UvA ECONOMETRICS

Discussion Paper: 2013/07

Identification and inference in moments based analysis of linear dynamic panel data models

Maurice J.G. Bun and Frank Kleibergen

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

Identification and inference in moments based analysis of linear dynamic panel data models

Maurice Bun* Frank Kleibergen†‡

September 17, 2014

Abstract

We show that Dif(ference), see Arellano and Bond (1991), Lev(el), see Arellano and Bover (1995) and Blundell and Bond (1998), or the N(on-)L(inear) moment conditions of Ahn and Schmidt (1995) do not separately identify the parameters of a first-order autoregressive panel data model when the autoregressive parameter is close to one and the variance of the initial observations is large. We, however, construct a new set of (robust) moment conditions that identifies the autoregressive parameter irrespective of the variance of the initial observations. These robust moment conditions are (non-linear) combinations of the System (Sys) and A(hn-)S(chmidt) moment conditions. We use them to construct the maximal attainable power curve under the worst case setting, which implies a quartic root convergence rate. It is identical for the AS and Sys moment conditions so assuming mean stationarity does not improve power in worst case settings. We compare the maximal attainable power curve under the worst case setting with the lower envelopes of the power curves of different GMM test procedures. These power envelopes show the lowest rejection frequencies of these test procedures. The power envelope of the K(leibergen) L(agrange) M(ultiplier) statistic of Kleibergen (2005) coincides with the maximal attainable power curve under the worst case setting so the KLM statistic is efficient both when the autoregressive parameter is one or less than one. Our results extend to general values of the autoregressive parameter for which identification fails when the variance of the individual specific effects becomes large.

*Department of Economics & Econometrics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands. Email: M.J.G.Bun@uva.nl. Homepage: http://www.uva.nl/profile/m.j.g.bun.
The research of the author has been funded by the NWO Vernieuwingsimpuls research grant 'Causal Inference with Panel Data'.

†Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States. Email: Frank_Kleibergen@brown.edu. Homepage: http://www.econ.brown.edu/fac/Frank_Kleibergen.

‡We thank Manuel Arellano, Richard Blundell, Steve Bond, Peter Boswijk, Geert Dhaene, Frank Windmeijer and participants of seminars at Bristol, CEMFI, CORE, the Cowles Summer Conference at Yale, the EC^2 meeting in Maastricht, Groningen, Leuven, Monash, the 19-th International conference on panel data in London, Oxford, the Tinbergen Institute in Amsterdam and UCL for helpful comments and stimulating discussion.

1 Introduction

It is common to estimate the parameters of linear dynamic panel data models using the Generalized Method of Moments (Hansen, 1982). The moment conditions for the linear dynamic panel data model either analyze it in first differences using lagged levels of the series as instruments, in levels using lagged first differences as instruments or using a combination of levels and first differences. We refer to the first set of moment conditions as Dif(ference) moment conditions, see Arellano and Bond (1991), the second set as Lev(el) moment conditions, see Arellano and Bover (1995), Blundell and Bond (1998) and the third set as N(on-)L(inear) moment conditions, see Ahn and Schmidt (1995).

The Dif, Lev and NL moment conditions can be used separately to identify the parameters of dynamic panel data models. In order to exhaust all information, however, two particular combinations of Dif, Lev and NL moment conditions have been proposed. We refer to the combination of the Dif and Lev moment conditions as the Sys(tem) moment conditions and the combination of the Dif and NL moment conditions as the A(hn-)S(chmidt) moment conditions.[1] The Sys moment conditions exhaust all information on the autoregressive parameter that is present under mean stationarity, see Arellano and Bover (1995) and Blundell and Bond (1998). The AS moment conditions exhaust all information whilst not assuming mean stationarity, see Ahn and Schmidt (1995).

We analyze the various moment conditions when the panel data are highly persistent. All moment conditions involve first differences of the series to remove the individual specific effects. The first difference operator removes information in the time series at the unit root value of the autoregressive parameter. It is well known that the Dif moment conditions therefore do not identify the autoregressive parameter when its true value is (close to) one, since lagged levels are then weak predictors of first differences. This has led to the development of the NL and Lev, and hence AS and Sys, moment conditions which originally were considered to identify the autoregressive parameter when the panel data are highly persistent.

We show that identification of the autoregressive parameter by Dif, NL or Lev moment conditions separately depends on the setting of the initial observations, individual specific effects and variances of the disturbances. We show that these affect the identification of the autoregressive parameter and that none of the previous moment conditions identifies the autoregressive parameter for all specifications of these nuisance parameters. For a range of relative convergences rates of the variance of the initial observations compared to the sample size, the Dif, Lev and NL sample moments and their derivatives diverge. Both the population moment and the Jacobian identification condition are then ill defined which implies that the autoregressive parameter is not identified. These results confirm and extend earlier findings

[1]Note that in a combination of all three sets of moments conditions, the NL moment conditions are redundant.

in Madsen (2003), Bond *et. al*. (2005), Hahn *et. al*. (2007) and Kruiniger (2009).

We construct a new combination of the Dif, Lev and NL moment conditions (other than AS and Sys) that does not depend on the initial observations. These novel robust moment conditions therefore identify the parameters irrespective of the setting of the nuisance parameters which obviously includes the case of highly persistent data. The novel robust moments are suboptimal under less severe settings of the nuisance parameters so we refrain from using them for estimation. Instead, we use them to obtain optimal inference procedures under the worst case settings of the nuisance parameters.

Our first main contribution is to show that the novel robust moment conditions are spanned by either the Sys or AS moments. Hence, the Sys and AS moments identify the parameters irrespective of the setting of the nuisance parameters when there are more than three time periods. It is remarkable that the AS moments always lead to identification since it implies that the assumption of mean stationarity is redundant. Despite that, the large sample distributions of one step and two step generalized method of moments (GMM) estimators are non-standard under worst case settings of the nuisance parameters. This explains their large biases and the size distortions of their corresponding t-statistics when the series are persistent, see *e.g.* Madsen (2003), Bond and Windmeijer (2005), Bond *et. al*. (2005), Dhaene and Jochmans (2012), Hahn *et. al*. (2007), Kruiniger (2009) and Bun and Windmeijer (2010).

We next determine which GMM test procedure, that remains size correct under worst case settings of the nuisance parameters, is optimal. This excludes one and two step t-statistics since they are obviously size distorted. The size correct GMM statistics that we therefore analyze are the GMM-A(nderson-)R(ubin) statistic of Anderson and Rubin (1949) and Stock and Wright (2000), the GMM-L(agrange-)M(ultiplier) statistic of Newey and West (1987) and the K(leibergen)LM statistic of Kleibergen (2005). Both GMM-LM and KLM statistics are LM or score statistics so they are optimal when the autoregressive parameter is less than one. To determine if any of these three test statistics is optimal under the worst case setting of the nuisance parameters, we construct the lower envelope of their power curves to which we refer as the power envelope. The power envelope shows the lowest attainable power which results from the worst case specification of the nuisance parameters.

Under the worst case specifications only the novel robust moments contain information on the autoregressive parameter. We therefore use them to construct the maximal attainable power curve under a worst case setting. It implies a quartic root convergence rate which further reflects the non-standard manner in which the autoregressive parameter is identified by the moment conditions. We compare the maximal attainable power curve with the power envelopes of the large sample distributions of the GMM-AR, GMM-LM and KLM statistics based on Sys or AS moments. In doing so, we provide an extension of Andrews *et. al*. (2006) from the linear instrumental variables regression model with one included endogenous parameter towards the panel autoregressive model of order one.

Our second main result is that the power envelope of the KLM statistic is on the maximal attainable power curve for all number of time periods. Thus the KLM statistic is optimal in worst case settings. Hence, since it is also optimal in all other settings, it is optimal in general for inference in the dynamic panel data model. This is a somewhat different conclusion compared with Andrews *et. al.* (2006), who do not recommend the KLM test for practical use.[2]

A final interesting outcome from the asymptotic power analysis is that the power envelopes that result for the AS and Sys moment conditions coincide. This shows that assuming mean stationarity is not only redundant for identification, but also does not add any further identifying information about the autoregressive parameter in worst case settings.

Throughout the analysis, we use an asymptotic sampling scheme in which we let both the variance of the initial observations and the number of cross section observations get large. In dynamic panel data models, the variance of the initial observations can be large due to a unit root value of the autoregressive parameter or because of a large variance of the individual specific effects. In the latter case, the identification of the autoregressive parameter also fails at values of the autoregressive parameter smaller than one. Although we, in our analysis, mainly focus on values of the autoregressive parameter close to one, all our results apply as well to the case of a large individual specific effect variance.

The paper is organized as follows. Section 2 introduces the linear dynamic panel data model and the different moment conditions we use to identify its parameters. In Section 3, we introduce our asymptotic sampling scheme, and as an illustration show that the Dif and Lev moment conditions with three time periods do not identify the autoregressive parameter. In Section 4, we use a representation theorem, akin to the cointegration representation theorem, see Engle and Granger (1987) and Johansen (1991), to pin down the identification properties of the different moment conditions for the general case. In Section 5, we construct the robust moments conditions and the maximal attainable power curve under worst case settings. We also briefly discuss the extension to a large individual specific effect variance. In Section 6, we construct the power envelopes of different GMM test procedures. The final section concludes. Proofs of theorems and definitions of test statistics are provided in Appendices A and B respectively. We use the following notation throughout the paper: $\underset{a}{=}$ means asymptotically equivalent, $\underset{p}{\rightarrow}$ indicates convergence in probability, and $\underset{d}{\rightarrow}$ indicates convergence in distribution.

[2] Compared with Andrews *et. al.* (2006) we didn't analyze the conditional LR test. For the panel data model, however, Newey and Windmeijer (2009) report in their simulation study that KLM and conditional LR statistics have similar power properties.

2 Moment conditions

We analyze the first-order dynamic panel data model

$$y_{it} = \quad c_i + \theta y_{it-1} + u_{it} \qquad i = 1, \ldots, N,\ t = 2, \ldots, T, \tag{1}$$

with T the number of time periods and N the number of cross section observations. For expository purposes, we analyze the simple dynamic panel data model in (1) which can be extended with additional lags of y_{it} and/or explanatory variables.[3] Estimation of the parameter θ by means of least squares leads to a biased estimator in samples with a finite value of T, see *e.g.* Nickell (1981). We therefore estimate it using GMM. We obtain the GMM moment conditions from the unconditional moment assumptions:

$$\begin{aligned} E[u_{it}u_{is}] = &\quad 0, \qquad s \neq t;\ t = 2, \ldots, T, \\ E[u_{it}c_i] = &\quad 0, \qquad t = 2, \ldots, T, \\ E[u_{it}y_{i1}] = &\quad 0, \qquad t = 2, \ldots, T. \end{aligned} \tag{2}$$

Under these assumptions, the moments of the $T(T-1)$ interactions of Δy_{it} and y_{it} :

$$E[\Delta y_{it} y_{ij}] \qquad j = 1, \ldots, T,\ t = 2, \ldots, T \tag{3}$$

can be used to construct functions which identify the parameter of interest θ. We do not use products of Δy_{it} to identify θ since we would need further assumptions, *i.e.* homoscedasticity or initial condition assumptions, see *e.g.* Han and Phillips (2010).

Two different sets of moment conditions, which are functions of the moments in (3), are commonly used to identify θ :

1. Difference (Dif) moment conditions:

$$E[y_{ij}(\Delta y_{it} - \theta \Delta y_{it-1})] = 0 \qquad j = 1, \ldots, t-2;\ t = 3, \ldots, T, \tag{4}$$

as proposed by *e.g.* Anderson and Hsiao (1981) and Arellano and Bond (1991). The Dif moment conditions solely result from the conditions in (2).

2. Level (Lev) moment conditions:

$$E[\Delta y_{it-1}(y_{it} - \theta y_{it-1})] = 0 \qquad t = 3, \ldots, T, \tag{5}$$

as proposed by Arellano and Bover (1995), see also Blundell and Bond (1998). Besides the conditions in (2), the Lev moment conditions use

$$E\left[\Delta y_{it} c_i\right] = 0, \tag{6}$$

[3] The extension to other explanatory variables would depend on the nature of these. For some settings such an extension would be trivial but for others not so.

which implies that the original data in levels have constant correlation over time with the individual-specific effects. This assumption implies the following for y_{i1}:

$$y_{i1} = \quad \mu_i + u_{i1}, \quad i = 1, \dots, N, \tag{7}$$

with $c_i = \mu_i(1 - \theta_0)$, which is often referred to as mean stationarity.

The Dif and Lev moments can be used separately or jointly to identify θ. When we use the moment conditions in (4) and (5) jointly, we refer to them as system (Sys) moment conditions,[4] see Arellano and Bover (1995) and Blundell and Bond (1998). Another set of nonlinear (NL) moment conditions, which just like the Dif moments only use the conditions in (2), results from Ahn and Schmidt (1995):

$$E[(y_{it} - \theta y_{it-1})(\Delta y_{it-1} - \theta \Delta y_{it-2})] = 0 \qquad t = 4, \dots, T. \tag{8}$$

The NL moments can be used separately or jointly with the Dif moments to identify θ. When we use the moment conditions in (4) and (8) jointly, we refer to them as Ahn-Schmidt (AS) moment conditions. Ahn and Schmidt (1995) show that these (combined) AS moment conditions exhaust the information on θ in the moment conditions (2) and are therefore complete. Mean stationarity adds one moment condition (6) to the moment conditions in (2). Hence, the complete set of moment conditions under (2) and (6) equals the AS moment conditions and (6). Upon rewriting we can show that these combined moment conditions are identical to the Sys moment conditions so they are complete under (2) and (6).

The Dif moment conditions do not identify θ when its true value is equal to one while the Lev moment conditions are supposed to do, see Arellano and Bover (1995) and Blundell and Bond (1998). Also the NL (and hence AS) moment conditions are considered to identify θ when its true value is one but since these moment conditions are quadratic in θ, they are less commonly used than the linear Dif, Lev, and Sys moment conditions, see Ahn and Schmidt (1995). The identification results in Blundell and Bond (1998) and Ahn and Schmidt (1995) are, however, silent about their sensitivity with respect to the initial observations. In the next Section we analyze the (non-) robustness with respect to the initial observations in more detail for $T = 3$ time periods, and show that identification by the Lev moment conditions is arbitrary in case of highly persistent panel data.

3 Identification: An illustrative example

In this Section we introduce our asymptotic sampling scheme which consists of drifting sequences for the autoregressive parameter and variance of the initial observation. Furthermore,

[4] We could extend the Lev moment conditions to $\frac{1}{2}(T-1)(T-2)$ sample moments by including additional interactions of Δy_{it-j} and $y_{it} - \theta y_{it-1}$, for $j = 2, \dots, t-2$. It can be shown, however, that all conditions on top of those in (5) can be constructed as linear combinations of the Dif conditions in (4) and the Lev conditions in (5).

we use it in an illustrative example to analyze the limiting behavior of Dif and Lev moment conditions when $T = 3$.[5]

The Dif and Lev moment conditions that we use to identify θ are semi-parametric with respect to the individual specific fixed effects, variances and initial observations so they identify θ for a variety of different specifications of them. These specifications, however, influence the identification of θ for persistent values of it, *i.e.* values that are close to one.[6] To exemplify this, we first consider the simplest setting which has T equal to three. We also note that, since we use the Lev moment conditions, we assume mean stationarity (6)-(7).

When there are three time series observations, the Dif and Lev moment conditions read:

$$\begin{array}{lrl} \text{Dif:} & E[y_{i1}(\Delta y_{i3} - \theta \Delta y_{i2})] & = 0 \\ \text{Lev:} & E[\Delta y_{i2}(y_{i3} - \theta y_{i2})] & = 0, \end{array} \tag{9}$$

with Jacobians:

$$\begin{array}{ll} \text{Dif:} & -E[y_{i1}\Delta y_{i2}] = -E((\mu_i + u_{i1})((\theta_0 - 1)u_{i1} + u_{i2}) \\ \text{Lev:} & -E[y_{i2}\Delta y_{i2}] = -E((c_i + \theta_0 y_{1i} + u_{i2})((\theta_0 - 1)u_{i1} + u_{i2})), \end{array} \tag{10}$$

where θ_0 is the true value of θ. For many data generating processes for the initial observations, the Jacobian of the Dif moment condition in (10) is equal to zero when θ_0 is equal to one.[7] The Dif moment condition does then not identify θ when θ_0 is equal to one for these DGPs. Under mean stationarity (6)-(7), the Jacobian of the Lev moment condition is such that

$$E(y_{i2}\Delta y_{i2}) = \quad (\theta_0 - 1)\theta_0 E(u_{i1}^2) + E(u_{i2}^2) \neq 0, \quad \text{when } \theta_0 = 1, \tag{11}$$

so the Lev moment condition seems to identify θ irrespective of the value of θ_0, see Arellano and Bover (1995) and Blundell and Bond (1998). There is a caveat though since for many data generating processes y_{i1} does not have a finite mean and/or variance when θ_0 is equal to one and, despite that y_{i1} and u_{i2} are uncorrelated, we then do not know the value of $E(y_{i1}u_{i2})$ which is both an element of the moment equation (upon recurrent substitution) in (9) and the Jacobian in (11). To ascertain the identification of θ by the Lev moment conditions when θ_0 is equal to one, we therefore consider a joint limit process where both θ_0 converges to one and the sample size goes to infinity. In order to do so, we first make a technical assumption about the variance of the idiosyncratic part of the initial observation under mean stationarity (6)-(7).

[5] We postpone discussion of the Sys moment conditions for $T = 3$ until the next Section, while the NL, and hence also AS, moment conditions are available for $T > 3$ only.

[6] In Section 5.1, we show that similar identification problems occur when θ is smaller than one, but the variance of the individual specific effects is large.

[7] Exceptions are when mean stationarity (6)-(7) does not hold, see *e.g.* Hayakawa (2009), or, for example, in case of covariance stationarity so $\text{var}(u_{i1}) = \frac{\sigma^2}{1-\theta_0^2}$.

Assumption 1. *The limit behavior of the variance of* $(1-\theta_0)u_{i1}$, *with* u_{i1} *the disturbance in the mean stationarity conditions (6)-(7), when* θ_0 *goes to one is such that*

$$E(\lim_{\theta_0\uparrow 1}((1-\theta_0)u_{i1})^2)=0. \tag{12}$$

Assumption 1 is necessary for the Dif and Lev moment conditions to hold when $\theta_0=1$ and mean stationarity (6)-(7) applies.

Furthermore, we make an assumption on the variance of the product of the initial observation y_{i1} and the disturbances u_{it} :

Assumption 2.

$$var(u_{it}y_{i1})=var(u_{it})var(y_{i1}),\ t=2,\ldots,T. \tag{13}$$

A condition under which this assumption holds is independence of u_{it} and y_{i1} but it can also hold under less stringent conditions. In the sequel, we analyze the identification of θ when the variance of the initial observations gets large compared to that of the subsequent disturbances. The assumption in (13) enables such settings.

We analyze the large sample behavior of the Lev sample moment, $\frac{1}{N}\sum_{i=1}^N\Delta y_{i2}(y_{i3}-\theta y_{i2})$, and its derivative, $-\frac{1}{N}\sum_{i=1}^N y_{i2}\Delta y_{i2}$, when θ_0 converges to one (we rule out explosive values of θ_0) and mean stationarity (6)-(7) applies. In order to do so we first list their relevant elements for the large sample behavior under some DGP for the initial observations:

$$\begin{array}{rl}\lim_{\theta_0\uparrow 1}\frac{1}{N}\sum_{i=1}^N\Delta y_{i2}(y_{i3}-\theta y_{i2})\approx & (1-\theta)\left\{\frac{1}{N}\sum_{i=1}^N u_{i2}^2+\lim_{\theta_0\uparrow 1}\frac{1}{N}\sum_{i=1}^N u_{i2}y_{i1}+\right.\\ & \left.\lim_{\theta_0\uparrow 1}\frac{1}{N}\sum_{i=1}^N(1-\theta_0)u_{i1}y_{i1}\right\}\\ \lim_{\theta_0\uparrow 1}\frac{1}{N}\sum_{i=1}^N y_{i2}\Delta y_{i2}\approx & \frac{1}{N}\sum_{i=1}^N u_{i2}^2+\lim_{\theta_0\uparrow 1}\frac{1}{N}\sum_{i=1}^N u_{i2}y_{i1}+\\ & \lim_{\theta_0\uparrow 1}\frac{1}{N}\sum_{i=1}^N(1-\theta_0)u_{i1}y_{i1}.\end{array} \tag{14}$$

We dropped all elements in (14) that do not affect the large sample behavior when θ_0 goes to one,[8] at least not under our drifting parameter sequences as defined below. What is left over are terms with non-zero mean and/or depending on the initial observations y_{i1}. Since u_{i2} and y_{i1} are uncorrelated and under (13), it holds that

$$h(\theta_0)\frac{1}{\sqrt{N}}\sum_{i=1}^N u_{i2}y_{i1}\underset{d}{\rightarrow}\psi_2, \tag{15}$$

with ψ_2 a normal random variable with mean zero and variance σ_2^2 =var(u_{i2}) and $h(\theta_0)^{-2}$ =var(y_{i1}).

We analyze a setting in which both θ_0 and the variance of the initial observations are functions of the sample size which we indicate by $\theta_{0,N}$ and $h_N(\theta_{0,N})$ respectively. When the sample size gets large, these sequences behave according to

$$\begin{array}{rl}\lim_{N\to\infty}\theta_{0,N}= & 1\\ \lim_{N\to\infty}h_N(\theta_{0,N})= & d,\end{array} \tag{16}$$

[8]This explains why we use the "≈" sign instead of the "=" sign.

with d a finite possibly zero constant. The sequences in (16) allow the variance of the initial observations to be large in combination with a large value for the autoregressive parameter. The limit sequences in (16) are such that (15) remains to hold so

$$h_N(\theta_{0,N})\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N u_{i2}y_{i1} \underset{d}{\rightarrow} \psi_2, \tag{17}$$

and which explains why $\frac{1}{N}\sum_{i=1}^N u_{i2}y_{i1}$ appears in (14).

When d in (16) equals zero, the rate at which $h_N(\theta_{0,N})$ goes to zero, or the variance of the initial observation goes to infinity, determines the behavior of the sample moments in (14). For example, when these sequences are such that

$$h_N(\theta_{0,N})\sqrt{N} \underset{N\rightarrow\infty}{\rightarrow} \infty, \tag{18}$$

it holds that

$$\begin{array}{rcl} \frac{1}{N}\sum_{i=1}^N y_{i2}\Delta y_{i2} & = & \frac{1}{N}\sum_{i=1}^N u_{i2}^2 + \frac{1}{h_N(\theta_{0,N})\sqrt{N}}\left[h_N(\theta_{0,N})\frac{1}{\sqrt{N}}\sum_{i=1}^N u_{i2}y_{i1}\right] + \\ & & \frac{1}{N}\sum_{i=1}^N(1-\theta_{0,N})u_{i1}y_{i1} \\ & \underset{p}{\rightarrow} & \sigma_2^2 + \lim_{N\rightarrow\infty} E((1-\theta_{0,N})u_{i1}^2), \end{array} \tag{19}$$

while when

$$h_N(\theta_{0,N})\sqrt{N} \underset{N\rightarrow\infty}{\rightarrow} 0, \tag{20}$$

the large sample behaviors of the Lev sample moment and its Jacobian are characterized by

$$\begin{array}{rcl} h_N(\theta_{0,N})\frac{1}{\sqrt{N}}\sum_{i=1}^N \Delta y_{i2}(y_{i3}-\theta y_{i2}) & = & (1-\theta)\left\{h_N(\theta_{0,N})\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^N u_{i2}^2\right] + \right. \\ & & h_N(\theta_{0,N})\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^N(1-\theta_{0,N})u_{i1}y_{i1}\right] + \\ & & \left.\left[h_N(\theta_{0,N})\frac{1}{\sqrt{N}}\sum_{i=1}^N u_{i2}y_{i1}\right]\right\} \\ & \underset{d}{\rightarrow} & (1-\theta)\psi_2 \\ h_N(\theta_{0,N})\frac{1}{\sqrt{N}}\sum_{i=1}^N y_{i2}\Delta y_{i2} & = & h_N(\theta_{0,N})\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^N u_{i2}^2\right] + \\ & & h_N(\theta_{0,N})\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^N(1-\theta_{0\,N})u_{i1}y_{i1}\right] + \\ & & h_N(\theta_{0\,N})\frac{1}{\sqrt{N}}\sum_{i=1}^N u_{i2}y_{i1} \\ & \underset{d}{\rightarrow} & \psi_2. \end{array} \tag{21}$$

Hence, when (20) holds:

$$\begin{array}{rl} \frac{1}{N}\sum_{i=1}^N \Delta y_{i2}(y_{i3}-\theta y_{i2}) = & \frac{1}{h_N(\theta_{0,N})\sqrt{N}}\frac{h_N(\theta_{0,N})}{\sqrt{N}}\sum_{i=1}^N \Delta y_{i2}(y_{i3}-\theta y_{i2}) \\ \underset{N\rightarrow\infty}{\rightarrow} & \infty \\ \frac{1}{N}\sum_{i=1}^N y_{i2}\Delta y_{i2} = & \frac{1}{h_N(\theta_{0,N})\sqrt{N}}\frac{h_N(\theta_{0,N})}{\sqrt{N}}\sum_{i=1}^N y_{i2}\Delta y_{i2} \\ \underset{N\rightarrow\infty}{\rightarrow} & \infty \end{array} \tag{22}$$

so the sample moments of the Lev population moment and Jacobian go to infinity when the sample size increases. It implies that the Lev population moment and Jacobian are not defined. In this case we conclude that θ is not identified.

Since any assumption about the convergence rates of the sample size and the variance of the initial observations is kind of arbitrary, also the identification of θ by the Lev moment conditions is arbitrary for DGPs for which θ_0 is close to one and the variance of the initial observations is infinite when θ_0 equals one. Some plausible DGPs, all of which accord with mean stationarity (6)-(7), for the initial observations belong to this category:

DGP 1. $\sigma_c^2 = \text{var}(c_i)$, $h(\theta_{0,N}) = (1-\theta_{0,N})/\sigma_c$.

DGP 2. $\sigma_c^2 = \text{var}(c_i)$, $\sigma_1^2 = \frac{\sigma^2}{1-\theta_{0,N}^2}$, $h(\theta_{0,N}) = (1-\theta_{0,N})/\sigma_c$.

DGP 3. $\sigma_\mu^2 = \text{var}(\mu_i)$, $\sigma_1^2 = \frac{\sigma^2}{1-\theta_{0,N}^2}$, $h(\theta_{0,N}) = \frac{1}{\sigma}\sqrt{1-\theta_{0,N}^2}$.

DGP 4. $\sigma_\mu^2 = \text{var}(\mu_i)$, $\sigma_1^2 = \sigma^2 \frac{1-\theta_{0,N}^{2(g+1)}}{1-\theta_{0,N}^2}$, $h(\theta_{0,N}) = \frac{1}{\sigma}\sqrt{\frac{1-\theta_{0,N}^2}{1-\theta_{0,N}^{2(g+1)}}}$.

DGP 5. $\sigma_c^2 = \text{var}(c_i)$, $\sigma_1^2 = \sigma^2 \frac{1-\theta_{0,N}^{2(g+1)}}{1-\theta_{0,N}^2}$, $h(\theta_{0,N}) = (1-\theta_{0,N})/\sigma_c$.

DGPs 4 and 5 characterize an autoregressive process of order one that has started g periods in the past while the initial observations that result from DGP 2 and 3 result from an autoregressive process that has started an infinite number of periods in the past. DGPs 2 and 3 are also used by Blundell and Bond (1998) while Arellano and Bover (1995) use DGP 2.

For DGPs 1-5 to accord with (20), the limiting sequence $\theta_{0,N}$ (16) is such that:

$$\begin{array}{llcll} \text{DGP } 1,\ 2,\ 5: & (1-\theta_{0,N})\sqrt{N} & \underset{N\to\infty}{\to} 0 & \text{or} & \theta_{0,N} = 1 - \frac{e}{N^{\frac{1}{2}(1+\epsilon)}} \\ \text{DGP } 3: & (1-\theta_{0,N}^2)N & \underset{N\to\infty}{\to} 0 & \text{or} & \theta_{0,N} = 1 - \frac{e}{N^{1+\epsilon}} \\ \text{DGP } 4: & \frac{N}{g} & \underset{N\to\infty,\ g\to\infty}{\to} 0, & & \end{array} \tag{23}$$

with e a constant and ϵ some real number larger than zero. In case of DGP 4, (23) implies that the process has been running longer than the sample size N. Kruiniger (2009) uses the above specification of DGP 3 with $\epsilon = 0$ and DGP 4 with N/g converging to a constant to construct local to unity asymptotic approximations of the distributions of two step GMM estimators that use the Dif, Lev and/or Sys moment conditions.

We do not confine ourselves to a specific DGP for the initial observations so we obtain results that apply generally. While the (non-) identification conditions for identifying θ that result from the above data generating processes might be (in)plausible, it is the arbitrariness of them which is problematic. Additionally, the identification condition might hold but it can still lead to large size distortions of Wald test statistics.

4 Identification from general moment conditions

We just showed that the Lev and Dif moment conditions do not identify θ when θ_0 is close to one and $T = 3$. To analyze the identification of θ by the different moment conditions for a general number of time periods T, we start out with a representation theorem. For the different moment conditions, it states the behavior of the sample moments and their derivatives under the previously defined limit sequences. Throughout we assume that the mean stationarity conditions (6)-(7) apply, so the Lev and Sys moment conditions are valid too.

Theorem 1 (Representation Theorem). *Under Assumptions 1 and 2, the conditions in (2), mean stationarity (6)-(7), finite fourth moments of* c_i *and* u_{it}, $i = 1, \ldots, N, t = 2, \ldots, T$, $T \geq 3$, *we can characterize the large sample behavior of the Dif, Lev, NL, AS and Sys sample moments and their derivatives for values* θ_0 *that drift to one according to (16) and (20) by:*

$$\begin{pmatrix} f_N^j(\theta) \\ q_N^j(\theta) \end{pmatrix} = \begin{pmatrix} A_f^j(\theta) \\ A_q^j(\theta) \end{pmatrix} \left[\frac{1}{h_N(\theta_{0,N})\sqrt{N}}\psi + \iota \left(\lim_{N\to\infty} E((\theta_{0,N} - 1)u_{i1}^2)\right)\right] + \begin{pmatrix} \mu_f^j(\theta, \sigma^2) \\ \mu_q^j(\theta, \sigma^2) \end{pmatrix} + \frac{1}{\sqrt{N}} \begin{pmatrix} B_f^j(\theta) \\ B_q^j(\theta) \end{pmatrix} \psi_{uu}, \tag{24}$$

with j = *Dif, Lev, NL, AS, Sys. Furthermore,* $\mu_f^j(\theta, \sigma^2)$ *and* $\mu_q^j(\theta, \sigma^2)$ *are constants,* $\sigma^2 = (\sigma_2^2 \ldots \sigma_T^2)$, ψ *and* ψ_{uu} *are mean zero finite variance normal random variables,* ι *is a vector of ones and the dimensions of* ι *and* ψ *are* $T - 1$ *for Dif, AS and Sys and* $T - 2$ *for Lev and NL.*
The specifications of $A_f^j(\theta)$, $A_q^j(\theta)$, $B_f^j(\theta)$, $B_q^j(\theta)$, $\mu_f^j(\theta, \sigma^2)$, $\mu_q^j(\theta, \sigma^2)$, ψ *and* ψ_{uu} *for values of* T *equal to 4 and 5 are stated in Appendix A.*

Proof. see Appendix A. ■

The representation theorem in Theorem 1 is reminiscent of the cointegration representation theorem, see *e.g.* Engle and Granger (1987) and Johansen (1991). Identical to that representation theorem, Theorem 1 shows that the behavior of the moment series changes over different directions.

The representation theorem shows that the sample moment and its derivative diverge in the direction of $\binom{A_f^j(\theta)}{A_q^j(\theta)}$ since the latter components get multiplied by $\frac{1}{h(\theta_{0,N})\sqrt{N}}\psi$, which under (20) goes to infinity when the sample size increases. The only identifying information for θ therefore results from that part of the sample moment which does not depend on ψ. Since ψ only affects the part of the sample moments spanned by $A_f^j(\theta)$, the sample moments are independent of ψ in the direction of the orthogonal complement of $A_f^j(\theta)$.

When we pre-multiply the sample moments by the orthogonal complement of $A_f^j(\theta)$, we obtain

$$A_f^j(\theta)'_\perp f_N^j(\theta) \approx \quad A_f^j(\theta)'_\perp \mu_f^j(\theta,\sigma^2) + \tfrac{1}{\sqrt{N}} A_f^j(\theta)'_\perp B_f^j(\theta)\psi_{cu}, \tag{25}$$

with $A_f^j(\theta)_\perp$ the orthogonal complement of $A_f^j(\theta)$, i.e. $A_f^j(\theta)'_\perp A_f^j(\theta) \equiv 0$. Compared with the expression in Theorem 1 (24), the elements multiplied by $A_f^j(\theta)$ have both dropped out since $A_f^j(\theta)'_\perp A_f^j(\theta) = 0$. The right hand side of (25) now contains all remaining identifying elements of the original moment conditions. From expression (25), it is seen that identification results only when $A_f^j(\theta)_\perp$ is a well defined matrix and, furthermore, $\mu_f^j(\theta,\sigma^2)$ is non-zero. We next briefly discuss what this implies for different moment conditions discussed previously.

Dif and Lev conditions When $T = 3$ or 4, the specifications of $\binom{\mu_f^j(\theta,\sigma^2)}{\mu_q^j(\theta,\sigma^2)}$ and $\binom{A_f^j(\theta)}{A_q^j(\theta)}$ for the Dif and Lev moment conditions, which are stated in the proof of Theorem 1 in Appendix A[9], are:

$$\begin{aligned}
\text{Dif:}\quad & T=3 \quad \binom{\mu_f^{Dif}(\theta,\sigma^2)}{\mu_q^{Dif}(\theta,\sigma^2)} = (0\ 0)',\ A_f^{Dif}(\theta) = (-\theta\ 1),\ A_q^{Dif}(\theta) = (-1\ 0).\\
& T=4 \quad \binom{\mu_f^{Dif}(\theta,\sigma^2)}{\mu_q^{Dif}(\theta,\sigma^2)} = (0\ldots 0)',\ A_f^{Dif}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \end{pmatrix},\ A_q^{Dif}(\theta) = -\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.\\
\text{Lev:}\quad & T=3 \quad \binom{\mu_f^{Lev}(\theta,\sigma^2)}{\mu_q^{Lev}(\theta,\sigma^2)} = \left(\sigma_2^2 \begin{pmatrix} 1-\theta \\ 1 \end{pmatrix}\right),\ A_f^{Lev}(\theta) = 1-\theta,\ A_q^{Lev}(\theta) = -1.\\
& T=4 \quad \binom{\mu_f^{Lev}(\theta,\sigma^2)}{\mu_q^{Lev}(\theta,\sigma^2)} = \left(\begin{pmatrix} 1-\theta \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \end{pmatrix}\right),\ A_f^{Lev}(\theta) = \begin{pmatrix} 1-\theta & 0 \\ 0 & 1-\theta \end{pmatrix},\\
& \qquad\quad A_q^{Lev}(\theta) = -I_2.
\end{aligned} \tag{26}$$

The expressions of $A_f^{Dif}(\theta)$ and $A_f^{Lev}(\theta)$ in (26) are all square or rectangular matrices. When T exceeds four, the expressions of $A_f^{Lev}(\theta)$ remain square matrices[10] so the Lev moment conditions do not identify θ since $A_f^{Lev}(\theta)_\perp$ does not exist. When $T = 3$ or $T > 4$, the expressions of $A_f^{Dif}(\theta)$ are not square so the orthogonal complement of $A_f^{Dif}(\theta)$, $A_f^{Dif}(\theta)_\perp$, is well defined. However, since $\mu_f^{Dif}(\theta,\sigma^2)$ equals zero, $A_f^{Dif}(\theta)'_\perp \mu_f^{Dif}(\theta,\sigma^2) = 0$ so the Dif moment conditions still do not identify θ for any value of T. Summarizing we have:

$$\begin{aligned}
\text{Dif, } T=4\text{:}\quad & A_f^{Dif}(\theta)_\perp \text{ does not exist.} && \text{No identification when } T=4.\\
\text{Dif, } T=3,\ T>4\text{:}\quad & A_f^{Dif}(\theta)'_\perp \mu_f^{Dif}(\theta,\sigma^2) = 0. && \text{No identification when } T=3,\ T>4.\\
\text{Lev:}\quad & A_f^{Lev}(\theta)_\perp \text{ does not exist.} && \text{No identification when } T\geq 3.
\end{aligned} \tag{27}$$

[9] The proofs in the Appendix A do not cover $T = 3$ since it straightforwardly results from $T = 4$.

[10] We refer to the proof of Theorem 1 for the expressions of $A_f^{Dif}(\theta)$ and $A_f^{Lev}(\theta)$ when $T = 5$.

NL condition The NL moment condition is not defined for $T = 3$. When $T = 4$, the expressions of $\binom{\mu_f^j(\theta,\sigma^2)}{\mu_q^j(\theta,\sigma^2)}$ and $\binom{A_f^j(\theta)}{A_q^j(\theta)}$ read

$$\text{NL:} \quad \binom{\mu_f^{NL}(\theta,\sigma^2)}{\mu_q^{NL}(\theta,\sigma^2)} = \binom{(1-\theta)\left(\sigma_3^2-\theta\sigma_2^2\right)}{2(\theta-1)\sigma_2^2-\sigma_3^2}, \; \binom{A_f^{NL}(\theta)}{A_q^{NL}(\theta)} = \begin{pmatrix} \theta(\theta-1) & 1-\theta \\ 2\theta-1 & -1 \end{pmatrix}. \tag{28}$$

Since there is only one sample moment, the specification of $A_f^{NL}(\theta)$ shows that there are more diverging components than sample moments so the NL moment condition does not identify θ.

The expression of $A_f^{NL}(\theta)$ for a larger number of time series observations[11] are also such that the number of divergent components exceeds the number of sample moments. Hence for larger values of T, the NL moment conditions also do not identify θ.

AS and Sys conditions The expressions of $\binom{\mu_f^j(\theta,\sigma^2)}{\mu_q^j(\theta,\sigma^2)}$ and $\binom{A_f^j(\theta)}{A_q^j(\theta)}$ when $T = 4$ for the AS and Sys moment conditions result from stacking those of the Dif and NL and Dif and Lev moment conditions respectively:

$$\text{AS:} \quad T=4 \quad \mu_f^{AS}(\theta,\sigma^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ (1-\theta)\left(\sigma_3^2-\theta\sigma_2^2\right) \end{pmatrix}, \; A_f^{AS}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \\ \theta(\theta-1) & 1-\theta & 0 \end{pmatrix},$$

$$\mu_q^{AS}(\theta,\sigma^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2(\theta-1)\sigma_2^2-\sigma_3^2 \end{pmatrix}, \; A_q^{AS}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 2\theta-1 & -1 & 0 \end{pmatrix}.$$

$$\text{Sys:} \quad T=3 \quad \mu_f^{Sys}(\theta,\sigma^2) = (1-\theta)\begin{pmatrix} 0 \\ \sigma_2^2 \end{pmatrix}, \; A_f^{Sys}(\theta) = \begin{pmatrix} -\theta & 1 \\ 1-\theta & 0 \end{pmatrix},$$

$$\mu_q^{Sys}(\theta,\sigma^2) = \begin{pmatrix} 0 \\ \sigma_2^2 \end{pmatrix}, \; A_q^{Sys}(\theta) = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}.$$

[11] The expression for $T = 5$ is stated in the proof of Theorem 1 in Appendix A.

$$\text{Sys:} \quad T=4 \quad \mu_f^{Sys}(\theta,\sigma^2)=(1-\theta)\begin{pmatrix}0\\0\\0\\\sigma_2^2\\\sigma_3^2\end{pmatrix},\ A_f^{Sys}(\theta)=\begin{pmatrix}-\theta & 1 & 0\\0 & -\theta & 1\\0 & -\theta & 1\\1-\theta & 0 & 0\\0 & 1-\theta & 0\end{pmatrix},$$

$$\mu_q^{Sys}(\theta,\sigma^2)=\begin{pmatrix}0\\0\\0\\\sigma_2^2\\\sigma_3^2\end{pmatrix},\ A_q^{Sys}(\theta)=\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 1 & 0\\-1 & 0 & 0\\0 & -1 & 0\end{pmatrix}. \tag{29}$$

When $T=3$, $A_f^{Sys}(\theta)$ is a square matrix so its orthogonal complement is not defined. It implies that the Sys moment conditions do not identify θ when $T=3$. When T equals 4, the specification of $A_f^j(\theta)$ is a rectangular matrix both for the AS and Sys moment conditions. It implies that the orthogonal complement of $A_f^j(\theta)$, $A_f^j(\theta)_\perp$, is a well defined matrix. Furthermore, the specification of $\mu_f^j(\theta,\sigma^2)$ for the AS and Sys moment conditions in (29) is such that $A_f^j(\theta)'_\perp\mu_f^j(\theta,\sigma^2)\neq 0$. It implies that although the AS and Sys sample moments diverge in the direction of $A_f^j(\theta)$, so that part cannot be used to identify θ, the AS and Sys sample moments identify θ by their part which is spanned by the orthogonal complement of $A_f^j(\theta)$. The expressions of $\mu_f^j(\theta,\sigma^2)$ and $A_f^j(\theta,\sigma^2)$ in the proof of Theorem 1 in Appendix A show that this argument extends to all values of T larger than three.

Corollary 1 (Identification of θ). *Under the assumptions of Theorem 1, θ is identified by the AS and Sys moment conditions when T exceeds three but is not identified by the Dif, Lev and NL moment conditions for any value of T.*

Corollary 1 shows that the identification issues for the Sys moment conditions with $T=3$ do not extend to more time series observations. Hence, θ is identified by the Sys moment conditions when there are more than three time periods. It also shows that θ is identified by the AS moment conditions. We used mean stationarity to construct the large sample behavior in Theorem 1 and Corollary 1. Unlike the Sys moment conditions, the AS moment conditions do, however, not need mean stationarity to hold. It shows that assuming mean stationarity for constructing moment conditions does not help to identify θ when $\theta_0=1$ since the same identification results are obtained from moment conditions that do not assume mean stationarity. When mean stationarity does not hold both the Dif and NL, and consequently the AS, moment conditions identify θ when $\theta_0=1$.

Corollary 1 shows that the AS and Sys moment conditions identify θ when T exceeds three. This does, however, not imply that GMM estimators based on these moment conditions behave in the manner that we are used to when estimating parameters which are identified by moment conditions. We distinguish two different cases:

1. The convergence rate accords with (18) so ψ vanishes from the large sample behavior of the moment conditions in Theorem 1. It leads to standard behavior of one and two step GMM estimators.

2. The convergence rate accords with (20). Only the part of the moment conditions in the direction of $A_f^j(\theta)_\perp$ (25) now identifies θ. One step and two step GMM estimators, however, use both the part of the sample moment that lies in the direction of $A_f^j(\theta)$, which diverges, and the part which lies in the direction of its orthogonal complement, which identifies θ. Usage of the first part results in an inconsistency so one and two step GMM estimators are inconsistent and have non-standard limiting distributions. To exemplify this, Corollary 2 states the limiting distribution of the one step estimator based on the Sys moment conditions which results in a straightforward manner from Theorem 1 since the Sys moment conditions are linear in θ.

Corollary 2. *Under the conditions of Theorem 1, the limiting behavior of the one step estimator for the Sys moment conditions is characterized by*

$$\hat{\theta}_{1s}^{Sys} \underset{d}{\rightarrow} 1 - (\psi' A_q^{Sys}(1)' A_q^{Sys}(1)\psi)^{-1}\psi' A_q^{Sys}(1)' A_f^{Sys}(1)\psi, \tag{30}$$

which is inconsistent since $A_f^{Sys}(1)$ does not equal zero.

Corollary 2 shows that the one step estimator based on the Sys moment conditions is inconsistent despite that the Sys moment conditions identify θ. It also shows that the limiting distribution of the one step estimator is non-standard. Similar results hold for the one step GMM estimator based on the AS moment conditions and the two step GMM estimator based on either the AS or Sys moment conditions. These are more involved to obtain since the AS moment conditions are a quadratic function of θ and we have to involve a covariance matrix estimator for the two step GMM estimators. For reasons of brevity, we therefore refrain from constructing these.

Corollary 2 shows that the identification of θ by the AS and Sys moment conditions when T exceeds three does not automatically lead to standard behavior of one and two step GMM estimators. It can be shown that, under the conditions of Theorem 1, the limiting behavior of one- and two-step GMM estimators is similar to the non-standard results in *e.g.* Madsen (2003) or Kruiniger (2009). Conducting inference based on these estimators exploiting the usual Wald or t statistic is therefore hard when θ_0 is close to one.

In order to fully exploit the identification for the AS and Sys moment conditions, we therefore make use of several identification robust GMM statistics, i.e. the GMM-AR statistic of Anderson and Rubin (1949) and Stock and Wright (2000), the GMM-LM statistic of Newey and West (1987) and the KLM statistic of Kleibergen (2005). These identification robust

GMM statistics, which are defined in Appendix B, are size correct for all values of θ_0.[12] Since θ is identified by the AS and Sys moment conditions, they also have discriminatory power.

For settings of θ_0 and the nuisance parameters for which no identification issues exist, both the GMM-LM and KLM statistics are efficient and more powerful than the GMM-AR statistic. This standard notion of efficiency does, however, not apply to values of θ_0 close to one which is also revealed by the inconsistency of the one and two step GMM estimators. To establish a sense of efficiency or optimality, we therefore in the next Section determine the maximal attainable power for testing values of θ under the worst case settings where its true value is one and the variance of the initial observations accords with (20).

5 Maximal attainable power in worst case setting

In this Section we construct the maximal attainable power curve under the worst case setting which results from the identifying part of the sample moments as in equation (25). As discussed previously, only the orthogonal complements with respect to $A_f^j(\theta)$ of the AS and Sys sample moments identify θ when T is larger than three. Expressions of the orthogonal complements of $A_f^j(\theta)$ for $T = 4$ and 5 for the AS and Sys moment conditions are stated in Appendix A. In general they can be specified as

$$A_f^j(\theta)_\perp = (G_{f,T}^j(\theta) \vdots G_{2,T}^j) \tag{31}$$

where T indicates the number of time periods and $G_{2,T}^j$ is such that $G_{2,T}^{j\prime}\mu_f^j(\theta,\sigma^2) = 0$. Furthermore, $G_{f,T}^j(\theta)$ is the only part of $A_f^j(\theta)_\perp$ that depends on θ. The orthogonal complements are such that the rotated robust AS and Sys moment conditions are quadratic in θ :

$$g_{f,T}(\theta) = A_f(\theta)_\perp^{j\prime} f_N(\theta) = a\theta^2 + b\theta + d, \tag{32}$$

with for

T=4: Sys $a = \frac{1}{N}\sum_{i=1}^N \binom{(\Delta y_{i2})^2}{0}$, $b = -\frac{1}{N}\sum_{i=1}^N \binom{(y_{i3}-y_{i1})^2}{\Delta y_{i2}\Delta y_{i3}}$, $d = \frac{1}{N}\sum_{i=1}^N \binom{(y_{i4}-y_{i1})\Delta y_{i3}}{\Delta y_{i3}\Delta y_{i4}}$.

AS $a = \frac{1}{N}\sum_{i=1}^N \binom{(y_{i3}-y_{i1})\Delta y_{i2}}{0}$, $b = -\frac{1}{N}\sum_{i=1}^N \binom{(y_{i3}-y_{i1})\Delta y_{i3}+(y_{i4}-y_{i1})\Delta y_{i2}}{\Delta y_{i2}\Delta y_{i3}}$,

$d = \frac{1}{N}\sum_{i=1}^N \binom{(y_{i4}-y_{i1})\Delta y_{i3}}{\Delta y_{i3}\Delta y_{i4}}$.

T=5: Sys $a = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (\Delta y_{i2})^2 \\ (y_{i3}-y_{i1})\Delta y_{i3} \\ (\Delta y_{i3})^2 \\ 0 \\ 0 \end{pmatrix}$, $b = -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (y_{i3}-y_{i1})^2 \\ (y_{i4}-y_{i1})(y_{i4}-y_{i2}) \\ (y_{i4}-y_{i2})^2 \\ \Delta y_{i2}\Delta y_{i4} \\ \Delta y_{i3}\Delta y_{i4} \end{pmatrix}$,

[12] The GMM-LM statistic is size correct in this case because under the null hypothesis no parameters are estimated.

$$d=\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}(y_{i4}-y_{i1})\Delta y_{i3}\\(y_{i5}-y_{i1})\Delta y_{i4}\\(y_{i5}-y_{i2})\Delta y_{i4}\\\Delta y_{i2}\Delta y_{i5}\\\Delta y_{i3}\Delta y_{i5}\end{pmatrix}.$$

AS $a=\sum_{i=1}^{N}\begin{pmatrix}(y_{i3}-y_{i1})\Delta y_{i2}\\(y_{i4}-y_{i1})\Delta y_{i3}\\(y_{i4}-y_{i2})\Delta y_{i3}\\0\\0\end{pmatrix}, b=-\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}(y_{i4}-y_{i1})\Delta y_{i2}+(y_{i3}-y_{i1})\Delta y_{i3}\\(y_{i4}-y_{i1})\Delta y_{i4}+(y_{i5}-y_{i1})\Delta y_{i3}\\(y_{i4}-y_{i2})\Delta y_{i4}+(y_{i5}-y_{i2})\Delta y_{i3}\\\Delta y_{i2}\Delta y_{i4}\\\Delta y_{i3}\Delta y_{i4}\end{pmatrix},$

$$d=\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}(y_{i4}-y_{i1})\Delta y_{i3}\\(y_{i5}-y_{i1})\Delta y_{i4}\\(y_{i5}-y_{i2})\Delta y_{i4}\\\Delta y_{i2}\Delta y_{i5}\\\Delta y_{i3}\Delta y_{i5}\end{pmatrix}.$$

The limiting sequence in (20) characterizes the worst case for identifying θ. The robust moment condition in (32) is the only part of the AS and Sys moment conditions that identifies θ in this worst case setting. We can therefore use it to obtain the maximal attainable power (MAP) of tests using the AS or SYS moment conditions of $\mathrm{H}_0 : \theta = \theta_0$ with 95% significance against the alternative (worst case) hypothesis $\mathrm{H}_1 : \theta = 1$ with nuisance parameters that are characterized by (20), see Lehmann and Romano (2005):

$$\begin{aligned} MAP^j(\theta_0|\mathrm{H}_1) = & \lim_{N\to\infty}\max_{TS}\Pr\left[ts(\theta_0) > ac_{ts(\theta_0)}|\mathrm{H}_1\right] & j = AS, Sys \\ = & \lim_{N\to\infty}\max_{TS}\min_{S}\Pr\left[ts(\theta_0) > ac_{ts(\theta_0)}|\theta = 1\right] & j = AS, Sys \end{aligned} \quad (33)$$

where TS is the set of statistics testing H_0, $ts(\theta_0)$ is an element of the set TS, *i.e.* a statistic that tests H_0, $ac_{ts(\theta_0)}$ the 95% asymptotic critical value of $ts(\theta_0)$ and S the set of processes for the initial observations that satisfy Assumption 1 and the mean stationarity conditions (6)-(7). We refer to the power function in (33) as the maximal attainable power curve. It just focuses on the alternative hypothesis $\mathrm{H}_1 : \theta = 1$ since the identification issues are most relevant at this value of θ. The standard optimality results, if any exist, do not apply under these conditions so we use the robust moments in (32) to establish them.

To construct the maximal attainable power curve, we first have to determine the slowest rate at which the hypothesized value of θ under H_0 can drift away from one, whilst the true value of θ equals one, such that the sample moment (32) converges to a random variable that is non-degenerate and remains finite with probability one. To determine this rate, we first state the probability limits of a, b and d when the true value of θ is one.

Theorem 2. *Under Assumptions 1 and 2, the conditions in (2), finite fourth moments of c_i and u_{it}, $i = 1, \ldots, N$, $t = 2, \ldots, T$, and $\omega = \lim_{\theta_0 \to 1} E((c_i - (1-\theta_0)y_{i1})^2)$, the limit behavior of the different components of $g_{f,T}(\theta)$ when the true value of θ is equal to one is characterized by:*

T=4, Sys: $a \underset{p}{\rightarrow} \binom{\omega+\sigma_2^2}{0}$, $b \underset{p}{\rightarrow} -\binom{(4\omega+\sigma_2^2+\sigma_3^2)}{\omega}$, $d \underset{p}{\rightarrow} \binom{3\omega+\sigma_3^2}{\omega}$.

T=4, AS: $a \underset{p}{\rightarrow} \binom{2\omega+\sigma_2^2}{0}$, $b \underset{p}{\rightarrow} -\binom{(5\omega+\sigma_2^2+\sigma_3^2)}{\omega}$, $d \underset{p}{\rightarrow} \binom{3\omega+\sigma_3^2}{\omega}$.

T=5, Sys:
$$a \underset{p}{\rightarrow} \begin{pmatrix} \omega+\sigma_2^2 \\ 2\omega^2+\sigma_3^2 \\ \omega^2+\sigma_3^2 \\ 0 \\ 0 \end{pmatrix}, \ b \underset{p}{\rightarrow} -\begin{pmatrix} 4\omega+\sigma_2^2+\sigma_3^2 \\ 6\omega+\sigma_3^2+\sigma_4^2 \\ 4\omega+\sigma_3^2+\sigma_4^2 \\ \omega \\ \omega \end{pmatrix}, \ d \underset{p}{\rightarrow} \begin{pmatrix} 3\omega+\sigma_3^2 \\ 4\omega+\sigma_4^2 \\ 3\omega+\sigma_4^2 \\ \omega \\ \omega \end{pmatrix}.$$

T=5, AS:
$$a \underset{p}{\rightarrow} \begin{pmatrix} 2\omega+\sigma_2^2 \\ 3\omega+\sigma_3^2 \\ 2\omega+\sigma_3^2 \\ 0 \\ 0 \end{pmatrix}, \ b \underset{p}{\rightarrow} -\begin{pmatrix} 5\omega+\sigma_2^2+\sigma_3^2 \\ 7\omega+\sigma_3^2+\sigma_4^2 \\ 5\omega+\sigma_3^2+\sigma_4^2 \\ \omega \\ \omega \end{pmatrix}, \ d \underset{p}{\rightarrow} \begin{pmatrix} 3\omega+\sigma_3^2 \\ 4\omega+\sigma_4^2 \\ 3\omega+\sigma_4^2 \\ \omega \\ \omega \end{pmatrix}.$$

Proof. see Appendix A. ■

The parameter ω in Theorem 2 reflects the deviation from mean stationarity. Mean stationarity corresponds with a zero value of ω. Theorem 3 states the convergence rate for the local to unity asymptotics that we employ to obtain the maximal attainable power curve.

Theorem 3. *Under the conditions of Theorem 2, the drifting sequence for θ under H_0 for the robust moments $g_{f,T}(\theta)$ is such that:*
1. *When $\omega = 0$, $\sigma_t^2 = \sigma^2$, $t = 2, \ldots T : \theta = 1 + \frac{e}{\sqrt[4]{N}}$,*
2. *When $\omega \neq 0$ or $\sigma_t^2 \neq \sigma^2$, for at least one value of t, $t = 2, \ldots T-1 : \theta = 1 + \frac{e}{\sqrt{N}}$,*
with $e < 0$ a finite constant.

Proof. see Appendix A. ■

The quartic root convergence rate in Theorem 3 results since the robust moment equation (32) is quadratic in θ. When we specify θ as $1 + \frac{e}{N^\xi}$ and $\omega = 0$, $\sigma_t^2 = \sigma^2$, $t = 2, \ldots T$, all elements which are linear in e cancel out in the limit. We are then left with the quadratic term in e and components that converge at the rate $\frac{1}{\sqrt{N}}$. A quartic root convergence rate, *i.e.* $\xi = \frac{1}{4}$, then makes all these components of the same order of magnitude in the sample size.

Instead of using the robust moment equation in (32) for testing hypotheses on θ, it can also be used to estimate θ. The estimator that results from it, is, however, a worst case estimator since it only does relatively well under worst case DGPs. For DGPs with values of θ less than one, estimators based on the other moment conditions outperform it. Alongside this suboptimality also its large sample distribution is, identical to the standard one-step and two-step GMM estimators, non-standard. This holds since the expected value of the discriminant of the quadratic equation (32) is equal to zero under the worst case DGPs, *i.e.* case 1 in Theorem 3. The estimator then has a quartic root convergence rate and a non-standard large sample distribution. Remarkably, a quartic root convergence rate is also found by Ahn and Thomas (2006) and Kruiniger (2013, Theorem 4) for random effects maximum likelihood estimators.

Theorem 3 shows that mean stationarity (6)-(7), under which $\omega = 0$, and variances that are constant over time lead to the slowest convergence rate for θ. It implies that jointly with (20) this setting provides the worst case data generating process under $\theta_0 = 1$. Therefore, to construct the maximal attainable power curve, we first obtain an approximation of the finite sample distribution of the GMM-AR statistic which tests $\mathrm{H}_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ just using the robust moments in (32) whilst the true value of θ is equal to one jointly with (20). This particular statistic can be written as:

$$\text{GMM-AR}(e) = N g_{f,T}(e)' \hat{V}_{gg}(e)^{-1} g_{f,T}(e), \tag{34}$$

with $g_{f,T}(e)$ the moments in (32) evaluated at $\theta = 1 + \frac{e}{\sqrt[4]{N}}$ and $\hat{V}_{gg}(e)$ the (Eicker-White) covariance matrix estimator of the covariance matrix of $g_{f,T}(e)$. The next Theorem states its large sample distribution.

Theorem 4. *Under the conditions of Theorem 2 and when the true value of θ is equal to one, $\omega = 0$, $\sigma_t^2 = \sigma^2$, $t = 2, \dots T$, the large sample distribution of the GMM-AR statistic (34) for testing the hypothesis $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$, is characterized by*

$$\chi^2(\delta, p_{\max}), \tag{35}$$

with $e^4 E(a)' \left[B(N)' V_{abd} B(N)\right]^{-1} E(a)$, $p_{\max}$ the number of elements $g_T(\theta)$, so when $T = 4$, $p_{\max} = 2$ or when $T = 5$, $p_{\max} = 5$, $\delta = (e\sigma)^4 \binom{\iota_p}{0}' (B(N)' V_{abd} B(N))^{-1} \binom{\iota_p}{0}$,

$$B(N) = (\iota_3 \otimes I_{p_{\max}}) + \tfrac{e}{\sqrt[4]{N}} \left[(2 + \tfrac{e}{\sqrt[4]{N}})(e_{1,3} \otimes I_{p_{\max}}) + (e_{2,3} \otimes I_{p_{\max}}) \right], \tag{36}$$

V_{abd} the covariance matrix of a, b and d, ι_3 a 3×1 dimensional vector of ones, $I_{p_{\max}}$ the $p_{\max} \times p_{\max}$ dimensional identity matrix, $e_{1,3}$ and $e_{2,3}$ the first and second 3×1 dimensional unity vectors and $\chi^2(\delta, p_{\max})$ a non-central χ^2 distribution with non-centrality parameter δ and degrees of freedom parameter $p_{\max}$.

Proof. see Appendix A. ■

The expression of the large sample distribution in Theorem 4 depends on the sample size. When the sample size goes to infinity, $\frac{e}{\sqrt[4]{N}}$ converges to zero so $B(N)$ converges to $(\iota_3 \otimes I_{p_{\max}})$. For most sample sizes, $\frac{e}{\sqrt[4]{N}}$ is, however, non-negligible and therefore important to incorporate in the expression of the large sample distribution to obtain an accurate approximation of the finite sample distribution of the GMM-AR statistic.

There are more moment conditions in $g_{f,T}(e)$ than the number of elements of θ, which is one, so they over identify θ. More powerful statistics for testing a point null hypothesis on θ can therefore be constructed using a weighted average of the moments $g_{f,T}(e)$ instead of all of them. We construct the maximal attainable power curve using the (infeasible) weighted average of the robust sample moments in the GMM-AR statistic (34) that leads to the largest value of the non-centrality parameter of the large sample distribution.

Theorem 5. *Under the conditions of Theorem 2 and when the true value of θ is equal to one, the maximal attainable power curve for testing $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ is*

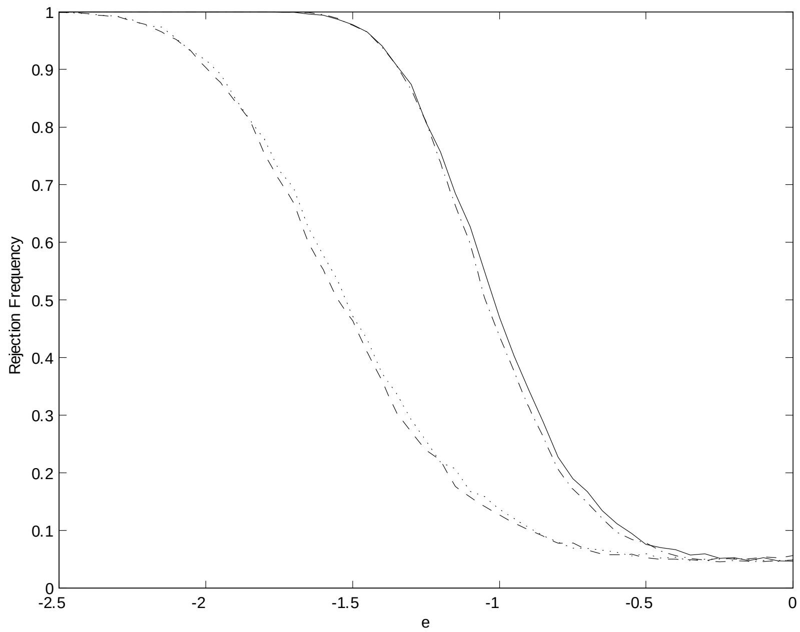
$$\chi^2(\delta, 1), \tag{37}$$

with $\delta = e^4 \binom{\iota_p}{0}' (B(N)' V_{abd} B(N))^{-1} \binom{\iota_p}{0}$, ι_p a $p \times 1$ dimensional vector of ones and p equals 1 when $T = 4$ and 3 when $T = 5$.

Proof. see Appendix A. ■

Figure 1 shows the maximal attainable power curves that result from the AS and Sys moment conditions when $T = 4$ and 5. Figure 1 shows that the maximal attainable power curves that result from the AS and Sys moment conditions are identical which is surprising. Apparently no power is lost by exploiting the AS moment conditions only. This novel result documents that imposing mean stationarity is not only unnecessary, but also superfluous. Furthermore, Figure 1 also shows that the maximal attainable power curves that result for a larger number of time series observations dominate those that result for smaller number of time series observations.

Figure 1. Power envelope for testing $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$



Note: 95% significance level, true value of θ is one, N=500, Sys & T=4 (dashed), AS & T=4 (dotted), Sys & T=5 (solid), AS & T=5 (dash-dotted).

5.1 Large individual fixed effect variance

Sofar we have focused on highly persistent panel data resulting from a large autoregressive parameter. However, the representation of the moment conditions and their derivatives in Theorem 1 applies to any setting where the variance of the initial observations gets large. The expression of the initial observation in (7) shows that its variance becomes large when either the variance of the initial disturbance term, u_{i1}, or the individual specific fixed effect, μ_i, become large. Theorem 1 and the resulting subsequent Theorems focus on a large variance that results from the initial disturbance term. This occurs when the initial observation results from the unconditional distribution of an AR(1) model and the autoregressive parameter is close to unity. Theorem 1 does, however, extend to the case where jointly with the sample size, the individual specific effect variance becomes large in such a manner that (20) holds. This drifting sequence applies to any value of the autoregressive parameter so the resulting identification issues are then no longer confined to the unit root value. Hence, they also apply to cases with only moderate autoregressive dynamics, but a large variance of the unobserved heterogeneity.

For Theorem 1 and subsequent Theorems to cover a large individual fixed effect variance, we have to minorly change the specification of $A_f(\theta, \theta_0)$, $A_q(\theta, \theta_0)$, $B_f(\theta, \theta_0)$, $B_q(\theta, \theta_0)$,

$\mu_f(\theta, \theta_0, \sigma^2)$ and $\mu_q(\theta, \theta_0, \sigma^2)$[13] accordingly. For example, the expressions of $A_f(\theta, \theta_0)$ when $T = 4$ for the AS and Sys moment conditions are:

$$\begin{aligned}
\text{AS:}\quad A_f^{AS}(\theta,\theta_0) &= \begin{pmatrix} \theta_0 - 1 - \theta & 1 & 0 \\ (\theta_0-\theta)(\theta_0-1) & \theta_0 - 1 - \theta & 1 \\ \theta_0(\theta_0-\theta)(\theta_0-1) & \theta_0(\theta_0-1-\theta) & \theta_0 \\ (1-\theta)(\theta_0-\theta-1) & 1-\theta & 0 \end{pmatrix}, \\
\text{Sys:}\quad A_f^{Sys}(\theta,\theta_0) &= \begin{pmatrix} \theta_0 - 1 - \theta & 1 & 0 \\ (\theta_0-\theta)(\theta_0-1) & \theta_0 - 1 - \theta & 1 \\ \theta_0(\theta_0-\theta)(\theta_0-1) & \theta_0(\theta_0-1-\theta) & \theta_0 \\ 1-\theta & 0 & 0 \\ (1-\theta)(\theta_0-1) & 1-\theta & 0 \end{pmatrix}.
\end{aligned} \tag{38}$$

The expressions in (38) are identical to those in (29) when $\theta_0 = 1$. Interestingly, the part of the orthogonal complement of $A_f(\theta, \theta_0)$ which depends on θ, *i.e.* $G_{f,T}^j(\theta)$ in (31) , remains unchanged:

$$\begin{aligned}
\text{AS:}\quad G_{f,T=4}^{AS}(\theta) &= \begin{pmatrix} -(1-\theta) \\ 0 \\ 0 \\ 1 \end{pmatrix},\ G_{2,T=4}^{AS} = \begin{pmatrix} 0 \\ -\theta_0 \\ 1 \\ 0 \end{pmatrix} \\
\text{Sys:}\quad G_{f,T=4}^{Sys}(\theta) &= \begin{pmatrix} -(1-\theta) \\ 0 \\ 0 \\ -\theta \\ 1 \end{pmatrix},\ G_{2,T=4}^{Sys} = \begin{pmatrix} 0 \\ -\theta_0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.
\end{aligned} \tag{39}$$

The robust moments which result from the orthogonal complement in (32) are therefore insensitive to two sets of nuisance parameters: the initial observations and the individual specific fixed effects. The robust moments still constitute a quadratic polynomial. Theorem 3 shows that the resulting convergence speed (of an estimator or a local-to-true-value hypothesized parameter value) is of a lower order in the sample size when the expected value of the discriminant of the quadratic polynomial equals zero. This occurs under a unit root value of the autoregressive parameter paired with mean-stationarity and constant variances over time. The expected value of the discriminant also equals zero when the autoregressive parameter is zero again paired with mean-stationarity and constant variances over time. Hence, the worst case data generating processes at a zero value of the autoregressive parameter also imply a quartic root convergence rate. For all other values of the autoregressive parameter besides zero and one, the expected value of the discriminant is not equal to zero so the convergence

[13] Instead of $A_f(\theta)$, $B_f(\theta), \ldots$ we now use $A_f(\theta, \theta_0)$, $B_f(\theta, \theta_0), \ldots$ since the results no longer apply to just the unit root value of θ_0.

rate is then equal to the square root of the sample size under the worst case data generating processes. The slower convergence rate at a zero value of the autoregressive parameter results because of the identification issues that occur for large values of the variance of the individual specific fixed effect. Since these are considered less pervasive then those that occur because of the unit root value, we have only covered them briefly here.

6 Power envelope and maximal attainable power curve

Theorems 1, 4 and the convergence rate of the maximal attainable power curve in Theorem 3 show that the limiting behavior of estimators is not uniform since it depends on the data generating process at hand. Wald statistics are then size distorted. Under the null hypothesis, the limiting distributions of the GMM-AR, GMM-LM and KLM statistics based on the AS or Sys moment conditions do not depend on nuisance parameters so they remain size correct irrespective of the data generating process. The recommended statistic to use amongst these is then the one which has the largest discriminatory power.

When the true value of θ is less than one, so θ is identified by all moment conditions, both the GMM-LM and KLM statistics are efficient and so are Wald statistics based on estimators that result from the moment conditions. When $\theta = 1$, it is, however, not obvious which statistic is optimal. We therefore construct the lower envelope of the power curves of the GMM-AR[14], GMM-LM and KLM statistics to determine which one, if any, coincides with the maximal attainable power curve. The lower envelope of power curves results from the worst case setting. The worst case large sample distributions of the GMM-AR, GMM-LM and KLM statistics are stated in Theorem 6.

Theorem 6. *Under the conditions from Theorem 2, the worst case large sample distributions, which apply under (20), mean stationarity (6)-(7) and $\sigma_t^2 = \sigma^2$, $t = 2, \ldots T$, of the GMM-AR, GMM-LM and KLM statistics for testing the hypothesis $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ whilst the true value of θ equals one are characterized by*

$$\begin{array}{rl} \text{GMM-AR}(e): & \chi^2(\delta_{GMM\text{-}AR}, p_{GMM\text{-}AR}) \\ \text{KLM}(e): & \chi^2(\delta_{KLM}, 1) \\ \text{GMM-LM}(e): & \chi^2(\delta_{GMM\text{-}LM}, 1), \end{array} \tag{40}$$

[14]We note that this is the GMM-AR statistic that is based on all sample moments which is defined in Appendix B. It therefore differs from the one in (34).

with $p_{GMM\text{-}AR} = \frac{1}{2}(T+1)(T-2)$ *for the Sys moment conditions,* $p_{GMM\text{-}AR} = \frac{1}{2}(T+1)(T-2)-1$ *for the AS moment conditions,*

$$\begin{array}{rll} \delta_{GMM\text{-}AR} = & (e\sigma)^4 \binom{\iota_p}{0}'(B(N)'V_{abd}B(N))^{-1}\binom{\iota_p}{0} & \\ \delta_{KLM} = & \delta_{GMM\text{-}AR} & \\ \delta_{GMM\text{-}LM} = & (e\sigma)^4\binom{\iota_p}{0}'\left(B(N)'V_{abd}B(N)\right)^{-\frac{1}{2}} P_{(B(N)'V_{abd}B(N))^{-\frac{1}{2}}\binom{G_f(e)'A_q\psi}{0}} & \\ & \left(B(N)'V_{abd}B(N)\right)^{-\frac{1}{2}}\binom{\iota_p}{0} & T>4 \\ = & \delta_{GMM\text{-}AR} & T=4, \end{array} \tag{41}$$

with p *equal to 1 when* $T=4$ *and* 3 *when* $T=5$, ψ *an independent normal* $(T-2)$*-dimensional random vector with mean zero and covariance matrix*

$$\lim_{N\to\infty} \operatorname{var}\left(h_N(\theta_{0,N}) \begin{bmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{bmatrix} \right). \tag{42}$$

Proof. see Appendix A. ∎

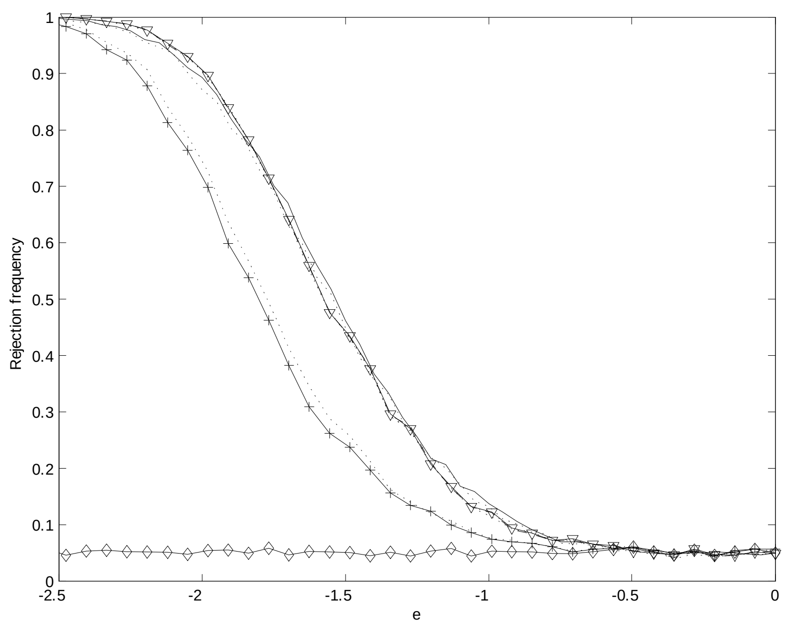
Theorem 6 shows that the lower power envelope of the KLM statistic coincides with the maximal attainable power curve as described in Theorem 5. For the GMM-LM statistic this only occurs when $T=4$. It shows that the KLM statistic is, in a sense, optimal when θ is equal to one. Since the KLM statistic is also efficient when θ is less than one, it is efficient both when θ is less than one or equal to one.

Figures 2 and 3 show the power envelopes of 95% significance tests using the GMM-AR, GMM-LM and KLM statistics for the AS and Sys moment conditions under a worst case DGP when T equals four and five respectively. The worst case DGP that we use results from DGP 1 in Section 3 with a large value of σ_c^2 (ten) compared to σ_t^2, $t=1,\ldots,T$ (one). The results of Theorems 5 and 6 follow from a quartic root convergence rate, and to provide a numerical assessment we therefore fix $N=2000$, a relatively large value. We next simulate for a wide range of values for θ, which together with N provides a mapping to the constant e in Figures 2 and 3 (horizontal axis).

Since the Sys moment conditions do not identify θ when its true value is equal to one and T equals three, all rejection frequencies under a worst case DGP are flat at 5% when T equals three. To reiterate that the Dif, Lev or non-linear part of the AS moment conditions by themselves do not identify θ when its true value is one, Figures 2 and 3 below also include the rejection frequencies that result from the GMM-AR statistic with Dif moment conditions. These rejection frequencies equal 5% for all values of θ which shows that the Dif moment conditions do not identify θ when its true value is equal to one. The same results are obtained when we use the Lev or NL moment conditions or instead of the GMM-AR statistic use the GMM-LM or KLM statistic.

Figures 2 and 3[15] provide a numerical proof of the main results from Theorems 5 and 6. Figure 2 shows that, when $T = 4$, the power envelopes of the KLM and GMM-LM statistics are on the maximal attainable power curve when we use the Sys or AS moment conditions as stated in Theorem 6. The power envelopes of the GMM-AR statistic are below this power curve. Figure 2 also shows that the power envelope of the GMM-AR statistic which uses the AS moment conditions is slightly above the one which results from the GMM-AR statistic that uses the Sys moment conditions. This results since, as stated in Theorem 6, the degrees of freedom parameter of the non-central χ^2 large sample distribution in case of the AS moment conditions is one less than the one which results for the Sys moment conditions while they have the same non-centrality parameter.

Figure 2: Power envelopes and maximal attainable power curve when $T = 4$



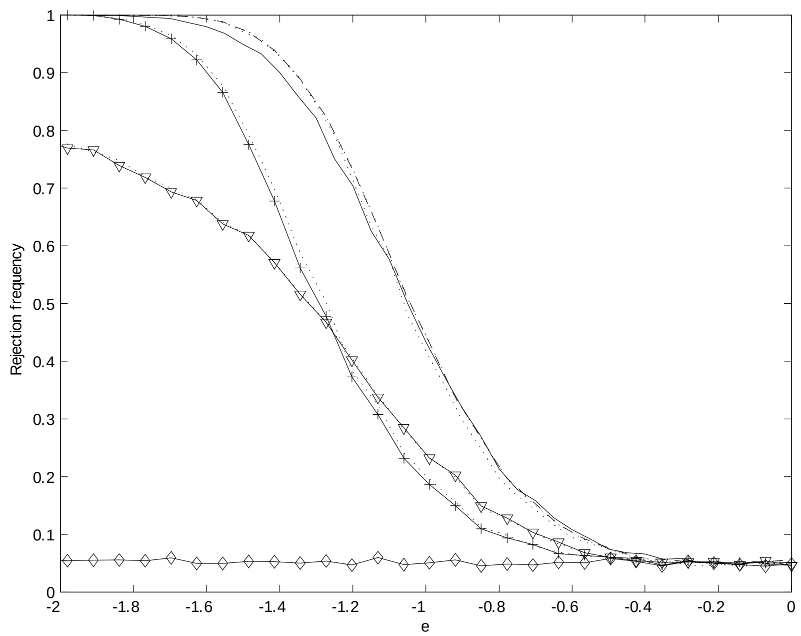
Note: Sys moment conditions: KLM statistic (dashed), GMM-AR (solid with plusses),

GMM-LM (solid with triangles), maximal attainable power curve (solid).

AS moment conditions (dotted lines); Dif moment conditions: GMM-AR (solid with diamonds). N=2000.

[15] For every statistic and the maximal attainable power curve, we use both the Sys and AS moment conditions. For all of these, the results from the AS moment conditions are reflected by a dotted line so there are four dotted lines. Most of these dotted lines are not visible since they are on top of some of the other lines in the figures.

Figure 3: Power envelopes and maximal attainable power curve when $T = 5$



Note: Sys moment conditions: KLM statistic (dashed), GMM-AR (solid with plusses),

GMM-LM (solid with triangles), maximal attainable power curve (solid).

AS moment conditions (dotted lines); Dif moment conditions: GMM-AR (solid with diamonds). N=2000.

Figure 3 shows that, when $T = 5$, the power envelopes that result from using the KLM statistic with either the AS or Sys moment conditions are on the maximal attainable power curve. Figure 3 also shows that the power envelopes which result from the GMM-LM and GMM-AR statistics are below this power curve which is in line with Theorem 6 since the simplification of the worst case large sample distribution of the GMM-LM statistic only applies to $T = 4$. The power envelopes that result from using either the AS or Sys moment conditions are the same for the KLM and GMM-LM statistics while those that result from the GMM-AR statistic using the AS moment conditions are slightly above the ones for the GMM-AR statistic using the Sys moment conditions. This again results from the smaller degrees of freedom parameter of the worst case non-central χ^2 limiting distribution of the GMM-AR statistic when we use the AS moment conditions compared to the Sys moment conditions while they have identical non-centrality parameters.

7 Conclusions

We show that the Dif, Lev and NL moment conditions separately do not identify the parameters in dynamic panel data models for a general number of time periods. This results from the divergence of the initial observations for some plausible data generating processes involving highly persistent panel data. When there are more than three time periods, however, the AS and Sys moment conditions do lead to identification.

Despite identification from four time periods onwards, GMM estimators based on the AS and Sys moment conditions behave in a non-standard manner so inference based on standard Wald statistics is difficult. We therefore use size correct GMM statistics to conduct inference. To recommend which one to use, we compare their worst case rejection frequencies with the maximal attainable power curve under worst case settings. The resulting rejection frequencies of the GMM-AR statistic are below this power curve whilst the rejection frequencies of the GMM-LM statistic are on it when there are four time periods and below it for more time periods. The rejection frequencies of the KLM statistic are always on the maximal attainable power curve for all number of time periods. This makes it our recommended statistic since it is also efficient for smaller values of the autoregressive parameter.

The maximal attainable power curves under worst case settings that result for the AS and Sys moment conditions coincide for all number of time periods. It shows that the additional assumption of mean stationarity made by the Sys moment conditions is not helpful for identification. This results since the worst case DGPs all satisfy the mean stationarity conditions.

The worst case DGPs imply a large variance of the initial observations which hampers identification. This large variance can either result from the disturbance term of the initial observation or from the individual specific fixed effect. In the first case, the identification issues are confined to the unit root value of the autoregressive parameter. In the second case, they can occur at any value of the autoregressive parameter. The first case is considered more pervasive so we primarily focus on it. Our results, however, extend in a straightforward manner to the second case as well.

We have analyzed identification in a worst case scenario setting. The identification issues of the autoregressive parameters at the unit root value are resolved when the mean-stationarity condition is violated (in which case we cannot use the Lev and Sys moment conditions) or when the variances of the disturbances are not constant over time.

Finally, for expository purposes we only have analyzed the first-order autoregressive panel data model. The extension to panel data models with multiple endogenous regressors, e.g. dynamic models with additional endogenous regressors, is an important area for future research.

Appendix A. Proofs

Proof of Theorem 1. T=4. We can write the Dif sample moments and their derivatives as

$$f_N^{Dif}(\theta) = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} y_{i1}\Delta y_{i3} - \theta y_{i1}\Delta y_{i2} \\ y_{i1}\Delta y_{i4} - \theta y_{i1}\Delta y_{i3} \\ y_{i2}\Delta y_{i4} - \theta y_{i2}\Delta y_{i3} \end{pmatrix}$$
$$= \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} \theta_{0,N}-1-\theta & 1 & 0 \\ (\theta_{0,N}-\theta)(\theta_{0,N}-1) & \theta_{0,N}-1-\theta & 1 \\ (\theta_{0,N}-\theta)\,\theta_{0,N}(\theta_{0,N}-1) & \theta_{0,N}(\theta_{0,N}-1-\theta) & \theta_{0,N} \end{pmatrix} \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} +$$
$$(\theta_{0,N}-\theta)(\theta_{0,N}-1)y_{i1}u_{i1}\begin{pmatrix} 1 \\ \theta_{0,N} \\ \theta_{0,N}^2 \end{pmatrix} + \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 0 \\ 0 \\ (u_{i2}+(\theta_0-1)u_{i1})(\Delta y_{i4}-\theta\Delta y_{i3}) \end{pmatrix},$$

$$q_N^{Dif}(\theta) = -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} y_{i1}\Delta y_{i2} \\ y_{i1}\Delta y_{i3} \\ y_{i2}\Delta y_{i3} \end{pmatrix}$$
$$= -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 1 & 0 & 0 \\ \theta_{0,N}-1 & 1 & 0 \\ \theta_{0,N}(\theta_{0,N}-1) & \theta_{0,N} & 0 \end{pmatrix}\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - (\theta_{0,N}-1)y_{i1}u_{i1}\begin{pmatrix} 1 \\ \theta_{0,N} \\ \theta_{0,N} \end{pmatrix}$$
$$-\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 0 \\ 0 \\ (u_{i2}+(\theta_{0,N}-1)u_{i1})\Delta y_{i3} \end{pmatrix}.$$

Under (16) and (20), these expressions are approximately equal to:

$$f_N^{Dif}(\theta) \approx \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \end{pmatrix}\left[\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 \lim_{N\to\infty} E((1-\theta_{0,N})u_{i1}^2\right] +$$
$$\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 0 \\ 0 \\ (u_{i2}+(\theta_{0,N}-1)u_{i1})(u_{i4}-\theta u_{i3}) \end{pmatrix},$$
$$q_N^{Dif}(\theta) \approx -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}\left[\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 \lim_{N\to\infty} E((1-\theta_{0,N})u_{i1}^2)\right]$$
$$-\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 0 \\ 0 \\ (u_{i2}+(\theta_{0,N}-1)u_{i1})u_{i3} \end{pmatrix}.$$

Since

$$\frac{h_N(\theta_{0,N})}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4}\end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix}\psi_{y_{i1}u_{i2}} \\ \psi_{y_{i1}u_{i3}} \\ \psi_{y_{i1}u_{i4}}\end{pmatrix} = \psi,$$

$$\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} u_{i1}u_{i2} \\ u_{i1}u_{i3} \\ u_{i1}u_{i4} \\ u_{i2}^2 \\ u_{i2}u_{i3} \\ u_{i2}u_{i4} \\ u_{i3}^2 \\ u_{i3}u_{i4}\end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ 0 \\ 0 \\ \sigma_3^2 \\ 0\end{pmatrix}\right] \underset{d}{\rightarrow} \begin{pmatrix}\psi_{u_{i1}u_{i2}} \\ \psi_{u_{i1}u_{i3}} \\ \psi_{u_{i1}u_{i4}} \\ \psi_{u_{i2}u_{i2}} \\ \psi_{u_{i2}u_{i3}} \\ \psi_{u_{i2}u_{i4}} \\ \psi_{u_{i3}u_{i3}} \\ \psi_{u_{i3}u_{i4}}\end{pmatrix} = \psi_{uu},$$

with ψ and ψ_{uu} normally distributed random variables, it is readily seen that

$$\begin{aligned}
&A_f^{Dif}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1\end{pmatrix},\ B_q^{Dif}(\theta) = \begin{pmatrix} 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0\end{pmatrix}, \\
&A_q^{Dif}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0\end{pmatrix},\ B_f^{Dif}(\theta) = \begin{pmatrix} 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&-\theta&1&0&0\end{pmatrix}, \\
&\mu_f^{Dif}(\theta,\sigma^2) = \begin{pmatrix} 0 \\ 0 \\ 0\end{pmatrix},\ \mu_q^{Dif}(\theta,\sigma^2) = \begin{pmatrix} 0 \\ 0 \\ 0\end{pmatrix}, \\
&C_f^{Dif}(\theta) = A_f^{Dif}(\theta)\iota_3,\ C_q^{Dif}(\theta) = A_q^{Dif}(\theta)\iota_3.
\end{aligned}$$

since Assumption 1 implies that $\lim_{\theta_{0,N}\uparrow 1} E((1-\theta_{0,N})u_{i1}u_{ij}) = 0,\ j = 2,3.$

We can write the Lev sample moments and their derivatives as

$$\begin{aligned}
&f_N^{Lev}(\theta) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\begin{pmatrix} y_{i3}\Delta y_{i2} - \theta y_{i2}\Delta y_{i2} \\ y_{i4}\Delta y_{i3} - \theta y_{i3}\Delta y_{i3}\end{pmatrix} \\
&= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\begin{pmatrix} 1-\theta & 0 & 0 \\ (1-\theta)(\theta_{0,N}-1) & 1-\theta & 0\end{pmatrix}\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4}\end{pmatrix} + (1-\theta)(\theta_{0,N}-1)y_{i1}u_{i1}\begin{pmatrix} 1 \\ \theta_{0,N}\end{pmatrix} + \\
&\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\begin{pmatrix} (1+(\theta_{0,N}-\theta))(\theta_{0,N}-1)u_{i1}\Delta y_{i2} + (\theta_{0,N}-\theta)u_{i2}\Delta y_{i2} + u_{i3}\Delta y_{i2} \\ (1+(\theta_{0,N}-\theta)(1+\theta_{0,N}))(\theta_{0,N}-1)u_{i1}\Delta y_{i3} + (\theta_{0,N}-\theta)u_{i3}\Delta y_{i3}\end{pmatrix} + \\
&\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\begin{pmatrix} 0 \\ (\theta_{0,N}-\theta)\theta_{0,N}u_{i2}\Delta y_{i3} + u_{i4}\Delta y_{i3}\end{pmatrix},
\end{aligned}$$

$$q_N^{Lev}(\theta) = -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} y_{i2}\Delta y_{i2} \\ y_{i3}\Delta y_{i3} \end{pmatrix}$$
$$= -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 1 & 0 & 0 \\ \theta_{0,N}-1 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - (\theta_{0,N}-1)y_{i1}u_{i1} \begin{pmatrix} 1 \\ \theta_{0,N} \end{pmatrix}$$
$$-\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (\theta_{0,N}-1)u_{i1}\Delta y_{i2} + u_{i2}\Delta y_{i2} \\ (1+\theta_{0,N})(\theta_{0,N}-1)u_{i1}\Delta y_{i3} + u_{i3}\Delta y_{i3} + \theta_{0,N}u_{i2}\Delta y_{i3} \end{pmatrix}.$$

Under (16) and (20), these expressions are approximately equal to:

$$f_N^{Lev}(\theta) = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 1-\theta & 0 & 0 \\ 0 & 1-\theta & 0 \end{pmatrix} \left[\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 \lim_{N\to\infty} E((1-\theta_{0,N})u_{i1}^2) \right] +$$
$$\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} (1-\theta)u_{i2}^2 + u_{i2}u_{i3} \\ (1-\theta)u_{i3}^2 + (1-\theta)u_{i2}u_{i3} + u_{i3}u_{i4} \end{pmatrix},$$

$$q_N^{Lev}(\theta) = -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} y_{i2}\Delta y_{i2} \\ y_{i3}\Delta y_{i3} \end{pmatrix}$$
$$= -\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \left[\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_{0,N}\uparrow 1}(1-\theta_{0,N})u_{i1}^2) \right]$$
$$-\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} u_{i2}^2 \\ u_{i3}^2 + u_{i2}u_{i3} \end{pmatrix},$$

so this implies that

$$A_f^{Lev}(\theta) = \begin{pmatrix} 1-\theta & 0 \\ 0 & 1-\theta \end{pmatrix},$$
$$B_f^{Lev}(\theta) = \begin{pmatrix} 0 & 0 & 0 & 1-\theta & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-\theta & 0 & 1-\theta & 1 \end{pmatrix},$$
$$A_q^{Lev}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$
$$B_q^{Lev}(\theta) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix},$$
$$\mu_f^{Lev}(\theta,\sigma^2) = (1-\theta)\begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \end{pmatrix},\ \mu_q^{Lev}(\theta,\sigma^2) = \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \end{pmatrix}.$$

We can write the NL sample moment and its derivative as

$$
\begin{aligned}
&f_N^{NL}(\theta) = \tfrac{1}{N}\textstyle\sum_{i=1}^N \left(y_{i4} - \theta y_{i3}\right)\left(\Delta y_{i3} - \theta \Delta y_{i2}\right) \\
&= \tfrac{1}{N}\textstyle\sum_{i=1}^N \begin{pmatrix} (1-\theta)(\theta_{0,N}-\theta-1) & (1-\theta) & 0 \end{pmatrix} \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} + (\theta_{0,N}-1)(\theta_{0,N}-\theta)(1-\theta)y_{i1}u_{i1} \\
&+\tfrac{1}{N}\textstyle\sum_{i=1}^N \left((\theta_{0,N}-1)(1+(\theta_{0,N}-\theta)(1+\theta_{0,N}))u_{i1} + \theta_{0,N}(\theta_{0,N}-\theta)u_{i2}\right)\left(\Delta y_{i3} - \theta \Delta y_{i2}\right), \\
&+\tfrac{1}{N}\textstyle\sum_{i=1}^N ((\theta_{0,N}-\theta)u_{i3} + u_{i4})\left(\Delta y_{i3} - \theta \Delta y_{i2}\right) \\
&q_N^{NL}(\theta) = -\tfrac{1}{N}\textstyle\sum_{i=1}^N \begin{pmatrix} \theta_{0,N}-2\theta & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} + (\theta_{0,N}-1)(1+\theta_{0,N}-2\theta)y_{i1}u_{i1} \\
&-\tfrac{1}{N}\textstyle\sum_{i=1}^N \left[(\theta_{0,N}-1)(1+(\theta_{0,N}-2\theta)(1+\theta_{0,N}))u_{i1} + \theta_{0,N}(\theta_{0,N}-2\theta)u_{i2} + (\theta_{0,N}-2\theta)u_{i3} + u_{i4}\right]\Delta y_{i2} \\
&-\tfrac{1}{N}\textstyle\sum_{i=1}^N \left[(1+\theta_{0,N})(\theta_{0,N}-1)u_{i1} + \theta_{0,N}u_{i2} + u_{i3}\right]\Delta y_{i3}.
\end{aligned}
$$

Under (16) and (20), these expressions are approximately equal to:

$$
\begin{aligned}
&f_N^{NL}(\theta) = \tfrac{1}{N}\textstyle\sum_{i=1}^N \begin{pmatrix} \theta(\theta-1) & 1-\theta & 0 \end{pmatrix} \left[\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_{0,N}\uparrow 1}(1-\theta_{0,N})u_{i1}^2)\right] + \\
&\tfrac{1}{N}\textstyle\sum_{i=1}^N \left((1-\theta)u_{i2} + (1-\theta)u_{i3} + u_{i4}\right)\left(u_{i3} - \theta u_{i2}\right),
\end{aligned}
$$

$$
\begin{aligned}
&q_N^{NL}(\theta) = -\tfrac{1}{N}\textstyle\sum_{i=1}^N \begin{pmatrix} 1-2\theta & 1 & 0 \end{pmatrix} \left[\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_{0,N}\uparrow 1}(1-\theta_{0,N})u_{i1}^2)\right] \\
&-\tfrac{1}{N}\textstyle\sum_{i=1}^N ((1-2\theta)\left(u_{i2}+u_{i3}\right) + u_{i4})u_{i2} \\
&-\tfrac{1}{N}\textstyle\sum_{i=1}^N \left[u_{i2} + u_{i3}\right]u_{i3},
\end{aligned}
$$

so this implies that:

$$
\begin{aligned}
A_f^{NL}(\theta) &= \begin{pmatrix} \theta(\theta-1) & 1-\theta \end{pmatrix}, \\
B_f^{NL}(\theta) &= \begin{pmatrix} 0 & 0 & 0 & -\theta(1-\theta) & (1-\theta)^2 & -\theta & 1-\theta & 1 \end{pmatrix}, \\
A_q^{NL}(\theta) &= \begin{pmatrix} 2\theta-1 & -1 \end{pmatrix}, \\
B_q^{NL}(\theta) &= \begin{pmatrix} 0 & 0 & 0 & 2\theta-1 & 2\theta-2 & -1 & -1 & 0 \end{pmatrix}, \\
\mu_f^{NL}(\theta,\sigma^2) &= (1-\theta)\left(\sigma_3^2 - \theta\sigma_2^2\right),\ \mu_q^{NL}(\theta,\sigma^2) = 2\left(\theta-1\right)\sigma_2^2 - \sigma_3^2.
\end{aligned}
$$

Finally, regarding AS and Sys moment conditions we simply have

$$A_f^{Sys}(\theta) = \begin{pmatrix} A_f^{Dif}(\theta) \\ A_f^{Lev}(\theta) \vdots 0 \end{pmatrix}, \; A_q^{Sys}(\theta) = \begin{pmatrix} A_q^{Dif}(\theta) \\ A_q^{Lev}(\theta) \vdots 0 \end{pmatrix},$$
$$B_f^{Sys}(\theta) = \begin{pmatrix} B_f^{Dif}(\theta) \\ B_f^{Lev}(\theta) \end{pmatrix}, \; B_q^{Sys}(\theta) = \begin{pmatrix} B_q^{Dif}(\theta) \\ B_q^{Lev}(\theta) \end{pmatrix},$$
$$\mu_f^{Sys}(\theta,\sigma^2) = \begin{pmatrix} \mu_f^{Dif}(\theta,\sigma^2) \\ \mu_f^{Lev}(\theta,\sigma^2) \end{pmatrix}, \; \mu_q^{Sys}(\theta,\sigma^2) = \begin{pmatrix} \mu_q^{Dif}(\theta,\sigma^2) \\ \mu_q^{Lev}(\theta,\sigma^2) \end{pmatrix}.$$

$$A_f^{AS}(\theta) = \begin{pmatrix} A_f^{Dif}(\theta) \\ A_f^{NL}(\theta) \vdots 0 \end{pmatrix}, \; A_q^{AS}(\theta) = \begin{pmatrix} A_q^{Dif}(\theta) \\ A_q^{NL}(\theta) \vdots 0 \end{pmatrix},$$
$$B_f^{AS}(\theta) = \begin{pmatrix} B_f^{Dif}(\theta) \\ B_f^{NL}(\theta) \end{pmatrix}, \; B_q^{AS}(\theta) = \begin{pmatrix} B_q^{Dif}(\theta) \\ B_q^{NL}(\theta) \end{pmatrix},$$
$$\mu_f^{AS}(\theta,\sigma^2) = \begin{pmatrix} \mu_f^{Dif}(\theta,\sigma^2) \\ \mu_f^{NL}(\theta,\sigma^2) \end{pmatrix}, \; \mu_q^{AS}(\theta,\sigma^2) = \begin{pmatrix} \mu_q^{Dif}(\theta,\sigma^2) \\ \mu_q^{NL}(\theta,\sigma^2) \end{pmatrix}.$$

T=5. Using similar calculations we can write:

$$\psi = \begin{pmatrix} \psi_{y_{i1}u_{i2}} \\ \psi_{y_{i1}u_{i3}} \\ \psi_{y_{i1}u_{i4}} \\ \psi_{y_{i1}u_{i5}} \end{pmatrix},$$

$$A_f^{Dif}(\theta) = \begin{pmatrix} -\theta & 1 & 0 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \end{pmatrix}, \; \mu_f^{Dif}(\theta,\sigma^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$A_f^{Lev}(\theta) = \begin{pmatrix} 1-\theta & 0 & 0 & 0 \\ 0 & 1-\theta & 0 & 0 \\ 0 & 0 & 1-\theta & 0 \end{pmatrix}, \; \mu_f^{Lev}(\theta,\sigma^2) = (1-\theta)\begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \\ \sigma_4^2 \end{pmatrix},$$

$$A_f^{NL}(\theta) = \begin{pmatrix} \theta(\theta-1) & 1-\theta & 0 & 0 \\ 0 & \theta(\theta-1) & 1-\theta & 0 \end{pmatrix}, \; \mu_f^{NL}(\theta,\sigma^2) = (1-\theta)\begin{pmatrix} \sigma_3^2 - \theta\sigma_2^2 \\ \sigma_4^2 - \theta\sigma_3^2 \end{pmatrix}.$$

General T. We have for linear moment conditions, i.e. $j = Dif, Lev, Sys$, that

$$\mu_f^j(\theta,\sigma^2) = (1-\theta)\,\mu_q^j(\theta,\sigma^2),$$

with k_j the number of moment conditions. Furthermore, due to linearity of the Dif, Lev and Sys moment conditions $\mu_q^j(\theta,\sigma^2)$ and $A_q^j(\theta)$ do not depend on θ.

Orthogonal complements of $A_f^{AS}(\theta)$ and $A_f^{Sys}(\theta)$ for $T=4$ and 5. We specify the orthogonal complements as in (31), which we repeat here for convenience:

$$A_f^j(\theta)_\perp = (G_{f,T}^j(\theta) \vdots G_{2,T}^j),$$

where T indicates the number of time periods and $G_{2,T}^j$ is such that $G_{2,T}^{j\prime}\mu_f^j(\theta,\sigma^2)=0$. This notation is used in the proofs of subsequent theorems.

T=4. From the expressions of $A_f^j(\theta)$ and $\mu_f^j(\theta,\sigma^2)$ in (29), $G_{f,T=4}^j(\theta)$ and $G_{2,T=4}^j$ for $j=AS$, Sys result as:

$$\begin{aligned} G_{f,T=4}^{AS}(\theta) &= \begin{pmatrix} -(1-\theta) \\ 0 \\ 0 \\ 1 \end{pmatrix}, \; G_{2,T=4}^{AS} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \\ G_{f,T=4}^{Sys}(\theta) &= \begin{pmatrix} -(1-\theta) \\ 0 \\ 0 \\ -\theta \\ 1 \end{pmatrix}, \; G_{2,T=4}^{Sys} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

From these expresssions it follows that $A_f^j(\theta)'_\perp\mu_f^j(\theta,\sigma^2)\neq 0$, for $j=AS$, Sys.

T=5. The expressions for $A_f^j(\theta)$, $\mu_f^j(\theta,\sigma^2)$, $G_{f,T=5}^j(\theta)$ and $G_{2,T=5}^j$ for $j=AS$, Sys are:

$$\begin{aligned} A_f^{AS}(\theta) &= \begin{pmatrix} -\theta & 1 & 0 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ \theta(\theta-1) & 1-\theta & 0 & 0 \\ 0 & \theta(\theta-1) & 1-\theta & 0 \end{pmatrix}, \; \mu_f^{AS}(\theta,\sigma^2) = (1-\theta)\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \sigma_3^2-\theta\sigma_2^2 \\ \sigma_4^2-\theta\sigma_3^2 \end{pmatrix}, \\ G_{f,T=5}^{AS}(\theta) &= \begin{pmatrix} -(1-\theta) & 0 & 0 \\ 0 & -(1-\theta) & 0 \\ 0 & 0 & -(1-\theta) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \; G_{2,T=5}^{AS} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

$$
A_f^{Sys}(\theta) = \begin{pmatrix} -\theta & 1 & 0 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 1-\theta & 0 & 0 & 0 \\ 0 & 1-\theta & 0 & 0 \\ 0 & 0 & 1-\theta & 0 \end{pmatrix}, \ \mu_f^{Sys}(\theta,\sigma^2) = (1-\theta) \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ \sigma_3^2 \\ \sigma_4^2 \end{pmatrix},
$$

$$
G_{f,T=5}^{Sys}(\theta) = \begin{pmatrix} -(1-\theta) & 0 & 0 \\ 0 & -(1-\theta) & 0 \\ 0 & 0 & -(1-\theta) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\theta & 0 & 0 \\ 1 & -\theta & -\theta \\ 0 & 1 & 1 \end{pmatrix}, \ G_{2,T=5}^{Sys} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},
$$

From these expresssions it follows that $A_f^j(\theta)'_\perp \mu_f^j(\theta,\sigma^2) \neq 0$, for $j = AS,\ Sys$.

The above specification of $A_f^j(\theta)_\perp$ as equal to $(G_{f,T}^j(\theta) \vdots G_{2,T}^j)$ is such that $(A_f^j(\theta) \vdots A_f^j(\theta)_\perp)$ is not invertible for the AS moment conditions both when $T = 4$ and 5. The invertibility of $(A_f^j(\theta) \vdots A_f^j(\theta)_\perp)$ is not needed for the construction of the maximal attainable power curve. It is, however, needed for obtaining the worst case large sample distributions of the GMM-AR, GMM-LM and KLM statistics. Instead of the current specfcation of $A_f^j(\theta)_\perp$, we then use a specification of $A_f^j(\theta)_\perp$:

$$
A_f^j(\theta)_\perp = (G_{f,T}^j(\theta) \vdots G_{2,T}^j)Q,
$$

with Q an identity matrix for the Sys moment conditions and a 2×1 matrix for the AS moment conditions when $T = 4$ and a 5×4 matrix when $T = 5$ which are such that

$$
\begin{aligned}
Q = & \begin{pmatrix} 1 \\ -(G_{2,T=4}^{AS\prime}\hat{V}_{ff}(\theta)G_{2,T=4}^{AS})^{-1}G_{2,T=4}^{AS\prime}\hat{V}_{ff}(\theta)G_{f,T=4}^{AS}(\theta) \end{pmatrix} & T = 4 \\
= & \begin{pmatrix} I_4 \\ -(\binom{1}{0}G_{2,T=4}^{AS\prime}\hat{V}_{ff}(\theta)G_{2,T=4}^{AS}\binom{1}{0})^{-1}\binom{1}{0}'G_{2,T=4}^{AS\prime}\hat{V}_{ff}(\theta)(G_{f,T=4}^{AS}(\theta) \vdots G_{2,T}^{j}\binom{0}{1}) \end{pmatrix} & T = 5.
\end{aligned}
$$

The specification of Q intends to economize on notation for the proof of Theorem 9. The proof of Theorem 9 constructs the worst case large sample distributions of the GMM-AR, GMM-LM and KLM statistics. The specification of Q stated above implies that the same expressions can be used when these statistics use either the Sys or AS sample moments.

Proof of Theorem 2. Since

$$
\begin{aligned}
\Delta y_{i2} = &\quad c_i - (1-\theta_0)y_{i1} + u_{i2} \\
\Delta y_{i3} = &\quad \theta_0(c_i - (1-\theta_0)y_{i1}) + (\theta_0 - 1)u_{i2} + u_{i3} \\
\Delta y_{i4} = &\quad \theta_0(c_i - (1-\theta_0)y_{i1}) + \theta_0(\theta_0 - 1)u_{i2} + (\theta_0 - 1)u_{i3} + u_{i4} \\
\Delta y_{i5} = &\quad \theta_0(c_i - (1-\theta_0)y_{i1}) + \theta_0(\theta_0 - 1)u_{i2} + \theta_0(\theta_0 - 1)u_{i3} + (\theta_0 - 1)u_{i4} + u_{i5} \\
y_{i3} - y_{i1} = &\quad (1+\theta_0)(c_i - (1-\theta_0)y_{i1}) + \theta_0 u_{i2} + u_{i3} \\
y_{i4} - y_{i1} = &\quad (1+\theta_0+\theta_0)(c_i - (1-\theta_0)y_{i1}) + \theta_0 u_{i2} + \theta_0 u_{i3} + u_{i4} \\
y_{i4} - y_{i2} = &\quad (\theta_0+\theta_0)(c_i - (1-\theta_0)y_{i1}) + (\theta_0 - 1)u_{i2} + \theta_0 u_{i3} + u_{i4} \\
y_{i5} - y_{i1} = &\quad (1+\theta_0+\theta_0+\theta_0)(c_i - (1-\theta_0)y_{i1}) + \theta_0 u_{i2} + \theta_0 u_{i3} + \theta_0 u_{i4} + u_{i5} \\
y_{i5} - y_{i2} = &\quad (\theta_0+\theta_0+\theta_0)(c_i - (1-\theta_0)y_{i1}) + (\theta_0 - 1)u_{i2} + \theta_0 u_{i3} + \theta_0 u_{i4} + u_{i5}
\end{aligned}
$$

it holds that for

T=4, Sys:

$$
\begin{aligned}
a \underset{p}{\rightarrow} &\quad \binom{E((c_i-(1-\theta_0)y_{i1})^2)+\sigma_2^2}{0} \\
b \underset{p}{\rightarrow} &\quad \binom{-(1+\theta_0)^2E((c_i-(1-\theta_0)y_{i1})^2)-\theta_0^2\sigma_2^2-\sigma_3^2}{-\theta_0^2E((c_i-(1-\theta_0)y_{i1})^2)} \\
d \underset{p}{\rightarrow} &\quad \binom{\theta_0(1+\theta_0+\theta_0^2)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^2(\theta_0-1)\sigma_2^2+\theta_0\sigma_3^2}{\theta_0^3E((c_i-(1-\theta_0)y_{i1})^2)}.
\end{aligned}
$$

T=4, AS:

$$
\begin{aligned}
a \underset{p}{\rightarrow} &\quad \binom{(1+\theta_0)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0\sigma_2^2}{0} \\
b \underset{p}{\rightarrow} &\quad \binom{-((1+\theta_0)^2+1)E((c_i-(1-\theta_0)y_{i1})^2)-\theta_0(2\theta_0-1)\sigma_2^2-\sigma_3^2}{-\theta_0^2E((c_i-(1-\theta_0)y_{i1})^2)} \\
d \underset{p}{\rightarrow} &\quad \binom{\theta_0(1+\theta_0+\theta_0^2)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^2(\theta_0-1)\sigma_2^2+\theta_0\sigma_3^2}{\theta_0^3E((c_i-(1-\theta_0)y_{i1})^2)}.
\end{aligned}
$$

T=5, Sys:

$$a \underset{p}{\rightarrow} \begin{pmatrix} E((c_i-(1-\theta_0)y_{i1})^2)+\sigma_2^2 \\ (1+\theta_0)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0(\theta_0-1)\sigma_2^2+\sigma_3^2 \\ \theta_0^2E((c_i-(1-\theta_0)y_{i1})^2)+(\theta_0-1)^2\sigma_2+\sigma_3^2 \\ 0 \\ 0 \end{pmatrix}$$

$$b \underset{p}{\rightarrow} -\begin{pmatrix} (1+\theta_0)^2E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^2\sigma_2^2+\sigma_3^2 \\ (1+\theta_0+\theta_0^2)(\theta_0+\theta_0^2)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^2(\theta_0^2-1)\sigma_2^2+\theta_0^2\sigma_3^2+\sigma_4^2 \\ (\theta_0+\theta_0^2)^2(\theta_0+\theta_0^2)E((c_i-(1-\theta_0)y_{i1})^2)+(\theta_0^2-1)^2\sigma_2^2+\theta_0^2\sigma_3^2+\sigma_4^2 \\ \theta_0^2E((c_i-(1-\theta_0)y_{i1})^2) \\ \theta_0^3E((c_i-(1-\theta_0)y_{i1})^2) \end{pmatrix}$$

$$d \underset{p}{\rightarrow} \begin{pmatrix} \theta_0(1+\theta_0+\theta_0^2)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^2(\theta_0-1)\sigma_2^2+\theta_0\sigma_3^2 \\ \theta_0^2(1+\theta_0+\theta_0^2+\theta_0^3)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^4(\theta_0-1)\sigma_2^2+\theta_0^2(\theta_0-1)\sigma_3^2+\theta_0\sigma_4^2 \\ \theta_0^2(\theta_0+\theta_0^2+\theta_0^3)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0(\theta_0-1)(\theta_0^3-1)\sigma_2^2+\theta_0^2(\theta_0-1)\sigma_3^2+\theta_0\sigma_4^2 \\ \theta_0^3E((c_i-(1-\theta_0)y_{i1})^2) \\ \theta_0^4E((c_i-(1-\theta_0)y_{i1})^2) \end{pmatrix}.$$

T=5, AS:

$$a \underset{p}{\rightarrow} \begin{pmatrix} (1+\theta_0)E((c_i-(1-\theta_0)y_{i1})^2)+\sigma_2^2 \\ \theta_0(1+\theta_0+\theta_0^2)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^2(\theta_0-1)\sigma_2^2+\theta_0\sigma_3^2 \\ \theta_0^2(1+\theta_0)E((c_i-(1-\theta_0)y_{i1})^2)+(\theta_0^2-1)(\theta_0-1)\sigma_2+\theta_0\sigma_3^2 \\ 0 \\ 0 \end{pmatrix}$$

$$b \underset{p}{\rightarrow} -\begin{pmatrix} -((1+\theta_0)^2+1)E((c_i-(1-\theta_0)y_{i1})^2)-\theta_0(2\theta_0-1)\sigma_2^2-\sigma_3^2 \\ (\theta_0+2\theta_0^2(1+\theta_0+\theta_0^2))E((c_i-(1-\theta_0)y_{i1})^2)+2\theta_0^3(\theta_0-1)\sigma_2^2+\theta_0(2\theta_0-1)\sigma_3^2+\sigma_4^2 \\ \theta_0^2(1+2\theta_0(1+\theta_0))E((c_i-(1-\theta_0)y_{i1})^2)+(2\theta_0^3-\theta_0-1)(\theta_0-1)\sigma_2^2+\theta_0(2\theta_0-1)\sigma_3^2+\sigma_4^2 \\ \theta_0^2E((c_i-(1-\theta_0)y_{i1})^2) \\ \theta_0^3E((c_i-(1-\theta_0)y_{i1})^2) \end{pmatrix}$$

$$d \underset{p}{\rightarrow} \begin{pmatrix} \theta_0(1+\theta_0+\theta_0^2)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^2(\theta_0-1)\sigma_2^2+\theta_0\sigma_3^2 \\ \theta_0^2(1+\theta_0+\theta_0^2+\theta_0^3)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0^4(\theta_0-1)\sigma_2^2+\theta_0^2(\theta_0-1)\sigma_3^2+\theta_0\sigma_4^2 \\ \theta_0^2(\theta_0+\theta_0^2+\theta_0^3)E((c_i-(1-\theta_0)y_{i1})^2)+\theta_0(\theta_0-1)(\theta_0^3-1)\sigma_2^2+\theta_0^2(\theta_0-1)\sigma_3^2+\theta_0\sigma_4^2 \\ \theta_0^3E((c_i-(1-\theta_0)y_{i1})^2) \\ \theta_0^4E((c_i-(1-\theta_0)y_{i1})^2) \end{pmatrix}.$$

Proof of Theorem 3. The components a, b and d in (32) are all sample averages so we can characterize their large sample behavior by

$$a \underset{a}{=} E(a)+\tfrac{\varepsilon_a}{\sqrt{N}},\ b \underset{a}{=} E(b)+\tfrac{\varepsilon_b}{\sqrt{N}},\ d \underset{a}{=} E(d)+\tfrac{\varepsilon_d}{\sqrt{N}},$$

with ε_a, ε_b and ε_d converging to mean zero random variables and the expressions for $E(a)$, $E(b)$ and $E(d)$, when the true value of θ is one, are stated in Theorem 5. To determine the appropriate rate for the local to unity asymptotics regarding θ, we insert

$$\theta = 1 + \tfrac{e}{N^{\xi}}$$

in (32) and determine the appropriate value of ξ :

$$(1+\tfrac{e}{N^{\xi}})^2(E(a)+\tfrac{\varepsilon_a}{\sqrt{N}}) + (1+\tfrac{e}{N^{\xi}})(E(b)+\tfrac{\varepsilon_b}{\sqrt{N}}) + E(d) + \tfrac{\varepsilon_d}{\sqrt{N}}.$$

When $\omega = 0$, $\sigma_t^2 = \sigma^2$ and $\xi = \frac{1}{4}$:

$$\begin{array}{l} (1+\tfrac{e}{\sqrt[4]{N}})^2(E(a)+\tfrac{\varepsilon_a}{\sqrt{N}}) + (1+\tfrac{e}{\sqrt[4]{N}})(E(b)+\tfrac{\varepsilon_b}{\sqrt{N}}) + E(d) + \tfrac{\varepsilon_d}{\sqrt{N}} \quad = \\ E(a)+E(b)+E(d)+\tfrac{1}{\sqrt{N}}(\varepsilon_a+\varepsilon_b+\varepsilon_d+e^2E(a))+ \\ \tfrac{e}{\sqrt[4]{N}}(E(b)+2E(a)) + \tfrac{e}{\sqrt{N}\sqrt[4]{N}}(\varepsilon_b+2\varepsilon_a) + \tfrac{e^2\varepsilon_a}{N} \quad = \\ \tfrac{1}{\sqrt{N}}(\varepsilon_a+\varepsilon_b+\varepsilon_d+e^2E(a)) + \tfrac{e}{\sqrt{N}\sqrt[4]{N}}(\varepsilon_b+2\varepsilon_a) + \tfrac{e^2\varepsilon_a}{N} \end{array}$$

since $E(a)+E(b)+E(d) = 0$ and $E(b) = -2E(a)$ so the appropriate specification for θ follows from $\theta = 1 + \frac{e}{\sqrt[4]{N}}$.
When $\omega \neq 0$, or $\sigma_t^2 \neq \sigma_j^2$, for at least one $t \neq j$, and $\xi = \frac{1}{2}$:

$$\begin{array}{l} (1+\tfrac{e}{\sqrt{N}})^2(E(a)+\tfrac{\varepsilon_a}{\sqrt{N}}) + (1+\tfrac{e}{\sqrt{N}})(E(b)+\tfrac{\varepsilon_b}{\sqrt{N}}) + E(d) + \tfrac{\varepsilon_d}{\sqrt{N}} \quad = \\ E(a)+E(b)+E(d)+\tfrac{1}{\sqrt{N}}(\varepsilon_a+\varepsilon_b+\varepsilon_d+e(E(b)+2E(a)))+ \\ \tfrac{e}{N}(2\varepsilon_a+\varepsilon_b+eE(a)) + \tfrac{e^2\varepsilon_a}{N\sqrt{N}} \quad = \\ \tfrac{1}{\sqrt{N}}(\varepsilon_a+\varepsilon_b+\varepsilon_d-e(E(b)+2E(a))) - \tfrac{e}{N}(2\varepsilon_a+\varepsilon_b-eE(a)) + \tfrac{e^2\varepsilon_a}{N\sqrt{N}} \end{array}$$

since $E(a)+E(b)+E(d) = 0$ but $E(b) \neq -2E(a)$ so $\theta = 1 + \frac{e}{\sqrt{N}}$.

Proof of Theorem 4. Denote with $g_{f,T}(e)$ the moments in (32) evaluated at $\theta = 1 + \frac{e}{\sqrt[4]{N}}$.
When $\omega = 0$ and $\sigma_t^2 = \sigma^2$, $g_{f,T}(e)$ is characterized by

$$\begin{aligned} g_{f,T}(e) = & \ (1+\tfrac{e}{\sqrt[4]{N}})^2(E(a)+\tfrac{1}{\sqrt{N}}\varepsilon_a) + (1+\tfrac{e}{\sqrt[4]{N}})(E(b)+\tfrac{1}{\sqrt{N}}\varepsilon_b) + E(d) + \tfrac{1}{\sqrt{N}}\varepsilon_d \\ = & \ \tfrac{1}{\sqrt{N}}(\varepsilon_a+\varepsilon_b+\varepsilon_d+e^2E(a)) + \tfrac{e}{\sqrt{N}\sqrt[4]{N}}(\varepsilon_b+2\varepsilon_a) + \tfrac{e^2\varepsilon_a}{N}. \end{aligned}$$

Therefore, we have

$$\sqrt{N}g_{f,T}(e) = e^2E(a) + (\varepsilon_a(1+\tfrac{2e}{\sqrt[4]{N}}+\tfrac{e^2\varepsilon_a}{\sqrt{N}}) + \varepsilon_b(1+\tfrac{e}{\sqrt[4]{N}}) + \varepsilon_d,$$

and

$$\sqrt{N}g_{f,T}(e) \simeq N(e^2E(a), B(N)'V_{abd}B(N)),$$

with

$$B(N) = (\iota_3 \otimes I_{p_{\max}}) + \tfrac{e}{\sqrt[4]{N}}\left[(2+\tfrac{e}{\sqrt[4]{N}})(e_{1,3}\otimes I_{p_{\max}}) + (e_{2,3}\otimes I_{p_{\max}})\right],$$

and V_{abd} the covariance matrix of $(a' \vdots b' \vdots d')'$, ι_3 a 3×1 dimensional vector of ones, $I_{p_{\max}}$ the $p_{\max} \times p_{\max}$ dimensional identity matrix, $p_{\max}$ equals the number of elements of a and $e_{1,3}$ and $e_{2,3}$ the first and second 3×1 dimensional unity vectors.

The individual moments $g_{f,n}(e)$ $(g_{f,T}(e) = \sum_{n=1}^{N} g_{f,n}(e))$ can be specified as:

$$\begin{aligned}
g_{f,n}(e) = & \quad (1 + \tfrac{e}{\sqrt[4]{N}})^2 a_n + (1 + \tfrac{e}{\sqrt[4]{N}}) b_n + d_n \\
= & \quad (1 + \tfrac{e}{\sqrt[4]{N}})^2 [E(a) + \varepsilon_{a_n}] + (1 + \tfrac{e}{\sqrt[4]{N}})[E(b) + \varepsilon_{b_n}] + [E(d) + \varepsilon_{d_n}] \\
= & \quad (E(a) + E(b) + E(d)) + \tfrac{e}{\sqrt[4]{N}}(2E(a) + E(b)) + \tfrac{e^2}{\sqrt{N}} E(a) + \\
& \quad \varepsilon_{a_n} + \varepsilon_{b_n} + \varepsilon_{d_n} + \tfrac{e}{\sqrt[4]{N}}(2\varepsilon_{a_n} + \varepsilon_{b_n}) + \tfrac{e^2}{\sqrt{N}} \varepsilon_{a_n} \\
= & \quad \tfrac{e^2}{\sqrt{N}} E(a) + \varepsilon_{a_n} + \varepsilon_{b_n} + \varepsilon_{d_n} + \tfrac{e}{\sqrt[4]{N}}(2\varepsilon_{a_n} + \varepsilon_{b_n}) + \tfrac{e^2}{\sqrt{N}} \varepsilon_{a_n},
\end{aligned}$$

with $a = \frac{1}{N} \sum_{n=1}^{N} a_n$, $b = \frac{1}{N} \sum_{n=1}^{N} b_n$, $d = \frac{1}{N} \sum_{n=1}^{N} d_n$, $\varepsilon_{a_n} = a_n - E(a)$, $\varepsilon_{b_n} = b_n - E(b)$, $\varepsilon_{d_n} = d_n - E(d)$, so taking $g_{t,n}(e)$ is deviation from its sample average $g_{f,T}(e)$ results in

$$g_{f,n}(e) - g_{f,T}(e) = \quad \varepsilon_{a_n} + \varepsilon_{b_n} + \varepsilon_{d_n} + \tfrac{e}{\sqrt[4]{N}}(2\varepsilon_{a_n} + \varepsilon_{b_n}) - \tfrac{1}{\sqrt{N}}((\varepsilon_a(1 + \tfrac{2e}{\sqrt[4]{N}} + \tfrac{e^2 \varepsilon_a}{\sqrt{N}}) + \varepsilon_b(1 + \tfrac{e}{\sqrt[4]{N}}) + \varepsilon_d)$$

From the above, it then straightforwardly follows that

$$\hat{V}_{gg}(e) = \tfrac{1}{N} \textstyle\sum_{i=1}^{N} (g_{f,n}(e) - g_{f,T}(e))(g_{f,n}(e) - g_{f,T}(e))' \simeq B(N)' V_{abd} B(N),$$

so the large sample distribution of the GMM-AR statistic is characterized by

$$\chi^2(\delta, p_{\max}),$$

with $\delta = e^4 E(a)' [B(N)' V_{abd} B(N)]^{-1} E(a)$.

Proof of Theorem 5. When we instead of the full vector $g_{f,T}(e)$ use a linear combination of it, say $w' g_{f,T}(e)$ with w an orthonormal $p_{\max} \times 1$ vector, the approximating distribution of the GMM-AR statistic for testing $\mathrm{H}_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ that uses $w' g_{f,T}(e)$ as the moment vector reads

$$\chi^2(e^4 (w' E(a))' [w' B(N)' V_{abd} B(N) w]^{-1} (w' E(a)), 1).$$

The optimal combination w is the one that leads to the largest value of the non-centrality parameter. The non-centrality parameter can be specified as

$$e^4 (w' E(a))' [w' B(N)' V_{abd} B(N) w]^{-1} (w' E(a)) \quad = e^4 \tfrac{(w' E(a))^2}{w' B(N)' V_{abd} B(N) w}.$$

The maximal value of $\frac{(w' E(a))^2}{w' B(N)' V_{abd} B(N) w}$ results from the largest root of the generalized eigenvalue problem

$$\left| \lambda B(N)' V_{abd} B(N) - E(a) E(a)' \right| = 0$$

and the optimal value of w equals the eigenvector associated with the largest root. Since $E(a)$ is only a vector, just one root of the generalized eigenvalue problem is non-zero so it is also the largest one. This root results from using

$$w = \quad (B(N)'V_{abd}B(N))^{-1}E(a)$$

and the largest root then equals

$$\lambda_{\max} = \quad E(a)'(B(N)'V_{abd}B(N))^{-1}E(a)$$

so the maximal value of the non-centrality parameter is

$$\delta = e^4 E(a)'(B(N)'V_{abd}B(N))^{-1}E(a) = (e\sigma)^4 \binom{\iota_p}{0}'(B(N)'V_{abd}B(N))^{-1}\binom{\iota_p}{0}$$

since $E(a) = \sigma^2\binom{\iota_p}{0}$ with ι_p a $p \times 1$ dimensional vector of ones and p the number of columns of $G_{f,T}(\theta)$.

Proof of Theorem 6. GMM-AR statistic To construct the worst case limiting distribution of the GMM-AR statistic to test $\mathrm{H}_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ whilst the true value of θ equals one, we first specify the GMM-AR statistic as

$$\begin{aligned}
\text{GMM-AR}(e) = \quad & N f_N(e)'\hat{V}_{ff}(e)^{-1}f_N(e) \\
= \quad & \left[\sqrt{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'f_N(e)\right]' \\
& \left[(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)\right]^{-1} \\
& \left[\sqrt{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'f_N(e)\right].
\end{aligned}$$

We now determine the limit behavior of the different components under the limit sequence in (20). The specification of $A_f(e)_\perp$ that we use is such that

$$A_f(e)_\perp = \quad (G^j_{f,T}(e) \vdots G^j_{2,T})Q,$$

with Q equal to the identity matrix for the Sys moment conditions and equal to the specifications stated in the proof of Theorem 4 for the AS moment conditions. The large sample behavior of the different components of Q for the AS moment conditions are such that

$$\begin{aligned}
G^{AS\prime}_{2,T}\hat{V}_{ff}(e)G^{AS}_{2,T} \approx \quad & \binom{0}{I_{p_{\max}-p}}'B(N)'V_{abd}B(N)\binom{0}{I_{p_{\max}-p}} \\
G^{AS\prime}_{2,T}\hat{V}_{ff}(e)G^{AS}_{f,T}(\theta) \approx \quad & \binom{0}{I_{p_{\max}-p}}'B(N)'V_{abd}B(N)\binom{I_p}{0},
\end{aligned}$$

with p the number of columns of $G^{AS}_{f,T}(e)$ so the limit behavior of Q is

$$\begin{aligned}
Q \approx \quad & \binom{1}{-\left(\binom{0}{I_{p_{\max}-p}}'B(N)'V_{abd}B(N)\binom{0}{I_{p_{\max}-p}}\right)^{-1}\binom{0}{I_{p_{\max}-p}}'B(N)'V_{abd}B(N)\binom{I_p}{0}}, & T = 4 \\
\approx \quad & \binom{I_4}{-\left(\binom{1}{0}'\binom{0}{I_{p_{\max}-p}}'B(N)'V_{abd}B(N)\binom{0}{I_{p_{\max}-p}}\binom{1}{0}\right)^{-1}\binom{1}{0}\binom{0}{I_{p_{\max}-p}}'B(N)'V_{abd}B(N)\left(\binom{I_p}{0} \vdots \binom{0}{1}\right)}, & T = 5 \\
= \quad & \bar{Q},
\end{aligned}$$

and $\bar{Q}$ equals the identity matrix for the Sys moment conditions. Our specification of $A_f(e)_\perp$ is such that

$$\sqrt{N}A_f(e)'_\perp f_N(e) = Q'\left(\sqrt{N}g_{f,T}(e)\right),$$

so using the limit behavior of $\sqrt{N}g_{f,T}(e)$ stated in Theorem 7, we have that

$$\sqrt{N}A_f(e)'_\perp f_N(e) \underset{d}{\to} \bar{Q}'\left[e^2\sigma^2\binom{\iota_p}{0} + B(N)'\begin{pmatrix}\varepsilon_a\\ \varepsilon_b\\ \varepsilon_d\end{pmatrix}\right].$$

The limit behavior of $\sqrt{N}h_N(\theta_{0,N})A_f(e)'f_N(e)$ accords with

$$\sqrt{N}h_N(\theta_{0,N})A_f(e)'f_N(e) \underset{d}{\to} \quad A_f(e)'A_f(e)\psi,$$

so combining,

$$\left[\sqrt{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'f_N(e)\right] \underset{d}{\to} \quad \left[\begin{array}{c} A_f(e)'A_f(e)\psi \\ \bar{Q}'\left(e^2\sigma^2\binom{\iota_p}{0} + B(N)'\begin{pmatrix}\varepsilon_a\\ \varepsilon_b\\ \varepsilon_d\end{pmatrix}\right)\end{array}\right].$$

Under mean stationarity, $\omega = 0$ and $g_{f,T}(e)$ does not depend on the initial observations y_{i1}. This implies that the (normalized) covariance of $A_f(e)'f_N(e)$ and $A_f(e)'_\perp f_N(e)$ equals zero:

$$h_N(\theta_{0,N})A_f(e)'\hat{V}_{ff}(e)A_f(e)_\perp \underset{p}{\to} \quad 0.$$

Under the worst case setting (20) also

$$\begin{aligned} h_N(\theta_{0,N})^2A_f(e)'\hat{V}_{ff}(e)A_f(e) &\underset{p}{\to} \quad A_f(e)'A_f(e)\left[\lim_{N\to\infty}\mathrm{var}(h_N(\theta_{0,N})\begin{pmatrix}y_{1i}u_{i2}\\ \vdots\\ y_{1i}u_{iT}\end{pmatrix})\right] \\ &\qquad A_f(e)'A_f(e) \\ A_f(e)'_\perp\hat{V}_{ff}(e)A_f(e)_\perp &\underset{p}{\to} \quad \bar{Q}'B(N)'V_{abd}B(N)\bar{Q} \end{aligned}$$

so

$$(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp) \underset{p}{\to}$$
$$\begin{pmatrix} A_f(e)'A_f(e)\left[\lim_{N\to\infty}\mathrm{var}(h_N(\theta_{0,N})\begin{pmatrix}y_{1i}u_{i2}\\ \vdots\\ y_{1i}u_{iT}\end{pmatrix})\right]A_f(e)'A_f(e) & 0 \\ 0 & \bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\end{pmatrix}.$$

Because $A_f(e)'f_N(e)$ and $A_f(e)'_\perp f_N(e)$ are uncorrelated under (20),

$$\left[(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)\right]$$

is block diagonal and the limit behavior of the GMM-AR statistic consists of two components, one resulting from the diverging part of the sample moments and one which results from the stable/identifying part:

$$\begin{array}{rl}
\text{i.} & \left[\sqrt{N}h_N(\theta_{0,N})A_f(e)'f_N(e)\right]'\left[h_N(\theta_{0,N})^2A_f(e)'\hat{V}_{ff}(e)A_f(e)\right]^{-1}\left[\sqrt{N}h_N(\theta_{0,N})A_f(e)'f_N(e)\right] \\
& \underset{d}{\rightarrow} \psi'\left[\lim_{N\rightarrow\infty}\mathrm{var}(h_N(\theta_{0,N})\begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT}\end{pmatrix})\right]^{-1}\psi \sim \chi^2(p_{GMM\text{-}AR}-p_{\max}) \\
\text{ii.} & \left(\sqrt{N}A_f(e)'_\perp f_N(e)\right)'\left[A_f(e)'_\perp\hat{V}_{ff}(e)A_f(e)_\perp\right]^{-1}\left(\sqrt{N}A_f(e)'_\perp f_N(e)\right) \\
& \underset{d}{\rightarrow} \chi^2(\delta, p_{\max})
\end{array}$$

with $p_{GMM\text{-}AR} = \frac{1}{2}(T+1)(T-2)$ for the Sys moment conditions and $p_{GMM\text{-}AR} = \frac{1}{2}(T+1)(T-2)-1$ for the AS moment conditions and when $T=4: p=1$, $p_{\max}=1$ for AS and 2 for Sys, $T=5: p=3$, $p_{\max}=5$ for Sys and 4 for AS and

$$\begin{aligned}
\delta = & \; e^4\sigma^4\binom{\iota_p}{0}'\bar{Q}\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}\binom{\iota_p}{0} \\
= & \; e^4\sigma^4\binom{\iota_p}{0}'\left(B(N)'V_{abd}B(N)\right)^{-1}\binom{\iota_p}{0}.
\end{aligned}$$

The latter result can be shown using the partitioned inverse of $(B(N)'V_{abd}B(N))^{-1}$ since

$$\begin{gathered}
\binom{\iota_p}{0}'\left(B(N)'V_{abd}B(N)\right)^{-1}\binom{\iota_p}{0} = \\
\iota_p'\left[\binom{I_p}{0}'\left(B(N)'V_{abd}B(N)\right)\binom{I_p}{0} - \binom{I_p}{0}'B(N)'V_{abd}B(N)\binom{0}{I_{p_{\max}-p}}\right. \\
\left.\left[\binom{0}{I_{p_{\max}-p}}'B(N)'V_{abd}B(N)\binom{0}{I_{p_{\max}-p}}\right]^{-1}\binom{0}{I_{p_{\max}-p}}'B(N)'V_{abd}B(N)\binom{I_p}{0}\right]^{-1}\iota_p = \\
\iota_p'\left[\binom{I_p}{0}'\left(B(N)'V_{abd}B(N)\right)^{-\frac{1}{2}}M_{(B(N)'V_{abd}B(N))^{-\frac{1}{2}}\binom{0}{I_{p_{\max}-p}}}\left(B(N)'V_{abd}B(N)\right)^{-\frac{1}{2}}\binom{I_p}{0}\right]^{-1}\iota_p = \\
\binom{\iota_p}{0}'\bar{Q}\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}\binom{\iota_p}{0},
\end{gathered}$$

which uses that $\bar{Q}$ represents a weighted regression of columns of $G_{2,T}$ on $G_{f,T}(\theta)$ and the remaining columns of $G_{2,T}$ using $B(N)'V_{abd}B(N)$ as weight matrix. Taken alltogether, the worst case large sample distribution of the GMM-AR statistic test $\mathrm{H}_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ reads

$$\text{GMM-AR}(e) \underset{d}{\rightarrow} \chi^2(\delta, p_{GMM\text{-}AR}).$$

GMM-LM statistic To obtain the large sample behavior of the GMM-LM statistic to test $\mathrm{H}_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ when θ_0 is one and under the limiting sequence in (20), we determine the behavior of the different components of:

$$(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'q_N(e)$$

for which we use the representation of $q_N(e)$.

$h_N(\theta_{0,N})' A_f(e)' q_N(e)$: Under the worst case DGPs characterized by (20):

$$\sqrt{N} h_N(\theta_{0,N}) A_f(e)' q_N(e) \approx A_f(e)' \Big[A_q(e)\psi + h_N(\theta_{0,N})\sqrt{N}(\mu_q(e,\sigma^2) + A_q(e)\iota \left(\lim_{\theta_0 \uparrow 1} E((\theta_0-1)u_{i1}^2)\right)) + h_N(\theta_{0,N}) B_q(e)\psi_{cu} \Big] \approx A_f(e)' A_q(e)\psi,$$

since under (20):

$$\sqrt{N} h_N(\theta_{0,N})(\mu_q(e,\sigma^2) + A_q(e)\iota \left(\lim_{\theta_0 \uparrow 1} E((\theta_0-1)u_{i1}^2)\right)) \to 0, \ \ h_N(\theta_{0,N})\sqrt{N} \to 0.$$

$A_f(e)'_\perp q_N(e)$: We distinguish between the AS and Sys moment conditions. For the Sys moment conditions:

$$A_f(e)'_\perp q_N(e) = \bar{Q} \begin{pmatrix} G_{f,T}(e)' q_N(e) \\ G'_{2,T} q_N(e) \end{pmatrix} \approx \ \bar{Q}' \begin{pmatrix} \frac{1}{h_N(\theta_{0,N})\sqrt{N}} G_f(e)' A_q(e)\psi - \frac{e}{\sqrt[4]{N}}\sigma^2 \iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq} \\ \frac{1}{\sqrt{N}}\varepsilon_{bq} \end{pmatrix},$$

since for the Sys moment conditions $G'_{2,T} A_q(e) = 0$, $G'_{2,T}\mu(e,\sigma^2) = 0$, $G_{f,T}(e)' A_q(e)\iota_p = 0$, $G_{f,T}(e)'\mu(e,\sigma^2) = -\frac{e}{\sqrt[4]{N}}\sigma^2\iota_p$ and $\varepsilon_{aq} = G_f(e)' B_q(e)\psi_{cu}$ and $\varepsilon_{bq} = G_{2,T}' B_q(e)\psi_{cu}$ are mean zero normal random variables that capture the remaining random parts.

For the AS moment conditions:

$$A_f(e)'_\perp q_N(e) \approx \bar{Q}' \begin{pmatrix} \frac{1}{h_N(\theta_{0,N})\sqrt{N}} G_f(e)' A_q(e)\psi - \frac{e}{\sqrt[4]{N}}\iota_p \left[2\sigma^2 - \lim_{\theta_0 \uparrow 1} E((\theta_0-1)u_{i1}^2)\right] + \frac{1}{\sqrt{N}}\varepsilon_{aq} \\ \frac{1}{\sqrt{N}}\varepsilon_{bq} \end{pmatrix},$$

since for the AS moment conditions $G'_{2,T} A_q(e) = 0$, $G'_{2,T}\mu(e,\sigma^2) = 0$, $G_{f,T}(e)' A_q(e)\iota = \frac{e}{\sqrt[4]{N}}\iota_p$, $G_f(e)'\mu(\sigma^2) = -\frac{2e}{\sqrt[4]{N}}\sigma^2\iota_p$ and $\varepsilon_{aq} = G_f(e)' B_q(e)\psi_{cu}$ and $\varepsilon_{bq} = G'_{2,T} B_q(e)\psi_{cu}$ are mean zero normal random variables that capture the remaining random parts.

Overall, we can specify $A_f(e)'_\perp q_N(e)$ for both the AS and Sys moment conditions as

$$A_f(e)'_\perp q_N(e) \approx \bar{Q}' \begin{pmatrix} \frac{1}{h_N(\theta_{0,N})\sqrt{N}} G_f(e)' A_q(e)\psi - \frac{\bar{e}}{\sqrt[4]{N}}\iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq} \\ \frac{1}{\sqrt{N}}\varepsilon_{bq} \end{pmatrix},$$

with

$$\begin{aligned} \bar{e} = & \ e\sigma^2 && \text{Sys moment conditions} \\ = & \ e\left[2\sigma^2 - \lim_{\theta_0 \uparrow 1} E((\theta_0-1)u_{i1}^2)\right] && \text{AS moment conditions.} \end{aligned}$$

Combining:

$$\begin{aligned} & (h_N(\theta_{0,N}) A_f(e) \vdots A_f(e)_\perp)' q_N(e) = (h_N(\theta_{0,N}) A_f(e) \vdots (G_{f,T}(e) \vdots G_{2,T}) Q)' q_N(e) \\ \approx & \begin{pmatrix} \frac{1}{\sqrt{N}} A_f(e)' A_q \\ \bar{Q}' \binom{\frac{1}{h_N(\theta_{0,N})\sqrt{N}} G_f(e)' A_q}{0} \end{pmatrix} \psi + \begin{pmatrix} 0 \\ \bar{Q}' \binom{-\frac{\bar{e}}{\sqrt[4]{N}}\iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq}}{\frac{1}{\sqrt{N}}\varepsilon_{bq}} \end{pmatrix}. \end{aligned}$$

Using that (which results from the derivations for the GMM-AR statistic)

$$\sqrt{N}((h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp))^{-1}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'f_N(e)$$
$$\underset{d}{\rightarrow} \begin{pmatrix} (A_f(e)'A_f(e))^{-1} \left[\lim_{N\to\infty} \mathrm{var}(h_N(\theta_{0,N}) \begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{pmatrix}) \right]^{-1} \psi \\ \left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1} \bar{Q}' \left(e^2\sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \end{pmatrix},$$

we now construct the limit behavior of $h_N(\theta_{0,N})Nq_N(e)'\hat{V}_{ff}(e)^{-1}f_N(e)$:

$$h_N(\theta_{0,N})Nq_N(e)'\hat{V}_{ff}(e)^{-1}f_N(e) = h_N(\theta_{0,N})\sqrt{N} \left[(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'q_N(e) \right]'$$
$$((h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp))^{-1}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\sqrt{N}f_N(e) \approx$$
$$\left[\begin{pmatrix} h_N(\theta_{0,N})A_f(e)'A_q(e) \\ \bar{Q}'\binom{G_f(e)'A_q}{0} \end{pmatrix} \psi + h_N(\theta_{0,N})\sqrt{N} \begin{pmatrix} 0 \\ \bar{Q}'\binom{-\frac{\bar{e}}{\sqrt[4]{N}}\sigma^2\iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq}}{\frac{1}{\sqrt{N}}\varepsilon_{bq}} \end{pmatrix} \right]'$$
$$\begin{pmatrix} (A_f(e)'A_f(e))^{-1} \left[\lim_{N\to\infty} \mathrm{var}(h_N(\theta_{0,N}) \begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{pmatrix}) \right]^{-1} \psi \\ \left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1} \bar{Q}' \left(e^2\sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \end{pmatrix} \underset{d}{\rightarrow}$$
$$\begin{pmatrix} G_f(e)'A_q(e)\psi \\ 0 \end{pmatrix}' \bar{Q}' \left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1} \bar{Q}' \left(e^2\sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) =$$
$$\begin{pmatrix} G_f(e)'A_q(e)\psi \\ 0 \end{pmatrix}' (B(N)'V_{abd}B(N))^{-1} \left(e^2\sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right),$$

where $\bar{Q}$ drops out for the same reason as discussed for the GMM-AR statistic. The elements multiplied by $h_N(\theta_{0,N})$ or $h_N(\theta_{0,N})\sqrt{N}$ are under (20) of a smaller order of magnitude and therefore drop out.

The limit behavior of $h_N(\theta_{0,N})^2 N q_N(e)'\hat{V}_{ff}(e)^{-1}q_N(e)$ results in a similar manner:

$$h_N(\theta_{0,N})^2 N q_N(e)'\hat{V}_{ff}(e)^{-1}q_N(e) = h_N(\theta_{0,N})^2 N \left[(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'q_N(e)\right]'$$
$$((h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp))^{-1}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'q_N(e)$$
$$\underset{d}{\to} \begin{pmatrix} G_f(e)'A_q(e)\psi \\ 0 \end{pmatrix}' \bar{Q}\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}' \begin{pmatrix} G_f(e)'A_q(e)\psi \\ 0 \end{pmatrix} =$$
$$\begin{pmatrix} G_f(e)'A_q(e)\psi \\ 0 \end{pmatrix}' (B(N)'V_{abd}B(N))^{-1} \begin{pmatrix} G_f(e)'A_q(e)\psi \\ 0 \end{pmatrix},$$

where again $\bar{Q}$ drops out for the same reason as discussed for the GMM-AR statistic.

Combining everything, we obtain the limit behavior of the GMM-LM statistic to test $\mathrm{H}_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ under (20):

$$\text{GMM-LM}(e) \underset{d}{\to} \eta' P_{(B(N)'V_{abd}B(N))^{-\frac{1}{2}}\binom{G_f(e)'A_q(e)\psi}{0}}\eta$$

with

$$\eta \sim N(e^2\sigma^2\,(B(N)'V_{abd}B(N))^{-\frac{1}{2}}\tbinom{\iota_p}{0}, I_{p_{\max}})$$

and independent from ψ, so

$$\text{GMM-LM}(e) \underset{d}{\to} \chi^2(\delta(\psi), 1),$$

with $\delta(\psi) = e^4\sigma^4\tbinom{\iota_p}{0}'\,(B(N)'V_{abd}B(N))^{-\frac{1}{2}}\,P_{(B(N)'V_{abd}B(N))^{-\frac{1}{2}}\binom{G_f(e)'A_q(e)\psi}{0}}\,(B(N)'V_{abd}B(N))^{-\frac{1}{2}}\,\tbinom{\iota_p}{0}$. The simplification of the non-centrality parameter for the GMM-LM statistic when $T = 4$ results since $G_f(e)'A_q\psi$ is then a scalar so $\tbinom{G_f(e)'A_q\psi}{0} = \tbinom{\iota_p}{0}G_f(e)'A_q\psi$ and $G_f(e)'A_q\psi$ cancels out of the expression of the non-centrality parameter.

KLM statistic. To obtain the large sample distribution of the KLM statistic to test $\mathrm{H}_0 : \theta = 1+\frac{e}{\sqrt[4]{N}}$ when θ_0 is equal to one and under the limiting sequence in (20), we first determine the behavior of

$$(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{\theta f}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp) \underset{p}{\to}$$
$$\begin{pmatrix} A_f(e)'A_q \\ \bar{Q}'\binom{\frac{1}{h_N(\theta_{0,N})}G_f(e)'A_q}{0} \end{pmatrix} \left[\lim_{N\to\infty} \operatorname{var}(h_N(\theta_{0,N}) \begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{pmatrix})\right] \begin{pmatrix} A_f(e)'A_f(e) \\ 0 \\ 0 \end{pmatrix}' +$$
$$\begin{pmatrix} 0 & 0 \\ 0 & \bar{Q}'\binom{V_{aq,abd}B(N)}{V_{bq,abd}B(N)}\bar{Q} \end{pmatrix},$$

with $V_{aq,abd}$, $V_{aq,abd}$ the covariance between ε_{aq} and $(\varepsilon_a' \vdots \varepsilon_b' \vdots \varepsilon_d')'$ and ε_{bq} and $(\varepsilon_a' \vdots \varepsilon_b' \vdots \varepsilon_d')'$ respectively.

Combining with the limit behavior of $\sqrt{N}((h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp))^{-1}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'f_N(e)$:

$$
\begin{array}{l}
\sqrt{N}((h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp))^{-1}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'f_N(e) \\
\underset{d}{\rightarrow} \left(\begin{array}{c} (A_f(e)'A_f(e))^{-1} \left[\lim_{N\rightarrow\infty} \operatorname{var}(h_N(\theta_{0,N}) \left(\begin{array}{c} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{array} \right)) \right]^{-1} \psi \\
\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}' \left(e^2\sigma^2\binom{\iota_p}{0} + B(N)' \left(\begin{array}{c} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{array} \right) \right) \end{array} \right)
\end{array}
$$

so

$$
\begin{array}{c}
\sqrt{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{\theta f}(e)\hat{V}_{ff}(e)^{-1}f_N(e) = \\
\sqrt{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{\theta f}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp) \\
((h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp))^{-1}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'f_N(e) \\
\underset{d}{\rightarrow} \left(\begin{array}{c} A_f(e)'A_q \\ \bar{Q}'\binom{\frac{1}{h_N(\theta_{0,N})}G_f(e)'A_q}{0} \end{array} \right)\psi + \left(\begin{array}{c} 0 \\ \bar{Q}'\binom{V_{aq,abd}B(N)}{V_{bq,abd}B(N)}\bar{Q} \end{array} \right) \\
\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}' \left(e^2\sigma^2\binom{\iota_p}{0} + B(N)' \left(\begin{array}{c} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{array} \right) \right).
\end{array}
$$

Upon combining with the limit behavior of $\sqrt{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'q_N(e)$, the convergence behavior of $\sqrt[4]{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{D}_N(e)$ then results as

$$
\begin{array}{c}
\sqrt[4]{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{D}_N(e) = \frac{1}{\sqrt[4]{N}}\left\{ \left[\sqrt{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'q_N(e) - \right.\right. \\
\left. \sqrt{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{\theta f}(e)\hat{V}_{ff}(e)^{-1}f_N(e) \right] \approx \left(\begin{array}{c} 0 \\ \bar{Q}'\left(\binom{\iota_p}{0}\bar{e} + \frac{1}{\sqrt[4]{N}}\nu\right) \end{array} \right) \\
\underset{d}{\rightarrow} \left(\begin{array}{c} 0 \\ \bar{Q}'\binom{\iota_p}{0} \end{array} \right)\bar{e},
\end{array}
$$

where we have rescaled since all the higher order terms have dropped out and

$$
\begin{array}{c}
\nu = \left(\bar{Q}'\binom{V_{aq,abd}B(N)}{V_{bq,abd}B(N)}\bar{Q}\right)\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}'\binom{\iota_p}{0} + \\
\bar{Q}'\left[\left(\begin{array}{c} \varepsilon_{aq} \\ \varepsilon_{bq} \end{array} \right) - \bar{Q}'\binom{V_{aq,abd}B(N)}{V_{bq,abd}B(N)}\bar{Q}\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}'B(N)' \left(\begin{array}{c} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{array} \right) \right].
\end{array}
$$

We obtain the limit behavior of $\sqrt{N}\hat{D}_N(e)'\hat{V}_{ff}(e)^{-1}D_N(e)$ from:

$$\begin{array}{c} \sqrt{N}\hat{D}_N(e)'\hat{V}_{ff}(e)^{-1}\hat{D}_N(e) = \left[\sqrt[4]{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{D}_N(e)\right]' \\ ((h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp))^{-1}\left[\sqrt[4]{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{D}_N(e)\right] \\ \approx \left[\binom{\iota_p}{0}\bar{e} + \frac{1}{\sqrt[4]{N}}\nu\right]'\bar{Q}\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}'\left[\binom{\iota_p}{0}\bar{e} + \frac{1}{\sqrt[4]{N}}\nu\right] \\ \underset{p}{\to} \bar{e}^2\binom{\iota_p}{0}'\bar{Q}\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}'\binom{\iota_p}{0} \end{array}$$

and

$$\begin{array}{c} N^{\frac{3}{4}}\hat{D}_N(e)'\hat{V}_{ff}(e)^{-1}f_N(e) = \left[\sqrt[4]{N}(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{D}_N(e)\right]' \\ ((h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp))^{-1}\sqrt{N}\left[(h_N(\theta_{0,N})A_f(e) \vdots A_f(e)_\perp)'f_N(e)\right] \\ \approx \left[\left(\bar{Q}B(N)'V_{abd}B(N)\bar{Q}\right)^{-\frac{1}{2}}\bar{Q}'\left[\binom{\iota_p}{0}\bar{e} + \frac{1}{\sqrt[4]{N}}\nu\right]\right]' \\ \left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-\frac{1}{2}}\bar{Q}'\left(e^2\sigma^2\binom{\iota_p}{0} + B(N)'\begin{pmatrix}\varepsilon_a \\ \varepsilon_b \\ \varepsilon_d\end{pmatrix}\right) \\ \underset{d}{\to} \binom{\iota_p}{0}'\bar{Q}\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-1}\bar{Q}'\left(e^2\sigma^2\binom{\iota_p}{0} + B(N)'\begin{pmatrix}\varepsilon_a \\ \varepsilon_b \\ \varepsilon_d\end{pmatrix}\right) \end{array}$$

Upon combining everything, we obtain the limit behavior of the KLM statistic to test $\mathrm{H}_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ under (20):

$$\mathrm{KLM}(e) \underset{d}{\to} \eta' P_{\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-\frac{1}{2}}\bar{Q}\binom{\iota_p}{0}\bar{e}}\eta$$

with

$$\eta \sim N(\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-\frac{1}{2}}\bar{Q}\binom{\iota_p}{0}, I_{p_{\max}})$$

so since $\bar{e}$ is a scalar it cancels out and

$$\mathrm{KLM}(e) \underset{d}{\to} \quad \chi^2(\delta_{KLM}, 1)$$

because

$$\begin{aligned} \delta_{KLM} = \quad & (e\sigma)^4\binom{\iota_p}{0}'\bar{Q}\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-\frac{1}{2}}P_{\left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-\frac{1}{2}}\bar{Q}\binom{\iota_p}{0}} \\ & \left(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q}\right)^{-\frac{1}{2}}\bar{Q}\binom{\iota_p}{0} \\ = \quad & (e\sigma)^4\binom{\iota_p}{0}'\bar{Q}(\bar{Q}'B(N)'V_{abd}B(N)\bar{Q})^{-1}\bar{Q}\binom{\iota_p}{0} \\ = \quad & (e\sigma)^4\binom{\iota_p}{0}'(B(N)'V_{abd}B(N))^{-1}\binom{\iota_p}{0}, \end{aligned}$$

where the last equality has been shown for the GMM-AR statistic.

Appendix B. Definitions

In GMM, we consider a k-dimensional vector of moment conditions, see Hansen (1982):

$$E[f_i(\theta_0)] = 0, \qquad i = 1, \ldots, N, \tag{43}$$

which are a function of observed data and the unknown parameter vector θ. The moment conditions are only satisfied at the true value of the p-dimensional vector θ, θ_0, and k is at least as large as p. We analyze the first-order autoregressive panel data model so $p = 1$. The population moments in (43) are estimated using the average sample moments,

$$f_N(\theta) = \tfrac{1}{N} \textstyle\sum_{i=1}^{N} f_i(\theta). \tag{44}$$

The $k \times p$ dimensional matrix $q_N(\theta)$ contains the derivative of $f_N(\theta)$ with respect to θ :

$$q_N(\theta) = \tfrac{\partial}{\partial \theta'} f_N(\theta) = \tfrac{1}{N} \textstyle\sum_{i=1}^{N} q_i(\theta), \tag{45}$$

with $q_i(\theta) = \frac{\partial}{\partial \theta'} q_i(\theta)$.

The two step estimator results by minimizing the objective function:

$$Q(\theta, \theta^1) = N f_N(\theta)' \hat{V}_{ff}(\hat{\theta}^1)^{-1} f_N(\theta), \tag{46}$$

with $\hat{V}_{ff}(\theta)$ the Eicker-White covariance matrix estimator:

$$\hat{V}_{ff}(\theta) = \tfrac{1}{N} \textstyle\sum_{i=1}^{N} (f_i(\theta) - f_N(\theta))(f_i(\theta) - f_N(\theta))'. \tag{47}$$

The two step estimator, $\hat{\theta}_{2step}$, uses the one step estimator $\hat{\theta}_{1step}$ which equals the minimizer of (46) when we replace $\hat{V}_{ff}(\theta)^{-1}$ by the identity matrix.

The expressions of the different statistics to test $\mathrm{H}_0 : \theta = \theta_0$ that we use read:

1. Two step Wald statistic:

$$W_{2step}(\theta_0) = N(\hat{\theta}_{2step} - \theta_0)' q_N(\hat{\theta}_{2step})' \hat{V}_{ff}(\hat{\theta}_{2step})^{-1} q_N(\hat{\theta}_{2step})(\hat{\theta}_{2step} - \theta_0). \tag{48}$$

2. The GMM-LM statistic of Newey and West (1987):

$$\begin{aligned} LM(\theta_0) = \; & N f_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} q_N(\theta_0) \left[q_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} q_N(\theta_0) \right]^{-1} \\ & q_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} f_N(\theta_0). \end{aligned} \tag{49}$$

3. The KLM statistic of Kleibergen (2005):

$$\begin{aligned} KLM(\theta_0) = \; & N f_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} \hat{D}_N(\theta_0) \left[\hat{D}_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} \hat{D}_N(\theta_0) \right]^{-1} \\ & \hat{D}_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} f_N(\theta_0), \end{aligned} \tag{50}$$

with $\hat{D}_N(\theta)$ a $k \times p$ dimensional matrix,

$$\text{vec}(\hat{D}_N(\theta)) = \text{vec}(q_N(\theta)) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1} f_N(\theta) \tag{51}$$

and

$$\hat{V}_{qf}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \text{vec}[q_i(\theta) - q_N(\theta)](f_i(\theta) - f_N(\theta))'. \tag{52}$$

4. The GMM extension of the Anderson-Rubin statistic, see Anderson and Rubin (1949) and Stock and Wright (2000):

$$GMM\text{-}AR(\theta) = N f_N(\theta)'\hat{V}_{ff}(\theta)^{-1} f_N(\theta) = Q(\theta, \theta). \tag{53}$$

We use these four statistics for five different sets of moment conditions (labeled Dif, Lev, NL, AS and Sys, see Section 2). For the Dif moment conditions in (4), k_{Dif} equals $\frac{1}{2}(T-2)(T-1)$ and the specifications of $f_i^{Dif}(\theta)$ and $q_i^{Dif}(\theta)$ read

$$\begin{array}{rl} f_i^{Dif}(\theta) = & Z_i^{Dif}\varphi_i^{Dif}(\theta) \\ q_i^{Dif}(\theta) = & -Z_i^{Dif}\Delta y_{-1,i}, \end{array} \tag{54}$$

with $\varphi_i^{Dif}(\theta) = (\Delta y_{i3} - \theta\Delta y_{i2} \ldots \Delta y_{iT} - \theta\Delta y_{iT-1})'$, $\Delta y_{-1,i} = (\Delta y_{i2} \ldots \Delta y_{iT-1})'$ and

$$Z_i^{Dif} = \begin{pmatrix} y_{i1} & 0\ldots0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\ldots0 & \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT-2} \end{pmatrix} \end{pmatrix} : \tfrac{1}{2}(T-1)(T-2) \times (T-2). \tag{55}$$

For the Lev moment conditions in (5), k_{Lev} equals $T-2$ while the moment functions can be specified as

$$\begin{array}{rl} f_i^{Lev}(\theta) = & Z_i^{Lev}\varphi_i^{Lev}(\theta) \\ q_i^{Lev}(\theta) = & Z_i^{Lev} y_{-1,i}, \end{array} \tag{56}$$

with $\varphi_i^{Lev}(\theta) = (y_{i3} - \theta y_{i2} \ldots y_{iT} - \theta y_{iT-1})'$, $y_{-1,i} = (y_{i2} \ldots y_{iT-1})'$, and

$$Z_i^{Lev} = \begin{pmatrix} \Delta y_{i2} & 0\ldots0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\ldots0 & \Delta y_{iT-1} \end{pmatrix} : (T-2) \times (T-2). \tag{57}$$

For the NL moment conditions in (8), k_{NL} equals $T-3$ while the moment functions can be specified as

$$\begin{array}{rl} f_i^{NL}(\theta) = & Z_i^{NL}\varphi_i^{NL}(\theta) \\ q_i^{NL}(\theta) = & Z_i^{NL}\frac{\partial}{\partial\theta}\varphi_i^{NL}(\theta), \end{array} \tag{58}$$

with $\varphi_i^{NL}(\theta) = (u_{i4}(u_{i3} - u_{i2}) \dots u_{iT}(u_{iT-1} - u_{iT-2})'$ and

$$Z_i^{NL} = \begin{pmatrix} 1 & 0\dots0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\dots0 & 1 \end{pmatrix} : (T-3) \times (T-3). \tag{59}$$

The specification of the moment functions for the AS moment conditions results by stacking the moment conditions in (54) and (58) so k_{AS} equals $\frac{1}{2}(T-1)(T-2)+T-3$. The specification of the Sys moment conditions results by stacking the moment conditions in (54) and (56) so k_{Sys} equals $\frac{1}{2}(T+1)(T-2)$.

References

Ahn, S.C. and G.M. Thomas (2006). Likelihood-based inference for dynamic panel data models. Unpublished manuscript, Arizona State University.

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

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 H. Rubin (1949). Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics* **21**, 570–582.

Andrews, D.W.K., M.J. Moreira and J.H. Stock (2006). Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica* **74**, 715–752.

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

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

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.

Bond, S., C. Nauges and F. Windmeijer (2005). Unit Roots: Identification and Testing in Micro Panels. *CEMMAP Working Paper CWP07/05*.

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.

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

Engle, R.F. and C.W.J. Granger (1987). Co-integration and error correction : Representation, estimation and testing. *Econometrica* **55**, 251–276.

Hahn, J., J. Hausman and G. Kuersteiner (2007). Long DIfference Instrumental Variable Estimation for Dynamic Panel Models with fixed efects. *Journal of Econometrics* **140**, 574–617.

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.

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

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

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551–1580.

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

Kruiniger, H. (2009). GMM Estimation and Inference in Dynamic Panel Data Models with Persistent Data. *Econometric Theory* **25**, 1348–1391.

Kruiniger, H. (2013). Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. *Journal of Econometrics* **173**, 175–188.

Lehmann, E.L. and J.P. Romano (2005). *Testing Statistical Hypotheses* (Third Edition ed.). Springer.

Madsen, E. (2003). GMM Estimators and Unit Root Tests in the AR(1) Panel Data Model. *Working Paper, Centre for Applied Micro Econometrics, University of Copenhagen*.

Newey, W.K. and K.D. West (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review* **28**, 777–787.

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

Stock, J.H. and J.H. Wright (2000). GMM with Weak Identification. *Econometrica* **68**, 1055–1096.