UNIVERSITÄT PASSAU

Wirtschaftswissenschaftliche Fakultät

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

Markus Fritsch, Andrew Adrian Yu Pua, Joachim Schnurbus

Diskussionsbeitrag Nr. B-39-19

Betriebswirtschaftliche Reihe ISSN 1435-3539

PASSAUER DISKUSSIONSPAPIERE

Herausgeber:
Die Gruppe der betriebswirtschaftlichen Professoren
der Wirtschaftswissenschaftlichen Fakultät
der Universität Passau
94030 Passau

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

Markus Fritsch, Andrew Adrian Yu Pua, Joachim Schnurbus

Diskussionsbeitrag Nr. B-39-19

Betriebswirtschaftliche Reihe ISSN 1435-3539

Adresse des Autors/der Autoren:

Markus Fritsch
Wirtschaftswissenschaftliche Fakultät
Universität Passau
94030 Passau

Telefon: +49 851 509 2565
Telefax: +49 851 509 2562
E-Mail: markus.fritsch@uni-passau.de

UNIVERSITÄT PASSAU

Wirtschaftswissenschaftliche Fakultät

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

Markus Fritsch, Andrew Adrian Yu Pua, Joachim Schnurbus

Diskussionsbeitrag Nr. B-39-19

Betriebswirtschaftliche Reihe ISSN 1435-3539

PASSAUER DISKUSSIONSPAPIERE

Herausgeber:
Die Gruppe der betriebswirtschaftlichen Professoren
der Wirtschaftswissenschaftlichen Fakultät
der Universität Passau
94030 Passau

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

Markus Fritsch, Andrew Adrian Yu Pua, Joachim Schnurbus

Diskussionsbeitrag Nr. B-39-19

Betriebswirtschaftliche Reihe ISSN 1435-3539

Adresse des Autors/der Autoren:

Markus Fritsch
Wirtschaftswissenschaftliche Fakultät
Universität Passau
94030 Passau

Telefon: +49 851 509 2565
Telefax: +49 851 509 2562
E-Mail: markus.fritsch@uni-passau.de

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

Markus Fritsch[1], Andrew Adrian Yu Pua[2], Joachim Schnurbus[3]

September 20, 2019

Abstract. pdynmc is an R-package for GMM estimation of linear dynamic panel data models that are based on linear and nonlinear moment conditions as proposed by Anderson and Hsiao (1982), Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bover (1995), and Ahn and Schmidt (1995). This paper describes the functionality of the package and the options regarding instrument type, estimation methodology, general configuration, specification testing and inference from the perspective of an applied statistician. The description of the functionality is based on replicating the results on a publicly available panel data set. Additionally, we link our implementation to other software and packages for GMM estimation of linear dynamic panel data models.

Keywords. panel data, linear dynamic model, generalized method of moments, linear moment conditions, nonlinear moment conditions, R.

JEL codes. C23, C87.

[1]University of Passau, Department of Statistics, email: markus.fritsch@uni-passau.de
[2]Wang Yanan Institute for Studies in Economics, Xiamen University, email: andrewypua@gmail.com
[3]Chair of Statistics, University of Passau, email: joachim.schnurbus@uni-passau.de

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, 2018) 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, Newey, and Rosen (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, Newey, and Rosen (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 $\varepsilon_{i,t}$.[4] The initial time period is denoted by $t = 1$. Combining the Equations (1) and (2) yields the single equation form of the model

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

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

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

[4]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 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).

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

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

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, Newey, and Rosen (1988) (hereafter HNR) propose the linear (in parameters) moment conditions

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

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):

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

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).[5]

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[6] $\mathbf{M}_i$ to be available for estimation:

[5]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.

[6]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).

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

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 $t = 3, \dots, T$ 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[7].

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:

[7]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.

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

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

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

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, Bond, and Windmeijer (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'[8] assumption, the additional $T-2$ Arellano and Bover (1995) (hereafter AB) linear moment conditions can be derived:

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

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:

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

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:

[8] 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.

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

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

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[9].

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[10]. 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 $\hat{\boldsymbol{\theta}}$.

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

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

[10]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).

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

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 $\widehat{\boldsymbol{W}}$ 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, Bond, and Windmeijer, 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 $\boldsymbol{W}$ 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 $\widehat{\boldsymbol{W}}_1$ of the one-step GMM estimator (GMM1S) is

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

The structure of the matrix $\boldsymbol{H}$ 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

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

The tridiagonal matrix $\boldsymbol{H}_{HNR}$ 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 $\boldsymbol{H}$ often encountered in practice is the identity matrix with $T-2$ diagonal elements

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

For the AS moment conditions, an identity matrix with $T-3$ diagonal elements is frequently used as $\boldsymbol{H}_{AS}$ (see, e.g., Blundell, Bond, and Windmeijer, 2001; Kripfganz, 2018).

Finally, when two different sets of moment conditions are employed, a general representation of $\boldsymbol{H}$ is

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

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

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

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]. Multistarting is used to avoid local minima. By default, GMM1S is calculated for 3 different parameter starting value combinations. The number of multistarts can be adjusted by the user. GMM2S then employs the parameter estimates obtained by GMM1S as starting values and performs no multistarting. For the optimization procedure, we rely on the `R`-package `optimr` (Nash and Varadhan, 2016). All optimization routines implemented in `optimr` 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 `optimr`[11] and serves as the default procedure in `pdynmc`. This may, however, change in future versions of `pdynmc` depending on prospective results and insights – this is especially the case for the nonlinear moment conditions, which cannot be compared to closed form results. 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 (see, e.g., Hansen, Heaton, and Yaron, 1996). In `pdynmc` all three different estimation procedures are available.

[11]For more details and references on the available optimization methods in `optimr` see the package documentation and Nash and Varadhan (2011).

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[12] from Equation (3) and $\widehat{\boldsymbol{W}}_1$ is the estimated one-step weighting matrix from Equation (12). In order to calculate the two-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 observations contained in the sample 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 one-step GMM estimation (see Doornik, Arellano, and Bond, 2012). As stated in Windmeijer (2005), robust one-step standard errors are available from

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

while asymptotic two-step standard errors can be computed from

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

[12]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.

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:

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

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 $\hat{\sigma}^2_{1s}$. 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[13], no higher order autocorrelation should prevail. The serial correlation test of Arellano

[13]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).

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 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, Arellano, and Bond (2012), the corresponding scaled autocovariance of the residuals can be calculated by

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

where $\hat{\boldsymbol{s}}_t$ and $\hat{\boldsymbol{s}}_{t-j}$ 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 $\hat{\boldsymbol{s}}_{t-j}$ 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

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

Note that the vectors of residuals $\hat{\boldsymbol{s}}$, $\hat{\boldsymbol{s}}_{-j}$ and the matrices $\widehat{\boldsymbol{W}}$, $\widehat{\boldsymbol{\Omega}}(\hat{\boldsymbol{s}})$, 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, Arellano, and Bond (2012), the test statistic of the Sargan test can be computed from

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

Under suitable conditions, which ensure asymptotic normality of the GMM estimator[14] 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 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

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

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.

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

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[15], 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.

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, Arellano, and Bond, 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, 2018) enables the user to include nonlinear moment conditions into the analysis. The command does not allow 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

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

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`[16] 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 United Kingdom which are observed over a maximum of $T = 9$ time periods. 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:

```
+ varname.i = "firm", varname.t = "year",
```

[16]All results contained in this section are reproducible with the functions provided in the accompanying `R`-code. The estimation function is approximately 2,000 lines of code and provides a flexible implementation of the estimation of linear dynamic panel data models with linear and nonlinear moment conditions. Most of the options described in the text are already fully implemented. Additionally, functions to carry out the tests in Section 4.1 are also provided.

```
+ varname.y = "n", lagTerms.y = 2,
+ fur.con = TRUE, varname.reg.fur = c("w", "k", "ys"),
+ lagTerms.reg.fur = c(1,2,2),
+ include.dum = TRUE, varname.dum = "year"
```

The arguments in lines one and two set the individual (‘`varname.i`’) and time series dimension (‘`varname.t`’), specify the dependent variable (‘`varname.y`’), and the number of lags of the dependent variable to be included as explanatory variables (‘`lagTerms.y`’). The third line denotes that the model contains further non-lagged dependent explanatory variables (‘`fur.con`’) and gives the names of these variables (‘`varname.reg.fur`’) – while the fourth line specifies their respective lag structure (‘`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. Line five includes time dummies into the model (‘`include.dum`’) and indicates the variable in the data set from which the dummies shall be derived (‘`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`’.

Including the following arguments in the function call

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

ensures 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 (‘`include.x = FALSE`’). The latter argument implies that the non-lagged dependent explanatory variables in the model are assumed to be exogenous and instrument themselves.

Non-lagged dependent explanatory variables and time dummies can be incorporated by

```
+ fur.con.diff = TRUE, dum.diff = TRUE
```

into the equations in first differences. The former argument includes the non-lagged dependent explanatory variables and the latter the time dummies.

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 setting:

```
+ w.mat = "iid.err", estimation = "onestep"
```

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`’.

The table contains the estimation results when 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).

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)

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

	Estimate	Std.Err.corr	z.corr	Pr(>|z.corr|)
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)

Changing the argument '`estimation`' to '`twostep`' yields the two-step GMM coefficient estimates 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"
```

Alternatively, this option can be set to '`onestep`', which reports one-step standard errors for the two-step coefficient estimates. Computing robust (for GMM1s) or Windmeijer-corrected (for GMM2s) standard errors requires setting '`std.err`' to '`corrected`' (the default in `pdynmc`).

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"
```

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), Sargan tests, 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 yields:

```
        Serial correlation test of degree 2

data:  GMM Estimation; H0: no serial correlation of order 2 in epsilon
normal = -0.35166, p-value = 0.7251
```

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 identical to `xtabond2` and `pgmm`.

Computing the Hansen J-test of the overidentifying restrictions described in Section 4.3 gives:

```
        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[17]. 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 Sargan test we get:

```
        Sargan Test

data:  GMM Estimation; H0: overidentifying restrictions valid
chisq = 54.756, df = 25, p-value = 0.0005297
```

Contrary to the J-Test of Hansen, the Sargan test rejects the null hypothesis and raises doubts about the instrument set. The results of both tests should, however, be interpreted with caution, as the Sargan test statistic is inconsistent when heteroscedasticity is present and the power of the J-Test is weakened by the presence of many instruments (see Roodman, 2009a). Comparing the result to `xtabond2` reveals differences besides the degrees of freedom and the corresponding p-value: The test statistics are not identical. The differences stem from the calculation requiring a correction of the degrees of freedom. While we correct the number of observations available in estimation for missing values in the case of an unbalanced panel data set, `xtabond2` does not. Additionally, `xtabond2` seems to use the number of instruments employed in estimation to correct the degrees of freedom without adjusting for the time dummies dropped in estimation, while we make this adjustment.

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 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

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

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.

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, Bond, and Windmeijer, 2001; Bun and Sarafidis, 2015).

In `pdynmc`, the AB moment conditions from equations in levels can be incorporated by:

```
+ use.mc.diff = TRUE, 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 non-lagged dependent explanatory variables, this equivalence does not hold.[18] 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`',

[18]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.

'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[19]. 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.

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
L0.w	-0.44169**	0.15175	-2.91100	0.00360
L1.w	0.42159**	0.15528	2.71500	0.00663
L0.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
L0.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)

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,

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

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:

```
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 – but not identical. The reason for the differences seems to be the instrument set, as xtabond2 reports a lower instrument count. When contrasting the instrument matrices used by xtabond2 and pdynmc, though, it appears that both functions employ the exact same instruments. It is currently unclear to us from where the difference in the results emerges.

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, Han, and Xue, 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 not available 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
```

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 the following features in the future:

- Provide a formula syntax as an alternative to specify the linear dynamic panel data model and the instruments to be used in estimation.
- 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.
- Facilitate choosing an adequate dynamic specification by lag selection techniques.
- Include moment selection capabilities based on an appropriate criterion into GMM estimation which allow to remove weak instruments/moment conditions.
- 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.
- Allow further types of moment conditions; an example are moment conditions derived from assumptions about second (alternatively, or additionally: third, fourth, ...) moments of the $y_{i,t}$ process (e.g., homoscedasticity as mentioned in Ahn and Schmidt, 1995)
- Enable the user to choose time period T instead of t as reference period for all moment conditions.
- Implement the IV estimator solely based on the nonlinear moment conditions proposed by Pua, Fritsch, and Schnurbus (2019a) and Pua, Fritsch, and Schnurbus (2019b).

References

Ahn, SC and P Schmidt (1995). "Efficient estimation of models for dynamic panel data". In: *Journal of Econometrics* 68.1, pp. 5–27.

Anderson, T and C Hsiao (1982). "Formulation and estimation of dynamic models using panel data". In: *Journal of Econometrics* 18.1, pp. 47–82.

Arellano, M (2003). *Panel Data Econometrics.* Oxford University Press.

Arellano, M and S Bond (1988). *Dynamic Panel Data Estimation Using DPD - A Guide for Users.* IFS Working Paper Series 88/15.

Arellano, M and S Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". In: *The Review of Economic Studies* 58.2, pp. 277–297.

Arellano, M and O Bover (1995). "Another look at the instrumental variable estimation of error-components models". In: *Journal of Econometrics* 68.1, pp. 29–51.

Blundell, R, S Bond, and F Windmeijer (2001). "Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator". In: *Nonstationary Panels, Panel Cointegration, and Dynamic Panels.* Ed. by BH Baltagi, TB Fomby, and RC Hill. Vol. 15. Advances in Econometrics. Emerald Group Publishing, pp. 53–91.

Blundell, R and S Bond (1998). "Initial conditions and moment restrictions in dynamic panel data models". In: *Journal of Econometrics* 87.1, pp. 115–143.

Bun, MJG and F Kleibergen (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.

Bun, MJG and V Sarafidis (2015). "Chapter 3 – Dynamic Panel Data Models". In: *The Oxford Handbook of Panel Data.* Ed. by BH Baltagi. Oxford University Press, pp. 76–110.

Croissant, Y and G Millo (2008). "Panel Data Econometrics in R: The plm Package". In: *Journal of Statistical Software* 27.2, pp. 1–43.

Doornik, JA, M Arellano, and S Bond (2012). *Panel Data estimation using DPD for Ox.* Version 1.25. URL: `http://www.doornik.com/download.html`.

Fletcher, R (1970). "A new approach to variable metric algorithms". In: *The Computer Journal* 13.3, pp. 317–322.

Fritsch, M (2019). *On GMM-estimation of linear dynamic panel data models.* University of Passau Working Papers in Business Administration B-36-19.

Gorgens, T, C Han, and S Xue (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". In: *Econometrica* 50.4, pp. 1029–1054.

Hansen, LP, J Heaton, and A Yaron (1996). "Finite-Sample Properties of Some Alternative GMM Estimators". In: *Journal of Business & Economic Statistics* 14.3, pp. 262–280.

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

Holtz-Eakin, D, K Newey Whitney, and HS Rosen (1988). "Estimating Vector Autoregressions with Panel Data". In: *Econometrica* 56.6, pp. 1371–1395.

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: *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis.* Ed. by GDA Phillips and E Tzavalis. Cambridge University Press, pp. 282–318.

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.

Kripfganz, S (2018). *XTDPDGMM: Stata module to perform generalized method of moments estimation of linear dynamic panel data models.* Version 1.1.1. URL: `http://EconPapers.repec.org/RePEc:boc:bocode:s458395`.

Nash, J and R Varadhan (2016). *optimr: A Replacement and Extension of the 'optim' Function.* Version 2016-8.16. URL: `https://cran.r-project.org/web/packages/optimr`.

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

Nash, JC and R Varadhan (2011). "Unifying Optimization Algorithms to Aid Software System Users: optimx for R". In: *Journal of Statistical Software* 43.9, pp. 1–14.

Newey, WK and DL McFadden (1994). "Chapter 36 – Large sample estimation and hypothesis testing". In: ed. by RF Engle and DL MacFadden. Vol. 4. Handbook of Econometrics. Elsevier, pp. 2111–2245.

Pua, A, M Fritsch, and J Schnurbus (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.

Pua, A, M Fritsch, and J Schnurbus (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.

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 (2009a). "How to do xtabond2: An Introduction to Difference and System GMM in Stata". In: *Stata Journal* 9.1, pp. 86–136.

Roodman, D (2009). "A Note on the Theme of Too Many Instruments". In: *Oxford Bulletin of Economics and Statistics* 71.1, pp. 135–158.

Sargan, JD (1958). "The Estimation of Economic Relationships Using Instrumental Variables". In: *Econometrica* 26.3, pp. 393–415.

Sigmund, M and R Ferstl (2019). "Panel vector autoregression in R with the package panelvar". In: *The Quarterly Review of Economics and Finance*, to appear.

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". In: *Journal of Econometrics* 126.1, pp. 25–51.

Betriebswirtschaftliche Reihe der Passauer Diskussionspapiere

Bisher sind erschienen:

B-1-98	Jochen Wilhelm, A fresh view on the Ho-Lee model of the term structure from a stochastic discounting perspective
B-2-98	Bernhard Nietert und Jochen Wilhelm, Arbitrage, Pseudowahrscheinlichkeiten und Martingale - ein didaktisch einfacher Zugang zur elementaren Bewertungstheorie für Derivate
B-3-98	Bernhard Nietert, Dynamische Portfolio-Selektion unter Berücksichtigung von Kurssprüngen
B-4-99	Jochen Wilhelm, Option Prices with Stochastic Interest Rates – Black/Scholes and Ho/Lee unified
B-5-99	Anja Tuschke, Stock-based Incentive Schemes and the Managerial Labor Market
B-6-00	Jochen Wilhelm, Das Gaußsche Zinsstrukturmodell – Eine Analyse auf der Basis von Wahrscheinlichkeitsverteilungen
B-7-01	Bernhard Nietert und Jochen Wilhelm, Some Economic Remarks on Arbitrage Theory
B-8-01	Jochen Wilhelm, Option Prices with Stochastic Interest Rates – Black/Scholes and Ho/Lee unified
B-9-02	Jochen Wilhelm, Risikoabschläge, Risikozuschläge und Risikoprämien – Finanzierungstheoretische Anmerkungen zu einem Grundproblem der Unternehmensbewertung
B-10-03	Jochen Wilhelm, Unternehmensbewertung – Eine finanzmarkttheoretische Untersuchung
B-11-04	Bernhard Nietert und Jochen Wilhelm, Non-Negativity of Nominal and Real Riskless Rates, Arbitrage Theory, and the Null-Alternative Cash
B-12-06	Armin Dolzer und Bernhard Nietert – Portfolio selection with Time Constraints and a Rational Explanation of Insufficient Diversification and Excessive Trading
B-13-08	Josef Schosser - Bewertung ohne "Kapitalkosten": ein arbitragetheoretischer An satz zu Unternehmenswert, Kapitalstruktur und persönlicher Besteuerung
B-14-14	Mathias Eickholt, Oliver Entrop, Marco Wilkens, Individual Investors and Suboptimal Early Exercises in the Fixed-Income Market
B-15-14	Mathias Eickholt, Oliver Entrop, Marco Wilkens, What Makes Individual Investors Exercise Early? Empirical Evidence from the Fixed-Income Market

B-16-14 Mathias Eickholt, Behavioral Financial Engineering in the Fixed-Income Market: The Influence of the Coupon Structure

B-17-16 Caroline Baethge, Performance in the Beauty Contest: How Strategic Discussion Enhances Team Reasoning

B-18-16 Caroline Baethge, Marina Fiedler, Aligning Mission Preferences: Does Self-Selection Foster Performance in Working Groups?

B-19-16 Caroline Baethge, Marina Fiedler, All or (Almost) Nothing? The Influence of Information Cost and Training on Information Selection and the Quality of Decision-Making.

B-20-16 Caroline Baethge, Marina Fiedler, Ernan Haruvey, In It to Win It: Experimental Evidence on Unique Bid Auctions

B-21-16 Markus Grottke, Maximilian Kittl, First the stick, then the carrot? A cross-country evaluation of the OECD's initiative against harmful tax competition

B-22-16 Heike Diller, Stephen Jeffrey, Marina Fiedler, Searching for the silver linings of techno-invasion

B-23-16 Stephen Jeffrey, Heike Diller, Marina Fiedler, How does intensification and mobile rearrangement affect employee commitment

B-24-16 Heike Diller, Life is tough so you gotta be rough – How resilience impacts employees' attitude towards ICT use

B-25-16 Stephen Jeffrey, Heike Diller, Marina Fiedler, Closing the Strategy-Performance Gap: The Role of Communication Fit and Distraction

B-26-17 S. Baller, O. Entrop, A. Schober, M. Wilkens, What drives Performance in the Speculative Market of Short-Term Exchange-Traded Retail Products?

B-27-17 S. Baller, Risk Taking in the Market of Speculative Exchange-Traded Retail Products: Do Socio-Economic Factors matter?

B-28-17 L. Frey, L. Engelhard, Review on Tax Research in Accounting: Is the information given by U.S. GAAP income taxes also provided by IFRS?

B-29-17 J. Lorenz, M. Diller, Do Tax Information Exchange Agreements Curb Transfer Pricing-Induced Tax Avoidance?

B-30-17 J. Lorenz, M. Grottke, Tax Consultants' Incentives – A Game-Theoretic Investigation into the Behavior of Tax Consultants, Taxpayers and the Tax Authority in a Setting of Tax Complexity

B-31-18 Oliver Entrop, Matthias F. Merkel, "Exchange Rate Risk" within the European Monetary Union? Analyzing the Exchange Rate Exposure of German Firms

B-32-18 Oliver Entrop, Matthias F. Merkel, Manager's Research Education, the Use of FX Derivatives and Corporate Speculation

B-33-18 Matthias F. Merkel, Foreign Exchange Derivative Use and Firm Value: Evidence from German Non-Financial Firms

B-34-19 Oliver Entrop, Georg Fischer, Hedging Costs and Joint Determinants of Premiums and Spreads in Structured Financial Products

B-35-19 Georg Fischer, How Dynamic Hedging Affects Stock Price Movements: Evidence from German Option and Certificate Markets

B-36-19 Markus Fritsch, On GMM estimation of linear dynamic panel data models

B-37-19 Adrew Adrian Yu Pua, Markus Fritsch, Joachim Schnurbus, Large sample properties of an IV estimator based on the Ahn and Schmidt moment conditions

B-38-19 Andrew Adrian Yu Pua, Markus Fritsch, Joachim Schnurbus, Practical aspects of using quadratic moment conditions in linear dynamic panel data models

Betriebswirtschaftliche Reihe der Passauer Diskussionspapiere

Bisher sind erschienen:

B-1-98 Jochen Wilhelm, A fresh view on the Ho-Lee model of the term structure from a stochastic discounting perspective

B-2-98 Bernhard Nietert und Jochen Wilhelm, Arbitrage, Pseudowahrscheinlichkeiten und Martingale - ein didaktisch einfacher Zugang zur elementaren Bewertungstheorie für Derivate

B-3-98 Bernhard Nietert, Dynamische Portfolio-Selektion unter Berücksichtigung von Kurssprüngen

B-4-99 Jochen Wilhelm, Option Prices with Stochastic Interest Rates – Black/Scholes and Ho/Lee unified

B-5-99 Anja Tuschke, Stock-based Incentive Schemes and the Managerial Labor Market

B-6-00 Jochen Wilhelm, Das Gaußsche Zinsstrukturmodell – Eine Analyse auf der Basis von Wahrscheinlichkeitsverteilungen

B-7-01 Bernhard Nietert und Jochen Wilhelm, Some Economic Remarks on Arbitrage Theory

B-8-01 Jochen Wilhelm, Option Prices with Stochastic Interest Rates – Black/Scholes and Ho/Lee unified

B-9-02 Jochen Wilhelm, Risikoabschläge, Risikozuschläge und Risikoprämien – Finanzierungstheoretische Anmerkungen zu einem Grundproblem der Unternehmensbewertung

B-10-03 Jochen Wilhelm, Unternehmensbewertung – Eine finanzmarkttheoretische Untersuchung

B-11-04 Bernhard Nietert und Jochen Wilhelm, Non-Negativity of Nominal and Real Riskless Rates, Arbitrage Theory, and the Null-Alternative Cash

B-12-06 Armin Dolzer und Bernhard Nietert – Portfolio selection with Time Constraints and a Rational Explanation of Insufficient Diversification and Excessive Trading

B-13-08 Josef Schosser - Bewertung ohne "Kapitalkosten": ein arbitragetheoretischer An satz zu Unternehmenswert, Kapitalstruktur und persönlicher Besteuerung

B-14-14 Mathias Eickholt, Oliver Entrop, Marco Wilkens, Individual Investors and Suboptimal Early Exercises in the Fixed-Income Market

B-15-14 Mathias Eickholt, Oliver Entrop, Marco Wilkens, What Makes Individual Investors Exercise Early? Empirical Evidence from the Fixed-Income Market

B-16-14 Mathias Eickholt, Behavioral Financial Engineering in the Fixed-Income Market: The Influence of the Coupon Structure

B-17-16 Caroline Baethge, Performance in the Beauty Contest: How Strategic Discussion Enhances Team Reasoning

B-18-16 Caroline Baethge, Marina Fiedler, Aligning Mission Preferences: Does Self-Selection Foster Performance in Working Groups?

B-19-16 Caroline Baethge, Marina Fiedler, All or (Almost) Nothing? The Influence of Information Cost and Training on Information Selection and the Quality of Decision-Making.

B-20-16 Caroline Baethge, Marina Fiedler, Ernan Haruvey, In It to Win It: Experimental Evidence on Unique Bid Auctions

B-21-16 Markus Grottke, Maximilian Kittl, First the stick, then the carrot? A cross-country evaluation of the OECD's initiative against harmful tax competition

B-22-16 Heike Diller, Stephen Jeffrey, Marina Fiedler, Searching for the silver linings of techno-invasion

B-23-16 Stephen Jeffrey, Heike Diller, Marina Fiedler, How does intensification and mobile rearrangement affect employee commitment

B-24-16 Heike Diller, Life is tough so you gotta be rough – How resilience impacts employees' attitude towards ICT use

B-25-16 Stephen Jeffrey, Heike Diller, Marina Fiedler, Closing the Strategy-Performance Gap: The Role of Communication Fit and Distraction

B-26-17 S. Baller, O. Entrop, A. Schober, M. Wilkens, What drives Performance in the Speculative Market of Short-Term Exchange-Traded Retail Products?

B-27-17 S. Baller, Risk Taking in the Market of Speculative Exchange-Traded Retail Products: Do Socio-Economic Factors matter?

B-28-17 L. Frey, L. Engelhard, Review on Tax Research in Accounting: Is the information given by U.S. GAAP income taxes also provided by IFRS?

B-29-17 J. Lorenz, M. Diller, Do Tax Information Exchange Agreements Curb Transfer Pricing-Induced Tax Avoidance?

B-30-17 J. Lorenz, M. Grottke, Tax Consultants' Incentives – A Game-Theoretic Investigation into the Behavior of Tax Consultants, Taxpayers and the Tax Authority in a Setting of Tax Complexity

B-31-18 Oliver Entrop, Matthias F. Merkel, "Exchange Rate Risk" within the European Monetary Union? Analyzing the Exchange Rate Exposure of German Firms

B-32-18 Oliver Entrop, Matthias F. Merkel, Manager's Research Education, the Use of FX Derivatives and Corporate Speculation

B-33-18 Matthias F. Merkel, Foreign Exchange Derivative Use and Firm Value: Evidence from German Non-Financial Firms

B-34-19 Oliver Entrop, Georg Fischer, Hedging Costs and Joint Determinants of Premiums and Spreads in Structured Financial Products

B-35-19 Georg Fischer, How Dynamic Hedging Affects Stock Price Movements: Evidence from German Option and Certificate Markets

B-36-19 Markus Fritsch, On GMM estimation of linear dynamic panel data models

B-37-19 Andrew Adrian Yu Pua, Markus Fritsch, Joachim Schnurbus, Large sample properties of an IV estimator based on the Ahn and Schmidt moment conditions

B-38-19 Andrew Adrian Yu Pua, Markus Fritsch, Joachim Schnurbus, Practical aspects of using quadratic moment conditions in linear dynamic panel data models