}.$$

For the AS moment conditions, an identity matrix with $T - 3$ diagonal elements is frequently used as H_{AS} (see, e.g., Blundell *et al.* 2001; Kripfganz 2019).

Finally, when two different sets of moment conditions are employed, a general representation of H is

$$H = \begin{pmatrix} A & B \\ B' & C \end{pmatrix},$$

where the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are chosen depending on the particular moment conditions employed in GMM estimation. Note that $\boldsymbol{A}$ and $\boldsymbol{C}$ correspond to the expected variance covariance properties within a set of moment conditions, while $\boldsymbol{B}$ corresponds to the expected covariance across the different sets of moment conditions and $\boldsymbol{B}'$ is the transpose.

An estimate for the weighting matrix $\widehat{\boldsymbol{W}}_2$ of the two-step GMM estimator (GMM2S) is

$$\widehat{\boldsymbol{W}}_2 = \left(\frac{1}{n} \cdot \boldsymbol{Z}' \hat{\boldsymbol{s}}_1 \hat{\boldsymbol{s}}_1' \boldsymbol{Z} \right)^{-1}, \tag{13}$$

where $\hat{\boldsymbol{s}}_1$ denotes the residuals from one-step estimation. When the nonlinear moment conditions are used, nonlinear optimization techniques are required to obtain coefficient estimates. Per default, GMM estimation by **pdynmc** is based on numerical optimization. To initialize the optimization procedure, starting values are drawn for all parameter estimates from a uniform distribution over the interval [-1, 1]. GMM2S then employs the parameter estimates obtained by GMM1S as starting values. For the optimization procedure, we rely on the R-package **optimx** (Nash and Varadhan 2011; Nash 2014). All optimization routines implemented in **optimx** are available in **pdynmc**. From our experience, the Variable Metric method (Fletcher 1970; Nash 1990) seems to work satisfactory in the estimation of linear dynamic panel data models. In all settings encountered while programming the package, the results from this method were close to closed form results for GMM estimation based on linear moment conditions. The Variable Metric method is named BFGS in **optimx**[8] and serves as the default procedure in **pdynmc**. Note that for GMM estimation based on linear moment conditions, the closed form results are computed and stored along with the optimization results.

Alternatively to one-step and two-step procedures, GMM estimation can be carried out with the continuously updating estimator (GMMCU). The GMMCU is an iterative procedure, where the weighting matrix, the corresponding parameter estimates, and the residuals are updated until either one of two stopping criteria is attained: The procedure stops, when the change in coefficient estimates from one estimation step to the next does not exceed a certain pre-specified threshold z_{tol}. Otherwise, GMMCU stops after a pre-specified number of maximum iterations h_{iter}. Asymptotically, one-step, two-step, and multiple-step GMM estimation are equivalent – though, differences occur in finite samples and Monte Carlo evidence exists that the finite sample performance may improve when using GMMCU (see, e.g., Hansen, Heaton, and Yaron 1996). In **pdynmc** all three different estimation procedures are available.

3.3. Closed form solution

When estimating the linear dynamic panel data model in Equation (3) by GMM based on linear moment conditions only, numerical optimization methods are not required to obtain coefficient estimates. One-step estimates $\hat{\boldsymbol{\theta}}_1$ are available from:

$$\hat{\boldsymbol{\theta}}_1 = (\boldsymbol{X}' \boldsymbol{Z} \widehat{\boldsymbol{W}}_1 \boldsymbol{Z}' \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{Z} \widehat{\boldsymbol{W}}_1 \boldsymbol{Z}' \boldsymbol{y}. \tag{14}$$

The matrix $\boldsymbol{X}$ contains all right-hand side variables[9] from Equation (3) and $\widehat{\boldsymbol{W}}_1$ is the estimated one-step weighting matrix from Equation (12). In order to calculate the two-

---

[8]For more details and references on the available optimization methods in **optimx** see the package documentation and Nash and Varadhan (2011).

[9]Note that – depending on the moment conditions employed in estimation – all matrices and vectors given in the following may contain observations in levels and/or first differences.

step coefficient estimates $\hat{\boldsymbol{\theta}}_2$, $\widehat{\boldsymbol{W}}_1$ needs to be replaced by the estimated two-step weighting matrix $\widehat{\boldsymbol{W}}_2$:

$$\hat{\boldsymbol{\theta}}_2 = (\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_2\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_2\boldsymbol{Z}'\boldsymbol{y}. \tag{15}$$

Two comments on the closed form expressions may be helpful here. First, recall from Equations (12) and (13) that the one-step- and two-step weighting matrices depend on the number of cross sectional units in the data set n. Considering the closed form for the coefficient estimates in greater detail reveals that for Equations (14) and (15), the factor n cancels from both expressions. Second, as mentioned in the previous section, updating the weighting matrix and computing coefficient estimates does not necessarily have to stop at the second step. The procedure can be iterated until either one of the stopping criteria is reached.

4. Standard errors and inference

4.1. Standard errors

Asymptotic one-step standard errors for the estimated coefficients can be obtained by taking the square root of the main diagonal elements of the estimated one-step variance covariance matrix

$$\widehat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}_1) = n \cdot (\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_1\boldsymbol{Z}'\boldsymbol{X})^{-1}\hat{\sigma}_1^2, \qquad \text{with} \qquad \hat{\sigma}_1^2 = \hat{\boldsymbol{s}}_1'\hat{\boldsymbol{s}}_1 \cdot \frac{1}{N-p}. \tag{16}$$

In the formula, N is the number of observations available for estimation (i.e., the cross section dimension times the time series dimension minus the number of missing observations), p denotes the number of estimated coefficients, and $\hat{\boldsymbol{s}}_1$ are residuals from GMM1S (see Doornik, Arellano, and Bond 2012). As stated in Windmeijer (2005), robust one-step standard errors are available from

$$\widehat{\boldsymbol{\Omega}}_r(\hat{\boldsymbol{\theta}}_1) = n \cdot (\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_1\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_1\widehat{\boldsymbol{W}}_2^{-1}\widehat{\boldsymbol{W}}_1\boldsymbol{Z}'\boldsymbol{X} \cdot \tag{17}$$
$$(\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_1\boldsymbol{Z}'\boldsymbol{X})^{-1},$$

while asymptotic two-step standard errors can be computed from

$$\widehat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}_2) = n \cdot (\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_2\boldsymbol{Z}'\boldsymbol{X})^{-1}. \tag{18}$$

Since asymptotic two-step GMM standard errors for the estimated coefficients exhibit a downward bias in small samples, they can, however, be substantially lower than one-step GMM standard errors (see, e.g., Arellano and Bond 1991). Windmeijer (2005) relates the bias to the dependence of the two-step weighting matrix on parameter estimates (the one-step estimates) and proposes an analytic correction of the two-step standard errors based on a first order Taylor-series expansion:

$$\widehat{\boldsymbol{\Omega}}_c(\hat{\boldsymbol{\theta}}_2) = \boldsymbol{F} + \boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\widehat{\boldsymbol{W}}_2}\boldsymbol{F} + \boldsymbol{F}\boldsymbol{D}'_{\hat{\boldsymbol{\theta}}_2,\widehat{\boldsymbol{W}}_2} \tag{19}$$
$$+ \boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\widehat{\boldsymbol{W}}_2}\widehat{\boldsymbol{\Omega}}_r(\hat{\boldsymbol{\theta}}_1)\boldsymbol{D}'_{\hat{\boldsymbol{\theta}}_2,\widehat{\boldsymbol{W}}_2},$$

where the expression $\boldsymbol{F}$ is defined as

$$\boldsymbol{F} = n \cdot (\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_2\boldsymbol{Z}'\boldsymbol{X})^{-1}.$$

This expression is equivalent to the estimated uncorrected two-step variance covariance matrix of the coefficient estimates $\widehat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}_2)$ in Equation (18). The computation of the correction $\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\widehat{\boldsymbol{W}}_2}$ is involved when multiple parameters are estimated. For a single parameter, it equals

$$\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\widehat{\boldsymbol{W}}_2} = -\frac{1}{n} \cdot \boldsymbol{F}\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_2 \left.\frac{\partial \widehat{\boldsymbol{W}}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_1} \widehat{\boldsymbol{W}}_2\boldsymbol{Z}'\hat{\boldsymbol{s}}_2.$$

The vector $\hat{\boldsymbol{s}}_2$ denotes the two-step residuals and the first derivative of the weighting matrix for two-step GMM estimation evaluated at $\hat{\boldsymbol{\theta}}_1$ can be calculated from

$$\left.\frac{\partial \widehat{\boldsymbol{W}}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_1} = -\frac{1}{n} \cdot \boldsymbol{Z}'(\boldsymbol{X}\hat{\boldsymbol{s}}_1' + \hat{\boldsymbol{s}}_1\boldsymbol{X}')\boldsymbol{Z}.$$

Two remarks on the computation of the standard errors might be helpful here: First, similar to the computation of the closed form expressions for the coefficient estimates, the formulas in the Equations (17), (18), and (19) do not depend on the number of cross sectional units in the data set n, as the term cancels out with the $1/n$ from $\widehat{\boldsymbol{W}}_1$ and $\widehat{\boldsymbol{W}}_2$. Equation (16) depends on the total number of observations available in the data set by the degrees of freedom correction in the calculation of $\hat{\sigma}_{1s}^2$. This affects the calculation of the standard errors when there are missing observations. Second, note that an alternative to the analytic correction of the two-step standard errors proposed by Arellano and Bond (1991) is to replace the standard errors of the second estimation step by those from the first step.

4.2. Specification testing

Arellano and Bond (1991) suggest a test for second order serial correlation in the idiosyncratic remainder components. The test is generalized to higher orders j by Arellano (2003) and can be used as a specification test in the estimation of linear dynamic panel data models. The reasoning is that, although, first order serial correlation is present in the idiosyncratic remainder components for GMM estimation based on first differenced equations[10], no higher order autocorrelation should prevail. The serial correlation test of Arellano and Bond (1991) boils down to checking if the deviation of the covariance of the residuals of period t with the residuals of period $t-j$ from zero is large enough to indicate that j-th order serial correlation might be present in the idiosyncratic remainder components. The null hypothesis of the test is that there is no serial correlation in the $\varepsilon_{i,t}$. The corresponding test statistics are defined as

$$T_{m_j} = \frac{\hat{r}_j}{\hat{\sigma}_{\hat{r}_j}}, \qquad \text{with} \qquad T_{m_j} \overset{a}{\sim} \mathcal{N}(0,1),$$

where $\hat{\sigma}_{\hat{r}_j}$ is the standard error of the j-th order autocovariance of the residuals $\hat{r}_j$. For the linear dynamic panel data model specified in Equation (3), this autocovariance of the residuals

---

[10]First order serial correlation in the $\varepsilon_{i,t}$ is introduced by first differencing. Even when the $\varepsilon_{i,t}$ in levels are i.i.d., the first differenced $\varepsilon_{i,t}$ are correlated (for a derivation see, e.g., Fritsch 2019).

is the sample equivalent to

$$r_j = \frac{1}{T - 3 - j} \cdot \sum_{t=4+j}^{T} r_{t,j}, \qquad \text{with} \qquad r_{t,j} = E(\Delta s_{i,t} \Delta s_{i,t-j}),$$

the average j-th order autocovariance of the idiosyncratic remainder components (see Arellano 2003). As detailed by Arellano and Bond (1991) and Doornik *et al.* (2012), the corresponding scaled autocovariance of the residuals can be calculated by

$$\hat{r}_{t,j} = \frac{1}{\sqrt{n}} \cdot \hat{s}'_t \; \hat{s}_{t-j},$$

where $\hat{s}_t$ and $\hat{s}_{t-j}$ are column vectors which contain the residuals from one-step, two-step, or multiple-step GMM estimation for all cross sectional units at the respective time period; the index at $\hat{s}_{t-j}$ indicates that the corresponding residuals are lagged j time periods. According to Arellano and Bond (1991), the estimated variance of the j-th order autocovariance of the residuals is available from

$$\hat{\sigma}^2_{\hat{r}_j} = \frac{1}{n} \cdot \hat{s}'_{t-j} \; \widehat{\boldsymbol{\Omega}}(\hat{s}) \hat{s}_{t-j} - 2 \cdot \hat{s}'_{t-j} \; \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}\,\boldsymbol{Z}'\widehat{\boldsymbol{\Omega}}(\hat{s})\hat{s}_{t-j} +$$
$$\hat{s}'_{t-j} \; \boldsymbol{X}\widehat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}})\boldsymbol{X}'\hat{s}_{t-j}.$$

Note that the vectors of residuals $\hat{s}$, $\hat{s}_{-j}$ and the matrices $\widehat{\boldsymbol{W}}$, $\widehat{\boldsymbol{\Omega}}(\hat{s})$, and $\widehat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}})$ depend on the actual estimation step and the latter two matrices also depend on the estimated type of variance covariance matrix (i.e., robust or asymptotic for one-step estimation; Windmeijer-corrected or asymptotic for two-step estimation). All corresponding indices are dropped here in order to provide one general formula instead of the four specific ones.

4.3. Overidentifying restrictions testing

When the system of equations from which the model parameters are estimated by GMM is overidentified (i.e., when the number of moment conditions exceeds the number of parameters to be estimated), it is possible to assess the validity of the overidentifying restrictions by the Sargan test (Sargan 1958). The presumed null hypothesis is that the overidentifying restrictions are valid. According to Arellano and Bond (1991) and Doornik *et al.* (2012), the test statistic of the Sargan test can be computed from

$$T_S = n \cdot \hat{s}'_1 \boldsymbol{Z}\widehat{\boldsymbol{W}}_1 \boldsymbol{Z}'\hat{s}_1 \cdot \hat{\sigma}_1^{-2}, \quad \text{with} \quad T_S \overset{a}{\sim} \chi^2(m - p).$$

Under suitable conditions, which ensure asymptotic normality of the GMM estimator[11] and the additional assumption of conditional homoscedasticity, the test statistic is asymptotically χ^2-distributed with $m - p$ degrees of freedom; m equals the number of instruments employed in estimation (see, e.g., Hayashi 2000).

An alternative test statistic – where a finite fourth moments assumption is imposed instead of conditional homoscedasticity – is the J-test (Hansen 1982). The J-test statistic results from

---

[11]For Theorems, Propositions, and extensive discussions for GMM estimators see Newey and McFadden (1994); Hayashi (2000). A discussion of GMM estimation of linear dynamic panel data models and the underlying assumptions is provided by Fritsch (2019).

replacing $\widehat{\boldsymbol{W}}_1$ in the above formula by $\widehat{\boldsymbol{W}}_2$, the one-step residuals by the two-step residuals, and dropping the multiplication with $\hat{\sigma}_1^{-2}$:

$$T_J = n \cdot \hat{\boldsymbol{s}}_2' \boldsymbol{Z} \widehat{\boldsymbol{W}}_2 \boldsymbol{Z}' \hat{\boldsymbol{s}}_2, \quad \text{with} \quad T_J \overset{a}{\sim} \chi^2(m-p).$$

The idea underlying the test statistics T_S and T_J is, that when the moment conditions are valid, the sample analogues of these conditions should be close to zero. A large value of the test statistic indicates that some of the moment conditions may be invalid, that some of the model assumptions may be incorrect, or both (see, e.g., Hayashi 2000).

Variations of the two tests allow to check the validity of subsets of moment conditions. These tests are also referred to as 'difference-in-Hansen'/'difference-in-Sargan' tests (see, e.g., Roodman 2009), 'incremental Hansen'/'incremental Sargan' tests (see, e.g., Arellano 2003), or C-statistics (see, e.g., Hayashi 2000). The test statistic is obtained by carrying out the unrestricted estimation and the estimation under the null hypothesis, computing the desired test statistic for both estimations, and then taking the difference of the two test statistics. This difference is asymptotically χ^2-distributed with $T_{J_{H_1}} - T_{J_{H_0}}$ degrees of freedom, where $T_{J_{H_1}}$ are the degrees of freedom of the unrestricted model and $T_{J_{H_0}}$ those of the restricted one (see Hayashi 2000).

4.4. Testing linear hypotheses

The Wald test is one possibility to test general linear hypotheses of the form

$$H_0 : \boldsymbol{R}\boldsymbol{\theta} = \boldsymbol{r},$$

where the matrix $\boldsymbol{R}$ is a $c \times p$ matrix, which selects the elements of the $p \times 1$ vector of population parameters $\boldsymbol{\theta}$ required to express the left-hand side of the c equations of the null hypothesis (i.e., the restrictions under the null) and the vector $\boldsymbol{r}$ is a $c \times 1$ vector that states the right-hand side of the equations. Tests of three different standard null hypotheses are currently available in **pdynmc**: (a) all population parameters corresponding to the right-hand side variables of the linear dynamic panel data model are zero jointly, (b) all population parameters corresponding to the lagged-dependent and non-lagged dependent explanatory variables are zero jointly, and (c) all population parameters corresponding to the time dummies are zero jointly.

In case of one-step GMM estimation, the Wald statistic can be obtained from

$$T_W = n \cdot (\boldsymbol{R}\hat{\boldsymbol{\theta}}_1 - \boldsymbol{r})' \left(\boldsymbol{R} \, \widehat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}_1) \, \boldsymbol{R}' \right)^{-1} (\boldsymbol{R}\hat{\boldsymbol{\theta}}_1 - \boldsymbol{r}), \quad \text{with} \quad T_W \overset{a}{\sim} \chi^2(c).$$

In order to calculate the Wald statistic for two-step GMM estimation, the vector of parameter estimates $\hat{\boldsymbol{\theta}}_1$ and the estimated variance covariance matrix of the parameter estimates $\widehat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}_1)$ need to be replaced by their equivalents from two-step estimation. Under suitable conditions[12], the Wald statistic T_W is asymptotically χ^2-distributed with c degrees of freedom (see Hayashi 2000). The estimated variance covariance matrix of coefficient estimates to be used in both calculations may be either the non-robust versions stated in Equations (16) and (18) or the robust/corrected versions of the matrix from Equations (17) and (19). The equivalent matrices need to be chosen in multiple-step GMM estimation to obtain the corresponding Wald statistic. As usual, a large value of the Wald statistic casts doubt on the null hypothesis.

---

[12]See Newey and McFadden (1994); Hayashi (2000) and the discussion in Fritsch (2019).

5. Related software and R-packages

GMM estimation of linear dynamic panel data models based on linear moment conditions is available in a number of software environments and packages such as Gauss, Ox, R, and Stata. We highlight the particularities of a few selected implementations here.

The Gauss and Ox implementations, which are both named **DPD** (Arellano and Bond 1988; Doornik *et al.* 2012), represent an important reference for later software. The packages include the computation of one-step and two-step closed form GMM estimators and standard specification testing such as overidentifying restrictions tests, serial correlation tests, and Wald tests. Some estimators for static panel data models like the within estimator and feasible generalized least squares estimation are also available.

In Stata (e.g., Stata Corporation 2011), the command **xtabond2** (Roodman 2009a) is a popular choice for GMM estimation of linear dynamic panel data models based on linear moment conditions. The command calculates the closed form solution for the estimators and is accompanied by extended model diagnostics to assess the validity of certain subsets of moment conditions and the overall specification. Employing nonlinear moment conditions and GMMCU are not supported. The recently contributed command **xtdpdgmm** (Kripfganz 2019) enables the user to include nonlinear moment conditions into the analysis. The command also allows for GMMCU and the numerical optimization of the GMM objective function is based on a Gauss-Newton technique.

In R (R Core Team 2019), the packages **plm** (Croissant and Millo 2008) and **panelvar** (Sigmund and Ferstl 2019) implement the functionality available in **xtabond2** with some additional features. For example, the package **panelvar** allows the user to perform lag selection based on information criteria, structural analysis based on impulse response functions, the computation of corresponding bootstrapped confidence intervals, and allows for GMMCU. The package **plm** provides a variety of functions for the estimation of static and linear panel models such as the within estimator, different random effects estimators, feasible generalized least squares estimation, and a number of different specification tests. The function **pgmm** is specifically designed to estimate linear dynamic panel data models – GMMCU is not implemented. Both R packages do not allow to incorporate nonlinear moment conditions into the analysis.

6. Sample session

The functionality of **pdynmc** is illustrated by replicating some of the empirical results in Arellano and Bond (1991). Additionally, we show how to incorporate the linear AB and the nonlinear AS moment conditions into the analysis. We explain all arguments which need to be set to reproduce the results and point out some alternative options. We also draw comparisons between **pdynmc**, the Stata implementations **xtabond2**, **xtdpdgmm**, and the **pgmm** function in the R-package **plm** – where we are aware of differences between the implementations.

The data set employed in Arellano and Bond (1991) is an unbalanced panel of $n = 140$ firms located in the UK observed over $T = 9$ time periods and is available from package R-**plm**:

```
data(EmplUK, package = "plm")
dat                   <- EmplUK
dat[,c(4:7)]          <- log(dat[,c(4:7)])
names(dat[,c(4:7)])   <- c("n", "w", "k", "ys")
```

The authors investigate employment equations and consider the dynamic specification

$$
\begin{aligned}
n_{i,t} =& \alpha_1 n_{i,t-1} + \alpha_2 n_{i,t-2}+ \\
& \beta_1 w_{i,t} + \beta_2 w_{i,t-1} + \beta_3 k_{i,t} + \beta_4 k_{i,t-1} + \beta_5 k_{i,t-2} + \beta_6 ys_{i,t} + \beta_7 ys_{i,t-1} + \beta_8 ys_{i,t-2}+ \\
& \gamma_3 d_3 + \cdots + \gamma_T d_T + \eta_i + \varepsilon_{i,t}, \qquad i = 1,...,n; \ t = 3,...,T.
\end{aligned}
\tag{20}
$$

In the equation, i denotes the firm and t is the time series dimension. The natural logarithm of employment n is explained by its first two lags and the further explanatory variables natural logarithm of wage w, natural logarithm of capital k, natural logarithm of output ys, and their lags of order up to one (for w) or two (for k and ys). The variables $d_3, \ldots, d_T$ are time dummies with corresponding coefficients $\gamma_3, \ldots, \gamma_T$; the unobserved individual-specific effect is represented by η, and ε is an idiosyncratic remainder component. The goal is to estimate the lag parameters α_1 and α_2 and the coefficients of the further explanatory variables β_j, with $j = 1, \ldots, 8$, while controlling for (unobserved) time effects and accounting for unobserved individual-specific heterogeneity.

6.1. GMM estimation with HNR moment conditions

When reproducing the results in Table 4 on p.290 of Arellano and Bond (1991) with **pdynmc**, the model structure of the underlying Equation (20) can be specified by:

```
m1 <- pdynmc(
  dat = dat, varname.i = "firm", varname.t = "year",
  use.mc.diff = TRUE, use.mc.lev = FALSE, use.mc.nonlin = FALSE,
  include.y = TRUE, varname.y = "emp", lagTerms.y = 2,
  fur.con = TRUE, fur.con.diff = TRUE, fur.con.lev = FALSE,
  varname.reg.fur = c("wage", "capital", "output"),
  lagTerms.reg.fur = c(1,2,2),
  include.dum = TRUE, dum.diff = TRUE, dum.lev = FALSE, varname.dum = "year",
  w.mat = "iid.err", std.err = "corrected",
  estimation = "onestep", opt.meth = "none")
```

The first arguments relate to the data set (`dat`), the individual (`varname.i`), and time series dimension (`varname.t`). Next, the moment conditions are defined – here, only moment conditions from equations in differences are used. The moment conditions are derived for the dependent variable (`varname.y`), and for the corresponding number of lags of the dependent variable to be included as explanatory variables (`lagTerms.y`). Further non-lagged dependent explanatory variables (`fur.con`) are included in the equations in differences, not in the equations in levels (compare `fur.con.diff = TRUE`, `fur.con.lev = FALSE`), their variable names are stated in (`varname.reg.fur`), their lag structure is specified in (`lagTerms.reg.fur`). Note that the first element of the vector denoting the lag structure corresponds to the first element of the vector with the variable names, the second element to the second, and so on. Also note that all names given in the vectors that refer to variables in the data set need to have the same names as in the data set. The time dummies are included by (`include.dum`) into the equations in differences not in the equations in levels (`dum.diff = TRUE`, `dum.lev = FALSE`), the dummy indicator variable is (`varname.dum`). Note that time dummies can be constructed from one or multiple variables by **pdynmc** by simply passing a scalar or vector

with the respective variable names in the data to `varname.dum`. Specifying the matrix $\boldsymbol{H}$ in Equation (12), which governs the structure of the one-step weighting matrix, and carrying out one-step estimation can be achieved by the commands in the last two codelines. Choosing the option `iid.err` uses the matrix $\boldsymbol{H}_{HNR}$ proposed by Arellano and Bond (1991). Alternatively, an identity matrix can be employed for $\boldsymbol{H}$ by the option `identity`. Code `std.err = "corrected"` yields robust standard errors for one-step estimation (in case of a two-step estimation, the correction of (Windmeijer 2005) was employed by this argument).

One-step, two-step, and GMMCU (accessible by setting the argument `estimation` to `cue`) estimation in **pdynmc** is carried out by numerical optimization of the GMM objective function given in Equation (11). Since a closed form solution exists for the estimator when employing only linear moment conditions, numerical optimization is not required and can be switched off by setting `opt.meth = "none"`.

The standard output can be accessed via `summary(m1)` and yields the estimation results when

Table 1: Column (a1) of Table 4 in Arellano and Bond (1991)

	Estimate	Std.Err.rob	z.rob	Pr(>|z.rob|)
L1.n	0.68623***	0.14459	4.74600	< 0.001
L2.n	-0.08536	0.05602	-1.52400	0.12751
w	-0.60782***	0.17821	-3.41100	< 0.001
L1.w	0.39262*	0.16799	2.33700	0.01944
k	0.35685***	0.05902	6.04600	< 0.001
L1.k	-0.05800	0.07318	-0.79300	0.42778
L2.k	-0.01995	0.03271	-0.61000	0.54186
ys	0.60851***	0.17253	3.52700	< 0.001
L1.ys	-0.71116**	0.23172	-3.06900	0.00215
L2.ys	0.10580	0.14120	0.74900	0.45386
1979	0.00955	0.01029	0.92900	0.35289
1980	0.02202	0.01771	1.24300	0.21387
1981	-0.01177	0.02951	-0.39900	0.68989
1982	-0.02706	0.02928	-0.92400	0.35549
1983	-0.02132	0.03046	-0.70000	0.48393
1984	-0.00770	0.03141	-0.24500	0.80646

Equations in first differences: $L(2/8).n, D.w, L.D.w, D.k,$
$L.D.k, L2.D.k, D.ys, L.D.ys, L2.D.ys, D.1979 - D.1984$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (refers to t-test of the null
  that the coefficient is equal to zero)

specifying all arguments as stated in this section and reproduces the results in Table 4, column (a1) on p.290 of Arellano and Bond (1991).

Changing the argument `estimation` to `twostep` yields the two-step GMM coefficient estimates (the `pdynmc`-output object is assigned to `m2`) from Table 4, column (a2) on p.290 of Arellano and Bond (1991). Note that the standard errors presented in Table 2 are based on the Windmeijer-correction and deviate from the conventional standard errors reported in Arellano and Bond (1991). The standard errors from the original analysis can be reproduced by setting `std.err = "unadjusted"`.

Regarding the arguments

```
use.mc.diff = TRUE, include.y = TRUE, include.x = FALSE
```

it has to be noted that the HNR moment conditions (`use.mc.diff`) derived from the lagged dependent variable (`include.y`) are employed, while none are derived from the further non-lagged dependent explanatory variables (by default `include.x = FALSE`). The latter argu-

Table 2: Column (a2) of Table 4 in Arellano and Bond (1991)

	Estimate	Std.Err.rob	z.rob	Pr(>|z.rob|)
L1.n	0.62871**	0.19341	3.25100	0.00115
L2.n	-0.06519	0.04505	-1.44700	0.14790
w	-0.52576***	0.15461	-3.40100	< 0.001
L1.w	0.31129	0.20300	1.53300	0.12528
k	0.27836***	0.07280	3.82400	< 0.001
L1.k	0.01410	0.09246	0.15200	0.87919
L2.k	-0.04025	0.04327	-0.93000	0.35237
ys	0.59192***	0.17309	3.42000	< 0.001
L1.ys	-0.56599*	0.26110	-2.16800	0.03016
L2.ys	0.10054	0.16110	0.62400	0.53263
1979	0.01122	0.01168	0.96000	0.33706
1980	0.02307	0.02006	1.15000	0.25014
1981	-0.02136	0.03324	-0.64200	0.52087
1982	-0.03112	0.03397	-0.91600	0.35967
1983	-0.01799	0.03693	-0.48700	0.62626
1984	-0.02337	0.03661	-0.63800	0.52347

Equations in first differences: $L(2/8).n, D.w, L.D.w, D.k,$
$L.D.k, L2.D.k, D.ys, L.D.ys, L2.D.ys, D.1979 - D.1984$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (refers to t-test of the null
that the coefficient is equal to zero)

ment implies that the non-lagged dependent explanatory variables in the model are assumed to be exogenous and instrument themselves.

Different capabilities for testing hypotheses about the population parameters are available in **pdynmc**. Among them are the tests for serial correlation in the idiosyncratic remainder components proposed by Arellano and Bond (1991), Hansen tests, and Wald tests. In the following, carrying out these tests and interpreting the results is briefly illustrated based on the two-step GMM estimation results presented in Table 2.

Employing the test for second order serial correlation of Arellano and Bond (1991) described in Section 4.2 by `m.test(m2, t.order = 2)` yields:

```
        Serial correlation test of degree 2

data:  GMM Estimation; H0: no serial correlation of order 2 in epsilon
normal = -0.36744, p-value = 0.7133
```

The test does not reject the null hypothesis at any plausible significance level and does not provide any indication that the model specification might be inadequate. The test statistic and p-value are similar to **xtabond2** and **pgmm**.

Computing the Hansen J-test of the overidentifying restrictions described in Section 4.3 by `j.test(m2)` yields:

```
        J-Test of Hansen

data:  GMM Estimation; H0: overidentifying restrictions valid
chisq = 31.381, df = 25, p-value = 0.1767
```

The test does not reject the overidentifying restrictions and does not provide any indications that the validity of the instruments employed in estimation may be in doubt. Comparing the results to **xtabond2** shows that the degrees of freedom and the p-value differ. We consider

25 degrees of freedom to be the appropriate number here, as 41 instruments are employed in estimation to obtain 16 coefficient estimates. It seems that the function **xtabond2** does not correct the degrees of freedom for the number of dummies dropped in estimation[13]. The difference in the *p*-value is due to the differences in the degrees of freedom. Our results are equivalent to the results of **pgmm** for the overidentifying restrictions test. In **pgmm**, the above test is referred to as 'Sargan test'.

For the Wald test illustrated in Section 4.4, consider the null hypothesis that the population parameters of all coefficients included in the model are zero jointly (`wald.fct(param = "all", object = m2)`):

```
        Wald test

data:  GMM Estimation; H0: beta = 0; tested model parameters: all
chisq = 1104.7, df = 16, p-value < 2.2e-16
```

The test rejects the null hypothesis. Comparing the test result to the implementation of the test in **xtabond2** – again – reveals differences concerning the degrees of freedom. We consider 16 to be the appropriate number of degrees of freedom here, since this corresponds to the number of estimated parameters. As noted previously, the differences seem to stem from **xtabond2** not adjusting the degrees of freedom for the dummies dropped in estimation. Alternative hypotheses that can be tested via the Wald test in **pdynmc** are that all slope parameters are zero jointly and that all parameters corresponding to the time dummies are zero jointly (`param = "time.dum"` only tests the time dummies, while `param = "slope"` only tests the slope parameters).

6.2. GMM estimation with HNR and AB moment conditions

When the 'constant correlated effects' assumption stated in Equation (9) holds, the HNR moment conditions from equations in differences employed in Section 6.1 can be extended by the AB moment conditions from equations in levels. The AB moment conditions are particularly useful for data generating processes, which are highly persistent (Blundell and Bond 1998). In this case, identification by the HNR moment conditions from equations in levels may fail and GMM estimation based on HNR moment conditions is documented to possess poor finite sample performance (see, e.g., Blundell and Bond 1998; Blundell *et al.* 2001; Bun and Sarafidis 2015).

In **pdynmc**, the AB moment conditions from equations in levels can be (additionally) incorporated by :

```
use.mc.lev = TRUE
```

In principle, both, the time dummies and the further explanatory variables can be included in the equations in first differences and the level equations. It is recommended, though, to include the dummies only in one of the equations, as it can be shown that incorporating them in both equations renders one set of dummies redundant for estimation – while for the

---

[13]Dummies are dropped by the estimation routine in case of high collinearity.

non-lagged dependent explanatory variables, this equivalence does not hold.[14] The arguments that govern accommodating non-lagged dependent explanatory variables and time dummies which instrument themselves in the levels equations are:

```
fur.con.lev = TRUE, dum.lev = TRUE
```

Using these arguments together with the earlier specified ones – except for setting 'dum.diff = FALSE' – leads to the time dummies being included in the level equations and the further explanatory variables being included in both equations.

In order to obtain coefficient estimates, a decision about the matrix $\boldsymbol{H}$ in the one-step weighting matrix is required. When using the HNR and AB moment conditions, the decision about $\boldsymbol{H}$ effectively involves specifying the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ in the general structure given in Section 3.2. As mentioned, the diagonal elements $\boldsymbol{A}$ and $\boldsymbol{C}$ reflect the expected variance covariance properties within a set of moment conditions, while $\boldsymbol{B}$ reflects the expected covariances across different sets of moment conditions. In the given setting, $\boldsymbol{A}$ corresponds to the variance covariance properties of the HNR moment conditions, $\boldsymbol{C}$ to those of the AB moment conditions, and $\boldsymbol{B}$ to those across the HNR and AB moment conditions. Three different options are currently available in **pdynmc** to set up the weighting matrix w.mat: iid.err, identity, and zero.cov. The first option leads to $\boldsymbol{H}_{HNR}$ being used for $\boldsymbol{A}$, an identity for $\boldsymbol{C}$, and a matrix $\boldsymbol{B}$, such that $\boldsymbol{B}\boldsymbol{B}' = \boldsymbol{H}_{HNR}$. Setting w.mat to identity leads to an identity matrix being used for the diagonal matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ and an adequately dimensioned matrix $\boldsymbol{B}$ with 1 on the diagonal[15]. When using the option zero.cov, the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ are as for option iid.err, but $\boldsymbol{B}$ is set to a null matrix. In case nonlinear moment conditions are used, the part of $\boldsymbol{H}$ which corresponds to the nonlinear moment conditions is set to an identity for all choices of w.mat. All elements of the matrices containing the expected covariance properties of the nonlinear moment conditions with other moment conditions are always set to zero.

The results presented in Table 3 are the two-step estimates of column (a2) of Table 4 in Arellano and Bond (1991) extended by the AB moment conditions. All arguments are specified as described above. Including the AB moment conditions into the analysis leads to substantial changes in the coefficient estimates of the first lag of the dependent variable. Note that the results indicate a markedly higher persistence of employment and render including two lags of the dependent variable questionable (Blundell and Bond 1998, e.g., estimate a version of the equation which contains only one lag of all explanatory variables). Note that the coefficient estimates of the explanatory variables, besides the first lag of the dependent variable, appear to be similar across estimations.

Equivalent results to Table 3 can be obtained from the **pgmm** function in the **plm**-package – besides some minor numerical differences at the fifth digit. When replicating the results with **xtabond2**, differences in the implementations become obvious: The instrument set for the AB moment conditions is extended in similar fashion to the HNR moment conditions in **xtabond2**, while this is not the case in **pgmm**. An argument is available in **pdynmc** to extend the instrument set as in **xtabond2**:

---

[14]Note that this is the case in balanced panels. The results may also not be numerically identical across function calls for different choices of the one-step weighting matrix. For a discussion, see https://www.statalist.org/forums/forum/general-stata-discussion/general/1357268-system-gmm-time-dummies.

[15]Note that the matrix $\boldsymbol{B}$ is not necessarily a quadratic matrix.

Table 3: Arellano and Bond (1991) estimation with AB moment conditions

	Estimate	Std.Err.corr	z.corr	Pr($>$|z.corr|)
L1.n	1.11650***	0.05192	21.50500	< 0.001
L2.n	-0.11352*	0.04764	-2.38300	0.01717
w	-0.44169**	0.15175	-2.91100	0.00360
L1.w	0.42159**	0.15528	2.71500	0.00663
k	0.28618***	0.04751	6.02400	< 0.001
L1.k	-0.16474*	0.06589	-2.50000	0.01242
L2.k	-0.12321**	0.04250	-2.89900	0.00374
ys	0.55793**	0.17651	3.16100	0.00157
L1.ys	-0.67392**	0.21707	-3.10500	0.00190
L2.ys	0.13372	0.14344	0.93200	0.35134
1978	-0.05313	0.35746	-0.14900	0.88155
1979	-0.03697	0.35698	-0.10400	0.91717
1980	-0.01933	0.35429	-0.05500	0.95614
1981	-0.05791	0.34696	-0.16700	0.86737
1982	-0.04334	0.34512	-0.12600	0.89973
1983	-0.01818	0.34583	-0.05300	0.95773
1984	-0.02815	0.34914	-0.08100	0.93544

Equations in first differences: $L\,(2/8)\,.n, D.w, L.D.w, L2.D.w,$
$D.k, L.D.k, L2.D.k, D.ys, L.D.ys, L2.D.ys$
Equations in levels: $L\,(1/7)\,.D.n, w, L.w, L2.w, k, L.k, L2.k,$
$ys, L.ys, L2.ys, 1978 - 1984$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (refers to t-test of the null
that the coefficient is equal to zero)

```
inst.stata = TRUE
```

Due to the reasons described in Section 2.3, this argument is set to `FALSE` per default. When setting the option to `TRUE`, the results from **xtabond2** and **pdynmc** are very close to our results.

6.3. GMM estimation with HNR and AS moment conditions

Recall, that the linear AB moment conditions from equations in levels comprise the nonlinear AS moment conditions and render them redundant for estimation (Blundell and Bond 1998; a derivation is provided in Fritsch 2019). Both sets of moment conditions may be useful in GMM estimation when the lag parameter is close to unity and it can be shown that extending the HNR moment conditions by either the AB- or the AS moment conditions may identify the lag parameter – even when the individual moment conditions fail to do so (Bun and Kleibergen 2014; Gorgens *et al.* 2016). The AB moment conditions require the 'constant correlated effects' assumption, while the AS moment conditions only require standard assumptions to hold. Therefore, the latter may be useful in situations where the 'constant correlated effects' assumption is in doubt and the statistician aims to investigate a highly persistent dynamic process with a structure similar to Equation (3). In **pdynmc**, including nonlinear moment conditions into the analysis is available via:

```
use.mc.nonlin = TRUE
```

When extending the analysis of Arellano and Bond (1991) by the nonlinear AS moment conditions, the results differ substantially from Table 2 and are very similar to the coefficient estimates shown in Table 3. This casts doubt on the HNR moment conditions and may be a hint that there is high persistence in the employment process – as high persistence

Table 4: Estimation in Arellano and Bond (1991) extended by AS moment conditions

	Estimate	Std.Err.corr	z.corr	Pr(>|z.corr|)
L1.n	1.10900***	0.06624	16.74100	< 0.001
L2.n	-0.06960	0.06290	-1.10700	0.26829
L0.w	-0.43855**	0.14709	-2.98100	0.00287
L1.w	0.43632**	0.14695	2.96900	0.00299
L0.k	0.30913***	0.05093	6.07000	< 0.001
L1.k	-0.19133**	0.06997	-2.73400	0.00626
L2.k	-0.14991**	0.04806	-3.11900	0.00181
L0.ys	0.63172***	0.17203	3.67200	< 0.001
L1.ys	-0.67326**	0.21243	-3.16900	0.00153
L2.ys	0.05287	0.14523	0.36400	0.71586
1978	-0.13575	0.38199	-0.35500	0.72259
1979	-0.11641	0.38167	-0.30500	0.76037
1980	-0.09147	0.37873	-0.24200	0.80878
1981	-0.12311	0.37084	-0.33200	0.73989
1982	-0.11354	0.36750	-0.30900	0.75732
1983	-0.09785	0.36719	-0.26600	0.79024
1984	-0.10241	0.37359	-0.27400	0.78408

Equations in first differences: $L(2/8).n, u, D.w, L.D.w, L2.D.w,$
$D.k, L.D.k, L2.D.k, D.ys, L.D.ys, L2.D.ys$
Equations in levels: $w, L.w, L2.w, k, L.k, L2.k, ys, L.ys, L2.ys,$
$1978 - 1984$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (refers to t-test of the null
that the coefficient is equal to zero)

leads to the lag parameters not being identified by the HNR moment conditions (Bun and Kleibergen 2014; Gorgens *et al.* 2016). Note that since the unobservable error term u (which can be expressed in terms of observable model components and parameters) is included in the instrument set for the equations in first differences, nonlinear moment conditions are employed in estimation. Also note that the coefficient estimates in Table 4 are very close to coefficient estimates obtained from **xtdpdgmm**.

6.4. Iterated GMM

Iterated GMM can be used, by specifying the following commands:

```
estimation = "iterative", max.iter = NULL, iter.tol = NULL,
```

When `estimation = "iterative"` is used, `max.iter` specifies the maximum number of iterations, `iter.tol` the search tolerance w.r.t. convergence. Unless changed by the user, `max.iter = 100` and `iter.tol = 0.01` are employed.

6.5. Starting values

If numeric optimization techniques are used, the starting for all parameters are drawn from the uniform distribution on an interval [-1, 1] by the following commands (set as default):

```
start.val.lo = -1, start.val.up = 1, seed.input = 42
```

As usual, the `seed.input` ensures reproducibility. The starting values can be varied by the user via arguments `start.val.lo` and `start.val.up` and setting `custom.start.val` to TRUE.

7. Concluding remarks

The R-package **pdynmc** provides a function to estimate linear dynamic panel data models. The implementation allows for general lag structures of the explanatory variables, which may encompass lags of the dependent variable and further non-lagged dependent explanatory variables. For estimation, linear and nonlinear moment conditions are derived from the model assumptions; further controls and external instruments (if available) may also be added. Estimation is carried out by numerical optimization of the GMM objective function. Corresponding closed form solutions are computed – where possible – and stored besides the results from numerical optimization. The estimation routine can handle balanced and unbalanced panel data sets and provides one-step-, two-step-, and continuously updating estimation. Accounting for (unobserved) time-specific effects is possible by including time dummies; alternatively, both sides of the equation can be transformed such that the time-specific heterogeneity is partialled out. The partialling out option is experimental at the moment. We plan to investigate the effects and implications of this way of dealing with unobserved time-specific heterogeneity in greater detail in the future. Different choices for the weighting matrix, which guides the aggregation of moment conditions in one-step GMM estimation are available. Concerning the computation of standard errors for the coefficient estimates, the following options are currently available in **pdynmc**: non-robust one- and two-step standard errors and robust one-step- and Windmeijer-corrected two-step standard errors. Some standard hypothesis and specification tests are also available. Among them are Wald tests, overidentifying restrictions tests and a test for serial correlation in the idiosyncratic remainder components.

We plan to extend the package by additional features in the future. Next steps involve:

- Incorporate further diagnostics and tests to assess the validity of the estimated specifications and the underlying moment conditions and assumptions (e.g., testing the 'constant correlated effects' assumption and testing for structural breaks).

- Add computation of confidence and prediction intervals.

- Expand the possible choices for the one-step weighting matrix by, e.g., the proposition in Kiviet (2007b) for GMM estimation based on linear HNR- and AB moment conditions.

- Implement the IV estimator solely based on the nonlinear moment conditions proposed by Pua, Fritsch, and Schnurbus (2019a,b).

References

Ahn SC, Schmidt P (1995). "Efficient estimation of models for dynamic panel data." *Journal of Econometrics*, **68**(1), 5–27. doi:https://doi.org/10.1016/0304-4076(94)01641-C.

Anderson T, Hsiao C (1982). "Formulation and estimation of dynamic models using panel data." *Journal of Econometrics*, **18**(1), 47–82. doi:https://doi.org/10.1016/0304-4076(82)90095-1.

Arellano M (2003). *Panel Data Econometrics*. Oxford University Press. doi:https://doi.org/10.1093/0199245282.001.0001.

Arellano M, Bond S (1988). "Dynamic Panel Data Estimation Using DPD - A Guide for Users." *IFS Working Paper Series 88/15*, Institute for Fiscal Studies. URL https://www.cemfi.es/~arellano/#dpd.

Arellano M, 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. doi:https://doi.org/10.2307/2297968.

Arellano M, Bover O (1995). "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, **68**(1), 29–51. doi:https://doi.org/10.1016/0304-4076(94)01642-D.

Blundell R, Bond S (1998). "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, **87**(1), 115–143. doi:http://dx.doi.org/10.1016/S0304-4076(98)00009-8.

Blundell R, Bond S, Windmeijer F (2001). "Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator." In BH Baltagi, TB Fomby, RC Hill (eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, volume 15 of *Advances in Econometrics*, pp. 53–91. Emerald Group Publishing. doi:https://doi.org/10.1016/S0731-9053(00)15003-0.

Bun MJG, Kleibergen F (2014). "Identification and inference in moments based analysis of linear dynamic panel data models." *UvA-Econometrics Discussion Paper 2013-07*, Universiteit van Amsterdam, Dept. of Econometrics. URL http://hdl.handle.net/11245/1.399892.

Bun MJG, Sarafidis V (2015). "Chapter 3 – Dynamic Panel Data Models." In BH Baltagi (ed.), *The Oxford Handbook of Panel Data*, pp. 76–110. Oxford University Press. doi:https://doi.org/10.1093/oxfordhb/9780199940042.013.0003.

Croissant Y, Millo G (2008). "Panel Data Econometrics in R: The plm Package." *Journal of Statistical Software*, **27**(2), 1–43. doi:https://doi.org/10.18637/jss.v027.i02.

Doornik JA, Arellano M, Bond S (2012). *Panel Data estimation using DPD for Ox*. URL http://www.doornik.com/download.html.

Fletcher R (1970). "A new approach to variable metric algorithms." *The Computer Journal*, **13**(3), 317–322. doi:https://doi.org/10.1093/comjnl/13.3.317.

Fritsch M (2019). "On GMM-estimation of linear dynamic panel data models." *University of Passau Working Papers in Business Administration B-36-19*, University of Passau. URL https://www.wiwi.uni-passau.de/en/research/working-papers/.

Gorgens T, Han C, Xue S (2016). "Moment restrictions and identification in linear dynamic panel data models." *ANU Working Papers in Economics and Econometrics 2016-633*, Australian National University, College of Business and Economics, School of Economics.

Hansen LP (1982). "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, **50**(4), 1029–1054. doi:https://doi.org/10.2307/1912775.

Hansen LP, Heaton J, Yaron A (1996). "Finite-Sample Properties of Some Alternative GMM Estimators." *Journal of Business & Economic Statistics*, **14**(3), 262–280. doi:https://doi.org/10.1080/07350015.1996.10524656.

Hayashi F (2000). *Econometrics.* Princeton University Press.

Holtz-Eakin D, Newey Whitney K, Rosen HS (1988). "Estimating Vector Autoregressions with Panel Data." *Econometrica*, **56**(6), 1371–1395. doi:https://doi.org/10.2307/1913103.

Hsiao C (2014). *Analysis of Panel Data.* Cambridge University Press.

Kiviet JF (2007a). "Chapter 11 – Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models." In GDA Phillips, E Tzavalis (eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, pp. 282–318. Cambridge University Press. doi:https://doi.org/10.1017/CBO9780511493157.

Kiviet JF (2007b). "On the optimal weighting matrix for the GMM system estimator in dynamic panel data models." *UvA-Econometrics Discussion Paper 2007/08*, Universiteit van Amsterdam, Dept. of Econometrics. URL https://ase.uva.nl/content/research-programmes/uva-econometrics/discussion-papers/discussion-papers-uva-econometrics.html#anker-discussion-papers-2007.

Kripfganz S (2019). "XTDPDGMM: Stata module to perform generalized method of moments estimation of linear dynamic panel data models." URL http://EconPapers.repec.org/RePEc:boc:bocode:s458395.

Nash JC (1990). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation.* CRC Press.

Nash JC (2014). "On Best Practice Optimization Methods in R." *Journal of Statistical Software*, **60**(2), 1–14. doi:https://doi.org/10.18637/jss.v060.i02.

Nash JC, Varadhan R (2011). "Unifying Optimization Algorithms to Aid Software System Users: optimx for R." *Journal of Statistical Software*, **43**(9), 1–14. doi:https://doi.org/10.18637/jss.v043.i09.

Newey WK, McFadden DL (1994). "Chapter 36 – Large sample estimation and hypothesis testing." In RF Engle, DL MacFadden (eds.), *Handbook of Econometrics*, volume 4, pp. 2111–2245. Elsevier. doi:https://doi.org/10.1016/S1573-4412(05)80005-4.

Pua A, Fritsch M, Schnurbus J (2019a). "Large sample properties of an IV estimator based on the Ahn and Schmidt moment conditions." *University of Passau Working Papers in Business Administration B-37-19*, University of Passau. URL https://www.wiwi.uni-passau.de/en/research/working-papers/.

Pua A, Fritsch M, Schnurbus J (2019b). "Practical aspects of using quadratic moment conditions in linear AR(1) panel data models." *University of Passau Working Papers in Business Administration B-38-19*, University of Passau. URL https://www.wiwi.uni-passau.de/en/research/working-papers/.

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Roodman D (2009). "A Note on the Theme of Too Many Instruments." *Oxford Bulletin of Economics and Statistics*, **71**(1), 135–158. doi:https://doi.org/10.1111/j.1468-0084.2008.00542.x.

Roodman D (2009a). "How to do xtabond2: An Introduction to Difference and System GMM in Stata." *Stata Journal*, **9**(1), 86–136. URL https://www.stata-journal.com/article.html?article=st0159.

Sargan JD (1958). "The Estimation of Economic Relationships Using Instrumental Variables." *Econometrica*, **26**(3), 393–415. doi:https://doi.org/10.2307/1907619.

Sigmund M, Ferstl R (2019). "Panel vector autoregression in R with the package panelvar." *The Quarterly Review of Economics and Finance*, p. forthcoming. doi:https://doi.org/10.1016/j.qref.2019.01.001.

Stata Corporation (2011). *Stata Statistical Software: Release 12.* StataCorp LP, College Station, TX.

Windmeijer F (2005). "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of Econometrics*, **126**(1), 25–51. doi:https://doi.org/10.1016/j.jeconom.2004.02.005.

Affiliation:

Markus Fritsch
Chair of Statistics and Data Analytics
School of Business, Economics, and Information Systems
University of Passau
94032 Passau, Germany
E-mail: markus.fritsch@uni-passau.de
URL: https://www.wiwi.uni-passau.de/en/statistics/team/