UvA ECONOMETRICS

Discussion Paper: 2013/01

Dynamic panel data models

Maurice J.G. Bun and Vasilis Sarafidis

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Dynamic Panel Data Models*

Maurice J.G. Bun† and Vasilis Sarafidis‡

this version: 1 July 2013

JEL-code: C13; C23

Keywords: panel data, dynamics, endogeneity, GMM, mean stationarity, constant correlated effects

Abstract

This Chapter reviews the recent literature on dynamic panel data models with a short time span and a large cross-section. Throughout the discussion we consider linear models with additional endogenous covariates. First we give a broad overview of available inference methods placing emphasis on GMM. We next discuss in more detail the assumption of mean stationarity underlying the system GMM estimator. We discuss causes of deviations from mean stationarity, their consequences and tests for mean stationarity.

1. introduction

This Chapter reviews the recent literature on dynamic panel data models. Economic relationships usually involve dynamic adjustment processes. In time series regression models it is common practice to deal with these by including in the specification lagged values of the covariates, the dependent variable, or both. The inclusion of lags of the dependent variable seems to provide an adequate characterization of many economic dynamic adjustment

*Prepared for the Oxford Handbook on Panel Data (editor: Badi H. Baltagi), Oxford University Press. We would like to thank Artūras Juodis, two anonymous referees and the Editor for helpful comments and suggestions.

†Tinbergen Institute and Amsterdam School of Economics, University of Amsterdam. E-mail: m.j.g.bun@uva.nl. The author wants to thank Monash University for hospitality while working on this paper. The research of the author has been funded by the NWO Vernieuwingsimpuls research grant 'Causal Inference with Panel Data'.

‡Department of Econometrics & Business Statistics, Monash University. E-mail: vasilis.sarafidis@monash.edu.

processes. However, in panel data analysis with a small number of time periods there often appear to be inference problems, such as small sample bias in coefficient estimation and hypothesis testing.

We consider a class of linear dynamic panel data models allowing for endogenous covariates. Sometimes it can be argued that the covariates are exogenous, at least conditional on individual- and time-specific effects, e.g. when these covariates reflect natural phenomena. However, in many areas of economic inquiry this is often not the case. For instance, in empirical analysis of policy interventions, policy variables are most likely not strictly exogenous but simultaneously determined with the outcome variable of interest (e.g. Besley and Case, 2000). Even if one is willing to assume that the covariates are not simultaneously determined, they may still be influenced by past values of the outcome variable.

Due to the various endogeneity problems mentioned above, least squares based inference methods, i.e. fixed effects or random effects estimators, are biased and inconsistent. Hence, it has become standard practice nowadays to use Instrumental Variables (IV) methods or the Generalized Method of Moments (GMM), which produce consistent parameter estimates for a finite number of time periods, T, and a large cross-sectional dimension, N (see e.g. Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). Within this class of methods, the system GMM estimator (Blundell and Bond, 1998) has become increasingly popular. We do not intend to provide a detailed overview of specific applications, but in labor economics (minimum wage effects, labor supply, returns to schooling, job training), development economics (effectiveness of foreign aid, transition economics), health economics (health expenditures, organization of health care, aging, addiction, insurance), industrial organization (mergers & acquisitions, evaluation of competition policy), international economics (effects of trade policy and economic integration), macroeconomics (economic growth, optimal currency areas) and finance (banking regulation) GMM inference methods have been applied extensively.

One main reason for their popularity in empirical research is that the GMM estimation approach may provide asymptotically efficient inference employing a relatively minimal set of statistical assumptions. However, despite its optimal asymptotic properties, the finite sample behaviour of the GMM estimator and corresponding test statistics can be rather poor due to weakness and/or abundance of moment conditions and dependence on crucial nuisance parameters. As a result, several alternative inference methods have been proposed, often requiring different and more stringent assumptions. Here we will survey some of the most recent contributions.

In addition, an issue that has recently attracted further attention is the mean stationarity assumption that underlies the system GMM estimator. Roodman (2009) points out that this assumption is not trivial, which seems to be underappreciated in applied research. The effect of deviations from mean stationarity are analysed theoretically by Hayakawa (2009) and Hayakawa and Nagata (2012). Kiviet (2007), Everaert (2012) and Juodis (2013) also explore this issue by Monte Carlo simulation. Consequently, in this Chapter we will focus on mean stationarity in more detail and analyse the main arguments.

This is not the first review study on linear dynamic models for panel data; Blundell, Bond and Windmeijer (2001) and Roodman (2009) provide excellent summaries of the GMM methodology. Arellano and Honore (2003) also provide a very comprehensive analysis, including results for nonlinear models. Specific chapters in books on panel data also pay ample attention to dynamic panel data modeling (Arellano, 2003a; Hsiao, 2003; Baltagi, 2008; Mátyás and Sevestre, 2008).

There are of course many other interesting and related topics that we don't cover in this Chapter. We do not discuss: (1) slope parameter heterogeneity; (2) cross-sectional dependence; (3) nonlinear models. Also we mainly focus on GMM inference methods but we briefly mention likelihood-based alternatives in Section 2. A discussion of some of these topics, however, can be found in other chapters of this volume.

2. review of the literature

Suppose the relation between the dependent variable y_{it} and a single covariate x_{it} can be modeled by the following dynamic specification:

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \eta_i + \varepsilon_{it}; \;\; i = 1, ...N, t = 1, ...T, \tag{2.1}$$

where η_i denotes unobserved time-invariant heterogeneity and ε_{it} is the idiosyncratic error component.[1] We assume that y_{i0} and x_{i0} are observed. The dynamic panel data model in (2.1) permits the distinction between the long run, or equilibrium, relationship and the short-run dynamics. Note that x_{it} could also be a vector, containing both contemporaneous and lagged values of explanatory variables. It can be seen in this case that the above specification encompasses several important other specifications, i.e. static models, distributed lag or first-differenced specifications.

Often the individual-specific effect, η_i, is thought to be correlated with x_{it}. Furthermore, by construction the lagged dependent variable is correlated with the individual specific effect, i.e. $E(\eta_i | y_{i,t-1}) \neq 0$. Additionally, the covariate may also exhibit a nonzero correlation with the contemporaneous or lagged idiosyncratic errors, such that $E(\varepsilon_{it} | x_{is}) \neq 0$ for $t \leq s$. All these endogeneity issues imply that least squares based estimators may be inconsistent. To this end, several alternative estimators have been proposed. In this Chapter we focus on GMM estimators, although at the end of the section we briefly describe relative merits of other procedures, especially likelihood-based inference methods.

[1] Time-specific effects can also be included explicitly or controlled for by cross-sectional demeaning of the data prior to estimation. We will discuss an example of a process with time-specific effects later.

We consider models where idiosyncratic errors obey the following conditional moment restriction:

$$E\left[\varepsilon_{it}|\mathbf{y}_i^{t-1}, \mathbf{x}_i^s, \eta_i\right] = 0;\ t = 1, ..., T, \tag{2.2}$$

where $\mathbf{y}_i^{t-1} = (y_{i0}, y_{i1}, ..., y_{i,t-1})'$ and $\mathbf{x}_i^s = (x_{i0}, x_{i1}, ..., x_{is})'$. Assumption (2.2) rules out serial correlation in ε_{it}, which is a base for constructing unconditional moments. However, it does not restrict the relationship between η_i and $\mathbf{x}_i^s$. Regarding the regressor x_{it} we distinguish between (i) strict exogeneity, $s = T$; (ii) predeterminedness, $s = t$; and (iii) endogeneity, $s < t$. That is, depending on s, equation (2.2) permits instantaneous or lagged feedback from y to x.

Based on assumption (2.2), the model can be expressed in first differences as

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \beta \Delta x_{it} + \Delta \varepsilon_{it}, \tag{2.3}$$

for which the following (DIF) unconditional moment conditions are available:

$$E\left[\mathbf{y}_i^{t-2}\Delta\varepsilon_{it}\right] = 0; \qquad E\left[\mathbf{x}_i^{s-1}\Delta\varepsilon_{it}\right] = 0;\ t = 2, ..., T, \tag{2.4}$$

with s depending on the exogeneity status of x_{it}. Lagged levels of the endogenous variables can be used as instruments for current changes. Simple IV estimators of this type were first proposed by Anderson and Hsiao (1981, 1982) for the first order autoregressive AR(1) model and in a multivariate setting and GMM framework by Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991).

Assumption (2.2) also rules out any correlations between ε_{it} and η_i.[2] This provides an additional set of $T - 3$ nonlinear moment conditions available for the model in first differences, as suggested by Ahn and Schmidt (1995):

$$E\left[(\eta_i + \varepsilon_{it})\,\Delta\varepsilon_{i,t-1}\right] = 0, \qquad t = 3, ...T. \tag{2.5}$$

Thus, under assumption (2.2) efficiency gains may occur by using (2.5) in addition to (2.4). Ahn and Schmidt (1995) show that the GMM estimator (labeled AS hereafter) that makes use of (2.4) and (2.5) is efficient in the class of estimators that make use of second moment information. They also report substantial efficiency gains when comparing asymptotic variances for the AR(1) model. Especially when the series is highly persistent, the additional quadratic moment conditions become relatively informative compared with the moment conditions in (2.4) as can be seen from the calculations in Ahn and Schmidt (1995).

It is well known (see e.g. Blundell and Bond, 1998) that the GMM estimator of the first-differenced model can have poor finite sample properties in terms of bias and precision

[2] This is not very restrictive, because in autoregressive models any nonzero correlation between individual effects and idiosyncratic errors tends to vanish over time (Arellano, 2003, p82).

when the series are persistent. One reason for this is that in this case lagged levels are weak predictors of the first differences. Blundell and Bond (1998) advocated the use of extra moment conditions that rely on certain stationarity restrictions on the time series properties of the data, as suggested by Arellano and Bover (1995). For the multivariate model in (2.1) these amount to assuming

$$E\left[\Delta y_{it}|\eta_i\right] = 0, \qquad E\left[\Delta x_{it}|\eta_i\right] = 0, \tag{2.6}$$

which imply that the original series in levels have *constant correlation* over time with the individual-specific effects.[3] Assumption (2.6) leads to the following additional moment conditions for the model in levels (2.1) (hereafter, LEV):

$$E\left[\Delta \mathbf{y}_i^{t-1}(\eta_i + \varepsilon_{it})\right] = 0, \qquad E\left[\Delta \mathbf{x}_i^{s}(\eta_i + \varepsilon_{it})\right] = 0, \tag{2.7}$$

for $t = 2, ..., T$, where $\Delta \mathbf{y}_i^{t-1} = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{i,t-1})'$ and so on. In words, with regards to endogenous variables, lagged changes can be used as instruments for current levels.

Notice that a subset of the moment conditions in (2.7) is redundant because it can be expressed as a linear combination of the moments in (2.4) (see e.g. Kiviet, Pleus and Poldermans, 2013, for a proof of this result). Therefore, the complete set of non redundant linear moment conditions in levels can be specified as

$$E\left[\Delta y_{i,t-1}(\eta_i + \varepsilon_{it})\right] = 0, \ t = 2, ..., T, \tag{2.8}$$

together with

$$E\left[\Delta x_{i,t-1}(\eta_i + \varepsilon_{it})\right] = 0, \ t = 2, ..., T, \tag{2.9}$$

in case of endogenous x_{it}, or

$$E\left[\Delta x_{it}(\eta_i + \varepsilon_{it})\right] = 0, \ t = 1, ..., T, \tag{2.10}$$

in case of predetermined or strictly exogenous x_{it}. Combining (2.4) with (2.8) and either (2.9) or (2.10) leads to the system GMM estimator (labeled SYS). It should also be noted that (2.7) render the nonlinear moment conditions in (2.5) redundant. Hence under assumption (2.6) SYS is asymptotically efficient. Blundell and Bond (1998) argued that SYS performs better than the DIF GMM estimator because the instruments in the LEV model remain good predictors for the endogenous variables even when the series are highly persistent.

[3] Assumption (2.6) is often labeled the 'mean stationarity' assumption. Some authors (e.g. Kiviet, 2007) prefer to label it as 'effect stationarity', because it is an expectation conditional on the individual specific effect η_i. In the next section we use the term 'constant correlated effects' to describe this assumption. We believe this is more precise because the additional moment conditions do not require mean stationarity, as it will become clear.

Notwithstanding the popularity of the GMM methodology in applied economic research, producing accurate statistical inferences for panel data models using instrumental variables has not been a straightforward exercise. In particular, the desirable asymptotic properties of the estimators do not safeguard their performance in finite samples. In what follows, we summarise some of the issues that may arise in finite samples.

2.1. asymptotic standard errors

As it has already been shown in the Monte Carlo study in Arellano and Bond (1991), estimated asymptotic standard errors of two step GMM estimators can be severely downward biased, suggesting more precision than is actually justified. Windmeijer (2005) showed that this is due to the fact that the weight matrix used in the second stage is based on initial parameter estimates, which themselves are subject to sampling variability that is not accounted for. Using asymptotic expansion techniques the author proposed a variance correction, leading to improved inference using the Wald test. In a rather extensive Monte Carlo study Bond and Windmeijer (2005) confirm the poor performance of the standard Wald test based on two step GMM. They find that using Wald statistics based on either one step GMM or the variance-corrected two step GMM or exploiting the LM statistic produces reliable inferences when identification is not too weak.

2.2. many instruments

Since dynamic panels are often largely overidentified, another important practical issue is how many moment conditions to use. Again, traditional first order asymptotics are not very helpful in answering this question as they imply 'the more the merrier'. In practice, however, it is well documented that numerous instruments can overfit endogenous variables in finite samples (see e.g. Bekker, 1994), resulting in a trade off between bias and efficiency. To gain some insight, consider a standard IV regression with one endogenous covariate; the R^2 coefficient of the first stage regression takes the value of one when the number of instruments is equal to the number of observations. Thus, the instrumental variable is perfectly correlated with the endogenous variable and the IV estimator is numerically identical to the (biased) OLS estimator.

There is substantial theoretical work on the overfitting bias of GMM estimators in panel data models. For example, Koenker and Machado (1999) establish that a sufficient condition for the usual limiting distribution of the GMM estimator to remain valid under instrument proliferation is $m = o\left(N^{1/3}\right)$, where m denotes the number of instruments. Arellano (2003b) shows that in models with predetermined variables, such as a pure AR model, the bias as a result of overfitting is of order $O\left(m/N\right)$, while for models with endogenous variables the bias is of order $O\left(mT/N\right)$. Similarly, Alvarez and Arellano (2003) analyse a panel autoregressive model of order one, and show that although GMM remains consistent for $T/N \to c$, so long as $0 \leq c \leq 2$, for $c > 0$ the estimator exhibits a bias in its asymptotic distribution that is of order $1/N$. Bun and Kiviet (2006) show that in comparison with GMM estimators that employ all available instruments, reducing the set of instruments by order T also decreases the bias by an order smaller in magnitude by a factor T. Ziliak (1997) examines the bias/efficiency trade off issue using bootstraping in an empirical application to life cycle labor supply under uncertainty. He shows that the bias of 2SLS and GMM estimators becomes larger as the number of instruments increases, and furthermore that GMM is biased downwards relative to 2SLS, arguably due to the nonzero correlation between the estimated weight matrix and the sample moments. Results from Monte Carlo simulation experiments vary, depending on the simulation design, the degree of overidentification in conjunction with the techniques employed for reducing the number of instruments, and finally the method employed in estimation. Windmeijer (2005) reported that for the two step DIF GMM, using only two lags of the dependent variable as instruments appeared to decrease the average bias by 40% relative to the estimator that made use of the full set of instruments, although the standard deviation of the estimator increased by about 7.5%. Roodman (2009) compared two popular approaches for limiting the number of instruments: (i) the use of (up to) certain lags instead of all available lags and (ii) combining instruments into smaller sets. His results show that the bias in SYS GMM based on the first approach is similar to the bias when using the full set of instruments. However, there is clear bias reduction under the second approach. On the other hand, Hayakawa (2009) shows that in panels with large unobserved heterogeneity the bias in DIF GMM can actually be larger when using a smaller set of instruments.

2.3. dependence on nuisance parameters

Various studies (e.g. Binder, Hsiao and Pesaran, 2005; Bun and Kiviet, 2006; Bun and Windmeijer, 2010) show that the finite sample properties of GMM estimators depend heavily on crucial nuisance parameters, especially the ratio of the variances of the individual-specific effects and the idiosyncratic errors ($\sigma_\eta^2/\sigma_\varepsilon^2$). Binder, Hsiao and Pesaran show that the asymptotic variance of the DIF GMM estimator increases with the variance of the individual-specific effects. Using asymptotic expansion techniques Bun and Kiviet (2006) approximate the bias of various one step GMM estimators. The asymptotic expansions provide analytic evidence on how the bias of the various GMM estimators depends on, among other things, the size of the variance of the individual effects and the correlation between regressors and individual effects. Bun and Windmeijer (2010) analyze the bias of DIF, LEV and SYS 2SLS estimators relative to bias in corresponding OLS estimators. They conclude that, although absolute bias of the LEV and SYS 2SLS estimators tends to be small for persistent panel data, this bias is an increasing function of $\sigma_\eta^2/\sigma_\varepsilon^2$. Furthermore, relative biases of LEV and SYS 2SLS estimators are smaller and the associated Wald tests perform better than those of DIF when $\sigma_\eta^2 < \sigma_\varepsilon^2$. The reverse is the case when σ_η^2 is larger then σ_ε^2. By Monte Carlo simulation these results are shown to extend to the panel data setting when estimating the model by GMM.

2.4. weak instruments

When instruments are weak, i.e. only lowly correlated with the endogenous variables, IV and GMM estimators can perform poorly in finite samples, see e.g. Bound, Jaeger and Baker (1995), Staiger and Stock (1997), Stock and Wright (2000) and Stock, Wright and Yogo (2002). With weak instruments, IV or GMM estimators for panel data models are biased in the direction of the least squares estimator, and their distributions are non-normal (Wansbeek and Knaap, 1999; Hahn, Hausman and Kuersteiner, 2007; Kruiniger, 2009; Bun and Kleibergen, 2013), affecting inference using standard t or Wald testing procedures.

To illustrate the weak instrument problem in dynamic panel data models, consider the the case of an AR(1) model, i.e. impose $\beta = 0$ in (2.1), and $T = 2$. The DIF and LEV models, i.e. (2.3) and (2.1), are now the following cross-sectional models:

$$\text{DIF} \;:\; \Delta y_{i2} = \alpha \Delta y_{i1} + \Delta \varepsilon_{i2}, \tag{2.11}$$

$$\text{LEV} \;:\; y_{i2} = \alpha y_{i1} + \eta_i + \varepsilon_{i2}. \tag{2.12}$$

The moment conditions for both models are:

$$\text{DIF} \;:\; E[y_{i0}(\Delta y_{i2} - \alpha \Delta y_{i1})] = 0, \tag{2.13}$$

$$\text{LEV} \;:\; E[\Delta y_{i1}(y_{i2} - \alpha y_{i1})] = 0, \tag{2.14}$$

hence for $T = 2$ simple IV estimators result:

$$\hat{\alpha}_{DIF} = \frac{\sum_{i=1}^N y_{i0} \Delta y_{i2}}{\sum_{i=1}^N y_{i0} \Delta y_{i1}}, \quad \hat{\alpha}_{LEV} = \frac{\sum_{i=1}^N y_{i2} \Delta y_{i1}}{\sum_{i=1}^N y_{i1} \Delta y_{i1}}. \tag{2.15}$$

Assuming mean stationarity, i.e. $y_{i0} = \frac{\eta_i}{1-\alpha} + \varepsilon_{i0}$, the resulting covariance between regressor and instrument is:

$$\text{DIF:} \quad E[y_{i0} \Delta y_{i1}] = (\alpha - 1)\, E[\varepsilon_{i0}^2], \tag{2.16}$$

$$\text{LEV:} \quad E[y_{i1} \Delta y_{i1}] = \alpha\, (\alpha - 1)\, E[\varepsilon_{i0}^2] + E[\varepsilon_{i1}^2]. \tag{2.17}$$

Because $E[\varepsilon_{i1}^2] \neq 0$ the LEV moment condition always seems to identify α even for true values close to one, see Arellano and Bover (1995) and Blundell and Bond (1998). There is a caveat, however, because identification using LEV moment conditions is affected by the model for the initial observations.

Bond, Nauges and Windmeijer (2005) show how identification of α depends on the variance of the initial observations. The LEV first stage regression is:

$$y_{i1} = \pi_l \Delta y_{i1} + l_i, \tag{2.18}$$

with l_i being the reduced form error. When $\alpha = 1$ we have $\pi_l = 1$ and $l_i = y_{i0}$. Therefore, weak identification does not originate from $\pi_l \to 0$, but from $Var(l_i) = Var(y_{i0})$ being large. When the number of time periods that the process has been in existence before the sample is drawn is fixed, then $Var(y_{i0}) < \infty$. In this case the LEV (and hence SYS) moment conditions identify α even when its true value is one. For many DGPs, however, $Var(y_{i0}) \to \infty$ when α approaches one leading to identification failure. An example of such a DGP is that of covariance stationarity.

Kruiniger (2009) also shows that weakness of DIF and LEV moment conditions can manifest itself in different ways depending on the model for the initial observations. Following Han and Phillips (2006) sample moment conditions can be decomposed in "signal" and "noise". Conventional asymptotics assume a strong signal, while noise is eliminated asymptotically. For the dynamic panel data model Kruiniger (2009) shows that, depending on the initial conditions, in some cases the signal becomes weak, while in other situations noise is dominating. For example, assuming covariance stationarity we have that $E[\varepsilon_{i0}^2] = \frac{\sigma_\varepsilon^2}{1-\alpha^2}$ and $E[\varepsilon_{i1}^2] = \sigma_\varepsilon^2$, hence (2.16) and (2.17) become:

$$\text{DIF:} \qquad E[y_{i0}\Delta y_{i1}] = -\frac{\sigma_\varepsilon^2}{1+\alpha}, \tag{2.19}$$

$$\text{LEV:} \qquad E[y_{i1}\Delta y_{i1}] = \frac{\sigma_\varepsilon^2}{1+\alpha}. \tag{2.20}$$

These expressions suggest a strong "signal" for both DIF and LEV moment conditions, even when α is (close to) one. However, at the same time the variance of the DIF and LEV moment conditions is proportional to $\frac{1}{1-\alpha}$ implying explosive behavior when α goes to one. In this case the noise in the moment equation dominates the signal and weak identification results for both DIF and LEV moment conditions.

Bun and Windmeijer (2010) show the weakness of DIF and LEV moment conditions in yet another way by calculating concentration parameters for DIF and LEV models assuming covariance stationarity. For a simple cross-sectional linear IV model, the concentration parameter is a measure of the information content of the instruments. When $T = 2$ and assuming covariance stationarity they are equal for both models:

$$\sigma_\varepsilon^2 \frac{(1-\alpha)^2}{1-\alpha^2+2(1+\alpha)\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}}. \tag{2.21}$$

This suggests a weak identification problem in the LEV model too when $\alpha \to 1$ (and/or $\frac{\sigma_\eta^2}{\sigma_\varepsilon^2} \to \infty$).

Bun and Kleibergen (2013) emphasize the arbitrariness of identification by the LEV moment condition by considering a joint limit process where both α converges to one and N goes to infinity. Specifying the function $h(\alpha)$ such that $h(\alpha)^{-2} \propto Var(y_{i0})$ they show that when $h(\alpha)\sqrt{N} \underset{N\to\infty,\ \alpha\uparrow 1}{\longrightarrow} \infty$ the derivative of the LEV moment condition converges to a nonzero constant. However, when $h(\alpha)\sqrt{N} \underset{N\to\infty,\ \alpha\uparrow 1}{\to} 0$, it is the case that

$$h(\alpha)\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i1}\Delta y_{i1} \underset{N\to\infty,\ \theta_0\uparrow 1}{\overset{d}{\longrightarrow}} N(0, Var(\varepsilon_{i1})). \tag{2.22}$$

This result shows identification failure since the derivative of the LEV moment condition converges to a random limit with mean zero.[4] Since any assumption on convergence rates of α and N is arbitrary, identification by LEV moment conditions is arbitrary. Assuming $h(\alpha)\sqrt{N} \underset{N\to\infty,\ \alpha\uparrow 1}{\to} 0$ Bun and Kleibergen (2013) show that 2-step DIF, LEV and SYS GMM estimators and associated Wald statistics have non-standard large sample distributions, which results are qualitatively similar to those in Kruiniger (2009). They also show, however, that for $T > 2$ it is possible to achieve identification of α even when $h(\alpha)\sqrt{N} \underset{N\to\infty,\ \alpha\uparrow 1}{\to} 0$ by combining SYS or AS moment conditions with the Lagrange multiplier GMM statistic proposed by Newey and West (1987), or with identification robust GMM statistics proposed by Stock and Wright (2000) and Kleibergen (2005).

Summarizing, whether the various sets of moment conditions identify the parameters of dynamic panel data models with persistent data depends on what seems reasonable to assume for the initial observations. In many microeconometric panel data a finite number of start-up periods may be a realistic scenario. In those cases identification issues are less severe, but this is not known on beforehand. Note that all above studies exploit mean stationarity and hence validity of LEV moment conditions. Strength of identification by the DIF (and also AS) moment conditions, however, may change substantially when we deviate from mean stationarity as we will discuss in Section 3.3 below.

2.5. alternative procedures

The dependence of finite sample distributions on the number and type of moment conditions as well as important nuisance parameters can be detrimental to the use of conventional GMM estimators in applied work. Hence, recent contributions propose to exploit alternative and possibly nonlinear moment conditions derived from inconsistent least squares procedures or likelihood based methods.

A central theme in linear dynamic panel data analysis is the fact that the fixed effects maximum likelihood (ML) estimator is inconsistent for a fixed number of time periods, as the number of cross-sectionalal units tends to infinity. This inconsistency is referred to as 'Nickell bias', due to Nickell (1981), and is an example of the incidental parameters problem (the number of parameters increasing with the sample size), analyzed first by Neyman and Scott (1948). This has led to an interest in likelihood-based methods that correct for the incidental parameters problem. Some of these methods are based on modifications

[4] A similar result holds for the DIF moment condition.

of the profile likelihood, see Lancaster (2002) and Dhaene and Jochmans (2012). Other methods start from the likelihood function of the first differences, see Hsiao, Pesaran and Tacmiscioglu (2002), Binder, Hsiao and Pesaran (2005) and Hayakawa and Pesaran (2012).

Well known transformations to remove individual-specific effects in panel data models are the within transformation and first differences. Kiviet (1995) and Bun and Carree (2005) exploit the possibility to correct the inconsistency of the fixed effects estimator, while Han and Phillips (2010) and Han, Phillips and Sul (2010) recently developed efficient GMM methods based on alternative moment conditions arising from the model in first differences. However, the models considered in these studies are mainly autoregressive of nature (possibly with additional exogenous regressors) which currently limits their practical use.

A common advantage of these alternative likelihood based inference procedures is that they are largely invariant to the model parameters because unobserved heterogeneity is a priori transformed away. In comparison with the GMM approach, a limitation is that they impose exogeneity restrictions on the covariates and time series homoskedasticity, which may be violated in practice. Especially endogeneity with respect to the idiosyncratic errors is a common scenario in many applied studies. As mentioned by Hayakawa and Pesaran (2012), in principle it is feasible to exploit likelihood-based estimators in case of endogeneity too, however this requires supplementing the structural dynamic equation (2.1) with a reduced form equation for the endogenous regressors. Estimates of the parameters of interest could be retrieved from the resulting panel VAR coefficients. This is still a matter of future research.

3. revisiting the issue of initial conditions

The popular system GMM estimator depends on (2.6), which is certainly satisfied if all variables are assumed to be mean stationary. A number of authors (see e.g. Roodman, 2009) have critically assessed the credibility of mean stationarity in applied economic research. In this section we discuss this issue in more detail. Furthermore, we describe consequences of departures from this assumption and statistical procedures to test it. Throughout the discussion the focus is on GMM inference methods.

3.1. constant-correlated effects

The issue of initial conditions in models with fixed T has attracted considerable attention in the dynamic panel data literature since its infancy. For instance, Anderson and Hsiao (1982) and Bhargava and Sargan (1983) analyse the asymptotic properties of various maximum likelihood and instrumental variable type procedures under a large variety of assumptions about the initial conditions of the processes being studied.[5] One possibility is to assume that the initial condition is such that the process is mean stationary. The growing concern about the properties of dynamic panel estimators in finite samples may have contributed to placing large emphasis on this assumption, both in terms of theoretical developments, as well as in empirical applications.

In particular, mean stationarity has been employed for deriving additional moment conditions and developing new estimators (e.g. Arellano and Bover, 1995; Blundell and Bond, 1998). Given its mathematical convenience, it has also become a standard assumption in the many/weak instruments literature (e.g. Alvarez and Arellano, 2003; Bun and Windmeijer, 2010). Moreover, it is fair to say that in a large part of the literature during the last fifteen years or so, which reports results on the performance of GMM estimators based on Monte Carlo experiments, mean stationarity is either assumed from the outset, or it is effectively imposed as a byproduct of the simulation design. In the former case this is achieved by drawing the initial observations from a covariance stationary distribution (e.g. Blundell, Bond and Windmeijer, 2001). In the latter case the design entails generating $T + S$ time series observations, with S equal to 50 or more, but using only T observations for estimation purposes. The first S observations are not considered in estimation, in order to 'minimize the effect of initial conditions' (e.g. Bun and Kiviet, 2006). Although this practice is rather innocuous in panels with T large, it can have important consequences in panels with small T.

Another point that is easily discernible on selective reading of a huge empirical literature utilising panel data, is that the GMM estimator proposed by Ahn and Schmidt (1995), utilising (2.4) and (2.5), is rarely used in practice. This is despite the fact that this is the efficient estimator under a relatively minimal set of assumptions, excluding mean stationarity, and that using these moment conditions identification is achieved even for persistent panel data (Bun and Kleibergen, 2013). Instead, in a substantial body of applied work the estimation strategy appears to involve the use of either DIF (which is not efficient under mean nonstationarity) or SYS (for which a sufficient condition for consistency is mean stationarity), or often both, without providing much theoretical justification for the implications of the underlying assumptions that validate the use of SYS specifically. The tendency to bypass AS is not surprising perhaps, given that both DIF and SYS are easy to compute and are readily available in several econometric packages of widespread use. On the contrary, so far as we know, AS is not yet part of a standard routine.

From a statistical perspective, and since most dynamic panel data models are typically overidentified, violations from mean stationarity are in principle detectable based on Sargan's or Hansen's test of overidentifying restrictions. However, it is now well known that these tests can have very low power, especially when the number of instruments used is

[5] See also Hsiao (2003, Ch. 4) for an excellent discussion.

relatively large (see e.g. Bowsher, 2002; Roodman, 2009). This could be partially mitigated by computing an incremental test based on AS and SYS, which involves a smaller number of degrees of freedom compared to an incremental test based on DIF and SYS. This is rarely implemented in practice.

In what follows, we revisit the conditions under which the LEV moment conditions hold true. We elaborate on what we call the "constant-correlated effects" assumption, which, for a limited lifespan of the time series, is a necessary and sufficient condition for the consistency of LEV and SYS GMM estimators. Since it is a rather intuitive concept to grasp, it has the benefit that, once one is prepared to motivate what unobserved heterogeneity is likely to capture in one's model, it becomes relatively straightforward to form an idea about how restrictive the condition appears to be on a specific application. If it does, the efficient estimator is AS and more effort should be made to apply it. Furthermore, we summarise some of the (limited) results existing and attempt to provide some guidance.

Recall that the LEV moment conditions (2.7) imply that the first difference of y_{it} and x_{it} are both uncorrelated with η_i, i.e.

$$E\left(\Delta y_{it}\eta_i\right) = 0, \tag{3.1}$$

$$E\left(\Delta x_{it}\eta_i\right) = 0, \tag{3.2}$$

for $t = 1, ..., T$. Thus, the moment conditions above imply that the first-differenced variables are free from the individual effects, which requires that the correlation between y_{it} (x_{it}) and η_i is *constant over time*. We phrase this high level condition as the "constant-correlated effects" (cce) assumption, which can be expressed as

$$E\left(y_{it}\eta_i\right) = c_y, \tag{3.3}$$

$$E\left(x_{it}\eta_i\right) = c_x, \tag{3.4}$$

for all t. The issue of whether the variables of the model exhibit a constant correlation over time with unobserved time-invariant heterogeneity depends on the application in mind. Below we consider a few applications where GMM estimators have been popular. While the discussion should not be interpreted as indicative of a general pattern, it does suggest that the cce assumption is often taken too lightly by empirical researchers.

Suppose that (2.1) represents an earnings determination equation (see also Hause, 1980; Arellano, 2003) with wage on the left hand side and experience on the right hand side (along with lagged wage and other variables, such as education and tenure). It is commonly viewed in this case that η_i captures, among other things, the effect of innate ability, or skills, which are unobserved to the econometrician and in any case hard to quantify. Consider the following scenario: the sample includes workers at different phases of their career; some of them are close to retirement and some are new starters, having entered the labor market only recently for the first time, or having made a career change soon prior to the beginning of the sampling period. An argument could be made that the subgroup of new starters who are highly skilled, and therefore are employed in knowledge-intensive jobs, is likely to accumulate proportionally more experience as time progresses, and indeed receive higher salaries for this reason, relative to those individuals within the same group who have lower skills. This systematic relationship over time between unobserved skills and experience, or wage, is ruled out by the cce assumption.

Alternatively, one can draw from the literature of the estimation of production functions, in which η_i may capture the effect of technical inefficiency and unobserved managerial practices. Additionally, short-run dynamics may originate from autoregressive productivity shocks (Blundell and Bond, 2000). One might argue that within new firms, or at least new entrants in a particular market, those which are more efficient are likely to be able to produce proportionally more output towards the end of the sampling period compared with inefficient firms, as the former group is able to learn better from past practices. Again, this scenario is ruled out by the cce assumption.

To obtain some insight about what the cce condition entails in our model, consider model (2.1) again, which is replicated below for ease of exposition

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \eta_i + \varepsilon_{it}. \tag{3.5}$$

We can express y_{it} recursively as follows:

$$y_{it} = \alpha^t \left(y_{i0} - \frac{\eta_i}{1-\alpha} \right) + \beta \sum_{s=0}^{t-1} \alpha^s x_{i,t-s} + \frac{\eta_i}{1-\alpha} + \sum_{s=0}^{t-1} \alpha^s \varepsilon_{i,t-s}. \tag{3.6}$$

It is immediately clear from the above expression that it is very unlikely that the correlation between y_{it} and η_i is constant over time, i.e. $E(y_{it}\eta_i) = c_y$, when the correlation between x_{it} and η_i is not. This is because y_{it} depends not only on the current but also on all lagged values of x_{it}, albeit their impact is declining with distance. To make further progress, let x_{it} form an AR(1) process such that

$$x_{it} = \rho x_{i,t-1} + \tau \eta_i + \upsilon_{it} = \rho^t \left(x_{i0} - \frac{\tau \eta_i}{1-\rho} \right) + \frac{\tau \eta_i}{1-\rho} + \sum_{s=0}^{t-1} \rho^s \upsilon_{i,t-s}, \tag{3.7}$$

where we assume that $-1 < \rho < 1$. As a result, (3.6) becomes

$$\begin{aligned} y_{it} &= \alpha^t \left(y_{i0} - \frac{\eta_i}{1-\alpha} \right) + \beta \sum_{s=0}^{t-1} \alpha^s \left[\rho^{t-s} \left(x_{i0} - \frac{\tau \eta_i}{1-\rho} \right) + \frac{\tau \eta_i}{1-\rho} + \sum_{j=0}^{t-1-s} \rho^j \upsilon_{i,t-s-j} \right] \\ &\quad + \frac{\eta_i}{1-\alpha} + \sum_{s=0}^{t-1} \alpha^s \varepsilon_{i,t-s} \\ &= \alpha^t \left(y_{i0} - \frac{1-\rho+\beta\tau}{(1-\alpha)(1-\rho)} \eta_i \right) + \beta \sum_{s=0}^{t-1} \alpha^s \rho^{t-s} \left(x_{i0} - \frac{\tau \eta_i}{1-\rho} \right) + \frac{1-\rho+\beta\tau}{(1-\alpha)(1-\rho)} \eta_i \\ &\quad + \beta \sum_{s=0}^{t-1} \alpha^s \sum_{j=0}^{t-1-s} \rho^j \upsilon_{i,t-s-j} + \sum_{s=0}^{t-1} \alpha^s \varepsilon_{i,t-s}. \end{aligned} \tag{3.8}$$

The first (second) right hand side term within the brackets is the deviation of the initial in-sample observation on y (x) from its steady state path, or its long run mean conditional on η_i. Eventually, assuming that the process for y and x is not altered, these deviations will die out because $|\alpha| < 1$ and $|\rho| < 1$. However, in series with a limited lifespan, which is typically the case in microeconometrics, these quantities are non-negligible, especially when the autoregressive coefficients are close to the value of one. Thus, the cce assumption suggests that any deviations from steady state behaviour need to be uncorrelated with η_i. We may also express this in an alternative form, as follows:

$$E\left[\left(x_{i0} - \frac{\tau\eta_i}{1-\rho}\right)\frac{\tau\eta_i}{1-\rho}\right] = 0, \tag{3.9}$$

and

$$E\left[\left(y_{i0} - \frac{1-\rho+\beta\tau}{(1-\alpha)(1-\rho)}\eta_i\right)\frac{1-\rho+\beta\tau}{(1-\alpha)(1-\rho)}\eta_i\right] = 0. \tag{3.10}$$

Both equations state effectively that *deviations* of the initial conditions from the steady state behaviour are not systematically related to the *level* of the steady state itself. Under our hypothesised scenario in the earnings determination example, the expectation in (3.9) is likely to be negative as workers with higher innate ability (i.e. whose η_i value is relatively large) accumulate proportionately more experience, and thereby deviate to a greater extent from their steady state path of experience in the beginning of the sample, than workers with a small η_i value. Likewise, high skilled workers will systematically have lower wage in the beginning of the sample relative to their steady state earnings, in comparison with low skilled workers. It is clear that in order for the LEV moment conditions in (2.7) to be valid, one typically requires that *all* distinct covariates in a particular model satisfy a condition like (3.9) and the dependent variable satisfies (3.10).

The cce assumption is less strong than assuming that the series have a stationary mean. In other words, one can think of initial condition processes where the latter is not true but deviations from the steady state path remain uncorrelated with unobserved heterogeneity. It is useful to illustrate an example of such process with an application. Consider the empirics of growth models using country level data. The GMM methodology has been a popular estimation approach in this field. The Solow model takes the following form:

$$y_{it} - y_{i,t-1} = (\alpha - 1)\, y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + \eta_i + \lambda_t + \varepsilon_{it}, \tag{3.11}$$

where $y_{it} - y_{i,t-1}$ is the log difference in per capita GDP over a five year interval (t), $y_{i,t-1}$ denotes the logarithm of per capita GDP at the start of that period, and $\mathbf{x}$ is a vector that contains variables such as the logarithm of the investment rate and the population growth rate, while in its augmented form various measures of human capital are included. Among other things, η_i reflects differences in the level of initial endowment of physical capital and natural resources across countries, as well as geographical location and topography, while λ_t reflects changes in productivity that are common to all countries. An equivalent representation of (3.11) arises by adding $y_{i,t-1}$ on both sides, which resembles the standard dynamic panel data formulation as in (2.1). As we have already discussed, a sufficient condition for the first difference of per capita GDP, Δy_{it}, to be uncorrelated with η_i is mean stationarity of the level of per capita GDP, y_{it}, which also requires mean stationarity of the covariates used in the model. However, as Bond, Hoeffler and Temple (2001) point out, while the Solow model is consistent with stationary conditional means of investment rates and population growth rates, this is clearly not the case for the per capita GDP series. One possibility is to assume that the conditional mean of y_{it} shifts intertemporally in some arbitrary way due to common technological progress. This is in fact what is already implied in equation (3.11) by the inclusion of common time effects, λ_t. Because this procedure is equivalent to transforming the series in terms of deviations from time-specific averages, we may consider instead the transformed model

$$\underline{y}_{it} = \alpha \underline{y}_{i,t-1} + \boldsymbol{\beta}' \underline{\mathbf{x}}_{it} + \underline{\eta}_i + \underline{\varepsilon}_{it}, \tag{3.12}$$

where $\underline{y}_{it} = y_{it} - \overline{y}_t$, $\overline{y}_t = N^{-1} \sum_{i=1}^{N} y_{it}$, and so on. The effect of common technological progress has been eliminated. Thus, any arbitrary pattern in the conditional mean of per capita GDP over time that is due to technological progress would be consistent with the cce assumption, provided that this is satisfied for the transformed series.

Nevertheless, the discussion above hinges on the assumption that the two way error components formulation is adequate in explaining deviations from steady state behaviour. One might object that what drives changes in the conditional mean of per capita GDP over time is the extent to which countries manage to absorb advances in technology available. Since this is likely to be different across i, depending on existing constraints and the production capacity that each country faces, among other considerations, a factor structure in the error term may be more appropriate to deal with this problem. It is worth emphasising that a factor structure implies that changes in productivity are not common to all countries, and thereby deviations from steady state behaviour are not identical across i, which can be an empirically relevant scenario. GMM type methods for estimating dynamic panel data models with a factor structure in the residuals and short T, have been developed by Ahn, Lee and Schmidt (2010) and Robertson and Sarafidis (2013). Sarafidis and Wansbeek (2012) provide a recent overview of these methods. Panel data models with a factor structure are also discussed in Chapter 2 of this volume.

3.2. deviations of initial conditions from steady state behaviour

Consider the following initial condition processes for x and y respectively:

$$x_{i0} = \delta_x \frac{\tau\eta_i}{1-\rho} + w_{i0}, \tag{3.13}$$

$$y_{i0} = \delta_y \frac{(1-\rho+\beta\tau)\,\eta_i}{(1-\alpha)(1-\rho)} + e_{i0}, \tag{3.14}$$

which can be motivated by (3.7) and (3.8). The conditional mean of x at the in-sample start-up period is $E(x_{i0}|\eta_i) = \delta_x \frac{\tau\eta_i}{1-\rho}$ and the conditional mean of y is $E(y_{i0}|\eta_i) = \delta_y \frac{(1-\rho+\beta\tau)\eta_i}{(1-\alpha)(1-\rho)}$. Both δ_x and δ_y are meaningful in economic terms.[6] In particular $0 < \delta_x < 1$ implies that the conditional mean of the initial in-sample observation is closer to zero than its steady state path. Therefore, assuming $\tau > 0$, if $\eta_i > 0$ the series approaches its steady state from below and if $\eta_i < 0$ then it approaches from above, with the rate of convergence depending on ρ. Similarly, $\delta_x > 1$ implies that the value of the initial observation lies further away from zero than the series' long run conditional mean. Thus, if $\eta_i > 0$ ($\eta_i < 0$) the series converges from above (below). When $\delta_x = 1$ the series is mean stationary, i.e. its conditional mean is constant over time throughout the sampling period. In this case one can readily check that (3.9) is satisfied. When $\delta_y = 1$ as well, y is also mean stationary and (3.10) is fulfilled.

Under our hypothetical earnings determination scenario, one would expect that $0 \leq \delta_x < 1$ since experience increases gradually over time and $\eta_i > 0$. On the other hand, suppose that the initial model in (3.5) represents a cost function with y_{it} denoting total cost, x denoting output, together with input prices, and η_i capturing the effect of cost inefficiency (so one would anticipate $\eta_i \geq 0$). In this case one might expect that $\delta_y \geq 1$, i.e. firms' conditional expected cost in the beginning of the sample is at most equal to its long run mean, but not less. Under the hypothesis that firms adopt new work practices over time as an effort to cut expenditure, those firms which are economically more efficient are likely to be able to reduce total cost by a larger proportion. Hence, δ_y would be strictly larger than one in this case and the series would approach its steady-state level from above.

One can provide an alternative interpretation of δ_x and δ_y when the initial conditions are perfectly correlated with the steady state levels. In particular, setting $var(w_{i0}) = 0$ and $var(e_{i0}) = 0$, we have

$$\delta_x = \frac{\sqrt{var(x_{i0})}}{\tau\sigma_\eta/(1-\rho)}, \tag{3.15}$$

and

$$\delta_y = \frac{\sqrt{var(y_{i0})}}{(1-\rho+\beta\tau)\,\sigma_\eta/[(1-\alpha)(1-\rho)]}. \tag{3.16}$$

It can be seen that δ_x and δ_y equal the ratio of the standard deviation of the initial observations on x and y, respectively, over the standard deviation of the corresponding

[6] A deviation from mean stationarity may occur also as the result of a finite number of start-up periods.

steady state levels. When there is more dispersion in the initial conditions than in the distribution of the steady state levels, δ_x and δ_y will be larger than one. In the economic growth literature, for example, this property is known as sigma convergence.

3.3. consequences of departures from steady state behaviour

The magnitude of δ_y and δ_x turns out to be very important for the finite sample properties of various GMM estimators. For instance, for $\delta_y = 1$ the correlation between Δy_{it-1} and y_{is}, $s < t-1$ and $t = 2, ...T$, converges to zero when the variance of the η_i component of the error grows large. This is due to the fact that the total variation in y_{is} is dominated in this case by the variation in η_i, which, however, Δy_{it-1} is free from. This can have adverse consequences for GMM estimators that use lagged values of y in levels as instruments for first-differenced regressors when there is large unobserved heterogeneity present in the data. Hayakawa (2009), based on a pure AR(1) model, shows that the situation can be starkly different when $\delta_y \neq 1$. To see this, consider again for simplicity the case of an AR(1) model, i.e. impose $\beta = 0$ in (2.1), and $T = 2$, and consider the DIF moment condition given in (2.13). Assuming time series homoskedasticity for idiosyncratic errors, the covariance between Δy_{i1} and y_{i0} is then:

$$\begin{aligned} \operatorname{cov}(\Delta y_{i1}, y_{i0}) &= E\left[(\alpha y_{i0} + \eta_i + \varepsilon_{i1} - y_{i0})\, y_{i0}\right] \\ &= -(1-\alpha) E\left[y_{i0}^2\right] + E\left[\eta_i y_{i0}\right] \\ &= -(1-\alpha)\left[\delta_y^2 \frac{\sigma_\eta^2}{(1-\alpha)^2} + \frac{\sigma_\varepsilon^2}{1-\alpha^2}\right] + \delta_y \frac{\sigma_\eta^2}{1-\alpha} \\ &= -\frac{\sigma_\varepsilon^2}{1+\alpha} + \frac{\sigma_\eta^2}{1-\alpha}\delta_y(1-\delta_y). \end{aligned} \tag{3.17}$$

The first term is always negative while the sign of the second term depends on δ_y.[7] For $\delta_y > 1$ or $\delta_y < 0$ the second term is always negative and thereby the correlation between Δy_{i1} and y_{i0} that is due to η_i adds up to the correlation that is due to the idiosyncratic component. Thus, the instruments should become stronger. This is not necessarily true when $0 < \delta_y < 1$, since in this case the second term is positive and the total effect on the correlation between Δy_{i1} and y_{i0} depends on the relative magnitude of σ_η^2 and σ_ε^2, for a given value of α. Thus, for $\sigma_\eta^2/\sigma_\varepsilon^2 \to \infty$ any deviation from mean stationarity is likely to improve dramatically the performance of GMM estimators that do not require mean stationarity at first place, such as DIF and AS.

[7] Notice that when $\delta_y = 1$, the expression above depends only on σ_ε^2, for given α, which confirms that in this case Δy_{i1} is free from η_i. In this case we have the earlier result (2.19).

3.4. testing for constant-correlated effects

Moment conditions can be tested using the Sargan (1958)/Hansen (1982) overidentifying restrictions (OIR) statistic, which equals N times the value for the GMM objective function evaluated at the efficient two step GMM estimates. Asymptotically the OIR test statistic is chi-squared distributed with degrees of freedom equal to the number of overidentifying restrictions. It is clear that when there is no mean stationarity the linear moment conditions in (2.7) cannot be exploited. And an OIR test based on optimal system GMM should detect any deviations from assumption (2.6).

It has been shown, however, that the OIR test may be subject to low power due to many instruments (Bowsher, 2002; Windmeijer, 2005). Therefore, it is often suggested to use incremental or difference OIR tests. For example, assumption (2.6) implies extra moment conditions on top of those derived from (2.2). Hence, the difference between OIR SYS and DIF GMM statistics can be used, which is expected to have more discriminatory power compared with the SYS OIR test. Alternatively, the difference between the OIR SYS and AS GMM can be used, which is expected to have even better power properties. In the next section we will investigate by Monte Carlo simulation to what extent these and other predictions hold in finite samples.

4. simulation results

In this section we first set out our Monte Carlo design, which is inspired by those of Blundell, Bond and Windmeijer (2001), Bun and Kiviet (2006) and Hayakawa and Nagata (2012). We allow for deviations from mean stationarity and pay special attention to some of the rules described by Kiviet (2007, 2012) for enhancing the scope of a simulation study and the interpretation of simulation results. For example, many existing Monte Carlo designs in the dynamic panel data literature do not obey any orthogonalization of the parameter space, which may hamper the interpretation of simulation results across experiments. Next, we discuss existing Monte Carlo studies simulating under deviations from mean stationarity. Finally, we report new simulation results investigating the impact of deviations from mean stationarity on various GMM coefficient estimators, corresponding Wald tests and Sargan statistics.

4.1. Monte Carlo design

The data generating process (dgp) is given by (3.5) and (3.7), which we replicate here for convenience:

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \eta_i + \varepsilon_{it}, \ |\alpha| < 1, \tag{4.1}$$

with

$$x_{it} = \rho x_{i,t-1} + \tau \eta_i + \upsilon_{it}, \ |\rho| < 1, \qquad (4.2)$$
$$\upsilon_{it} = \nu_{it} + \phi_0 \varepsilon_{it} + \phi_1 \varepsilon_{i,t-1},$$

such that

$$\sigma_\upsilon^2 \equiv var(\upsilon_{it}) = \sigma_\nu^2 + \left(\phi_0^2 + \phi_1^2\right)\sigma_\varepsilon^2. \qquad (4.3)$$

The long run coefficient of x on y equals $\frac{\beta}{1-\alpha}$. The initial condition for x in (3.13) is specified as

$$x_{i0} = \delta_x \xi \eta_i + w_{i0}, \ w_{i0} \sim i.i.d.\left(0, \sigma_w^2\right), \ E\left(w_{i0}|\eta_i\right) = 0, \qquad (4.4)$$

where $\xi = \frac{\tau}{1-\rho}$ and $\sigma_w^2 = \frac{1}{1-\rho^2}\sigma_\upsilon^2$. $\xi\eta_i$ is the long run conditional mean, or steady state path, of x_{it} given η_i. Let r_x denote the correlation between the deviation of the initial condition of x from its long run steady state path and the level of the steady state path itself:

$$r_x = corr\left(x_{i0} - \xi\eta_i, \xi\eta_i\right) = \frac{\xi\left(\delta_x - 1\right)\sigma_\eta^2}{\sqrt{\left[\xi^2\left(\delta_x - 1\right)^2\sigma_\eta^2 + \sigma_w^2\right]\sigma_\eta^2}}. \qquad (4.5)$$

Solving for δ_x yields

$$\delta_x = \frac{r_x}{\left(1 - r_x^2\right)^{1/2}} \frac{\sigma_w}{\sigma_\eta} \frac{1}{\xi} + 1. \qquad (4.6)$$

Thus, instead of setting an arbitrary value of δ_x in order to investigate departures from steady state behaviour, as it is common practice in the literature, we can set δ_x according to r_x, which is more meaningful. For a fixed value of r_x, different values of σ_η^2 change δ_x. Similarly, larger values of σ_ν^2, and hence of σ_w^2, increase the signal-to-noise ratio of the model and so δ_x changes accordingly. When $r_x = 0$, $\delta_x = 1$ under the current design.[8]

We further specify the initial condition for y as

$$y_{i0} = \delta_y\left(\beta\xi + 1\right)\mu_i + e_{i0}, \ e_{i0} \sim i.i.d.\left(0, \sigma_e^2\right), \ E\left(e_{i0}|\eta_i\right) = 0, \qquad (4.7)$$

where $\mu_i = \eta_i/(1-\alpha)$; $(\beta\xi + 1)\mu_i$ is the long run conditional mean, or steady state path, of y_{it} given η_i. Thus, the process for y_{it} can be written as follows:

$$y_{it} = \varsigma_t \eta_i + \varpi_{it}, \qquad (4.8)$$

where

$$\varsigma_t = \frac{1}{1-\alpha}\left[\alpha^t\left(\delta_y - 1\right)\left(\beta\xi + 1\right) + \left(\beta\xi + 1\right)\right] + \beta\xi\left(\delta_x - 1\right)\sum_{s=0}^{t-1}\alpha^s\rho^{t-s}, \qquad (4.9)$$

[8] Notice that setting r_x equal to a fixed value, say $r_x = c$, also captures the case where the x process is mean-stationary ($\delta_x = 1$) for a proportion of individuals only. For example, if the proportion of individuals that satisfy $\delta_x = 1$ is .5, $r_x = 0$ for those individuals and $r_x = 2c$ for the remaining ones, provided that $|c| \leq 0.5$.

and

$$\varpi_{it} = \beta \sum_{s=0}^{t-1} \alpha^s w_{i,t-s} + \alpha^t e_{i0} + \sum_{s=0}^{t-1} \alpha^s \varepsilon_{i,t-s}. \tag{4.10}$$

Let r_y denote the correlation coefficient between the deviation of the initial condition of the y process from its long run mean and the level of its long run mean:

$$r_y = corr\left(y_{i0} - (\beta\xi + 1)\mu_i, (\beta\xi + 1)\mu_i\right) = \frac{(\delta_y - 1)\frac{1}{1-\alpha}(\beta\xi + 1)\sigma_\eta^2}{\sqrt{\left[(\delta_y - 1)^2(\beta\xi + 1)^2\left(\frac{1}{1-\alpha}\right)^2\sigma_\eta^2 + \sigma_\varpi^2\right]\sigma_\eta^2}}, \tag{4.11}$$

where $var(\varpi_{it}) = \sigma_\varpi^2 = \sigma_\nu^2 c_\nu^2 + \sigma_\varepsilon^2 c_\varepsilon^2$.[9] Solving for δ_y yields

$$\delta_y = \frac{r_y}{\left(1 - r_y^2\right)^{1/2}} \frac{\sigma_\varpi}{\sigma_\eta} \frac{1-\alpha}{\beta\xi + 1} + 1. \tag{4.12}$$

Thus, similarly to δ_x, δ_y is set according to r_y. Clearly large values of δ_y imply a high value for r_y, ceteris paribus.

Finally, as described in Kiviet (1995) and Bun & Kiviet (2006), the variances σ_v^2 and σ_η^2 are major determinants of the signal-to-noise ratio and the relative strength of the error components, respectively. The variance of y_{it} is equal to

$$\begin{aligned} var(y_{it}) &= \varsigma_t^2\sigma_\eta^2 + \sigma_\varpi^2 \\ &= \varsigma_t^2\sigma_\eta^2 + \sigma_\nu^2 c_\nu^2 + \sigma_\varepsilon^2 c_\varepsilon^2, \end{aligned} \tag{4.13}$$

A relationship between σ_η^2 and σ_ε^2 can be defined such that the cumulative impact on the average of $var(y_{it})$ over time of the two error components η_i and ε_{it} is equal to the 'variance ratio' (VR):

$$VR = \frac{\overline{\varsigma^2}\sigma_\eta^2}{\sigma_\varepsilon^2 c_\varepsilon^2}, \tag{4.14}$$

where $\overline{\varsigma^2} = T^{-1}\sum_{t=1}^T \varsigma_t^2$. Solving for σ_η^2 yields

$$\sigma_\eta^2 = \frac{\sigma_\varepsilon^2 c_\varepsilon^2 VR}{\overline{\varsigma^2}}. \tag{4.15}$$

Both c_ε^2 and $\overline{\varsigma^2}$ depend on the design parameters. Therefore, changes in these parameters will also affect the value of σ_η^2 for a fixed variance ratio VR.

[9] We have

$$c_\nu^2 = \frac{(1+\alpha\rho)\beta^2}{(1-\rho^2)(1-\alpha^2)(1-\alpha\rho)},$$

and

$$c_\varepsilon^2 = \frac{(1+\alpha\rho)}{(1-\rho^2)(1-\alpha^2)(1-\alpha\rho)}(1+\beta\phi_0)^2 + (\beta\phi_1 - \rho)^2 + 2\frac{(\alpha+\rho)}{1+\alpha\rho}(\beta\phi_1 - \rho)(1+\beta\phi_0).$$

Thus, similarly to the process for x, we have assumed that the idiosyncratic component in y, ϖ_{it}, is covariance-stationary.

Consider the variance of the signal of the model at time t, conditionally on η_i, which can be written as

$$\sigma^2_{signal,t} = var\left(y_{it}|\eta_i\right) - var\left(\varepsilon_{it}\right) = \sigma^2_{\varpi} - \sigma^2_{\varepsilon}. \tag{4.16}$$

The signal-to-noise ratio, defined conditionally on η_i, is now simply

$$SNR = \frac{\sigma^2_{\varpi} - \sigma^2_{\varepsilon}}{\sigma^2_{\varepsilon}}. \tag{4.17}$$

SNR depends on the value of σ^2_{ϖ}, which in turn is a function of σ^2_{ν}. Hence, we may set σ^2_{ν} such that SNR is controlled across experiments.

It should be noted that the proposed reparametrization of the parameter space to enhance the interpretability of Monte Carlo results is not unique. However, it follows closely the rules described in Kiviet (2007, 2012), notably the advice to reparametrize into an orthogonal autonomous design parameter space.

Setting $\phi_1 = 0$ in (4.2), the dgp in (4.1) and (4.2) is equal to that of Blundell, Bond and Windmeijer (2001). Setting $\phi_0 = 0$ in (4.2), the dgp in (4.1) and (4.2) is equal to one of the schemes analyzed in Bun and Kiviet (2006). One can choose the vector of parameters $(\delta_x, \delta_y, \sigma^2_{\eta}, \sigma^2_{\nu})$ by choosing values for (r_x, r_y, VR, SNR) or vice versa. The advantage of fixing (r_x, r_y, VR, SNR) is that we control some important model characteristics across experiments. The remaining parameters in the design are $(\alpha, \beta, \rho, \phi_0, \phi_1, \tau, \sigma^2_{\varepsilon})$ and the dimensions are (T, N).

4.2. existing results

Blundell and Bond (1998) already showed the vulnerability of system GMM to a deviation of the initial conditions from steady state behaviour for the AR(1) model. The specification of their initial condition is such that the implied $r_y \approx -0.65$ in their Table 6, which seems a quite strong deviation from cce. As a result the SYS GMM estimator of α has a large upward bias, while DIF GMM is virtually unbiased and compared with the mean stationary case gets a much smaller Monte Carlo standard deviation. SYS Wald statistics are heavily size distorted, while rejection frequencies of DIF Wald statistics are close to nominal significance levels. Finally, SYS Sargan has power to detect the violation of the cce assumption.

Hayakawa (2009) and Hayakawa and Nagata (2012) also provide simulation results for the AR(1) model. In Hayakawa (2009) only coefficient bias for the DIF GMM estimator is investigated, while Hayakawa and Nagata (2012) analyze other estimators as well and investigate finite sample properties of Sargan tests too. They find favorable behaviour of the DIF GMM estimator when $\delta_y \neq 1$. It should be noted, however, that these results are partly driven by the set-up of their Monte Carlo design. In particular, the variance of the individual effects, σ^2_{η}, instead of the variance ratio, VR, is fixed across experiments. This implies that when α is close to one the correlation between the endogenous regressor and the instrument suddenly becomes very large when $\delta_y \neq 1$; see result (3.17). In other words, minor deviations of initial conditions from steady state behaviour have a huge impact on the relevance of the instruments.

Regarding the dynamic panel data model with additional regressors[10] Everaert (2012) provides simulation results (coefficient bias and t test) for a model with an additional strictly exogenous covariate. In other words, $\phi_0 = \phi_1 = 0$ in (4.2). The only deviation from mean stationarity investigated is to set $y_{i0} = 0$. As a result SYS GMM becomes heavily biased, as expected.

Hayakawa and Nagata (2012) provide simulation results on coefficient bias for a model with an additional endogenous covariate. Also Sargan and incremental Sargan tests are investigated. They closely follow the design of Blundell, Bond and Windmeijer (2001), i.e. $\phi_0 \neq 0$ and $\phi_1 = 0$ in (4.2). DIF GMM shows favorable results when cce is violated, but the Monte Carlo design is again such that the individual effects dominate the idiosyncratic disturbances when persistence is high. In other words, important model characteristics like the variance ratio can achieve rather extreme values.

In Kiviet (1995) and Bun and Kiviet (2006) it has been shown that a proper comparison of simulation results over different parameter values requires control over basic model characteristics like goodness of fit and relative strength of error components. In the above design these are quantified by SNR and VR respectively, which in turn determine the values of the variances σ_ν^2 and σ_η^2. It is instructive to analyze what implied values for VR and SNR other studies have chosen. Blundell, Bond and Windmeijer (2001) choose $\beta = 1, \tau = 0.25, \phi_0 = -0.1, \phi_1 = 0, \sigma_\eta^2 = 1, \sigma_\varepsilon^2 = 1$ and $\sigma_\nu^2 = 0.16$. Furthermore, they consider four designs with α and ρ either 0.5 or 0.95. Choosing $\alpha = \rho = 0.5$ implies $SNR = 0.48$ and $VR = 9$. Increasing ρ to 0.95 results in $SNR = 6.36$ and $VR = 119$, a large proportional increase in both signal-to-noise ratio and variance ratio. Setting α equal to 0.95 (but keeping $\rho = 0.5$) results in $SNR = 11.88$ and $VR = 134$, again a large increase in both signal and relative strength of individual effects. Finally, increasing both α and ρ to 0.95 results in $SNR = 337.17$ and $VR = 1478$, a huge increase in both variance ratio and signal-to-noise ratio. It is clear from these calculations that changing the autoregressive dynamics has substantial consequences for both explained variation and unobserved heterogeneity in the model. For proper comparison across experiments it is therefore necessary to control at least VR and SNR, but preferably other model characteristics as well. Similar calculations can be made for the Monte Carlo design of Hayakawa and Nagata (2012), which is basically that of Blundell, Bond and Windmeijer (2001), but choosing σ_η^2, δ_y and δ_x different from 1 too.

[10] Juodis (2013) provides simulation results under mean nonstationarity for panel VAR models.

4.3. new simulation results

We report results for the within group or Least Squares Dummy Variable (LSDV) estimator, DIF, AS and SYS estimators. We report coefficient bias (bias), standard deviation (sd) and root mean squared error (rmse) as well as rejection frequencies (rf) of nominal 5% Wald significance tests and overidentifying restrictions (OIR) tests. The GMM Wald tests use the variance correction of Windmeijer (2005), since it is well known (Arellano and Bond, 1991) that two step GMM variance estimators are heavily downward biased. We also apply this finite sample variance correction to the nonlinear moment conditions of Ahn and Schmidt (1995). It can be expected that this leads to an improvement in the estimation of variances, although theoretically that is only the case for linear moment conditions (Windmeijer, 2005). The incremental OIR tests are based on either the difference between SYS and AS, or between SYS and DIF.[11]

One problem with existing simulation results is that a comparison across experiments is hampered by the fact that typically more than one model characteristic is changed. Furthermore, it seems that the chosen design parameters often imply rather extreme values for VR, SNR, r_y and r_x. Therefore, we control for these four model characteristics across experiments.

Regarding the error components we specify η_i and ε_{it} *i.i.d.* $N\left(0,\sigma_\eta^2\right)$ and $N\left(0,\sigma_\varepsilon^2\right)$ respectively, with $\sigma_\varepsilon^2 = 1$ and σ_η^2 determined by (4.15). We consider $VR = \{3, 100\}$, and we set $SNR = 3$. We have also experimented with $SNR = 9$ and generally the precision of all GMM estimators improves substantially for sufficiently high values of SNR. For $VR = 3$, we report simulation results for four parameter configurations: (1) $\alpha = 0.2, \rho = 0.5$; (2) $\alpha = 0.2, \rho = 0.95$; (3) $\alpha = 0.8, \rho = 0.5$; (4) $\alpha = 0.8, \rho = 0.95$. For $VR = 100$, we report results only for (4) $\alpha = 0.8, \rho = 0.95$, in order to save space.[12]

We set $\beta = 1 - \alpha$ across all experiments, so that the long run effect of x on y is one. Following Blundell, Bond and Windmeijer (2001) we fix $\tau = 0.25$, $\phi_0 = -0.1$ and $\phi_1 = 0$. We have also experimented with other types of endogeneity: (1) $\tau = 0$, i.e. no correlation between η_i and x_{it}; (2) $\phi_0 = 0$, $\phi_1 = -0.1$, i.e. weak exogeneity. The results are qualitatively similar for these cases. Finally, we only report results for $T = 3$ and $N = 500$. For smaller N larger biases are seen for all GMM estimators. A larger value of T introduces instrument proliferation issues, as discussed in Section 2.

The pattern of the 9 columns within each table is: (1) baseline of cce, i.e. $r_y = r_x = 0$; (2) $r_y = 0.5$; (3) $r_y = -0.5$; (4) $r_x = 0.5$; (5) $r_x = -0.5$; (6) $r_y = 0.5, r_x = 0.5$; (7) $r_y = 0.5, r_x = -0.5$; (8) $r_y = -0.5, r_x = 0.5$; (9) $r_y = -0.5, r_x = -0.5$. Hence, columns (2)-(9) investigate all possible combinations of deviations from the cce assumption.

[11]For LSDV, DIF and SYS estimation we use the DPD for Ox package (Doornik et al., 2006). AS estimation is based on our own Ox code.

[12]The results for the remaining configurations are available on the author's website.

The following observations can be made regarding bias and precision of coefficient estimators:

1. LSDV coefficient bias is negative.

2. Unless it is negligible, DIF GMM coefficient bias is almost always negative.

3. SYS GMM coefficient bias can have either sign, but tends to be positive in most cases.

4. Under cce, coefficient biases for all GMM estimators are larger for $VR = 100$ showing their lack of invariance to σ_η^2.

5. SYS GMM coefficient bias for α is always larger under deviations from the cce assumption. In a few cases, however, it happens that coefficient bias is smaller for β, most notably in Table 5, where both y and x are highly persistent, unless r_y, r_x are both negative.

6. DIF and AS GMM coefficient biases are affected under deviations from the cce assumption, but there is no clear pattern in the simulation results. Hence, benefits for the location of the DIF GMM estimator may or may not occur depending on the particular parameter configuration.

7. AS GMM often performs equally well or better compared with SYS GMM. This somewhat remarkable result appears to hold even under cce. Actually the only case in which AS GMM is noticeably outperformed by SYS GMM in terms of bias is Table 3, column 1.

8. When $\alpha = 0.8$ (Tables 3-5) DIF GMM can have large coefficient bias or large standard deviation or both, indicating a weak instrument problem.

9. In Tables 1, 2 and 3 SYS GMM has often smaller standard deviation than AS GMM, which in turn has smaller dispersion than DIF GMM; this is unless there is moderate persistence, in which case DIF and AS GMM have similar or smaller standard deviation than SYS GMM.

10. Under cce, a weak instrument problem seems present for SYS GMM too when there is strong unobserved heterogeneity, i.e. $VR = 100$. Although coefficient bias seems limited, in relative terms (i.e. compared with the standard deviation of the estimator) it becomes large. This is consistent with the results in Bun and Windmeijer (2010), who show that also for moderate autoregressive dynamics LEV moment conditions may become less informative when VR gets large.

Regarding rejection frequencies of Wald statistics the following can be observed:

1. LSDV size distortions are large. For hypothesis testing on α the actual rejection frequency is always 1.

2. DIF GMM rejection frequencies under the null hypothesis are close to the nominal significance level of 5%. In those cases where there appears to be a size distortion, it is probably caused by the weak instrument problem, as documented above.

3. The same weak instruments problem appears to hold for SYS GMM. Note that even under cce size distortions can be large.

4. The vulnerability of SYS GMM to deviations from cce is obvious. Rejection frequencies under the null hypothesis can become 1.

5. AS GMM rejection frequencies under the null hypothesis are often close to the nominal significance level of 5%, but sometimes size distortions appear.

Finally, regarding OIR test statistics the following observations can be made:

1. The performance of DIF and AS OIR test statistics under the null hypothesis is satisfactory. We didn't examine power, but it can be expected that power is low when persistence is high.

2. The performance of the SYS OIR test statistic under the null hypothesis is satisfactory. Also its power is high in case of moderate persistence (Table 1). However, as can be seen from Tables 4-5, power is low when there is a lot of persistence in both y and x. In between, it depends on the particular deviation from cce.

3. Similar conclusions hold for incremental SYS-DIF and SYS-AS statistics, but they outperform SYS OIR statistic in terms of power.

4. No clear ranking exists between SYS-AS and SYS-DIF statistics although the former has often a slightly higher rejection frequency under the alternative hypothesis.

5. Perhaps surprisingly, sometimes OIR SYS, SYS-AS and SYS-DIF tests have complete lack of power against deviations from cce. Sometimes this is even the case when coefficient bias in the SYS GMM estimator is relatively large, e.g. Table 5, column 9, or Table 2, column 3.

5. concluding remarks

In this Chapter we have reviewed the literature on dynamic panel data models estimated by GMM. We have focused on the analysis of GMM estimators in dynamic models with additional endogenous regressors. We have discussed in detail the assumptions underlying the validity of, especially, the system GMM estimator. Furthermore, we have embarked on the consequences of violation of mean stationarity for several GMM estimators. In cases where the constant correlated effects assumption is violated, individual-specific unobserved heterogeneity is only partially removed by taking first differences. Obviously, lagged differenced instruments for the model in levels are then not exogenous anymore, therefore invalidating the system GMM estimator. Additionally, the relevance of the lagged level instruments for the first-differenced model changes in a nontrivial manner. Apart from mean stationarity we have discussed briefly a number of other practical issues when applying GMM inference methods, e.g. how to determine the optimal number of moment conditions.

Our simulation results indicate that no universal ranking exists among first-differenced (DIF), non-linear (AS) and system (SYS) GMM estimators. Some general observations can be made. First, DIF GMM has low precision and coefficient bias, especially when the series are persistent. Second, SYS GMM is vulnerable to nuisance parameters, and its performance deteriorates rapidly under deviations from cce. Even when absolute coefficient bias seems small, large size distortion can still occur. Third, the AS GMM estimator performs quite satisfactory in most experiments. It has higher precision than DIF GMM and only moderate coefficient bias and size distortion. Compared with SYS GMM, however, its root mean squared error is relatively large when the series are persistent. Fourth, in testing for cce all OIR tests appear to lack power in case of high persistence.

Summarizing, GMM estimators for dynamic panel data models can be vulnerable to important nuisance parameters and weak identification issues. Until recently, system GMM has been considered to be the solution to the first-differenced GMM estimator in case of persistent panel data. However, its additional restriction on the initial conditions has been criticized for being unrealistic precisely in case of persistent panel data. Additionally, tests for cce lack power when having persistent panel data and/or an abundance of moment conditions. This may lead to acceptance of the levels moment conditions when this is not appropriate. But even in case of mean stationarity inference based on system GMM may be inaccurate. A straightforward advise for practitioners regarding which method to prefer in small samples does not emerge, but the non-linear AS GMM estimator seems a relatively safe choice. It is robust to deviations from cce, and more efficient than first-differenced linear GMM.

Table 1: Simulation results for $\alpha = 0.2$, $\rho = 0.5$ and $VR = 3$

		1	2	3	4	5	6	7	8	9
	σ_η	0.922	0.879	0.961	0.770	1.067	0.723	1.028	0.814	1.102
	σ_v	1.698	1.698	1.698	1.698	1.698	1.698	1.698	1.698	1.698
	r_y	0.000	0.500	-0.500	0.000	0.000	0.500	0.500	-0.500	-0.500
	r_x	0.000	0.000	0.000	0.500	-0.500	0.500	-0.500	0.500	-0.500
bias α	lsdv	-0.153	-0.132	-0.132	-0.150	-0.150	-0.141	-0.122	-0.121	-0.141
	dif	-0.007	-0.003	-0.006	-0.004	-0.004	-0.004	-0.003	-0.001	-0.009
	as	-0.000	-0.001	0.001	-0.000	-0.002	-0.001	-0.002	-0.000	0.000
	sys	0.003	0.112	0.311	0.226	0.135	0.043	0.204	0.152	0.086
bias β	lsdv	-0.046	-0.045	-0.045	-0.048	-0.048	-0.039	-0.052	-0.052	-0.039
	dif	0.004	-0.001	0.000	-0.001	0.002	0.001	-0.001	-0.004	0.009
	as	-0.003	-0.002	-0.005	-0.004	0.000	-0.002	-0.001	-0.006	-0.001
	sys	-0.005	-0.012	-0.200	-0.139	-0.014	-0.126	-0.036	0.154	0.077
sd α	lsdv	0.021	0.019	0.019	0.020	0.020	0.020	0.019	0.018	0.020
	dif	0.065	0.037	0.058	0.050	0.044	0.052	0.028	0.028	0.072
	as	0.050	0.035	0.038	0.041	0.041	0.046	0.027	0.027	0.051
	sys	0.047	0.053	0.062	0.049	0.058	0.054	0.040	0.036	0.051
sd β	lsdv	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021
	dif	0.103	0.087	0.092	0.074	0.075	0.095	0.067	0.066	0.110
	as	0.093	0.085	0.085	0.072	0.073	0.088	0.067	0.066	0.092
	sys	0.085	0.121	0.138	0.110	0.116	0.097	0.111	0.099	0.092
rmse α	lsdv	0.154	0.134	0.133	0.152	0.152	0.143	0.123	0.122	0.142
	dif	0.065	0.037	0.059	0.050	0.044	0.053	0.028	0.028	0.073
	as	0.050	0.035	0.038	0.041	0.041	0.046	0.027	0.027	0.051
	sys	0.047	0.124	0.317	0.231	0.147	0.069	0.208	0.157	0.100
rmse β	lsdv	0.051	0.049	0.049	0.052	0.052	0.044	0.056	0.056	0.044
	dif	0.103	0.087	0.092	0.074	0.075	0.095	0.067	0.066	0.110
	as	0.093	0.085	0.086	0.072	0.073	0.088	0.067	0.066	0.092
	sys	0.085	0.121	0.243	0.177	0.117	0.160	0.116	0.183	0.120
rf α	lsdv	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	dif	0.044	0.047	0.045	0.045	0.054	0.044	0.057	0.047	0.043
	as	0.042	0.044	0.039	0.044	0.051	0.043	0.060	0.039	0.040
	sys	0.048	0.669	1.000	0.997	0.730	0.129	1.000	0.982	0.420
rf β	lsdv	0.599	0.563	0.563	0.630	0.631	0.459	0.696	0.700	0.464
	dif	0.044	0.050	0.040	0.040	0.047	0.041	0.051	0.051	0.044
	as	0.048	0.061	0.045	0.045	0.053	0.045	0.062	0.055	0.050
	sys	0.048	0.118	0.462	0.339	0.120	0.280	0.160	0.443	0.151
rf OIR	dif	0.051	0.055	0.049	0.052	0.056	0.052	0.057	0.052	0.052
	as	0.045	0.052	0.048	0.055	0.051	0.050	0.052	0.052	0.050
	sys	0.047	1.000	1.000	1.000	1.000	0.955	1.000	1.000	0.982
	sys-as	0.058	1.000	1.000	1.000	1.000	0.990	1.000	1.000	0.998
	sys-dif	0.052	1.000	1.000	1.000	1.000	0.986	1.000	1.000	0.995

Note: $\beta = 1 - \alpha$, $\sigma_\varepsilon^2 = 1$, $T = 3$, $N = 500$, $\phi_0 = -0.1$, $\phi_1 = 0$, $\tau = 0.25$ and $SNR = 3$.

Table 2: Simulation results for $\alpha = 0.2$, $\rho = 0.95$ and $VR = 3$

		1	2	3	4	5	6	7	8	9
	σ_η	0.268	0.256	0.279	0.130	0.406	0.119	0.394	0.141	0.417
	σ_v	0.552	0.552	0.552	0.552	0.552	0.552	0.552	0.552	0.552
	r_y	0.000	0.500	-0.500	0.000	0.000	0.500	0.500	-0.500	-0.500
	r_x	0.000	0.000	0.000	0.500	-0.500	0.500	-0.500	0.500	-0.500
bias α	lsdv	-0.333	-0.246	-0.242	-0.269	-0.273	-0.329	-0.158	-0.154	-0.328
	dif	-0.006	-0.004	-0.002	-0.003	-0.008	-0.003	-0.004	-0.002	-0.013
	as	0.004	0.001	0.002	0.000	0.002	0.002	-0.000	-0.000	0.007
	sys	-0.001	0.020	0.020	0.010	0.043	-0.002	0.038	0.005	0.014
bias β	lsdv	-0.221	-0.231	-0.226	-0.251	-0.254	-0.212	-0.273	-0.268	-0.209
	dif	-0.083	-0.064	-0.075	-0.053	-0.075	-0.044	-0.061	-0.057	-0.154
	as	-0.019	-0.008	-0.048	-0.046	0.001	-0.026	-0.012	-0.047	-0.004
	sys	0.007	0.166	0.154	0.063	0.270	-0.008	0.305	0.079	0.112
sd α	lsdv	0.028	0.026	0.025	0.026	0.027	0.028	0.021	0.020	0.028
	dif	0.052	0.037	0.036	0.039	0.046	0.047	0.029	0.026	0.067
	as	0.047	0.036	0.034	0.036	0.040	0.045	0.027	0.025	0.050
	sys	0.041	0.035	0.035	0.035	0.039	0.042	0.025	0.024	0.043
sd β	lsdv	0.068	0.068	0.068	0.068	0.068	0.067	0.068	0.068	0.067
	dif	0.505	0.427	0.470	0.365	0.439	0.350	0.399	0.377	0.668
	as	0.287	0.280	0.283	0.262	0.270	0.266	0.267	0.267	0.284
	sys	0.105	0.061	0.052	0.048	0.076	0.146	0.071	0.026	0.083
rmse α	lsdv	0.334	0.247	0.243	0.270	0.274	0.330	0.159	0.156	0.329
	dif	0.052	0.037	0.036	0.039	0.047	0.048	0.030	0.026	0.069
	as	0.047	0.036	0.034	0.036	0.040	0.045	0.027	0.025	0.050
	sys	0.042	0.041	0.040	0.036	0.058	0.042	0.045	0.024	0.045
rmse β	lsdv	0.231	0.241	0.236	0.260	0.263	0.223	0.282	0.277	0.219
	dif	0.511	0.432	0.476	0.369	0.445	0.353	0.404	0.381	0.686
	as	0.288	0.280	0.287	0.266	0.270	0.268	0.267	0.271	0.284
	sys	0.105	0.176	0.163	0.079	0.280	0.147	0.314	0.083	0.139
rf α	lsdv	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	dif	0.046	0.049	0.041	0.046	0.054	0.049	0.052	0.041	0.047
	as	0.047	0.039	0.039	0.043	0.037	0.049	0.046	0.047	0.049
	sys	0.056	0.087	0.083	0.057	0.211	0.051	0.359	0.047	0.067
rf β	lsdv	0.914	0.927	0.918	0.961	0.963	0.894	0.979	0.973	0.879
	dif	0.041	0.047	0.040	0.051	0.043	0.043	0.046	0.041	0.034
	as	0.156	0.146	0.154	0.132	0.140	0.131	0.138	0.136	0.162
	sys	0.043	0.819	0.843	0.254	0.973	0.043	0.996	0.856	0.260
rf OIR	dif	0.050	0.048	0.052	0.056	0.050	0.058	0.050	0.056	0.047
	as	0.069	0.064	0.066	0.070	0.068	0.068	0.065	0.071	0.071
	sys	0.048	0.223	0.199	0.075	0.606	0.049	0.579	0.085	0.066
	sys-as	0.048	0.332	0.279	0.090	0.741	0.049	0.736	0.102	0.066
	sys-dif	0.057	0.300	0.265	0.086	0.718	0.049	0.705	0.100	0.083

Note: see Table 1.

Table 3: Simulation results for $\alpha = 0.8$, $\rho = 0.5$ and $VR = 3$

		1	2	3	4	5	6	7	8	9
	σ_η	0.507	0.383	0.630	0.478	0.536	0.354	0.411	0.601	0.660
	σ_v	2.015	2.015	2.015	2.015	2.015	2.015	2.015	2.015	2.015
	r_y	0.000	0.500	-0.500	0.000	0.000	0.500	0.500	-0.500	-0.500
	r_x	0.000	0.000	0.000	0.500	-0.500	0.500	-0.500	0.500	-0.500
bias α	lsdv	-0.564	-0.536	-0.536	-0.561	-0.561	-0.551	-0.517	-0.517	-0.551
	dif	-0.049	-0.017	-0.286	-0.042	-0.038	-0.018	-0.017	-0.111	-0.102
	as	-0.011	-0.012	-0.016	-0.012	-0.014	-0.012	-0.012	-0.011	-0.010
	sys	-0.002	0.050	0.110	0.111	0.176	0.069	0.152	0.161	0.128
bias β	lsdv	-0.062	-0.060	-0.060	-0.065	-0.064	-0.055	-0.070	-0.070	-0.055
	dif	-0.010	-0.005	-0.059	-0.011	-0.008	-0.004	-0.007	-0.044	-0.006
	as	-0.004	-0.005	-0.006	-0.005	-0.004	-0.004	-0.006	-0.008	-0.002
	sys	-0.002	-0.027	0.014	-0.038	-0.042	-0.046	-0.028	-0.015	0.019
sd α	lsdv	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.030	0.031
	dif	0.153	0.087	0.383	0.149	0.129	0.097	0.082	0.240	0.227
	as	0.110	0.080	0.102	0.109	0.101	0.088	0.077	0.102	0.116
	sys	0.056	0.076	0.029	0.023	0.047	0.029	0.027	0.020	0.047
sd β	lsdv	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016
	dif	0.067	0.064	0.105	0.062	0.061	0.056	0.063	0.106	0.058
	as	0.061	0.062	0.060	0.056	0.057	0.055	0.060	0.062	0.054
	sys	0.045	0.045	0.046	0.046	0.052	0.047	0.043	0.046	0.055
rmse α	lsdv	0.565	0.537	0.537	0.562	0.562	0.551	0.518	0.518	0.551
	dif	0.160	0.089	0.478	0.155	0.134	0.098	0.084	0.265	0.248
	as	0.110	0.081	0.104	0.110	0.102	0.089	0.078	0.102	0.116
	sys	0.056	0.091	0.114	0.113	0.182	0.074	0.154	0.163	0.136
rmse β	lsdv	0.064	0.062	0.062	0.067	0.066	0.057	0.072	0.072	0.057
	dif	0.068	0.064	0.121	0.063	0.062	0.056	0.063	0.115	0.058
	as	0.061	0.062	0.061	0.057	0.057	0.055	0.060	0.062	0.054
	sys	0.045	0.052	0.048	0.059	0.067	0.066	0.051	0.048	0.058
rf α	lsdv	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	dif	0.068	0.058	0.150	0.067	0.064	0.058	0.060	0.081	0.094
	as	0.101	0.077	0.128	0.117	0.099	0.082	0.076	0.139	0.119
	sys	0.061	0.153	0.940	0.986	0.909	0.661	0.999	1.000	0.789
rf β	lsdv	0.975	0.965	0.965	0.981	0.979	0.932	0.989	0.988	0.932
	dif	0.052	0.055	0.088	0.050	0.055	0.049	0.055	0.082	0.046
	as	0.065	0.067	0.057	0.061	0.065	0.057	0.067	0.067	0.050
	sys	0.044	0.081	0.058	0.144	0.174	0.175	0.114	0.072	0.068
rf OIR	dif	0.059	0.052	0.065	0.055	0.056	0.055	0.052	0.059	0.059
	as	0.075	0.058	0.082	0.077	0.072	0.063	0.059	0.083	0.078
	sys	0.049	0.312	0.060	0.237	0.820	0.211	0.503	0.332	0.803
	sys-as	0.050	0.477	0.061	0.345	0.910	0.315	0.670	0.479	0.900
	sys-dif	0.050	0.416	0.074	0.305	0.896	0.279	0.622	0.431	0.883

Note: see Table 1.

Table 4: Simulation results for $\alpha = 0.8$, $\rho = 0.95$ and $VR = 3$

		1	2	3	4	5	6	7	8	9
	σ_η	0.268	0.200	0.336	0.237	0.299	0.169	0.229	0.304	0.367
	σ_v	0.438	0.438	0.438	0.438	0.438	0.438	0.438	0.438	0.438
	r_y	0.000	0.500	-0.500	0.000	0.000	0.500	0.500	-0.500	-0.500
	r_x	0.000	0.000	0.000	0.500	-0.500	0.500	-0.500	0.500	-0.500
bias α	lsdv	-0.657	-0.620	-0.615	-0.641	-0.645	-0.651	-0.573	-0.565	-0.648
	dif	-0.056	-0.021	-0.099	-0.033	-0.069	-0.019	-0.026	-0.039	-0.217
	as	-0.015	-0.010	-0.015	-0.015	-0.016	-0.011	-0.012	-0.016	-0.020
	sys	-0.005	0.037	0.040	0.031	-0.006	0.000	0.081	0.036	0.015
bias β	lsdv	-0.458	-0.457	-0.448	-0.476	-0.477	-0.442	-0.504	-0.494	-0.434
	dif	-0.302	-0.131	-0.696	-0.170	-0.310	-0.083	-0.154	-0.290	-0.925
	as	-0.063	-0.048	-0.086	-0.058	-0.062	-0.044	-0.063	-0.069	-0.085
	sys	0.005	0.087	0.081	0.025	0.088	-0.013	0.087	0.076	0.120
sd α	lsdv	0.032	0.032	0.031	0.032	0.032	0.032	0.031	0.031	0.032
	dif	0.148	0.090	0.201	0.117	0.158	0.097	0.090	0.113	0.305
	as	0.095	0.079	0.088	0.090	0.094	0.087	0.073	0.077	0.103
	sys	0.054	0.063	0.039	0.041	0.074	0.056	0.042	0.043	0.047
sd β	lsdv	0.076	0.077	0.076	0.076	0.076	0.076	0.078	0.077	0.075
	dif	0.833	0.578	1.424	0.603	0.771	0.457	0.583	0.824	1.368
	as	0.344	0.345	0.340	0.329	0.330	0.322	0.339	0.344	0.312
	sys	0.131	0.124	0.080	0.110	0.109	0.222	0.116	0.066	0.088
rmse α	lsdv	0.657	0.621	0.616	0.642	0.646	0.652	0.574	0.566	0.649
	dif	0.158	0.092	0.224	0.122	0.173	0.098	0.094	0.119	0.375
	as	0.096	0.079	0.090	0.091	0.095	0.087	0.074	0.079	0.105
	sys	0.054	0.073	0.056	0.051	0.074	0.056	0.091	0.056	0.050
rmse β	lsdv	0.464	0.464	0.454	0.482	0.483	0.449	0.509	0.500	0.440
	dif	0.886	0.593	1.585	0.627	0.832	0.465	0.603	0.874	1.652
	as	0.350	0.349	0.351	0.334	0.335	0.324	0.345	0.351	0.323
	sys	0.132	0.152	0.114	0.113	0.140	0.223	0.145	0.101	0.149
rf α	lsdv	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	dif	0.065	0.055	0.062	0.061	0.088	0.057	0.068	0.055	0.138
	as	0.067	0.047	0.074	0.079	0.066	0.051	0.069	0.100	0.095
	sys	0.048	0.089	0.196	0.127	0.058	0.047	0.534	0.144	0.074
rf β	lsdv	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	dif	0.063	0.055	0.057	0.059	0.079	0.052	0.061	0.048	0.113
	as	0.147	0.147	0.162	0.100	0.164	0.108	0.164	0.110	0.190
	sys	0.043	0.088	0.162	0.057	0.115	0.059	0.075	0.227	0.241
rf OIR	dif	0.052	0.057	0.042	0.053	0.059	0.053	0.053	0.049	0.045
	as	0.102	0.083	0.107	0.103	0.108	0.083	0.082	0.098	0.113
	sys	0.048	0.089	0.055	0.050	0.122	0.050	0.133	0.058	0.051
	sys-as	0.041	0.105	0.050	0.053	0.143	0.042	0.191	0.058	0.037
	sys-dif	0.062	0.108	0.090	0.063	0.159	0.051	0.174	0.075	0.099

Note: see Table 1.

Table 5: Simulation results for $\alpha = 0.8$, $\rho = 0.95$ and $VR = 100$

		1	2	3	4	5	6	7	8	9
	σ_η	1.549	1.480	1.617	1.518	1.579	1.450	1.511	1.586	1.647
	σ_v	0.438	0.438	0.438	0.438	0.438	0.438	0.438	0.438	0.438
	r_y	0.000	0.500	-0.500	0.000	0.000	0.500	0.500	-0.500	-0.500
	r_x	0.000	0.000	0.000	0.500	-0.500	0.500	-0.500	0.500	-0.500
bias α	lsdv	-0.655	-0.633	-0.600	-0.631	-0.652	-0.655	-0.591	-0.545	-0.641
	dif	-0.120	-0.145	-0.018	-0.161	-0.126	-0.093	-0.179	-0.051	-0.020
	as	0.006	-0.006	-0.002	-0.014	0.014	-0.001	-0.018	-0.013	0.009
	sys	0.015	0.067	0.043	0.051	0.053	0.035	0.067	0.047	0.020
bias β	lsdv	-0.450	-0.474	-0.418	-0.464	-0.474	-0.455	-0.523	-0.462	-0.408
	dif	-0.591	-0.812	-0.142	-0.997	-0.447	-0.501	-0.890	-0.517	-0.055
	as	0.057	-0.003	0.031	0.013	0.060	0.048	-0.077	-0.030	0.051
	sys	0.132	0.093	0.116	0.079	0.121	0.139	0.092	0.103	0.167
sd α	lsdv	0.032	0.032	0.031	0.032	0.032	0.032	0.031	0.031	0.032
	dif	0.235	0.235	0.102	0.245	0.246	0.192	0.252	0.142	0.120
	as	0.102	0.143	0.083	0.109	0.117	0.114	0.169	0.080	0.094
	sys	0.049	0.047	0.036	0.039	0.061	0.051	0.040	0.037	0.044
sd β	lsdv	0.076	0.076	0.076	0.076	0.075	0.076	0.077	0.078	0.075
	dif	1.190	1.357	0.735	1.491	0.901	1.034	1.289	1.285	0.437
	as	0.330	0.614	0.309	0.377	0.342	0.434	0.702	0.375	0.267
	sys	0.081	0.085	0.072	0.082	0.106	0.145	0.085	0.069	0.082
rmse α	lsdv	0.655	0.634	0.601	0.632	0.653	0.656	0.592	0.546	0.642
	dif	0.264	0.276	0.103	0.294	0.277	0.214	0.309	0.151	0.122
	as	0.102	0.143	0.083	0.109	0.117	0.114	0.170	0.081	0.094
	sys	0.051	0.082	0.056	0.064	0.081	0.062	0.078	0.060	0.048
rmse β	lsdv	0.457	0.480	0.425	0.471	0.480	0.461	0.528	0.469	0.414
	dif	1.329	1.581	0.749	1.794	1.006	1.149	1.566	1.385	0.441
	as	0.335	0.614	0.310	0.377	0.347	0.436	0.707	0.376	0.272
	sys	0.155	0.126	0.136	0.114	0.160	0.201	0.126	0.124	0.185
rf α	lsdv	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	dif	0.083	0.084	0.044	0.081	0.105	0.076	0.111	0.039	0.050
	as	0.047	0.054	0.053	0.069	0.044	0.034	0.093	0.116	0.060
	sys	0.087	0.305	0.256	0.282	0.145	0.128	0.418	0.283	0.102
rf OIR	dif	0.047	0.044	0.052	0.037	0.063	0.051	0.039	0.044	0.057
	as	0.073	0.068	0.073	0.071	0.073	0.068	0.071	0.060	0.069
	sys	0.053	0.076	0.056	0.053	0.109	0.060	0.085	0.062	0.052
	sys-as	0.045	0.085	0.059	0.062	0.137	0.064	0.105	0.079	0.055
	sys-dif	0.073	0.111	0.072	0.093	0.143	0.077	0.124	0.091	0.060

Note: see Table 1.

References

Ahn, S.C., Lee, Y.H. and P. Schmidt (2013), Panel Data Models with Multiple Time-Varying Individual Effects. *Journal of Econometrics* 174, 1-14.

Ahn, S.C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5-27.

Alvarez, J. and M. Arellano (2003). The time series and cross-sectional asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121-1159.

Anderson, T.W. and C. Hsiao (1981), Estimation of dynamic models with error components, *Journal of the American Statistical Association* 76, 598-606.

Anderson, T.W. and C. Hsiao (1982), Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47-82.

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

Arellano, M. (2003b). Modeling Optimal Instrumental Variables for Dynamic Panel Data Models, Working Paper 0310, Centro de Estudios Monetarios y Financieros, Madrid.

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

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68, 29-51.

Arellano, M. and B. Honoré (2001). Panel data models: some recent developments. In J. Heckman and E. Leamer (eds.): *Handbook of Econometrics*, Volume 5.

Baltagi, B.H. (2008). Econometric analysis of panel data. *Wiley.*

Bekker, P.A. (1994). Alternative approximations to the distributions of Instrumental Variable estimators. *Econometrica* 62, 657-681.

Besley, T. and A. Case (2000). Unnatural experiments? Estimating the incidence of endogenous policies. *The Economic Journal* 110, F672-F694.

Bhargava, A. and J.D. Sargan (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51, 1635-1659.

Binder, M., Hsiao, C. and M. H. Pesaran (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795–837.

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

Blundell, R. and S. Bond (2000). GMM Estimation with persistent panel data: an application to production functions. *Econometric Reviews* 19, 321-340.

Blundell, R., Bond, S. and F. Windmeijer (2001). Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. In: B.H. Baltagi, T.B. Fomby, R. Carter Hill (eds.), Nonstationary Panels, Panel Cointegration, and Dynamic

Panels. Advances in Econometrics, Volume 15, Emerald Group Publishing Limited, 53-91.

Bond, S., Hoefller, A. and J. Temple (2001). GMM estimation of empirical growth models. Economics group working paper 2001-W21, University of Oxford.

Bond, S. and F. Windmeijer (2005). Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews* 24, 1-37.

Bound, J., D.A. Jaeger and R.M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90, 443-450.

Bowsher, C. G. (2002). On testing overidentifying restrictions in dynamic panel data models. *Economics Letters* 77, 211–220.

Bun, M.J.G. and M.A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics* 23, 200-210.

Bun, M.J.G. and J.F. Kiviet (2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132, 409-444.

Bun, M.J.G. and F. Windmeijer (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal* 13, 95-126.

Bun, M.J.G. and F. Kleibergen (2013). Identification and inference in moments based analysis of linear dynamic panel data models. *UvA-Econometrics discussion paper* 2013/07, University of Amsterdam.

Dhaene, G. and K. Jochmans (2012). An adjusted profile likelihood for non-stationary panel data models with fixed effects. Working paper, KU Leuven.

Doornik, J.A., Arellano, M. and S. Bond (2006). Panel data estimation using DPD for Ox. mimeo, University of Oxford.

Everaert, G. (2013). Orthogonal to Backward Mean Transformation for Dynamic Panel Data Models. Forthcoming in *Econometrics Journal*.

Hahn, J., Hausman, J. and G. Kuersteiner (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics* 140, 574-617.

Han, C. and P.C.B. Phillips (2006). GMM with many moment conditions. *Econometrica* 74, 147-192.

Han, C. and P.C.B. Phillips (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory* 26, 119-151.

Han, C. Phillips, P.C.B. and D. Sul (2010). X-Differencing and dynamic panel model estimation. Cowles Foundation Discussion Papers 1747, Yale University.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054.

Hause, J.C. (1980). The fine structure of earnings and the on-the-job training hypothesis. *Econometrica* 48, 1013-1029.

Hayakawa, K. (2009). On the effect of mean-nonstationarity in dynamic panel data models. *Journal of Econometrics* 153, 133-135.

Hayakawa, K. and S. Nagata (2012). On the Behavior of the GMM Estimator in Persistent Dynamic Panel Data Models with Unrestricted Initial Conditions. Working paper, Hiroshima University.

Hayakawa, K. and M.H. Pesaran (2012). Robust standard errors in transformed likelihood estimation of dynamic panel data models. CWPE 1224, University of Cambridge.

Holtz-Eakin, D., Newey, W. and H.S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56, 1371–1395.

Hsiao, C. (2003). Analysis of panel data. *Cambridge University Press.*

Hsiao, C., Pesaran, M.H. and A.K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109,107–150.

Juodis, A. (2013). First difference transformation in panel VAR models: robustness, estimation and inference. *UvA-Econometrics discussion paper* 2013/06, University of Amsterdam.

Kiviet, J.F. (1995). On bias, inconsistency and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 1995, 68, 53-78.

Kiviet, J.F. (2007). Judging contending estimators by simulation: tournaments in dynamic panel data models. In: *The Refinement of Econometric Estimation and Test Procedures* (eds.: G.D.A. Phillips and E. Tzavalis), 282-318. Cambridge University Press.

Kiviet, J.F. (2012). Monte Carlo Simulation for Econometricians. *Foundations and Trends in Econometrics*: Vol. 5.

Kiviet, J.F., Pleus, M. and R. Poldermans (2013). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models. Mimeo, University of Amsterdam.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica* 73, 1103-1124.

Koenker, R. and J.A.F. Machado (1999). GMM inference when the number of moment conditions is large. Journal of Econometrics 93, 327-344.

Kruiniger, H. (2009). GMM estimation and inference in dynamic panel data models with persistent data. *Econometric Theory* 25, 1348-1391.

Lancaster, T. (2002), Orthogonal parameters and panel data, *Review of Economic Studies* 69, 647-666.

Mátyás, L. and P. Sevestre (2008). The Econometrics of panel data. *Springer.*

Moreira, M.J. (2009). A maximum likelihood method for the incidental parameter problem.

Annals of Statistics, 37, 3660–3696.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49, 1417–1426.

Newey, W.K. and K.D. West (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28, 777-787.

Neyman, J. and E.L. Scott (1948), Consistent estimates based on partially consistent observations. *Econometrica*, 16, 1–32.

Robertson, D. and V. Sarafidis (2013), IV Estimation of Panels with Factor Residuals. Mimeo.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71, 135-158.

Sarafidis, V. (2011). GMM estimation of short dynamic panel data models with error cross-sectionalal dependence. MPRA Paper No. 36154.

Sarafidis V., and T. Wansbeek (2012) cross-sectionalal Dependence in Panel Data Analysis. *Econometric Reviews* 31, 483-531.

Sargan, J.D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393-415.

Staiger, D. and J.H. Stock (1997). Instrumental Variables regression with weak instruments. *Econometrica* 65, 557-586.

Stock, J.H. and J.H. Wright (2000). GMM with weak identification. *Econometrica* 68, 1055-1096.

Stock, J.H., Wright, J.H. and M. Yogo (2002). A survey of weak instruments and weak identification in Generalized Method of Moments, *Journal of Business & Economic Statistics*, 518-529.

Wansbeek, T. J., and T. Knaap (1999). Estimating a dynamic panel data model with heterogenous trends. *Annales d' Economie et de Statistique* 55–56, 331–350.

Windmeijer, F. (2005), A finite sample correction for the variance of linear efficient two step GMM estimators. *Journal of Econometrics* 126, 25-517.

Ziliak, J.P. (1997). Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators. *Journal of Business & Economic Statistics* 15, 419-431.