UNIVERSITÄT PASSAU

Wirtschaftswissenschaftliche Fakultät

On GMM estimation of linear dynamic panel data models

Markus Fritsch

Diskussionsbeitrag Nr. B-36-19

Betriebswirtschaftliche Reihe ISSN 1435-3539

PASSAUER DISKUSSIONSPAPIERE

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

On GMM estimation of linear dynamic panel data models

Markus Fritsch

Diskussionsbeitrag Nr. B-36-19

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

On GMM estimation of linear dynamic panel data models[1]

Markus Fritsch[2]

September 20, 2019

Abstract. The linear dynamic panel data model provides a possible avenue to deal with unobservable individual-specific heterogeneity and dynamic relationships in panel data. The model structure renders standard estimation techniques inconsistent. Estimation and inference can, however, be carried out with the generalized method of moments (GMM) by suitably aggregating population orthogonality conditions directly deduced from the underlying modeling assumptions. Different variations of these assumptions are proposed in the literature – often lacking a thorough discussion of the implications for estimation and inference. This paper aims to enhance the understanding of the assumptions and their interplay by connecting the assumptions and the conditions required to establish identification and consistency, derive the asymptotic properties, and carry out inference for the GMM estimator.

Keywords. GMM, linear dynamic panel data model, identification, large sample properties, inference.

JEL codes. C10, C23.

[1]Valuable discussions, comments, and guidance by Harry Haupt, Andrew A.Y. Pua, and Joachim Schnurbus is appreciated. All remaining errors are mine.

[2]University of Passau, Department of Statistics, 94030 Passau, email: markus.fritsch@uni-passau.de

1 Introduction

Linear dynamic panel data models contain lags of the dependent variable as explanatory variables and an unobservable composite error term consisting of an individual-specific effect, a time-specific effect, and an idiosyncratic remainder component. Due to the presence of lagged dependent variables besides the (unobservable) individual-specific effects, conventional estimation techniques such as ordinary least squares (OLS) or within estimation do not generally produce consistent estimates. An alternative to account for individual-specific effects is to first difference the equation. However, this induces correlation between the lagged dependent explanatory variables and the idiosyncratic remainder components which also leads to inconsistent parameter estimates. An established way to account for unobservable individual-specific heterogeneity in the presence of lagged dependent explanatory variables is deriving instruments and corresponding population orthogonality conditions (or moment conditions) from the modeling assumptions (see, e.g., Anderson and Hsiao, 1981; Anderson and Hsiao, 1982; Holtz-Eakin, Newey, and Rosen, 1988; Ahn and Schmidt, 1995; Arellano and Bover, 1995) and employing the sample analogues of the population orthogonality conditions in estimation[3]. Typically, more population orthogonality conditions are available than there are parameters to be estimated and there is no unique solution when minimizing the objective function. Appropriately reducing the number of population orthogonality conditions provides a remedy. This may be achieved by selecting as many population orthogonality conditions as there are parameters to be estimated or by forming an equivalent number of linear combinations of the population orthogonality conditions. Such linear combinations may be obtained by the generalized method of moments, where the aggregation is governed by the so-called weighting matrix.

Several conditions are required to ensure identification, consistency, and asymptotic normality or carry out inference for the GMM estimator. In practice, theoretical and applied work on GMM estimation of linear dynamic panel data models commonly employs different versions of assumptions (or slight variations thereof) – often lacking a thorough discussion of their implications, limitations, and consequences for estimation and inference. The aim of the paper is to provide such a discussion and thereby enhance the understanding of the assumptions frequently encountered in GMM estimation of linear dynamic panel data models and their interplay. The assumptions are connected with and discussed in the context of the required conditions stated in general theorems and propositions which establish identification, consistency, and the asymptotic properties of the GMM estimator. Additionally, potential directions for future research are pointed out.

The structure of the paper is as follows: Section 2 briefly sketches the linear dynamic panel data model, elaborates on the underlying assumptions, and introduces estimation by the generalized method of moments. Section 3 discusses the assumptions in the context of the conditions required for identification, consistency, and asymptotic normality of the model parameters by GMM. Section 4 considers the consistent estimation of standard errors, specification testing, and testing of general hypotheses, while Section 5 concludes.

[3]Estimation here may be carried out by the method of moments. In the remainder of the paper, it will refer to OLS when the type of estimation is not expressed more accurately.

2 Modeling framework

2.1 Linear dynamic panel data model

The model structure underlying the linear dynamic panel data model can be represented by the two equations

$$Y_t = Y_{t-1}\rho_1 + \cdots + Y_{t-p}\rho_p + \boldsymbol{X}_t\boldsymbol{\beta}_0 + \cdots + \boldsymbol{X}_{t-q}\boldsymbol{\beta}_q + U_t, \qquad t \in \mathbb{Z}, \tag{1}$$

$$U_t = A + \lambda_t + \varepsilon_t. \tag{2}$$

The variable Y_t denotes the dependent variable and the right-hand side of the equation contains p lags of the dependent variable, the $J-1$ contemporaneous non-lagged dependent explanatory variables $\boldsymbol{X}_t = \{X_{t,1}, \ldots, X_{t,J-1}\}$, and up to q lags of the $\boldsymbol{X}_t$ as explanatory variables. The scalars $\rho_1, \ldots, \rho_p$ are the lag parameters, $\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_q$ are parameter vectors, U_t is an unobservable composite error term, A is an unobservable individual-specific effect, λ_t is an unobservable time-specific effect, and $\varepsilon_{i,t}$ is an idiosyncratic remainder component. Dependence of a variable of the time period is expressed by a t in the index – where t is an integer.

In the following a setup is considered where the variables $Y_t, \ldots, Y_{t-p}, U_t, A, \lambda_t$, and ε_t denote random variables and $\boldsymbol{X}_t = \{X_{t,1}, \ldots, X_{t,J-1}\}$ and its lags represent vectors of random variables.[4] The Equations (1) and (2) are typically derived from the following assumptions about the individual model components and their interplay:

(*A.1.1*) *Existence of a reduced form:* It is possible to solve the structural form

$$f(Y_t, \ldots, Y_{t-p}, \boldsymbol{X}_t, \ldots, \boldsymbol{X}_{t-q}, U_t) = 0$$

for the endogenous variable Y_t; this leads to the existence of the reduced form

$$Y_t = g(Y_{t-1}, \ldots, Y_{t-p}, \boldsymbol{X}_t, \ldots, \boldsymbol{X}_{t-q}, U_t).$$

(*A.1.2*) *Additive separability:* The reduced form may be decomposed into an observable model component $g_1(Y_{t-1}, \ldots, Y_{t-p}, \boldsymbol{X}_t, \ldots, \boldsymbol{X}_{t-q})$ and an unobservable composite error term U_t, which captures all unobservable influences on Y_t, such that

$$Y_t = g_1(Y_{t-1}, \ldots, Y_{t-p}, \boldsymbol{X}_t, \ldots, \boldsymbol{X}_{t-q}) + U_t.$$

The unobservable composite error term U_t may be expressed as a function $g_2(A, \lambda_t, \varepsilon_t)$, such that $U_t = g_2(A, \lambda_t, \varepsilon_t)$.

[4] Put differently, Y_t and $\boldsymbol{X}_t$ can be seen as resulting from time series processes with random starting values and are, therefore, also random variables. The variable A is constant over time and can be thought of as randomly assigned at the initial time period, while one common λ_t is drawn at each time period for all individuals. For a DGP without time-specific effects specified in this spirit see, e.g., Blundell, Bond, and Windmeijer (2001).

(*A.1.3*) *Linearity and additivity:* The functions $g_1(\cdot)$ and $g_2(\cdot)$ are additively separable as well. The function $g_1(\cdot)$ is additionally assumed to be linear in parameters, such that

$$g_1(Y_{t-1},\dots,Y_{t-p},\boldsymbol{X}_t,\dots,\boldsymbol{X}_{t-q}) = Y_{t-1}\rho_1 + \dots + Y_{t-p}\rho_p + \boldsymbol{X}_t\boldsymbol{\beta}_0 + \dots + \boldsymbol{X}_{t-q}\boldsymbol{\beta}_q$$
$$\text{and} \qquad g_2(A,\lambda_t,\varepsilon_t) = A + \lambda_t + \varepsilon_t.$$

Replacing the upper-case letters by lower-case letters for all model components and adding the index i indicating the individual yields the conventional representation of the linear dynamic panel data model in the literature.

$$y_{i,t} = y_{i,t-1}\rho_1 + \dots + y_{i,t-p}\rho_p + \boldsymbol{x}'_{i,t}\boldsymbol{\beta}_0 + \dots + \boldsymbol{x}'_{i,t-q}\boldsymbol{\beta}_q + u_{i,t}, \quad i = 1,\dots,N; t \in \mathbb{Z}, \tag{3}$$
$$u_{i,t} = a_i + \lambda_t + \varepsilon_{i,t}. \tag{4}$$

Plugging Equation (4) into Equation (3) yields the single equation form of the model

$$y_{i,t} = y_{i,t-1}\rho_1 + \dots + y_{i,t-p}\rho_p + \boldsymbol{x}'_{i,t}\boldsymbol{\beta}_0 + \dots + \boldsymbol{x}'_{i,t-q}\boldsymbol{\beta}_q + a_i + \lambda_t + \varepsilon_{i,t}. \tag{5}$$

This version of the linear dynamic panel data model is frequently employed in the empirical applications in Arellano and Bond (1991), Blundell and Bond (1998), Blundell and Bond (2000), Blundell, Bond, and Windmeijer (2001), and Bond (2002). Examples where adjusted variations of the model in (5) are employed are Lemmon, Roberts, and Zender (2008), Michaud and Van Soest (2008), and Cavalcanti, Mohaddes, and Raissi (2015): In the first paper the dependent variable is used in first differences and the right-hand side variables are considered in levels. The second paper includes interactions of selected explanatory variables with time and the latter paper employs the geometric average growth rate of real GDP as dependent variable and the natural logarithm of real GDP constitutes the lagged explanatory variable.

Theoretical work, review articles and discussions of particular properties in the estimation of linear dynamic panel data models often consider a simplified version of Equation (5) by incorporating only one lag of the dependent variable, excluding non-lagged dependent explanatory variables, and dropping time-specific effects λ_t

$$y_{i,t} = y_{i,t-1}\rho + a_i + \varepsilon_{i,t}, \qquad i = 1,\dots,N;\ t = 2,\dots,T. \tag{6}$$

The initial time period is denoted by $t = 1$. Examples are Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1998), Alvarez and Arellano (2003), Hayakawa (2009), Bun and Kleibergen (2014), Gorgens, Han, and Xue (2016a), and Gorgens, Han, and Xue (2016b) – where Ahn and Schmidt (1995), Arellano and Bover (1995), and Blundell and Bond (1998) discuss including non-lagged dependent explanatory variables and Gorgens, Han, and Xue (2016b) discuss including time-specific effects.

A more flexible specification contains contemporaneous non-lagged dependent explanatory variables

$$y_{i,t} = y_{i,t-1}\rho + \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + a_i + \varepsilon_{i,t}. \tag{7}$$

Examples where this specification is considered are Blundell, Bond, and Windmeijer (2001) and Bun and Sarafidis (2015). Hayakawa (2015) does not incorporate $\boldsymbol{x}_{i,t}$, but includes time-specific effects. All illustrations in the remainder of this paper are based on Equation (7) to simplify the notation and enhance the clarity of exposition.

Different modeling techniques have been proposed to estimate the model parameters of Equation (7) while accounting for unobservable individual-specific heterogeneity: Since the set of explanatory variables includes a lagged dependent variable, the within transformation – which amounts to subtracting the mean over time for all individuals from both sides of the equation – and subsequently estimating the parameters of the transformed equation generally does not yield consistent estimates. The same applies when simply ignoring the unobservable individual-specific effect a_i (for an exposition on both estimators see, e.g., Hsiao, 2014, p.82-86). Other approaches are based on first differencing (Anderson and Hsiao, 1982; Holtz-Eakin, Newey, and Rosen, 1988; Arellano and Bond, 1991), as the unobservable composite error term in Equation (7) then reduces to $\Delta u_{i,t} = \Delta\varepsilon_{i,t}$, models involving factor structures (Bai, 2013), and clustering-based techniques (Bonhomme and Manresa, 2015). The focus of the following exposition is on approaches based on first differencing.

First differencing equals expressing Equation (7) in first differences:

$$\Delta y_{i,t} = \Delta y_{i,t-1}\rho + \Delta \boldsymbol{x}_{i,t}'\boldsymbol{\beta} + \Delta\varepsilon_{i,t}. \tag{8}$$

The consequences of the first differencing, indicated by the Δ-operator, are twofold. First, it eliminates the individual-specific effect a_i. Second, first differencing induces correlation of the first differenced lagged dependent variable $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ with the first differenced unobservable composite error term $\Delta\varepsilon_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$. Consequently, estimates of ρ and $\boldsymbol{\beta}$ based on Equation (8) are inconsistent. This can be resolved by instrumenting the first differenced lagged dependent variable in Equation (8) or the lagged dependent variable in Equation (7) in estimation. Natural instruments may be derived from population orthogonality conditions which result from assumptions imposed on top of Assumptions (*A.1.1*)-(*A.1.3*).

Different sets of assumptions for the estimation of linear dynamic panel data models are proposed in the literature. Ahn and Schmidt (1997) provide an extensive discussion of several different sets of assumptions and their implications for the type and number of available population orthogonality conditions. The guise of the assumptions varies in the literature: Some authors use conditional moment restrictions (Arellano and Bover, 1995; Bun and Sarafidis, 2015); others employ unconditional moment restrictions (e.g., Ahn and Schmidt, 1995; Blundell and Bond, 1998; Blundell, Bond, and Windmeijer, 2001; Bun and Kleibergen, 2014; Kiviet, Pleus, and Poldermans, 2017). Ahn and Schmidt (1995) illustrate the assumptions in terms of restrictions on the covariance matrix of observable model components. One particular set of assumptions for the estimation of linear dynamic panel data models in Equation (7), which is often encountered in practice and which is imposed here, is

(*A.2.1*) $|\rho| \leq 1$,

(*A.2.2*) $\mathrm{E}[\varepsilon_{i,t}|y_{i,1}, y_{i,2}, \ldots, y_{i,t-1}, \boldsymbol{x}_{i,1}', \boldsymbol{x}_{i,2}', \ldots, \boldsymbol{x}_{i,r}', a_i] = 0, \qquad i = 1, \ldots, N;\; t = 2, \ldots, T;$
$r \leq T$,

(*A.2.3*) $\mathrm{E}[\Delta y_{i,t}|a_i] = 0, \qquad i = 1, \ldots, N;\ t = 2, \ldots, T,$

$(\mathrm{E}[\Delta \boldsymbol{x}'_{i,t}|a_i] = 0, \qquad i = 1, \ldots, N;\ t = 2, \ldots, T),$

(*A.2.4*) $y_{i,1}$ and $\boldsymbol{x}'_{i,1}$ are observable.

Assumption (*A.2.1*) restricts the range of possible values for the lag parameter ρ. The assumption prevents the $y_{i,t}$-process to exhibit explosive behaviour but includes the unit root case (for $\rho = 1$). Assumption (*A.2.2*) requires that the idiosyncratic remainder component $\varepsilon_{i,t}$ cannot be expressed as a function of any of the explanatory variables and the unobservable individual-specific effects. Additionally, it rules out serial correlation in $\varepsilon_{i,t}$ and allows the individual, non-lagged dependent explanatory variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{J-1}$ to be strictly exogenous ($r = T$), predetermined ($r = t$), or endogenous ($r < t$).[5] In short, Assumption (*A.2.2*) requires that the model is dynamically complete. Note that the assumption allows for arbitrary correlation between a_i and the explanatory variables. This correlation is restricted by Assumption (*A.2.3*), which requires the dependent variable and the individual-specific effect a_i to be constantly correlated over time (Bun and Sarafidis, 2015). Other expressions commonly used in the literature for this assumption are 'mean stationarity' (Arellano, 2003) and 'effect stationarity' (Bun and Kiviet, 2006). When the linear dynamic panel data model in Equation (7) and the Assumptions (*A.2.1*) and (*A.2.2*) hold and the dependent variable and the unobservable individual-specific effects fulfill the assumption of 'constant correlated effects', the assumption also needs to hold for the non-lagged dependent explanatory variables.

For an illustration, consider the linear dynamic panel data model in Equation (7) and suppose that the Assumptions (*A.2.1*) and (*A.2.2*) hold and that the $y_{i,t}$-process fulfills the 'constant correlated effects' assumption. By considering the conditional expectation of the equation given the unobservable individual-specific effects, the following can be shown:

$$\begin{aligned} E[y_{i,t}|a_i] &= E[y_{i,t-1}|a_i]\rho + E[\boldsymbol{x}'_{i,t}|a_i]\boldsymbol{\beta} + a_i + E[\varepsilon_{i,t}|a_i] \\ \Leftrightarrow E[y_{i,t}|a_i] - E[y_{i,t-1}|a_i]\rho &= E[\boldsymbol{x}'_{i,t}|a_i]\boldsymbol{\beta} + a_i + E[\varepsilon_{i,t}|a_i] \\ \Leftrightarrow E[y_{i,t}|a_i] - E[y_{i,t-1}|a_i]\rho &= E[\boldsymbol{x}'_{i,t}|a_i]\boldsymbol{\beta} + a_i. \end{aligned}$$

Rearranging the equation using that $E[y_{i,t}|a_i] = E[y_{i,t-1}|a_i]$ when the dependent variable and the unobservable individual-specific effects are constantly correlated results in

$$\begin{aligned} E[y_{i,t}|a_i](1-\rho) &= E[\boldsymbol{x}'_{i,t}|a_i]\boldsymbol{\beta} + a_i \\ \Leftrightarrow E[y_{i,t}|a_i] &= \frac{E[\boldsymbol{x}'_{i,t}|a_i]\boldsymbol{\beta} + a_i}{1-\rho}. \end{aligned}$$

Equivalently stating the process for time period $t-1$ and deducting this expression from the equation yields

$$E[y_{i,t}|a_i] - E[y_{i,t-1}|a_i] = \frac{E[\boldsymbol{x}'_{i,t}|a_i]\boldsymbol{\beta} + a_i - E[\boldsymbol{x}'_{i,t-1}|a_i]\boldsymbol{\beta} - a_i}{1-\rho}.$$

As the left-hand side of the equation is zero, $E[\boldsymbol{x}'_{i,t}|a_i] - E[\boldsymbol{x}'_{i,t-1}|a_i] = 0$ needs to hold to ensure that the right-hand side of the equation is zero as well. This requires that the $\boldsymbol{x}_{i,t}$-process is constantly correlated

[5] When the model includes lags of $\boldsymbol{X}$, r needs to be shifted backwards appropriately in the predetermined and endogenous case (by one period for each lag).

with the unobservable individual-specific effect. This derivation confirms Blundell, Bond, and Windmeijer (2001, p.69-70) who consider it 'very unlikely' that this is not the case – though the statement made in the previous sentence is somewhat stronger.[6]

Assumption (*A.2.4*) requires that the initial conditions of the $y_{i,t}$-process and the $\boldsymbol{x}'_{i,t}$-processes are contained in the data set employed for estimation (see, e.g., Gorgens, Han, and Xue, 2016b). In practice, strengthening this assumption is a common way to ensure that the 'constant correlated effects' assumption holds. Two different alternatives to strengthen Assumption (*A.2.4*) are: (1) explicitly assume that any deviations of the $y_{i,t}$- and the $\boldsymbol{x}_{i,t}$-processes from their long-term paths are not systematic over time or individuals (Blundell, Bond, and Windmeijer, 2001; Bun and Sarafidis, 2015); (2) assume a specific functional form for the $y_{i,t}$-process and the $\boldsymbol{x}'_{i,t}$-processes and further suppose that the processes run for a sufficiently long period of time prior to sampling such that the impact of the initial conditions on the first period contained in the sample disappears. The processes are then considered to be on their long-term paths (this is illustrated in Section 6.1 of the Appendix).

The Assumptions (*A.1.1*)-(*A.1.3*) and (*A.2.2*) allow deriving the following population orthogonality conditions (Holtz-Eakin, Newey, and Rosen, 1988) for Equation (8),

$$\begin{aligned} & \mathrm{E}[y_{i,s} \cdot \Delta u_{i,t}] = 0, \qquad t = 3, \ldots, T;\ s = 1, \ldots, t-2 \qquad (9) \\ \Leftrightarrow\ & \mathrm{E}[y_{is} \cdot (\Delta y_{i,t} - \Delta y_{i,t-1}\rho - \Delta \boldsymbol{x}'_{i,t}\boldsymbol{\beta})] = 0, \end{aligned}$$

where lags of the dependent variable – lagged at least two periods – can be used to instrument the endogenous regressor $\Delta y_{i,t-1}$. When further lags of the first differenced dependent variable are contained in Equation (8), the $y_{i,t}$ need to be lagged by the number of lags plus one period to be included as instruments.

Under the Assumptions (*A.1.1*)-(*A.1.3*) and (*A.2.2*), Ahn and Schmidt (1995) derive additional population orthogonality conditions for Equation (8), which are nonlinear in parameters,

$$\begin{aligned} & \mathrm{E}[u_{i,t} \cdot \Delta u_{i,t-1}] = 0, \qquad t = 4, \ldots, T \qquad (10) \\ \Leftrightarrow\ & \mathrm{E}[(y_{i,t} - y_{i,t-1}\rho - \boldsymbol{x}'_{i,t}\boldsymbol{\beta}) \cdot (\Delta y_{i,t-1} - \Delta y_{i,t-2}\rho - \Delta \boldsymbol{x}'_{i,t}\boldsymbol{\beta})] = 0. \end{aligned}$$

In contrast to the population orthogonality conditions in Equation (9), the population orthogonality conditions in Equation (10) depend on unobservable model components, but can be rewritten and represented in terms of observable model components. Rewriting the population orthogonality conditions reveals, that they are quadratic in the lag parameter ρ and the parameter vector $\boldsymbol{\beta}$. Note that, compared to Ahn and Schmidt (1995), the reference period of Equation (10) is changed from T to t as in Pua, Fritsch, and Schnurbus (2019a) and Pua, Fritsch, and Schnurbus (2019b) to use the same reference period for all population orthogonality conditions (see also Blundell, Bond, and Windmeijer, 2001; Bun and Kleibergen, 2014; Bun and Sarafidis, 2015).

Imposing the Assumptions (*A.1.1*)-(*A.1.3*) and (*A.2.2*)-(*A.2.3*) yields the further linear population orthog-

[6] Gratitude is owed to Andrew A.Y. Pua for pointing this out.

onality conditions (Arellano and Bover, 1995) for Equation (7)

$$\begin{aligned} &\mathrm{E}[\Delta y_{i,t-1} \cdot u_{i,t}] = 0, \qquad\qquad t = 3, \ldots, T \\ \Leftrightarrow\ &\mathrm{E}[\Delta y_{i,t-1} \cdot (y_{i,t} - y_{i,t-1}\rho - \boldsymbol{x}'_{i,t}\boldsymbol{\beta})] = 0, \end{aligned} \tag{11}$$

where instruments in differences $\Delta y_{i,t-1}$ are employed instead of the endogenous variable in levels. Two comments on the Arellano and Bover (1995) population orthogonality conditions are in order: First, when instead of the Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bover (1995) population orthogonality conditions, only the Arellano and Bover (1995) population orthogonality conditions are used in estimation, more population orthogonality conditions are available. Second, the linear population orthogonality conditions given in Equation (11) render the nonlinear population orthogonality conditions in Equation (10) redundant for estimation. The redundance of detailed in Section 6.2 of the Appendix.

Additional population orthogonality conditions for Equations (7) and (8) which are analogous to Equations (9) and (11) may arise from the non-lagged dependent explanatory variables – depending on the assumptions about the processes generating these variables one is willing to impose (see, e.g., Blundell, Bond, and Windmeijer, 2001). Other population orthogonality conditions proposed in literature which are not discussed here can be found in Ahn and Schmidt (1995), Hahn, Hausman, and Kuersteiner (2007), Han and Phillips (2010), and Han, Phillips, and Sul (2014). In practice, any arbitrary assumption about population moments can be imposed in order to derive suitable population orthogonality conditions.

2.2 Generalized method of moments

The brief discussion of the generalized method of moments and its properties in this section is tailored towards the estimation of linear dynamic panel data models. More thorough treatments of extremum-estimators in general and GMM estimation in particular are provided in a number of textbooks such as Davidson and MacKinnon (1993, chapter 17), Hayashi (2000, chapters 3, 4, and 7), Arellano (2003, Appendix A), and Wooldridge (2010, chapters 8, 11, and 14). A detailed summary of the historic and recent developments of the GMM methodology is provided by Hall (2015). Applications of GMM estimation of linear dynamic panel data models can be found in a number of different fields such as finance (Lemmon, Roberts, and Zender, 2008; Faulkender et al., 2012; Flannery and Watson Hankins, 2013), health economics (Michaud and Van Soest, 2008), industrial organization (Nickell, 1996), labor economics (Arellano and Bond, 1991), microeconomics (Banks, Blundell, and Lewbel, 1997; Blundell and Bond, 2000; Browning and Collado, 2007), and macroeconomics (Nason and Smith, 2008; Cavalcanti, Mohaddes, and Raissi, 2015). Software implementations are readily available in a number of different programming languages (see, e.g., Arellano and Bond, 1988 for `Gauss`, Doornik, Arellano, and Bond, 2012 for `Ox`, Roodman, 2009a and Kripfganz, 2018 for `Stata`, and Croissant and Millo, 2008 and Fritsch, Pua, and Schnurbus, 2019 for `R`).

GMM estimation uses the sample analogues of population orthogonality conditions derived from the model assumptions to obtain parameter estimates. In particular, GMM estimation utilizes that the expectation of the function $\delta(\cdot)$ of some random variables and the parameter θ is zero (Hayashi, 2000, p.204). Generalizing this statement to the estimation of a vector of parameters $\boldsymbol{\theta}$ and stacking multiple functions in $\boldsymbol{\delta}(\cdot)$ leads to the expression

$$\mathrm{E}\left[\boldsymbol{\delta}((Y, \boldsymbol{Z}, \boldsymbol{V}), \boldsymbol{\theta})\right] = \mathbf{0}, \qquad \text{with} \qquad \boldsymbol{\theta} \in \Theta. \tag{12}$$

In Equation (12), $\boldsymbol{\delta}(\cdot)$ denotes K functions of the J-dimensional parameter vector $\boldsymbol{\theta}$ taken from the set of admissible values Θ, the random variable Y, and the vectors containing multiple random variables $\boldsymbol{Z}$ and $\boldsymbol{V}$. The $K \times 1$ vector $\mathbf{0}$ is the zero vector. In the context of Section 2.1, $\mathrm{E}\left[\boldsymbol{\delta}(\cdot)\right] = \mathbf{0}$ are the population orthogonality conditions derived from the model assumptions given in Equations (9)-(11). The sample analogue of Equation (12) can then be used to estimate the parameters ρ and $\boldsymbol{\beta}$ of the linear dynamic panel data model in Equation (7). For a given sample, the uppercase letters are replaced by lowercase letters. The K instruments (which encompass the exogenous and the predetermined explanatory variables from $\boldsymbol{x}$) are collected in the vector $\boldsymbol{z}$ and the endogenous explanatory variables from $\boldsymbol{x}$ are assigned to the vector $\boldsymbol{v}$. The parameter vector $\boldsymbol{\theta}$ contains ρ and $\boldsymbol{\beta}$. Typically, more equations are available than there are parameters to be estimated and an aggregation and minimization scheme is required. The generalized method of moments provides such a scheme[7], where the aggregated Euclidean distance of the individual population orthogonality conditions from zero is minimized. The GMM estimator $\hat{\boldsymbol{\theta}}_N$ maximizes the objective function

$$\hat{Q}_N(\boldsymbol{\theta}) = -\, \boldsymbol{h}_N(\boldsymbol{\theta})' \widehat{\boldsymbol{W}}_N \boldsymbol{h}_N(\boldsymbol{\theta}), \tag{13}$$

where the $K \times K$ matrix $\widehat{\boldsymbol{W}}_N$ governs the aggregation of the individual sample orthogonality conditions and assigns a weight to each of the K conditions $\boldsymbol{h}_N(\boldsymbol{\theta})$. The hat on $\widehat{\boldsymbol{W}}_N$ indicates that the weighting matrix needs to be estimated beforehand, while the N index at the estimator $\hat{\boldsymbol{\theta}}_N$, the objective function $\hat{Q}_N(\boldsymbol{\theta})$, the sample orthogonality conditions $\boldsymbol{h}_N(\boldsymbol{\theta})$, and the weighting matrix denote that these quantities depend on the individual observations contained in the sample (see, e.g., Arellano, 2003, p.180). The sample orthogonality conditions are defined as

$$\boldsymbol{h}_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\delta}_i((y_i, \boldsymbol{z}_i', \boldsymbol{v}_i'), \boldsymbol{\theta}). \tag{14}$$

Under suitable assumptions, minimization of the GMM objective function in Equation (13) is equivalent to other well-known estimators: When using only population orthogonality conditions which are linear in parameters and simplifying the weighting matrix to an identity matrix, the GMM estimator is equivalent to the generalized instrumental variables (GIV) estimator. When additionally restricting the instrument set to one instrument per regressor to be instrumented, the GMM estimator is equivalent to the instrumental variables (IV) estimator. The OLS estimator results, when restricting the weighting matrix to an identity and when using the explanatory variables as instruments. Section 6.3 of the Appendix illustrates the connection of the estimators starting from the minimization of the objective functions.

[7] Alternative approaches are based on correcting the bias of the LSDV estimator or maximizing a pseudo log-likelihood function; analytic bias corrections are derived by Kiviet (1995) and Bun and Carree (2005); for approaches maximizing the log-likelihood function of the first differenced model, see Hsiao, Pesaran, and Tahmiscioglu (2002) and Hayakawa and Pesaran (2015); for a method based on setting up reduced form equations for the first and second moments of the observed parameters and imposing covariance restrictions in estimation, see Moral-Benito (2013).

3 Asymptotic properties

3.1 Identification

In GMM estimation of linear dynamic panel data models, point identification (or global identification) is sometimes assumed to establish consistency of the estimator (see, e.g., Newey and McFadden, 1994, Theorem 2.1). Point identification means that the objective function possesses a unique maximum at the true population parameter $\boldsymbol{\theta}_0$, while set identification refers to the case where maximizing the objective function leads to multiple solutions (for an excellent summary of identification concepts see, e.g., Lewbel, 2019). In the context of the GMM sample objective function denoted by Equation (13), overidentification is present when more population orthogonality conditions K are available than there are parameters to be estimated J. In the exactly identified setting, as many population orthogonality conditions are available as there are parameters ($K = J$) (see Arellano, 2003, p.180). In the underidentified case ($K < J$), the J model parameters cannot be recovered without imposing additional restrictions on the parameter vector. Therefore, a necessary (but not sufficient) condition for identification is that the number of population orthogonality conditions is at least as large as the number of model parameters ($K \geq J$). This condition is referred to as the 'order condition for identification' (see Hayashi, 2000, p. 202). Depending on the type of population orthogonality conditions available for estimation, sufficient conditions for identification can be stated based on the objective function (13) maximized in GMM estimation. This concept is utilized by Newey and McFadden (1994) and labelled 'extremum based identification' by Lewbel (2019).

For point identification of the parameters ρ and $\boldsymbol{\beta}$ in Equation (7), consider the Assumptions (*A.1.1*)-(*A.1.3*) on the model structure of the linear dynamic panel data model; further impose the Assumptions (*A.2.1*)-(*A.2.4*). All population orthogonality conditions available for estimation from these assumptions are linear in parameters. Additionally impose the further technical conditions.

(*A.3.1*) $\text{rank}(\text{E}[\boldsymbol{Z}'\boldsymbol{X}]) = J$.

(*A.3.2*) $\text{plim}_{N\to\infty} \widehat{\boldsymbol{W}}_N = \boldsymbol{W}$, with the $K \times K$ matrix $\boldsymbol{W}$ being symmetric and positive definite[8].

(*A.3.3*) The second moments $\sigma^2_{y_1}, \sigma^2_{x_1}, \sigma^2_a, \sigma^2_\varepsilon$ are assumed to be positive and finite; $\sigma_{y_1 a}$ and $\sigma_{x_1 a}$ are assumed to be finite.

(*A.3.4*) T fixed and $N \to \infty$, such that $T/N \to 0$.

(*A.3.5*) The data $(\boldsymbol{y}_i, \boldsymbol{z}_i, \boldsymbol{v}_i)$ are independently distributed across individuals.

Assumption (*A.3.1*) is the 'rank condition for identification' (see Hayashi, 2000, p.200-202). Since the largest rank a matrix can take is defined to be the minimum of its column and row rank, the assumption can only hold, when the 'order condition for identification' is fulfilled and there are at least as many

[8]When $\boldsymbol{l}'\widehat{\boldsymbol{W}}_N\boldsymbol{l} > 0$ holds for any arbitrary (real) non-zero column vector $\boldsymbol{l}$, all eigenvalues of $\widehat{\boldsymbol{W}}_N$ are strictly positive and the matrix is said to be positive definite (see Golub and Van Loan, 2012, p.159). Note that all vectors and matrices considered in this text are assumed to have real entries only.

population orthogonality conditions as there are parameters to be estimated. By Assumption (*A.3.2*), the weighting matrix $\widehat{\boldsymbol{W}}_N$ is assumed to converge to a nonstochastic symmetric positive definite matrix $\boldsymbol{W}$. Assumption (*A.3.3*) ensures that the second moments of the initial conditions of the $y_{i,t}$- and the $x_{i,t}$-process, the unobservable individual-specific effects, the idiosyncratic remainder component, and the cross moments of the initial conditions and the unobservable individual-specific effects are finite. The assumption is imposed, as identification failures may be attributable to these second moments (see, e.g., Bun and Sarafidis, 2015; Gorgens, Han, and Xue, 2016b). Limiting the time series dimension T to be finite while the cross section dimension tends to infinity by Assumption (*A.3.4*) provides the setting originally proposed for GMM estimation of linear dynamic panel data models by Arellano and Bond (1991) and Ahn and Schmidt (1995). The Assumption (*A.3.5*) that the data are independently distributed across individuals allows for dependence of the model components across time and (e.g.,) for the presence of unobservable time-specific effects when considering a more general setting than the one in Equation (7).

The implications of the Assumptions (*A.3.1*)-(*A.3.5*) can be illustrated in the context of GMM estimation: Assume that the Assumptions (*A.1.1*)-(*A.1.3*) and (*A.2.1*)-(*A.2.4*) hold and that the parameters ρ and $\boldsymbol{\beta}$ of the linear dynamic panel data model in Equation (7) are estimated by GMM. When using the Assumptions (*A.3.4*) and (*A.3.5*) and employing the law of large numbers, it can be shown that the sample orthogonality conditions stated in Equation (14) converge in probability to the true population orthogonality conditions (see Newey and McFadden, 1994, p.2126):

$$\operatorname*{plim}_{N\to\infty} \boldsymbol{h}_N(\boldsymbol{\theta}) = \boldsymbol{h}_0(\boldsymbol{\theta}) = \mathrm{E}[\boldsymbol{\delta}((Y, \boldsymbol{Z}, \boldsymbol{V}), \boldsymbol{\theta})],$$

Under Assumption (*A.3.2*), the sample objective function in Equation (13) converges in probability to the true objective function (see Newey and McFadden, 1994, p.2126):

$$\operatorname*{plim}_{N\to\infty} -\boldsymbol{h}_N(\boldsymbol{\theta})'\widehat{\boldsymbol{W}}_N\boldsymbol{h}_N(\boldsymbol{\theta}) = -\boldsymbol{h}_0(\boldsymbol{\theta})'\boldsymbol{W}\boldsymbol{h}_0(\boldsymbol{\theta}).$$

As illustrated in Section 6.3 of the appendix, computing the first derivative of the objective function and solving for $\boldsymbol{\theta}$ yields the closed form solution for the GMM estimator when only linear population orthogonality conditions are used in estimation

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

A unique solution to this expression exists, when the matrix product $\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X}$ is invertible. In Equation (15), the vector $\boldsymbol{y}$ contains the dependent variables, the matrix $\boldsymbol{X}$ collects the lagged dependent and non-lagged dependent explanatory variables, and the matrix $\boldsymbol{Z}$ contains the instruments. Depending on the population orthogonality conditions employed in estimation, all of these variables may be included in levels and/or first differences. Point identification of the parameters in Equation (7) by the GMM estimator employing only linear population orthogonality conditions can then be established with the following lemma.

Lemma 1 *(Point Identification of the GMM estimator)*
Assume the Assumptions (A.1.1)-(A.1.3), (A.2.1)-(A.2.4), and (A.3.1)-(A.3.5) hold: Maximization of the

GMM objective function in Equation (13) for the GMM estimator employing only linear population orthogonality conditions then yields a unique solution at the true population parameter.

Proof. By Assumption (*A.3.2*), the $K \times K$ weighting matrix $\widehat{\boldsymbol{W}}_N$ converges in probability to a symmetric positive definite matrix $\boldsymbol{W}$. Due to Assumption (*A.3.1*), the $K \times J$ matrix product $\boldsymbol{Z}'\boldsymbol{X}$ converges in probability to a matrix of full column rank. When utilizing the Slutzky theorem together with Theorem 4.2.1 of Golub and Van Loan (2012), it follows that the matrix product $\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X}$ converges in probability to a positive definite $K \times K$ matrix of full column rank. As a result, the probability limit of the matrix product $\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X}$ is invertible and the closed form of the GMM estimator given in Equation (15) possesses a unique solution (in probability).

Using the Assumptions (*A.1.1*)-(*A.1.3*), $\boldsymbol{y}$ can be substituted by $\boldsymbol{X}\boldsymbol{\theta}_0 + \boldsymbol{u}$ in Equation (15), where $\boldsymbol{\theta}_0$ is the vector of true population parameters and $\boldsymbol{u}$ may contain unobservable composite error terms in levels and/or first differences. Then

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= N(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X})^{-1}\ \frac{1}{N}\ \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'(\boldsymbol{X}\boldsymbol{\theta}_0 + \boldsymbol{u}) \\ \Leftrightarrow \quad &= \boldsymbol{\theta}_0 + N(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X})^{-1}\ \frac{1}{N}\ \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{u}. \end{aligned}$$

Under the Assumptions (*A.2.2*)-(*A.2.4*), it holds that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \cdot \boldsymbol{Z}'\boldsymbol{u} = 0.$$

From this result, it follows that

$$\operatorname*{plim}_{N\to\infty} \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0.$$

■

Non-orthogonality between the instruments in $\boldsymbol{Z}$ and the explanatory variables in $\boldsymbol{X}$ and orthogonality of the instruments and the composite error terms $\boldsymbol{u}$ holds due to Assumptions (*A.1.1*)-(*A.1.3*) and (*A.2.1*)-(*A.2.4*). Note that Lemma 1 departs from Lemma 2.3 of Newey and McFadden (1994) as it avoids the assumption that $\boldsymbol{h}_0(\boldsymbol{\theta}_0) = \boldsymbol{0}$ and $\boldsymbol{W}\boldsymbol{h}_0(\boldsymbol{\theta}) \neq \boldsymbol{0}$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Instead, similar to Hayashi (2000, p.206) and Windmeijer (2005), a more restrictive assumption is imposed on the weighting matrix: Assumption (*A.3.2*) requires that the weighting matrix is positive definite, while Newey and McFadden (1994) only impose positive semi-definiteness. The latter generality seems to have no immediate relevance for empirical applications.

Results on identification when linear and nonlinear population orthogonality conditions such as those stated in Equations (9)-(11) are employed in GMM estimation of linear dynamic panel data models are derived in two recent papers. Both papers consider a simplified version of the linear dynamic panel data model in Equation (7) and omit the non-lagged dependent explanatory variables. Bun and Kleibergen (2014) investigate a number of particular (mean stationary) worst-case DGPs and show that identification may fail under the population orthogonality conditions given in Equation (11). Corollary 1 of their paper illustrates that under additional assumptions, the population orthogonality conditions stated in Equations (9) and (10) or the population orthogonality conditions stated in Equations (9) and (11) identify the model parameter – even when the individual population orthogonality conditions fail to provide point identification. The additional assumptions are that the variance of the initial unobservable composite error term tends to zero as the true lag parameter tends to one and that the variance of the product of the initial $y_{i,t}$ and the unobservable composite error component $u_{i,t}$ is the variance of the product of the terms. Gorgens, Han, and Xue (2016b) utilize a set of assumptions which is similar to the assumptions considered in this paper. In particular, they impose the Assumptions (*A.1.1*)-(*A.1.3*), (*A.2.1*)-(*A.2.2*), (*A.2.4*), and (*A.3.4*)-(*A.3.5*). They also impose the Newey and McFadden (1994) version of Assumption (*A.3.2*) of a positive semi-definite weighting matrix $\boldsymbol{W}$, while no rank condition similar to Assumption (*A.3.1*) is stated. The authors investigate identification of the GMM estimator when the linear population orthogonality conditions given in Equation (9) and the nonlinear population orthogonality conditions from Equation (10) are employed in estimation. In their Theorem 1, they show that the linear population orthogonality conditions in Equation (9) fail to identify the lag parameter for some DGPs; they further show in their Theorem 2, that adding the nonlinear population orthogonality conditions in Equation (10) provides at least partial identification of the lag parameter – as long as the dependent variable exhibits variation over time. In their Theorem 3 they characterize particular settings (DGPs) in which the lag parameter is unidentified and partially identified. A potential avenue for future research is to generalize the results of Gorgens, Han, and Xue (2016b) to a setting which includes non-lagged dependent explanatory variables.

3.2 Consistency

Consistency of the GMM estimator requires that $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}_0$. Note that consistency of the GMM estimator is already established in the proof of Lemma 1. Nevertheless, since several consistency theorems of varying generality are available in the literature, a version is given here. Theorem 4.1.1 in Amemiya (1985) and Theorem 2.1 in Newey and McFadden (1994) treat extremum estimators, Theorem 2.6 in Newey and McFadden (1994) covers GMM estimators employing i.i.d. data, while Propositions 3.1 and 7.7 in Hayashi (2000) consider linear and nonlinear GMM estimators based on data which are jointly stationary and ergodic. Consistent estimation of the model parameters in Equation (7) by the GMM estimator can be established by the following theorem, which is similar to Theorem 2.6 in Newey and McFadden (1994).

Theorem 2 *(Consistency of the GMM estimator)*
Suppose the Assumptions (A.1.1)-(A.1.3), (A.2.1)-(A.2.4), and (A.3.1)-(A.3.5) hold. Further suppose:

(i) The population orthogonality conditions $\boldsymbol{h}_0(\boldsymbol{\theta}) = \mathrm{E}[\boldsymbol{\delta}((Y, \boldsymbol{Z}, \boldsymbol{V}), \boldsymbol{\theta})] = \boldsymbol{0}$, *for* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$*; for all* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\mathrm{E}[\boldsymbol{\delta}((Y, \boldsymbol{Z}, \boldsymbol{V}), \boldsymbol{\theta})] \neq \boldsymbol{0}$.

(ii) The set of admissible values Θ *is a compact subset of* $\mathbb{R}^J$, *where* $\boldsymbol{\theta}, \boldsymbol{\theta}_0 \in \Theta$.

(iii) The functions $\boldsymbol{\delta}((Y, \boldsymbol{Z}, \boldsymbol{V}), \boldsymbol{\theta})$ *are continuous and measurable functions in the parameter vector* $\boldsymbol{\theta}$ *for the data* $(y_i, \boldsymbol{z}_i, \boldsymbol{v}_i)$.

(iv) It holds that $\mathrm{E}[\sup_{\boldsymbol{\theta}\in\Theta} ||\boldsymbol{\delta}((Y, \boldsymbol{Z}, \boldsymbol{V}), \boldsymbol{\theta})||] < \infty$.

Under these conditions, $\mathrm{plim}_{N\to\infty} \hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta}_0$.

Proof. Condition (i) requires point identification of the parameter vector in Equation (7) by the GMM estimator, which holds under the given set of assumptions (see Lemma 1 and the corresponding proof). Employing the compactness condition for Θ in (ii), the continuity and measurability condition in (iii), and the 'dominance' condition (see Hayashi, 2000, p.468) in (iv), the sample orthogonality conditions $\boldsymbol{h}_N(\boldsymbol{\theta})$ are bounded by $\boldsymbol{h}_0(\boldsymbol{\theta})$ and it can be shown that:

$$\underset{N\to\infty}{\mathrm{plim}} \left(\sup_{\boldsymbol{\theta}\in\Theta} ||\boldsymbol{h}_N(\boldsymbol{\theta}) - \boldsymbol{h}_0(\boldsymbol{\theta})|| \right) = 0.$$

As a consequence, $\boldsymbol{h}_N(\boldsymbol{\theta})$ converge uniformly in probability to $\boldsymbol{h}_0(\boldsymbol{\theta})$ over Θ. Continuity of $\boldsymbol{h}_0(\boldsymbol{\theta})$ over Θ follows immediately and the sample objective function $\hat{Q}_N(\boldsymbol{\theta})$ in Equation (13) and the population objective function $Q_0(\boldsymbol{\theta})$ are also continuous and measurable functions in $\boldsymbol{\theta}$ for the data (see Lemma 2.4 in Newey and McFadden, 1994 and Lemma 7.2 in Hayashi, 2000).

It can then be shown by the Triangle and Cauchy-Schwarz Inequalities (see Newey and McFadden, 1994, p.2132) that the sample objective function converges uniformly in probability to the true objective function

$$\underset{N\to\infty}{\mathrm{plim}} \left(\sup_{\boldsymbol{\theta}\in\Theta} |\hat{Q}_N(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \right) = 0.$$

When point identification holds, the true population parameter $\boldsymbol{\theta}_0$ uniquely maximizes the population objective function and consistency of the GMM estimator follows immediately. ■

Condition (i) requires that the population orthogonality conditions hold exactly at $\boldsymbol{\theta}_0$ only. Condition (ii) requires that bounds for the vector $\boldsymbol{\theta}_0$ are known. This assumption is not uncommon in theoretical work (see, e.g., Arellano and Bond, 1991; Ahn and Schmidt, 1995; Alvarez and Arellano, 2003; Gorgens, Han, and Xue, 2016a) – in practice, though, knowledge about the bounds of the parameter vector is typically not available. Alternatively, consistency can be established without condition (ii) by using Theorem 2.7 in Newey and McFadden (1994): If the objective function is concave, the conditions (ii)-(iv) can be replaced by the condition that $\mathrm{E}[\boldsymbol{\delta}(\cdot)]$ exists and is finite. This is satisfied when all population orthogonality conditions employed in GMM estimation are linear in parameters. To provide an example, consider the second derivative of the objective function in Equation (13) with respect to $\boldsymbol{\theta}$: Under the Assumptions (*A.1.1*)-(*A.1.3*), (*A.2.1*)-(*A.2.4*), and (*A.3.1*)-(*A.3.5*), all entries of the matrix resulting from the matrix product $-2 \cdot \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X}$ are negative, rendering the objective function strictly concave with a unique maximum at $\boldsymbol{\theta}_0$.

For GMM estimators employing nonlinear population orthogonality conditions dropping the compactness condition (ii) complicates matters since the objective function may not be strictly concave (Hayashi, 2000, p.468). In practice, this can be resolved by simply assuming identification (see, e.g., Gorgens, Han, and Xue, 2016a) or by relying on local identification arguments. A sufficient condition for identification in a 'small enough' neighborhood around the true population parameter requires: First, the population orthogonality conditions are continuously differentiable in $\boldsymbol{\theta}$; second, the first derivative of the expectation of the population orthogonality conditions equals the expectation of the derivative of the population orthogonality conditions (i.e., $\partial/\partial\boldsymbol{\theta}\ E[\boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta})] = E[\partial/\partial\boldsymbol{\theta}\ \boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta})]$); third, the rank of the matrix of population orthogonality conditions aggregated by a suitable weighting matrix $\boldsymbol{W}$, $\text{rank}(\boldsymbol{W}\cdot E[\boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta})])$, is equal to the number of parameters that are estimated (see Rothenberg, 1971 and the summary by Newey and McFadden, 1994, p.2127).

Results on consistency of the GMM estimator without the Assumption (*A.3.4*) that the time series dimension is held fixed are derived by Alvarez and Arellano (2003), who consider a GMM estimator based on the linear population orthogonality conditions in Equation (9) and impose the additional assumption of finite moments up to order four for the unobservable idiosyncratic remainder component and the individual-specific effects. Under the rate condition $(\log T)^2/N \to \infty$ for the time series dimension and the cross section dimension, they establish consistency of the GMM estimator based on linear population orthogonality conditions. Hayakawa (2015) investigates the GMM estimator based on the linear population orthogonality conditions in Equations (9) and (11) and establishes consistency of the GMM estimator when N and T are large under the additional assumption of finite moments of the unobservable idiosyncratic remainder components and the individual-specific effects up to order eight. Gorgens, Han, and Xue (2016a) do not impose the 'constant correlated effects' assumption and derive the asymptotic distribution for a GMM estimator based on the linear and nonlinear population orthogonality conditions in Equations (9) and (10).

3.3 Asymptotic normality

When estimating $\boldsymbol{\theta}$ by GMM, asymptotic normality can be established based on a theorem, which is similar to Theorems 3.2 and 3.4 in Newey and McFadden (1994) and Proposition 7.10 in Hayashi (2000).

Theorem 3 *(Asymptotic normality of the GMM estimator)*
Suppose the Assumptions (A.1.1)-(A.1.3), (A.2.1)-(A.2.4), and (A.3.1)-(A.3.5) hold and the GMM estimator $\hat{\boldsymbol{\theta}}$ is identified and consistent. Further assume:

(i) The vector of true population parameters $\boldsymbol{\theta}_0$ is in the interior of Θ.

(ii) The functions $\boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta})$ are continuously differentiable in a neighborhood $\mathscr{N}$ of $\boldsymbol{\theta}_0$, with probability approaching one.

(iii) It holds that $E[\boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta}_0)] = 0$ and $E[||\boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta}_0)||^2] < \infty$.

(iv) It holds in a neighborhood $\mathscr{N}$ of $\boldsymbol{\theta}_0$ that $E[\sup_{\boldsymbol{\theta}\in\mathscr{N}} ||\partial/\partial\boldsymbol{\theta}\ \boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta})||] < \infty$.

(v) The Jacobian matrix of the population orthogonality conditions $\boldsymbol{G} = E[\partial/\partial\boldsymbol{\theta}\ \boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta}_0)]$ is of full column rank.

Then, the GMM estimator $\hat{\boldsymbol{\theta}}$ is asymptotically normal, with

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, (\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{W}\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}), \qquad \textit{where} \tag{16}$$

$$\boldsymbol{\Omega} = \text{E}[\boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta}_0)\cdot\boldsymbol{\delta}((Y,\boldsymbol{Z},\boldsymbol{V}),\boldsymbol{\theta}_0)']. \tag{17}$$

Proof. Calculating the first derivative of the GMM sample objective function in Equation (13) with respect to the parameter vector and dividing by 2 yields the first order conditions

$$\left.\frac{\partial \hat{Q}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\boldsymbol{G}_N(\hat{\boldsymbol{\theta}})'\boldsymbol{W}\boldsymbol{h}_N(\hat{\boldsymbol{\theta}}) \overset{!}{=} \mathbf{0}, \tag{18}$$

where $\boldsymbol{G}_N(\hat{\boldsymbol{\theta}})$ is the Jacobian matrix containing the first derivatives of the sample orthogonality conditions with respect to $\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$. By conditions (i)-(iii), the first order conditions hold with probability approaching one. The sample orthogonality conditions $\boldsymbol{h}_N(\hat{\boldsymbol{\theta}})$ can be expanded around the vector of true population parameters $\boldsymbol{\theta}_0$ with the Mean Value Theorem:

$$\boldsymbol{h}_N(\hat{\boldsymbol{\theta}}) = \boldsymbol{h}_N(\boldsymbol{\theta}_0) + \boldsymbol{G}_N(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \tag{19}$$

where the vector $\bar{\boldsymbol{\theta}}$ contains the element-wise means of the vectors $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ (see Hayashi, 2000, p.470-471 and p.479). Plugging the expansion in Equation (19) into the first order condition in Equation (18), multiplying both sides of the equation with $\sqrt{N}$, and solving for $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ yields

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -(\boldsymbol{G}_N(\hat{\boldsymbol{\theta}})'\boldsymbol{W}\boldsymbol{G}_N(\bar{\boldsymbol{\theta}}))^{-1} \cdot \boldsymbol{G}_N(\hat{\boldsymbol{\theta}})'\boldsymbol{W} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\delta}_i((y_i, \boldsymbol{z}_i, \boldsymbol{v}_i), \boldsymbol{\theta}_0). \tag{20}$$

Based on condition (iv), the first derivatives of the sample orthogonality conditions evaluated at $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$ converge in probability to the first derivative of the population orthogonality conditions evaluated at the vector of true population parameters, $\boldsymbol{G}$ (see Newey and McFadden, 1994, p.2148). Using the Slutzky theorem on the matrix product of all but the last term on the right-hand side of Equation (20) reveals that

$$\operatorname*{plim}_{N\to\infty} -(\boldsymbol{G}_N(\hat{\boldsymbol{\theta}})'\boldsymbol{W}\boldsymbol{G}_N(\bar{\boldsymbol{\theta}}))^{-1} \cdot \boldsymbol{G}_N(\hat{\boldsymbol{\theta}})'\boldsymbol{W} = -(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W},$$

while the last term converges in probability to the true objective function (see Section 3.1 and Newey and McFadden, 1994, p.2126). Taking expectations of the probability limit of Equation (20) yields the zero vector by the Assumptions (*A.1.1*)-(*A.1.3*), (*A.2.1*)-(*A.2.4*), and (*A.3.1*)-(*A.3.5*). Under these assumptions, identification, and consistency of the GMM estimator, a Central Limit theorem can be employed to establish the result in Equation (16), when the true variance covariance matrix of the population orthogonality conditions is defined as $\boldsymbol{\Omega}$. The matrix governs the variance covariance properties of the individual population orthogonality conditions employed in estimation – within and across the types. ■

The proof of this Theorem is similar to the one sketched below Theorems 3.2 and 3.4 in Newey and McFadden (1994) and the text preceding Proposition 7.10 in Hayashi (2000). Condition (i) requires that the vector of true population parameters $\boldsymbol{\theta}_0 \in \Theta$ is not located at the boundary of the set of admissible values Θ. Condition (ii) ensures that the first derivative of the population orthogonality conditions is defined in the neighborhood of the vector of true population parameters. Condition (iii) requires that the population orthogonality conditions hold at the vector of true population parameters and that the squared Euclidean distance of the population orthogonality conditions from zero is finite. Condition (iv) requires that the Euclidean distance of the first derivative of the population orthogonality conditions can be bounded in expectation in a neighborhood $\mathscr{N}$. Condition (v) establishes that the first derivative of the population orthogonality conditions possesses full column rank and ensures invertibility of $(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})$ (see Theorem 4.2.1 in Golub and Van Loan, 2012). When identification and consistency of the GMM estimator can be established, asymptotic normality follows when the conditions (i)-(v) are fulfilled (Newey and McFadden, 1994, p.2148).

A lower bound for the asymptotic variance covariance matrix of the GMM estimator can be derived based on the 'efficiency condition' $\widehat{\boldsymbol{W}}_N \xrightarrow{p} \boldsymbol{\Omega}^{-1}$ (see Hayashi, 2000, p.481). This condition requires that the estimated weighting matrix is a consistent estimator for the true variance covariance matrix of the population orthogonality conditions. In this case the formula for the asymptotic variance covariance matrix of the GMM estimator $\hat{\boldsymbol{\theta}}$ reduces to $(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}$. Alternatively, the formula can also be expressed as $(\boldsymbol{G}'\boldsymbol{\Omega}^{-1}\boldsymbol{G})^{-1}$. Calculating either expression for the asymptotic variance covariance matrix involves replacing $\boldsymbol{G}$ and $\boldsymbol{W}$ or $\boldsymbol{\Omega}^{-1}$ by suitable estimates.

Employing population orthogonality conditions in GMM estimation which are nonlinear in parameters may provide partial identification, where the linear moment conditions do not identify the lag parameter (Gorgens, Han, and Xue, 2016b). However, partial identification means that there are multiple possible solutions and resolving partial identification requires selecting among the different possible solutions. This selection may cause the GMM estimators to have nonstandard asymptotic distributions (Gorgens, Han, and Xue, 2016a). A selection rule to resolve partial identification that does not require using nonstandard asymptotics is provided by Pua, Fritsch, and Schnurbus (2019a) and Pua, Fritsch, and Schnurbus (2019b) for an IV estimator which employs only nonlinear moment conditions. Note that as Gorgens, Han, and Xue (2016a) and Pua, Fritsch, and Schnurbus (2019a) and Pua, Fritsch, and Schnurbus (2019b) investigate versions of Equation (7) without non-lagged dependent explanatory variables, generalizing their results to the functional form given in Equation (7) is a potential topic for future research.

4 Specification testing and inference

4.1 Standard errors

Closely related to establishing identification, consistency, and asymptotic normality for the linear dynamic panel data model are assessing the validity of the model specification and carrying out inference for the vector of population parameters. Both aspects require the computation of standard errors, which are available by taking the square root of the main diagonal of an estimate of the asymptotic variance covariance matrix in Equation (16), where the matrices $\boldsymbol{G}$ and $\boldsymbol{W}$ are replaced by suitable estimates. A readily available estimate for $\boldsymbol{G}$ is the corresponding sample analogue evaluated at the GMM estimate of the parameter vector $\boldsymbol{G}_N(\hat{\boldsymbol{\theta}})$. Different propositions exist in the literature for estimating the weighting matrix $\boldsymbol{W}$ which are directed towards reflecting the 'efficiency condition' mentioned in the preceding section. The suggestions for $\widehat{\boldsymbol{W}}_N$ are typically derived from the model assumptions and the population orthogonality conditions employed in GMM estimation (see, e.g., Arellano and Bond, 1991; Blundell, Bond, and Windmeijer, 2001; Kiviet, 2007) and the number of estimation steps. In one-step GMM estimation, the propositions aim at obtaining an initial consistent estimate of the inverse of the variance covariance matrix of the population orthogonality conditions $\widehat{\mathbf{\Omega}}^{-1}$, while the goal in two-step and multiple-step GMM estimation is to obtain more efficient estimates. Selected suggestions for estimating the weighting matrix $\widehat{\boldsymbol{W}}_N$ are sketched in Fritsch, Pua, and Schnurbus (2019).

Under the conditions stated in Theorem 4.5 in Newey and McFadden (1994), the asymptotic variance covariance matrix given in Equation (16) can be estimated consistently. A similar theorem is given here.

Theorem 4 *Suppose the conditions stated in Theorem 3 are fulfilled and the GMM estimator is identified, consistent, and asymptotically normal. This already implies the following conditions:*

(i) The functions $\boldsymbol{\delta}((Y, \boldsymbol{Z}, \boldsymbol{V}), \boldsymbol{\theta})$ *are continuous at* $\boldsymbol{\theta}_0$ *with probability one.*

(ii) In a neighborhood $\mathscr{N}$ *of* $\boldsymbol{\theta}_0$ *it holds that* $\mathrm{E}[\sup_{\boldsymbol{\theta} \in \mathscr{N}} ||\boldsymbol{\delta}((Y, \boldsymbol{Z}, \boldsymbol{V}), \boldsymbol{\theta})||^2] < \infty$.

Then the asymptotic variance covariance matrix of the GMM estimator $\hat{\boldsymbol{\theta}}$ *can be estimated consistently by*

$$\widehat{Avar(\hat{\boldsymbol{\theta}})} = (\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}})^{-1}\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\mathbf{\Omega}}\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}})^{-1}. \tag{21}$$

Proof. By applying Lemma 4.3 in Newey and McFadden (1994) to $\boldsymbol{\delta}(\cdot) \cdot \boldsymbol{\delta}(\cdot)'$, it can be established that $\mathrm{plim}_{N\to\infty} \widehat{\mathbf{\Omega}} = \mathbf{\Omega}$. Employing the result that $\boldsymbol{G}_N(\hat{\boldsymbol{\theta}})$ converges in probability to $\boldsymbol{G}$ as the sample size tends to infinity from Theorem 3 and the corresponding proof, Assumption (*A.3.2*) that $\mathrm{plim}_{N\to\infty} \widehat{\boldsymbol{W}}_N = \boldsymbol{W}$, and the Slutzky theorem it follows that

$$\operatorname*{plim}_{N\to\infty} (\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}})^{-1}\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\mathbf{\Omega}}\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}})^{-1} = (\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\mathbf{\Omega}\boldsymbol{W}\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}.$$

■

In finite samples, there may be pronounced differences between the asymptotic standard errors and the standard errors obtained by two- or multiple-step GMM estimation of the model parameters (Arellano and Bond, 1991). The reason for this is that the estimated variance covariance matrix of the population orthogonality conditions depends on estimated parameters which may lead to a substantial down-ward bias in the standard errors. An indication of a down-ward bias of the standard errors is when the standard errors obtained from one-step GMM estimation are substantially higher than standard errors obtained from two- or multiple-step GMM estimation. Finite sample corrections are available to adjust the standard errors of the parameter estimates when only linear moment conditions are employed in two-step estimation (Windmeijer, 2005) and multiple-step estimation (Windmeijer, 2000). An alternative to reporting corrected standard errors proposed by Arellano and Bond (1991) is to report the one-step standard errors instead of the two- or multiple-step standard errors.

The derivation of the correction of the standard errors by Windmeijer (2005) considers GMM estimation of linear dynamic panel data models employing only linear population orthogonality conditions. Nevertheless, the implementation of GMM estimation of linear dynamic panel data models based on nonlinear population orthogonality conditions available in standard software (Kripfganz, 2018) also utilizes the correction to adjust the standard errors. According to Windmeijer (2005), it is not clear wether the correction improves the estimation of the standard errors in this setting. Comparing the corrected two-step standard errors with one-step and bootstrapped alternatives for selected DGPs may be of practical relevance and is a possible topic for future research.

4.2 Specification testing

Different specification tests are available to assess the validity of the model specification after GMM estimation of the linear dynamic panel data model in Equation (7). The idea behind specification tests is to empirically check the assumptions imposed in estimation or their implications and derive a test statistic for a pre-specified confidence level under a distinct null hypothesis. Test results may then cast doubt on some of the assumptions used in estimation. The specification tests outlined in this section are based on a GMM estimator which is asymptotically normal.

Serial correlation tests A specification test which is often employed in practice are the m_2-statistics proposed by Arellano and Bond (1991) and generalized to higher orders j in Arellano (2003). The underlying idea is to check the implication of Assumption (*A.2.2*) that there should be no serial correlation above order one in the first differences of the unobservable idiosyncratic remainder components. First order serial correlation is introduced by first differencing Equation (7) to eliminate the unobservable individual-specific effects (for a derivation see Section 6.4 of the Appendix). The m_j-statistics are based on considering the suitably scaled average j-th order autocovariance of the unobservable idiosyncratic remainder components.

$$T_{m_j} = \frac{\hat{r}_j}{\hat{\sigma}_{\hat{r}_j}}, \quad \text{with} \quad T_{m_j} \overset{a}{\sim} \mathcal{N}(0,1), \tag{22}$$

where $\hat{r}_j$ is the estimated average j-th order autocovariance of the residuals and $\hat{\sigma}_{\hat{r}_j}$ is the corresponding standard error. For the linear dynamic panel data model in Equation (7), the former term can be obtained from

$$r_j = \frac{1}{T-3-j} \cdot \sum_{t=4+j}^{T} r_{t,j}, \quad \text{with} \quad r_{t,j} = E(\Delta s_{i,t} \Delta s_{i,t-j}),$$
$$\text{where} \quad \hat{r}_{t,j} = \frac{1}{\sqrt{N}} \cdot \hat{\boldsymbol{s}}_t' \, \hat{\boldsymbol{s}}_{t-j}.$$

The $\hat{r}_{t,j}$ are a scaled version of the estimated average j-th order autocovariance of the residuals (see Arellano and Bond, 1991; Arellano, 2003; Doornik, Arellano, and Bond, 2012) and the column vectors $\hat{\boldsymbol{s}}_t$ and $\hat{\boldsymbol{s}}_{t-j}$ contain the residuals obtained from GMM estimation. Note that the dependence on the estimation step is suppressed in this paragraph to keep the exposition general. Plugging in one-, two-, or multiple-step estimation results is straightforward. The standard error of the estimated j-th order autocovariance of the residuals is available by taking the square root of

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

where $\widehat{\mathbf{\Omega}}(\hat{\boldsymbol{s}})$ is the estimated variance covariance matrix of the residuals and $\widehat{\mathbf{\Omega}}(\hat{\boldsymbol{\theta}})$ is the corresponding matrix for the estimated parameters. The test statistic is distributed asymptotically standard normal under the null hypothesis that there is no serial correlation of a pre-specified order in the first differenced unobservable idiosyncratic remainder component (see Arellano and Bond, 1991, and Arellano, 2003, p.121-123). A large value of the test statistic leads to a rejection of the null hypothesis which may be interpreted as an indication that the model is misspecified.

An alternative test which allows to test for serial correlation in the first differenced idiosyncratic remainder component of order 2 up to a particular order jointly is mentioned in Arellano (2003, p.122) and proposed by Yamagata (2008).

Overidentifying restrictions tests A further specification test for GMM estimation of linear dynamic panel data models are the overidentifying restrictions tests proposed by Sargan (1958) and Hansen (1982), where the former Sargan test employs the one-step weighting matrix and residuals and the latter J-test uses their two-step counterparts. The tests originate from the fact, that there are typically substantially more population orthogonality conditions K than there are parameters to be estimated J, which allows to assess the validity of these $K - J$ overidentifying conditions. The null hypothesis is that the overidentifying restrictions are valid.

$$T_S = n \cdot \hat{\boldsymbol{s}}_1' \boldsymbol{Z} \widehat{\boldsymbol{W}}_1 \boldsymbol{Z}' \hat{\boldsymbol{s}}_1 \cdot \hat{\sigma}_1^{-2} \tag{23}$$

represents the test statistic of the Sargan test (see Sargan, 1958; Arellano and Bond, 1991; Doornik, Arellano, and Bond, 2012). When the GMM estimator $\hat{\boldsymbol{\theta}}$ is distributed asymptotically normal, when the idiosyncratic remainder components exhibit conditional homoscedasticity, and under the null, the test statistic is asymptotically χ^2-distributed with $K - J$ degrees of freedom (see Hayashi, 2000, p.227-228).

The test statistic corresponding to the $J-$test (see Hansen, 1982; Arellano and Bond, 1991; Doornik, Arellano, and Bond, 2012) is

$$T_H = n \cdot \hat{\boldsymbol{s}}_2' \boldsymbol{Z} \widehat{\boldsymbol{W}}_2 \boldsymbol{Z}' \hat{\boldsymbol{s}}_2. \tag{24}$$

Under the null, the test statistic is asymptotically χ^2-distributed with $K - J$ degrees of freedom when the GMM estimator $\hat{\boldsymbol{\theta}}$ exhibits asymptotic normality and when the conditions (i) and (ii) of Theorem 4 hold.

A rejection of the null hypothesis in the two overidentifying restriction tests casts doubt on the instrument set employed in estimation and/or the model assumptions. Two comments on the overidentifying restrictions tests are in order: First, both tests may have low power when the number of overidentifying restrictions is large (Bowsher, 2002; Windmeijer, 2005); Second, the Sargan test statistic is inconsistent when heteroscedasticity is present (see Roodman, 2009a).

Alternatively, certain subsets of overidentifying restrictions can be tested by computing the test statistic for each set of population orthogonality conditions K_1 and K_2 ($K_1 > K_2$ and $K_2 > J$) and differencing the test statistics. The difference is then asymptotically χ^2-distributed with $K_1 - K_2$ degrees of freedom. This latter version of the tests is also referred to as 'difference-in-Hansen'/'difference-in-Sargan' test (Roodman, 2009), 'incremental Hansen'/'incremental Sargan' test (Arellano, 2003), or C-statistic (Hayashi, 2000).

Underidentification tests A related type of specification test are tests for underidentification. In the underidentification test proposed by Arellano, Hansen, and Sentana (2012), the structural form is augmented and the restrictions obtained from augmenting the model are then tested with conventional overidentifying restrictions tests. When these overidentifying restrictions are not rejected, this is interpreted as an indication that the model parameters are not identified. Arellano, Hansen, and Sentana (2012) sketch two approaches to carry out the test. The first is based on evaluating parameters over a discrete grid of points, while the second involves fitting a spline. The details and the implementation of both approaches are not discussed in greater detail here, as Windmeijer (2017) proposes an alternative testing procedure for linear instrumental variable models which is straightforward to implement and simple to compute with the methodology already outlined in this paper. For GMM estimation of linear dynamic panel data models, the underidentification test of Windmeijer (2017) basically involves two GMM estimations. First, Equation (7) is estimated by GMM. Subsequently, a model is estimated by GMM in which one of the instruments employed in the first estimation is used as the dependent variable and all other instruments are used as explanatory variables. The J-test statistic for the second estimation represents the test statistic of the underidentification test. The null hypothesis of the test is that the model is underidentified and a rejection of the null hypothesis is interpreted as an indication that underidentification of the model parameters may not be an issue.

Tests for structural breaks Another type of specification test are tests for structural breaks. In linear dynamic panel data models as stated in Equation (7) the assumption is imposed that the unobservable individual-specific effects and the coefficients attributable to the explanatory variables remain constant over time (and individuals). When the slope coefficient and/or the individual-specific effect varies over time, this may be referred to as a structural break. The presence of a structural break is an indication that the linear dynamic panel data model is misspecified. A test to detect a single structural breakpoint is proposed by De Wachter and Tzavalis (2012), where the null hypothesis is that there is no breakpoint. Depending on if the breakpoint is known or unknown, two different versions of the test exist with different asymptotic distributions of the test statistic. Compared to the assumptions imposed in this paper, De Wachter and Tzavalis (2012) do not explicitly impose the 'constant correlated effects' assumption and exclude the unit root case (i.e., $|\rho| < 1$) in their derivation of the test statistic.

Further specification tests Other conditions besides the ones addressed by the previously mentioned tests also deserve attention in GMM estimation of the linear dynamic panel data model in Equation (7) and may be checked by specification tests. Among these are: (i) checking the positive definiteness of the weighting matrix (an empirical check may involve a singular value decomposition of $\widehat{\boldsymbol{W}}_N$); (ii) test for the presence of unobserved heterogeneity (see, e.g., Arellano, 2003, p.124-125 and Harris, Mátyás, and Sevestre, 2008); (iii) assessing the validity of the 'constant correlated effects' assumption (and deviations thereof), as the assumption ensures that all population orthogonality conditions employed in GMM estimation are linear in parameters (Blundell, Bond, and Windmeijer, 2001). The assumption of 'constant correlated effects' may be tested by a 'difference-in-Hansen' or a 'difference-in-Sargan' test (see, e.g., Arellano, 2003, p.123-124 and

Bun and Sarafidis, 2015), where the unrestricted model is based on the population orthogonality conditions in Equations (9) and (11) and the restricted model is based on those in Equation (9) only. Extending the specification tests available for linear dynamic panel data models and investigating tests in either of these directions which are straightforward to implement and easy to compute may constitute other worthwhile future research endeavours.

4.3 General hypothesis testing

The discussion on hypothesis testing in this section is kept brief and treats the Wald test only. Complementary overviews which also treat Likelihood Ratio and Lagrange Multiplier tests are provided in Newey and McFadden (1994, p.2217-2226), Hayashi (2000, p.211-214 and 487-495), and Davidson and MacKinnon (1993, p.617-619). Testing general and potentially nonlinear hypotheses about the population parameter vector after GMM estimation of the linear dynamic panel data model denoted by Equation (7) employs that the GMM estimator is distributed asymptotically normal and that the variance covariance matrix of $\hat{\boldsymbol{\theta}}$ can be estimated consistently. These properties can be established by the Assumptions (*A.1.1*)-(*A.1.3*), (*A.2.1*)-(*A.2.4*), (*A.3.1*)-(*A.3.5*), and the conditions stated in Theorem 3. A general null hypothesis and a corresponding alternative are

$$\begin{aligned} H_0 &: \ \boldsymbol{R}(\boldsymbol{\theta}_0) = \mathbf{0}, \\ H_1 &: \ \boldsymbol{R}(\boldsymbol{\theta}_0) \neq \mathbf{0}, \end{aligned} \tag{25}$$

where $\boldsymbol{R}(\boldsymbol{\theta}_0)$ are functions expressing the C restrictions imposed on the vector of true population parameters and $\mathbf{0}$ is a $C \times 1$ zero vector. The first derivative of the functions $\boldsymbol{R}(\boldsymbol{\theta}_0)$ with respect to $\boldsymbol{\theta}$ is required to possess a rank of C, which means that the hypotheses contain no redundant restrictions (see Newey and McFadden, 1994, p.2217-2218 and Hayashi, 2000, p.487-488).

The null hypothesis can be tested against the alternative with the Wald test. Hayashi (2000) derives the corresponding test statistic based on rewriting the mean value expansion of $N\boldsymbol{R}(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$ as a first-order Taylor series expansion and combines this expression with the first order Taylor series expansion of $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Additionally utilizing the requirement of asymptotic normality of the GMM estimator $\hat{\boldsymbol{\theta}}$ and replacing all terms referring to the population with their sample analogues yields the following expression for the Wald statistic

$$T_W = N\boldsymbol{R}(\hat{\boldsymbol{\theta}})' \left(\frac{\partial \boldsymbol{R}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \ (\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}})^{-1}\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{G}})^{-1} \ \frac{\partial \boldsymbol{R}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}' \right)^{-1} \boldsymbol{R}(\hat{\boldsymbol{\theta}}), \tag{26}$$

which – under the 'efficiency condition' – may be simplified to

$$N\boldsymbol{R}(\hat{\boldsymbol{\theta}})' \left(\frac{\partial \boldsymbol{R}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \ \widehat{\boldsymbol{G}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{G}} \ \frac{\partial \boldsymbol{R}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \right)^{-1} \boldsymbol{R}(\hat{\boldsymbol{\theta}}). \tag{27}$$

Under the null hypothesis stated in Equation (25), the test statistic in Equations (26) and the simplified version in Equation (27) can be shown to be asymptotically χ^2-distributed with C degrees of freedom under the null when the GMM estimator $\hat{\boldsymbol{\theta}}$ exhibits asymptotic normality and when a consistent estimator for $\hat{\boldsymbol{\Omega}}$ is available (see Hayashi, 2000, p.481 and 489-491).

Note that the derivation of the asymptotic distribution of the Wald test statistic depends on the assumption of asymptotic normality of the GMM estimator. When nonlinear population orthogonality conditions are employed in GMM estimation, the asymptotic distribution of the estimator may be affected. Gorgens, Han, and Xue (2016a) characterize the conditions under which standard and non-standard asymptotic results are available for a GMM estimator employing nonlinear moment conditions when the only explanatory variable is one lag of the dependent variable. It may be an interesting topic for future research to revisit the derivations and investigate if the conditions require adjustments or can be weakened in the presence of additional non-lagged dependent explanatory variables.

5 Concluding remarks

This paper illustrates the assumptions involved in GMM estimation of the model parameters of linear dynamic panel data models. The standard assumptions frequently used in the literature are outlined and their practical implications are discussed in the context of existing Lemmas and Theorems which establish identification, consistency, and asymptotic normality of the GMM estimator. Some particularities when linear and nonlinear moment conditions are employed in estimation are highlighted. Furthermore, the discussion is connected to different propositions for testing the modeling assumptions and testing of general (potentially nonlinear) hypothesis.

Areas for future research may involve reconsidering the identifying conditions stated in Gorgens, Han, and Xue (2016b) and the asymptotic distributions characterized in Gorgens, Han, and Xue (2016a) for the GMM estimator when linear dynamic panel data models are estimated based on linear and nonlinear population orthogonality conditions. One interesting question is if additional assumptions need to be imposed and/or if assumptions can be weakened when explanatory variables besides the lagged dependent variables are included in the model. Assessing the validity and the empirical performance of the finite sample correction proposed by Windmeijer (2005) in the presence of nonlinear moment conditions and work on specification tests to assess (e.g.,) the 'constant correlated effects' assumption are further potential topics for future research.

Other areas, besides the ones mentioned in this paper, could involve investigating the small sample properties of GMM estimators by Monte Carlo simulation to assess the impact of key modeling decisions such as the choice of weighting matrix, the power of selected specification tests to detect deviations from the model assumptions, and the performance of estimators which carry out a selection of instruments. Revisiting existing areas of application to assess the robustness of the conclusions derived from the modeling task may be another potential topic. Attempting to connect the literature on the linear dynamic panel data model with nonseparable models could also be a worthwhile endeavour.

References

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

Ahn, SC and P Schmidt (1997). "Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation". In: *Journal of Econometrics* 76.1–2, pp. 309–321.

Alvarez, J and M Arellano (2003). "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators". In: *Econometrica* 71.4, pp. 1121–1159.

Amemiya, T (1985). *Advanced Econometrics*. Harvard University Press.

Anderson, T and C Hsiao (1981). "Estimation of Dynamic Models with Error Components". In: *Journal of the American Statistical Association* 76.375, pp. 598–606.

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, LP Hansen, and E Sentana (2012). "Underidentification?" In: *Journal of Econometrics* 170.2, pp. 256–280.

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.

Bai, J (2013). "Fixed-Effects Dynamic Panel Models, a Factor Analytical Method". In: *Econometrica* 81.1, pp. 285–314.

Banks, J, R Blundell, and A Lewbel (1997). "Quadratic Engel Curves and Consumer Demand". In: *The review of economics and statistics* 79.4, pp. 527–539.

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.

Blundell, R and S Bond (2000). "GMM Estimation with persistent panel data: an application to production functions". In: *Econometric Reviews* 19.3, pp. 321–340.

Bond, SR (2002). "Dynamic panel data models: a guide to micro data methods and practice". In: *Portuguese Economic Journal* 1.2, pp. 141–162.

Bonhomme, S and E Manresa (2015). "Grouped Patterns of Heterogeneity in Panel Data". In: *Econometrica* 83.3, pp. 1147–1184.

Bowsher, CG (2002). "On testing overidentifying restrictions in dynamic panel data models". In: *Economics Letters* 77.2, pp. 211–220.

Browning, M and MD Collado (2007). "Habits and heterogeneity in demands: a panel data analysis". In: *Journal of Applied Econometrics* 22.3, pp. 625–640.

Bun, MJG and JF Kiviet (2006). "The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models". In: *Journal of Econometrics* 132.2, pp. 409–444.

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.

Bun, MJ and MA Carree (2005). "Bias-Corrected Estimation in Dynamic Panel Data Models". In: *Journal of Business & Economic Statistics* 23.2, pp. 200–210.

Cavalcanti, TVdV, K Mohaddes, and M Raissi (2015). "Commodity Price Volatility and the Sources of Growth". In: *Journal of Applied Econometrics* 30.6, pp. 857–873.

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

Davidson, R and JG MacKinnon (1993). *Estimation and Inference in Econometrics.* Oxford University Press.

De Wachter, S and E Tzavalis (2012). "Detection of structural breaks in linear dynamic panel data models". In: *Computational Statistics & Data Analysis* 56.11, pp. 3020–3034.

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`.

Faulkender, M, MJ Flannery, KW Hankins, and JM Smith (2012). "Cash flows and leverage adjustments". In: *Journal of Financial Economics* 103.3, pp. 632–646.

Flannery, MJ and K Watson Hankins (2013). "Estimating dynamic panel models in corporate finance". In: *Journal of Corporate Finance* 19, pp. 1–19.

Fritsch, M, A Pua, and J Schnurbus (2019). *pdynmc – An R-package for estimating linear dynamic panel data models based on nonlinear moment conditions.* University of Passau Working Papers in Business Administration B-39-19.

Golub, GH and CF Van Loan (2012). *Matrix Computations.* Vol. 4. John Hopkins University Press.

Gorgens, T, C Han, and S Xue (2016a). *Asymptotic distributions of the quadratic GMM estimator in linear dynamic panel data models.* ANU Working Papers in Economics and Econometrics 2016-635. Australian National University, College of Business and Economics, School of Economics.

Gorgens, T, C Han, and S Xue (2016b). *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.

Hahn, J, J Hausman, and G Kuersteiner (2007). "Long difference instrumental variables estimation for dynamic panel models with fixed effects". In: *Journal of Econometrics* 140.2, pp. 574–617.

Hall, AR (2015). "Econometricians Have Their Moments: GMM at 32". In: *Economic Record* 91.S1, pp. 1–24.

Han, C and PCB Phillips (2010). "GMM Estimation For Dynamic Panels With Fixed Effects And Strong Instruments At Unity". In: *Econometric Theory* 26.1, pp. 119–151.

Han, C, PC Phillips, and D Sul (2014). "X-Differencing And Dynamic Panel Model Estimation". In: *Econometric Theory* 30.1, pp. 201–251.

Hansen, LP (1982). "Large Sample Properties of Generalized Method of Moments Estimators". In: *Econometrica* 50.4, pp. 1029–1054.

Harris, MN, L Mátyás, and P Sevestre (2008). "Chapter 8 – Dynamic Models For Short Panels". In: *The Econometrics of Panel Data.* Ed. by L Mátyás and P Sevestre. Springer, pp. 249–278.

Hayakawa, K (2009). "On the effect of mean-nonstationarity in dynamic panel data models". In: *Journal of Econometrics* 153.2, pp. 133–135.

Hayakawa, K (2015). "The Asymptotic Properties Of The System GMM Estimator In Dyanmic Panel Data Models When Both N And T Are Large". In: *Econometric Theory* 31.3, pp. 647–667.

Hayakawa, K and MH Pesaran (2015). "Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity". In: *Journal of Econometrics* 188.1, pp. 111–134.

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.

Hsiao, C, MH Pesaran, and AK Tahmiscioglu (2002). "Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods". In: *Journal of Econometrics* 109.1, pp. 107–150.

Kiviet, JF (1995). "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models". In: *Journal of Econometrics* 68.1, pp. 53–78.

Kiviet, JF (2007). *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.

Kiviet, JF, M Pleus, and R Poldermans (2017). "Accuracy and Efficiency of Various GMM Inference Techniques in Dynamic Micro Panel Data Models". In: *Econometrics* 5.1, p. 14.

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`.

Lemmon, ML, MR Roberts, and JF Zender (2008). "Back to the Beginning: Persistence and the Cross-Section of Corporate Capital Structure". In: *The Journal of Finance* 63.4, pp. 1575–1608.

Lewbel, A (2019). "The Identification Zoo - Meanings of Identification in Econometrics". In: *Journal of Economic Literature*, to appear.

Michaud, PC and A Van Soest (2008). "Health and wealth of elderly couples: Causality tests using dynamic panel data models". In: *Journal of Health Economics* 27.5, pp. 1312–1325.

Moral-Benito, E (2013). "Likelihood-Based Estimation of Dynamic Panels With Predetermined Regressors". In: *Journal of Business & Economic Statistics* 31.4, pp. 451–472.

Nason, JM and GW Smith (2008). "Identifying the new Keynesian Phillips curve". In: *Journal of Applied Econometrics* 23.5, pp. 525–551.

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.

Nickell, SJ (1996). "Competition and Corporate Performance". In: *Journal of Political Economy* 104.4, pp. 724–746.

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.

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.

Rothenberg, TJ (1971). "Identification in Parametric Models". In: *Econometrica* 39.3, pp. 577–591.

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

Windmeijer, F (2000). *A Finite Sample Correction For The Variance Of Linear Two-Step GMM Estimators.* IFS Working Paper 00/19. Institute for Fiscal Studies.

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.

Windmeijer, F (2017). *A simple underidentification test for linear IV models, with an application to dynamic panel data models.* Working Paper.

Wooldridge, JM (2010). *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Yamagata, T (2008). "A joint serial correlation test for linear panel data models". In: *Journal of Econometrics* 146.1, pp. 135–145.

6 Appendix

6.1 Initial conditions restrictions to ensure 'constant correlated effects'

The 'constant correlated effects' assumption can be assured in two different, but related ways. The first approach amounts to specifying the processes of the dependent variable and the regressors, imposing random starting values and assuming that both processes run for a sufficiently long period of time prior to sampling. This results in the impact of the initial conditions to disappear. The second approach involves making an explicit assumption about the initial conditions. The following exposition is similar to Blundell, Bond, and Windmeijer (2001) and Bun and Sarafidis (2015) and illustrates the two alternatives. To begin with, consider the $y_{i,t}$-process with one further explanatory variable x.

$$
\begin{aligned}
y_{i,t} &= y_{i,t-1}\rho + x_{i,t}\beta + a_i + \varepsilon_{i,t}, \qquad i = 1, \ldots, N;\ t = 2, \ldots, T, \qquad \text{where} \\
x_{i,t} &= \gamma x_{i,t-1} + \tau a_i + e_{i,t} \\
e_{i,t} &= \zeta_1 \varepsilon_{i,t} + \zeta_2 \varepsilon_{i,t-1} + \epsilon_{i,t}, \qquad \text{with} \qquad \epsilon_{i,t} \sim (0, \sigma_\epsilon^2).
\end{aligned}
$$

The term $\epsilon_{i,t}$ is an unobservable idiosyncratic remainder component, while the parameter τ governs the correlation between the individual-specific effect a_i and $x_{i,t}$. The parameters ζ_1 and ζ_2 determine the covariance between the unobservable idiosyncratic remainder component $\varepsilon_{i,t}$ and the unobservable composite error term $e_{i,t}$ of the $x_{i,t}$-process; this causes the regressor $x_{i,t}$ to be strictly exogenous ($\zeta_1 = \zeta_2 = 0$; i.e., past, present, and future $x_{i,t}$ are orthogonal to $\varepsilon_{i,t}$), predetermined ($\zeta_1 = 0$ and $\zeta_2 \neq 0$; i.e., past and present $x_{i,t}$ are orthogonal to $\varepsilon_{i,t}$) or endogenous ($\zeta_1 \neq 0$ and $\zeta_2 = 0$; i.e., past $x_{i,t}$ are orthogonal to $\varepsilon_{i,t}$).

Considering a strictly exogenous regressor to simplify the notation and enhance the clarity of exposition and writing down the $x_{i,t}$-process starting at the initial period contained in the sample $x_{i,1}$ to p periods in the past yields

$$
\begin{aligned}
x_{i,1} &= \gamma x_{i,0} + \tau a_i + e_{i,1}, \\
x_{i,0} &= \gamma x_{i,-1} + \tau a_i + e_{i,0}, \\
&\vdots \\
x_{i,-p+1} &= \gamma x_{i,-p} + \tau a_i + e_{i,-p+1}.
\end{aligned}
$$

Rewriting the $x_{i,t}$-process by repeatedly plugging the equations for earlier time periods into the equations for later time periods (i.e., replacing $x_{i,-p+2}$ by the corresponding equation in the equation for $x_{i,-p+1}$,

$x_{i,-p+3}$ by the equation for $x_{i,-p+2}$, etc.) results in

$$\begin{aligned}
& x_{i,1} = \gamma^{p+1} x_{i,-p} + \tau a_i (1 + \gamma + \cdots + \gamma^p) + e_{i,1} + \gamma e_{i,0} + \cdots + \gamma^p e_{i,-p+1} \\
\Leftrightarrow\ & x_{i,1} = \gamma^{p+1} x_{i,-p} + \sum_{s=0}^{p} \gamma^s (\tau a_i + e_{i,1-s}) \\
\Leftrightarrow\ & x_{i,1} = \gamma^{p+1} x_{i,-p} + \tau a_i \cdot \frac{1-\gamma^{p+1}}{1-\gamma} + \sum_{s=0}^{p} \gamma^s e_{i,1-s} \\
\Leftrightarrow\ & x_{i,1} = \gamma^{p+1} \left(x_{i,-p} - \tau a_i \cdot \frac{1}{1-\gamma} \right) + \tau a_i \cdot \frac{1}{1-\gamma} + \sum_{s=0}^{p} \gamma^s e_{i,1-s}.
\end{aligned}$$

Assuming $|x_{i,-p}| < \infty$, $|\gamma| < 1$, and letting $p \to \infty$, the expression reduces to

$$x_{i,1} = \frac{\tau a_i}{1-\gamma} + \underbrace{\sum_{s=0}^{p} \gamma^s e_{i,1-s}}_{\nu_{i,1}}, \qquad \text{with} \qquad \mathrm{E}[\nu_{i,1}] = 0.$$

This implies that the process does not depend on its starting value $x_{i,-p}$. The 'constant correlated effects' assumption is fulfilled for the $x_{i,t}$-process under the initial condition:

$$\mathrm{E}\left[\left(x_{i,1} - \frac{\tau a_i}{1-\gamma} \right) \cdot \frac{\tau a_i}{1-\gamma} \right] = 0 \qquad \Leftrightarrow \qquad \mathrm{E}[\nu_{i,1} \cdot \frac{\tau a_i}{1-\gamma}] = 0.$$

As noted by Bun and Sarafidis (2015), this requires that the $x_{i,t}$-process is on its long term path from the first period contained in the sample on (i.e., $\mathrm{E}[x_{i,t}] = \mathrm{E}[x_{i,t-1}] = \cdots = \mathrm{E}[x_{i,1}]$ holds) and that any deviations of the process from its long term path are not systematic over time or individuals. For the $x_{i,t}$-process, the initial conditions restriction is fulfilled, when ν_i and the individual-specific effect a_i are orthogonal.

A similar restriction can be derived for the initial conditions of the $y_{i,t}$-process starting from the representation

$$\begin{aligned}
y_{i,1} &= y_{i,0}\rho + x_{i,1}\beta + a_i + \varepsilon_{i,1}, \\
y_{i,0} &= y_{i,-1}\rho + x_{i,0}\beta + a_i + \varepsilon_{i,0}, \\
&\vdots \\
y_{i,-p+1} &= y_{i,-p}\rho + x_{i,-p+1}\beta + a_i + \varepsilon_{i,-p+1}.
\end{aligned}$$

Again, plugging the process for earlier time periods into the process for later time periods yields

$$\begin{aligned}
y_{i,1} = {} & y_{i,-p}\rho^{p+1} + x_{i,1}\beta + \cdots + \rho^p x_{i,-p+1}\beta + a_i(1 + \cdots + \rho^p) + \\
& \varepsilon_{i,1} + \varepsilon_{i,0}\rho + \cdots + \varepsilon_{i,-p+1}\rho^p \\
\Leftrightarrow y_{i,1} = {} & \left(y_{i,-p}\rho^{p+1} - a_i \cdot \frac{\rho^{p+1}}{1-\rho} \right) + \beta \cdot \sum_{s=0}^{p} x_{i,1-s}\rho^s + a_i \cdot \frac{1}{1-\rho} + \sum_{s=0}^{p} \varepsilon_{i,1-s}\rho^s.
\end{aligned}$$

By utilizing the recursion of the $x_{i,t}$-process, the expression for $y_{i,1}$ can be rewritten as

$$\begin{aligned}
y_{i,1} &= \rho^{p+1}\left(y_{i,-p} - \frac{a_i}{1-\rho}\right) + \beta \cdot \sum_{s=0}^{p} \rho^s \left(\gamma^{p-s+1}\left(x_{i,-p} - \frac{\tau a_i}{1-\gamma}\right) + \frac{\tau a_i}{1-\gamma} + \right.\\
&\left. \sum_{j=0}^{p-s} \gamma^{p-j} e_{i,-p-s+j+1}\right) + \frac{a_i}{1-\rho} + \sum_{s=0}^{p} \rho^s \varepsilon_{i,1-s}\\
\Leftrightarrow y_{i,1} &= \rho^{p+1}\left(y_{i,-p} - \frac{1-\gamma+\tau\beta}{(1-\rho)(1-\gamma)} \cdot a_i\right) + \beta \cdot \sum_{s=0}^{p} \rho^s \gamma^{p-s+1}\left(x_{i,-p} - \frac{\tau a_i}{1-\gamma}\right) + \\
&\frac{1-\gamma+\tau\beta}{(1-\rho)(1-\gamma)} \cdot a_i + \beta \cdot \sum_{s=0}^{p} \rho^s \cdot \sum_{j=0}^{p-s} \gamma^{p-j} e_{i,-p-s+j+1} + \sum_{s=0}^{p} \rho^s \varepsilon_{i,1-s}
\end{aligned}$$

Assuming $|y_{i,-p}| < \infty$ and $|x_{i,t}| < \infty$, imposing $|\rho| < 1$ and $|\gamma| < 1$, and letting $p \to \infty$, the equation for $y_{i,1}$ reduces to

$$y_{i,1} = \frac{1-\gamma+\beta\tau}{(1-\rho)(1-\gamma)} \cdot a_i + \omega_{i,1}, \qquad \text{with}$$

$$\omega_{i,1} = \beta \cdot \sum_{s=0}^{p} \rho^s \left(\sum_{j=0}^{p-s} \gamma^{p-j} e_{i,-p-s+j+1}\right) + \sum_{s=0}^{p} \rho^s \varepsilon_{i,1-s}, \qquad \text{where} \qquad \mathrm{E}[\omega_{i,1}] = 0.$$

Similar to the $x_{i,t}$-process, the $y_{i,t}$-process does not depend on its starting value. From the equation, it is obvious that the 'constant correlated effects' assumption hinges on the following initial conditions restriction for the $y_{i,t}$-process:

$$\begin{aligned}
&\mathrm{E}\left[\left(y_{i,1} - \frac{1-\gamma+\beta\tau}{(1-\rho)(1-\gamma)} \cdot a_i\right) \cdot \frac{1-\gamma+\beta\tau}{(1-\rho)(1-\gamma)} \cdot a_i\right] = 0\\
\Leftrightarrow &\mathrm{E}\left[\omega_{i,1} \cdot \frac{1-\gamma+\beta\tau}{(1-\rho)(1-\gamma)} \cdot a_i\right] = 0.
\end{aligned}$$

This requires that the $y_{i,t}$-process is also on its long term path from the initial observation contained in the sample on and, hence, that $\mathrm{E}[y_{i,t}] = \mathrm{E}[y_{i,t-1}] = \cdots = \mathrm{E}[y_{i,1}]$ holds. The initial conditions restriction holds, when $\omega_{i,1}$ and the individual-specific effect a_i are orthogonal.

In practice, the processes generating $y_{i,t}$ and $x_{i,t}$ are typically unknown and investigating the properties of the processes is not feasible. An alternative is to explicitly impose the 'constant correlated effects' assumption and assess its plausibility based on economic considerations (examples for the implications are given in Blundell, Bond, and Windmeijer, 2001; Arellano, 2003; Bun and Sarafidis, 2015).

6.2 Redundance of population orthogonality conditions

In principle, more Arellano and Bover (1995)-type linear population orthogonality conditions are available than stated in Equation (11) when imposing 'constant correlated effects' on top of the Assumptions (*A.1.1*)-(*A.1.3*) and (*A.2.1*)-(*A.2.2*):

$$E[\Delta y_{i,v} \cdot u_{i,t}] = 0, \qquad t = 3, \ldots, T; v = 2, \ldots, t-1.$$

For the time periods $t = 3, 4, 5$, the set of available Arellano and Bover (1995) population orthogonality conditions can be represented by:

$$
\begin{array}{ccc}
t=3 & t=4 & t=5 \\
E[\Delta y_{i,2}\cdot u_{i,3}]=0 & E[\Delta y_{i,3}\cdot u_{i,4}]=0 & E[\Delta y_{i,4}\cdot u_{i,5}]=0 \\
 & E[\Delta y_{i,2}\cdot u_{i,4}]=0 & E[\Delta y_{i,3}\cdot u_{i,5}]=0 \\
 & (E[\Delta y_{i,2}\cdot u_{i,3}]=0) & (E[\Delta y_{i,3}\cdot u_{i,4}]=0) \\
 & & E[\Delta y_{i,2}\cdot u_{i,5}]=0 \\
 & & (E[\Delta y_{i,2}\cdot u_{i,4}]=0)
\end{array}
$$

All redundant population orthogonality conditions for the different time periods are stated in parentheses. Employing the Holtz-Eakin, Newey, and Rosen (1988) population orthogonality conditions

$$E[y_{i,s}\cdot \Delta u_{i,t}] = \;0, \qquad i=1,\dots,N; t=3,\dots,T; s=1,\dots,t-2,$$

which are available from the assumptions stated above anyway, the following population orthogonality conditions for the time periods $t=3,4,5$ result:

$$
\begin{array}{ccc}
t=3 & t=4 & \text{t=5} \\
E[y_{i,1}\cdot \Delta u_{i,3}]=0 & E[y_{i,1}\cdot \Delta u_{i,4}]=0 & E[y_{i,1}\cdot \Delta u_{i,5}]=0 \\
 & E[y_{i,2}\cdot \Delta u_{i,4}]=0 & E[y_{i,2}\cdot \Delta u_{i,5}]=0 \\
 & & E[y_{i,3}\cdot \Delta u_{i,5}]=0
\end{array}
$$

By forming linear combinations of the (zero expectation) Holtz-Eakin, Newey, and Rosen (1988) population orthogonality conditions, it can be shown that further Arellano and Bover (1995) population orthogonality conditions are redundant in estimation. For example, it can be shown for $t=4$ that

$$
\begin{aligned}
E[y_{i,2}\cdot(u_{i,4}-u_{i,3})]-E[y_{i,1}\cdot(u_{i,4}-u_{i,3})] &= E[\Delta y_{i,2}\cdot(u_{i,4}-u_{i,3})] \\
&= E[\Delta y_{i,2}\cdot u_{i,4}]-E[\Delta y_{i,2}\cdot u_{i,3}],
\end{aligned}
$$

and for $t=5$ that

$$
\begin{aligned}
E[y_{i,3}\cdot(u_{i,5}-u_{i,4})]-E[y_{i,2}\cdot(u_{i,5}-u_{i,4})] &= E[\Delta y_{i,3}\cdot(u_{i,5}-u_{i,4})] \\
&= E[\Delta y_{i,3}\cdot u_{i,5}]-E[\Delta y_{i,3}\cdot u_{i,4}]; \\
E[y_{i,2}\cdot(u_{i,5}-u_{i,4})]-E[y_{i,1}\cdot(u_{i,5}-u_{i,4})] &= E[\Delta y_{i,2}\cdot(u_{i,5}-u_{i,4})] \\
&= E[\Delta y_{i,2}\cdot u_{i,5}]-E[\Delta y_{i,2}\cdot u_{i,4}].
\end{aligned}
$$

Consequently, all but the most recent Arellano and Bover (1995)-type population orthogonality condition for each time period are redundant in estimation, when using both types of linear population orthogonality conditions.

Under the same set of assumptions, more redundancies arise and it can be shown that all of the Ahn and Schmidt (1995) nonlinear population orthogonality conditions become redundant in estimation. To illustrate this, consider the Arellano and Bover (1995) population orthogonality conditions

$$E[\Delta y_{i,t-1}\cdot(y_{i,t}-y_{i,t-1}\rho-\boldsymbol{x}'_{i,t}\boldsymbol{\beta})]=0, \qquad t=3,\dots,T$$

which result from the 'constant correlated effects' assumption (i.e., from $E[\Delta y_{i,t} \cdot a_i] = 0$) and use the assumption to rewrite the above equation:

$$
\begin{aligned}
E[\Delta y_{i,t-1} \cdot (y_{i,t} - y_{i,t-1}\rho - \boldsymbol{x}'_{i,t}\boldsymbol{\beta})] + E\left[(-\rho(y_{i,t-2} - y_{i,t-3}) - (\boldsymbol{x}'_{i,t-1} - \boldsymbol{x}'_{i,t-2})\boldsymbol{\beta})\right. \\
\left. \cdot (a_i + u_{i,t})\right] = 0 \\
\Leftrightarrow E[(y_{i,t} - y_{i,t-1}\rho - \boldsymbol{x}'_{i,t}\boldsymbol{\beta}) \cdot (\Delta y_{i,t-1} - \Delta y_{i,t-2}\rho - \Delta \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta})] = 0.
\end{aligned}
$$

The expression reveals that the Arellano and Bover (1995) population orthogonality conditions render the nonlinear Ahn and Schmidt (1995) population orthogonality conditions redundant for estimation.

6.3 Minimization problem and connection to (G)IV and OLS

When only linear (in parameters) population orthogonality conditions are employed in GMM estimation, minimization of the objective function given in Equation (13) can be represented as

$$
\arg\max_{\hat{\boldsymbol{\theta}}} \quad -\hat{\boldsymbol{s}}'\boldsymbol{Z} \cdot \widehat{\boldsymbol{W}}_N \cdot \boldsymbol{Z}'\hat{\boldsymbol{s}},
$$

where $\hat{\boldsymbol{s}}$ denotes the vector of residuals. Note that the use of population orthogonality conditions from equations in differences and levels in estimation could be reflected in the notation by stacking residuals in differences and levels in the vector $\hat{\boldsymbol{s}}$. The same comment applies to the matrix of instruments $\boldsymbol{Z}$ and the vector of dependent variables $\boldsymbol{y}$ (for details on the structure of the individual matrices and vectors see, e.g., Fritsch, Pua, and Schnurbus, 2019). Since this only renders the notation more cumbersome and does not change the essential results, it is not reflected in the following derivations.

Using only linear population orthogonality conditions in GMM estimation prevents the instruments in $\boldsymbol{Z}$ from depending on the estimated parameters $\hat{\boldsymbol{\theta}}$. Rewriting the minimization problem above results in

$$
-\boldsymbol{y}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{y} + 2 \cdot \boldsymbol{y}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{X}\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}'\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{X}\hat{\boldsymbol{\theta}}.
$$

Taking the first derivative and simplifying yields the following closed form expression of the GMM estimator:

$$
\begin{aligned}
\frac{\partial Q(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} &= +2 \cdot \boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{y} - 2 \cdot \boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{X}\hat{\boldsymbol{\theta}} \stackrel{!}{=} 0 \\
&\Leftrightarrow \boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{X}\hat{\boldsymbol{\theta}} = \boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{y} \\
&\Leftrightarrow \hat{\boldsymbol{\theta}} = (\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{W}}_N\boldsymbol{Z}'\boldsymbol{y}.
\end{aligned}
$$

Under additional assumptions, the closed form can be related to the (G)IV and the OLS estimator for specific choices of the instrument matrix $\boldsymbol{Z}$ and the weighting matrix, for which the general structure $\widehat{\boldsymbol{W}}_N = (\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z})^{-1}$ is assumed. Note that the following discussion solely concentrates on aligning the formulas for the estimators by suitable choosing the individual components of the closed form formula for the GMM estimator using only linear population orthogonality conditions.

When making the assumption that the variance of the idiosyncratic remainder components is conditionally homoscedastic, $\boldsymbol{H}$ in the general structure of $\widehat{\boldsymbol{W}}_N$ can be set to an identity. The weighting matrix then reduces to $\widehat{\boldsymbol{W}}_N = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ and the closed form for the GMM estimator given in Appendix 6.3 can be rewritten as

$$\tilde{\boldsymbol{\theta}} = (\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}.$$

This is essentially the formula to calculate the generalized instrumental variables estimator, where the regressors are projected into the column space spanned by the instruments (when the number of instruments exceeds the number of regressors). In case the number of instruments and the number of regressors are identical, the weighting matrix $\widehat{\boldsymbol{W}}_N$ reduces to an identity. In this case, the following expression results

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{y}\\ &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{y}\\ &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{y},\end{aligned}$$

which is equivalent to the formula for the instrumental variables estimator.

Assuming conditional homoscedasticity of the unobservable idiosyncratic remainder components, exact identification, and imposing the mean independence assumption for the regressors $\boldsymbol{X}$ yields the formula for the OLS estimator:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= (\boldsymbol{X}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{y}\\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.\end{aligned}$$

6.4 Serial correlation introduced by first differencing

Assume the idiosyncratic remainder components $\varepsilon_{i,t}$ are independently and identically distributed with $\mathrm{E}[\varepsilon_{i,t}] = 0$ and $\mathrm{Var}[\varepsilon_{i,t}] = \sigma^2_\varepsilon$, for all $i = 1, \dots, N; t = 2, \dots, T$. First differencing Equation (7) to eliminate the unobservable individual-specific effects affects the covariance of adjacent first differenced idiosyncratic remainder components as follows:

$$\begin{aligned}\mathrm{Cov}[\Delta\varepsilon_{i,t}, \Delta\varepsilon_{i,t-1}] &= \mathrm{E}[(\Delta\varepsilon_{i,t} - \mathrm{E}[\Delta\varepsilon_{i,t}])(\Delta\varepsilon_{i,t-1} - \mathrm{E}[\Delta\varepsilon_{i,t-1}])]\\ &= \mathrm{E}[(\varepsilon_{i,t} - \varepsilon_{i,t-1})(\varepsilon_{i,t-1} - \varepsilon_{i,t-2})]\\ &= \mathrm{E}[\varepsilon_{i,t}\varepsilon_{i,t-1}] - \mathrm{E}[\varepsilon_{i,t}\varepsilon_{i,t-2}] - \mathrm{E}[\varepsilon^2_{i,t-1}] + \mathrm{E}[\varepsilon_{i,t-1}\varepsilon_{i,t-2}]\\ &= -\,\mathrm{E}[\varepsilon^2_{i,t-1}] = -\,\mathrm{E}[(\varepsilon_{i,t-1} - \mathrm{E}[\varepsilon_{i,t-1}])^2] = -\sigma^2_\varepsilon.\end{aligned}$$

This leads to the following correlation of adjacent first differenced idiosyncratic remainder components:

$$\mathrm{Cor}[\Delta\varepsilon_{i,t}, \Delta\varepsilon_{i,t-1}] = \frac{\mathrm{Cov}[\Delta\varepsilon_{i,t}, \Delta\varepsilon_{i,t-1}]}{\sqrt{\mathrm{Var}[\Delta\varepsilon_{i,t}] \cdot \mathrm{Var}[\Delta\varepsilon_{i,t-1}]}} = \frac{-\sigma^2_\varepsilon}{2\sigma^2_\varepsilon} = -\frac{1}{2}$$

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