

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

Maurice Bun\*      Frank Kleibergen†

June 13, 2016

## Abstract

We show that the System (Sys), see Arellano and Bover (1995) and Blundell and Bond (1998), as well as the Ahn and Schmidt (1995) moment conditions (AS) 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 furthermore construct a new set of (robust) moment conditions that identify the autoregressive parameter irrespective of the variance of the initial observations. These robust moment conditions are (non-linear) combinations of the original Sys and AS moment conditions. Despite that they identify the parameters when the true value of the autoregressive parameter is one and under the worst case setting of nuisance parameters, the GMM estimator that results from them has in such cases a non-standard distribution and a quartic root convergence rate. Instead of using the novel robust moment conditions for estimation, we therefore use them to determine the maximal attainable power of GMM tests in worst case settings. We show it to be identical for the AS and Sys moment conditions so assuming mean stationarity does not improve the power of tests in worst case settings. We compare the maximal attainable power under the worst case setting with the lower envelopes of power curves of different identification robust GMM test procedures. These power envelopes show their lowest rejection frequencies. The power envelope of the Lagrange Multiplier statistic of Kleibergen (2005) coincides with the maximal attainable power curve under the worst case setting so this statistic is optimal. Our results extend to other values of the autoregressive parameter for which identification fails when the variance of the individual specific effects becomes large.

JEL codes: C12, C23, C26

---

\*Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: M.J.G.Bun@uva.nl. The research of the author has been funded by the NWO Vernieuwingsimpuls research grant ‘Causal Inference with Panel Data’.

†Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: f.r.kleibergen@uva.nl.

# 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. 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.<sup>1</sup> 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 nuisance parameters like the initial observations, individual specific effects and variances of the disturbances. None of these moment conditions identifies the autoregressive parameter for all specifications of them. 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 in Madsen (2003), Bond *et. al.* (2005), Hahn *et. al.* (2007) and Kruiniger (2009).

---

<sup>1</sup>Note that in a combination of all three sets of moments conditions, the NL moment conditions are redundant.

Our first main contribution is to show that, when there are more than three time periods, the Sys and AS moments identify the parameters irrespective of the setting of the nuisance parameters. It is remarkable that the AS moments always lead to identification since it implies that the assumption of mean stationarity is redundant. We furthermore 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 are spanned by either the Sys or AS moments and also identify the parameters irrespective of the setting of the nuisance parameters and including the case of highly persistent data.

Despite these positive identification results for the Sys and AS moments, the large sample distributions of corresponding one step and two step generalized method of moments (GMM) estimators are non-standard under worst case settings of the nuisance parameters.<sup>2</sup> 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), Krueger (2009) and Bun and Windmeijer (2010). We show that GMM estimators based on the novel robust moments also have non-standard asymptotic distributions under worst case settings of the nuisance parameters. Furthermore, the robust moments are suboptimal under less severe settings of the nuisance parameters, *e.g.* away from the unit root, in which case original Sys and AS moments are preferred.

If we are not advised to use the GMM estimator that results from the robust moments for inference, one could question the practical relevance of these robust moments. To address it and show relevance for applied purposes, we focus on GMM methods more robust to weak instruments than conventional GMM procedures. To fully exploit the positive identification results for the AS and Sys moment conditions, we therefore use weak identification robust GMM tests. Unlike GMM estimators and corresponding Wald t-statistics, weak identification robust GMM statistics have limiting distributions which remain unaltered near the unit root and under worst case settings of the nuisance parameters. The GMM statistics that we 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). These statistics are size correct, easy to implement and have been used in a variety of models analyzed using GMM.

We use the novel robust moments to determine the maximal attainable power of size correct GMM tests using the original Sys or AS moment conditions under worst case specifications of the nuisance parameters. In worst case settings only the novel robust moments contain information on the autoregressive parameter. Their identifying power is, however, affected by these worst case settings as reflected by a non-standard quartic root convergence

---

<sup>2</sup>We will see that worst case settings of the nuisance parameters occur in case of: (1) mean stationarity; (2) time series homoscedasticity.

rate.

Our second main result is that the maximal attainable power curves of the Sys and AS 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. This result invalidates the long standing view that mean stationarity and the resulting Sys moment conditions help to identify the autoregressive parameter when its true value is around the unit root.

We next use the maximal attainable power to determine the GMM testing procedure with optimal power properties under worst case settings of the nuisance parameters. 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 the GMM-AR, GMM-LM or KLM test statistics is optimal near the unit root and under the worst case setting of the nuisance parameters, we construct the lower envelope of their power curves under all settings of the nuisance parameters 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. We compare the maximal attainable power 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 regressor towards the panel autoregressive model of order one.

Our third main result is that the power envelope of the KLM statistic coincides with 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 linear 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.<sup>3</sup>

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 in our analysis we 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

---

<sup>3</sup>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.

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. This theorem also allows us to construct the novel robust moments. In Section 5 we discuss properties of GMM estimators. We show that non-standard limiting distributions may result near the unit root irrespective of whether AS, Sys or robust moments have been used in estimation. In Section 6, we use the novel robust moments conditions to construct the maximal attainable power of hypothesis tests under worst case settings. We also briefly discuss the extension to a large individual specific effect variance. In Section 7, we construct the power envelopes of the weak instrument robust 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:  $\text{vec}(A)$  stands for the (column) vectorization of the  $k \times n$  matrix  $A$ ,  $\text{vec}(A) = (a'_1 \dots a'_n)'$  for  $A = (a_1 \dots a_n)$ ,  $P_A = A(A'A)^{-1}A'$  is a projection on the columns of the full rank matrix  $A$  and  $M_A = I_N - P_A$  is a projection on the space orthogonal to  $A$ . Convergence in probability is denoted by “ $\xrightarrow{p}$ ”, convergence in distribution by “ $\xrightarrow{d}$ ” and “ $\xrightarrow{a}$ ” means asymptotically equivalent.

## 2 Moment conditions

We analyze the first-order linear dynamic panel data model

$$y_{it} = c_i + \theta y_{it-1} + u_{it} \quad i = 1, \dots, N, \quad t = 2, \dots, T, \quad (1)$$

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.<sup>4</sup> Estimation of the parameter  $\theta$  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:

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

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

$$E[\Delta y_{it}y_{ij}] \quad j = 1, \dots, T, \quad t = 2, \dots, T \quad (3)$$

---

<sup>4</sup>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.

can be used to construct functions which identify the parameter of interest  $\theta$ . We do not use products of  $\Delta y_{it}$  to identify  $\theta$  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  $\theta$  :

1. Difference (Dif) moment conditions:

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

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:

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

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

$$E[\Delta y_{it} c_i] = 0, \quad (6)$$

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

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

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  $\theta$ . When we use the moment conditions in (4) and (5) jointly, we refer to them as system (Sys) moment conditions,<sup>5</sup> 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):

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

The NL moments can be used separately or jointly with the Dif moments to identify  $\theta$ . When we use the moment conditions in (4) and (8) jointly, we refer to them as Ahn-Schmidt (AS) moment conditions.

---

<sup>5</sup>We could extend the Lev moment conditions to  $\frac{1}{2}(T-1)(T-2)$  sample moments by including additional interactions of  $\Delta y_{it-j}$  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).

Ahn and Schmidt (1995) show that their AS moment conditions exhaust the information on  $\theta$  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  $\theta$  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  $\theta$  when its true value is one but since these moment conditions are quadratic in  $\theta$ , 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.

Our novel robust moments are formally derived in Section 4, and we state them here for reference. They result as combinations of either the AS or Sys moment conditions and turn out to be quadratic in  $\theta$ :

$$E[a\theta^2 + b\theta + d] = 0, \tag{9}$$

where the explicit specifications of  $a$ ,  $b$  and  $d$  are listed in Section 4 below (32) for the AS and Sys moment conditions and different values of  $T$ . The expressions for  $a$ ,  $b$  and  $d$  are such that the robust moments in (9) are just functions of differences of the dependent variable  $y_{it}$ . In Section 4 we show that they result from the non-diverging part of the AS and Sys moment conditions in limit sequences where  $\theta$  goes to one jointly with the cross-sectional sample size and variance of the initial observations going off to infinity. Therefore, their limiting behavior does not depend on the initial observations and also does not need mean stationarity (6)-(7). The robust moments identify  $\theta$  for all values including the unit root case. However, they do not necessarily lead to the same convergence rate for all values of  $\theta$ .

### 3 Identification: An illustrative example

In this Section we introduce our asymptotic sampling scheme which consists of drifting sequences for the autoregressive parameter and the variance of the initial observation. Furthermore, we use it in an illustrative example with three time series observations to analyze the limiting behavior of Dif and Lev sample moments.<sup>6</sup>

The Dif and Lev moment conditions are semi-parametric with respect to the individual specific fixed effects, variances and initial observations so they identify  $\theta$  for a variety of different specifications of them. These specifications, however, influence the identification of

---

<sup>6</sup>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.

$\theta$  for persistent values of it, *i.e.* values that are close to one.<sup>7</sup> 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:

$$\begin{aligned} \text{Dif: } & E[y_{i1}(\Delta y_{i3} - \theta \Delta y_{i2})] = 0 \\ \text{Lev: } & E[\Delta y_{i2}(y_{i3} - \theta y_{i2})] = 0, \end{aligned} \tag{10}$$

with Jacobians:

$$\begin{aligned} \text{Dif: } & -E[y_{i1} \Delta y_{i2}] = -E((\mu_i + u_{i1})((\theta_0 - 1)u_{i1} + u_{i2})) \\ \text{Lev: } & -E[y_{i2} \Delta y_{i2}] = -E((c_i + \theta_0 y_{i1} + u_{i2})((\theta_0 - 1)u_{i1} + u_{i2})), \end{aligned} \tag{11}$$

where  $\theta_0$  is the true value of  $\theta$ . For many data generating processes for the initial observations, the Jacobian of the Dif moment condition in (11) is equal to zero when  $\theta_0$  is equal to one.<sup>8</sup> The Dif moment condition does then not identify  $\theta$  when  $\theta_0$  is equal to one for these DGPs. Under mean stationarity (6)-(7), the Jacobian of the Lev moment condition is such that

$$E(y_{i2} \Delta y_{i2}) = (\theta_0 - 1)\theta_0 E(u_{i1}^2) + E(u_{i2}^2) \neq 0, \quad \text{when } \theta_0 = 1, \tag{12}$$

so the Lev moment condition seems to identify  $\theta$  irrespective of the value of  $\theta_0$ , 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  $\theta_0$  is equal to one. 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 (10) and the Jacobian in (12). To ascertain the identification of  $\theta$  by the Lev moment conditions when  $\theta_0$  is equal to one, we therefore consider a joint limit process where both  $\theta_0$  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).

**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  $\theta_0$  goes to one is such that*

$$E(\lim_{\theta_0 \uparrow 1} ((1 - \theta_0)u_{i1})^2) = 0. \tag{13}$$

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  $u_{it}$  :

<sup>7</sup>In Section 6.1, we show that similar identification problems occur when  $\theta$  is smaller than one, but the variance of the individual specific effects is large.

<sup>8</sup>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 2.**

$$\text{var}(u_{it}y_{i1}) = \text{var}(u_{it})\text{var}(y_{i1}), \quad t = 2, \dots, T. \quad (14)$$

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  $\theta$  when the variance of the initial observations gets large compared to that of the subsequent disturbances. The assumption in (14) 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  $\theta_0$  converges to one (we rule out explosive values of  $\theta_0$ ) 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:

$$\begin{aligned} \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N \Delta y_{i2}(y_{i3} - \theta y_{i2}) &\approx (1 - \theta) \left\{ \frac{1}{N} \sum_{i=1}^N u_{i2}^2 + \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N u_{i2}y_{i1} + \right. \\ &\quad \left. \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N (1 - \theta_0)u_{i1}y_{i1} \right\} \\ \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N y_{i2} \Delta y_{i2} &\approx \frac{1}{N} \sum_{i=1}^N u_{i2}^2 + \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N u_{i2}y_{i1} + \\ &\quad \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N (1 - \theta_0)u_{i1}y_{i1}. \end{aligned} \quad (15)$$

We dropped all elements in (15) that do not affect the large sample behavior when  $\theta_0$  goes to one,<sup>9</sup> at least not under our drifting parameter sequences as defined below. What is left 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 (14), it holds that

$$h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2}y_{i1} \xrightarrow{d} \psi_2, \quad (16)$$

with  $\psi_2$  a normal random variable with mean zero and variance  $\sigma_2^2 = \text{var}(u_{i2})$  and  $h(\theta_0)^{-2} = \text{var}(y_{i1})$ .

We analyze a setting in which both  $\theta_0$  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

$$\begin{aligned} \lim_{N \rightarrow \infty} \theta_{0,N} &= 1 \\ \lim_{N \rightarrow \infty} h_N(\theta_{0,N}) &= d, \end{aligned} \quad (17)$$

with  $d$  a finite possibly zero constant. The sequences in (17) allow the variance of the initial observations to be large jointly with a large value for the autoregressive parameter. The limit sequences in (17) are such that (16) remains to hold so

$$h_N(\theta_{0,N}) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2}y_{i1} \xrightarrow{d} \psi_2, \quad (18)$$

and which explains why  $\frac{1}{N} \sum_{i=1}^N u_{i2}y_{i1}$  appears in (15).

---

<sup>9</sup>This explains why we use the “ $\approx$ ” sign instead of the “=” sign.

When  $d$  in (17) 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 (15). For example, when these sequences are such that

$$h_N(\theta_{0,N})\sqrt{N} \xrightarrow{N \rightarrow \infty} \infty, \quad (19)$$

it holds that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N y_{i2} \Delta y_{i2} &= \frac{1}{N} \sum_{i=1}^N u_{i2}^2 + \frac{1}{h_N(\theta_{0,N})\sqrt{N}} \left[ h_N(\theta_{0,N}) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2} y_{i1} \right] + \\ &\quad \frac{1}{N} \sum_{i=1}^N (1 - \theta_{0,N}) u_{i1} y_{i1} \\ &\xrightarrow{p} \sigma_2^2 + \lim_{N \rightarrow \infty} E((1 - \theta_{0,N}) u_{i1}^2), \end{aligned} \quad (20)$$

while when

$$h_N(\theta_{0,N})\sqrt{N} \xrightarrow{N \rightarrow \infty} 0, \quad (21)$$

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

$$\begin{aligned} h_N(\theta_{0,N}) \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta y_{i2} (y_{i3} - \theta y_{i2}) &= (1 - \theta) \left\{ h_N(\theta_{0,N})\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N u_{i2}^2 \right] + \right. \\ &\quad \left. h_N(\theta_{0,N})\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N (1 - \theta_{0,N}) u_{i1} y_{i1} \right] + \right. \\ &\quad \left. \left[ h_N(\theta_{0,N}) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2} y_{i1} \right] \right\} \\ &\xrightarrow{d} (1 - \theta) \psi_2 \\ h_N(\theta_{0,N}) \frac{1}{\sqrt{N}} \sum_{i=1}^N y_{i2} \Delta y_{i2} &= h_N(\theta_{0,N})\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N u_{i2}^2 \right] + \\ &\quad h_N(\theta_{0,N})\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N (1 - \theta_{0,N}) u_{i1} y_{i1} \right] + \\ &\quad h_N(\theta_{0,N}) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2} y_{i1} \\ &\xrightarrow{d} \psi_2. \end{aligned} \quad (22)$$

Hence, when (21) holds:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \Delta y_{i2} (y_{i3} - \theta y_{i2}) &= \frac{1}{h_N(\theta_{0,N})\sqrt{N}} \frac{h_N(\theta_{0,N})}{\sqrt{N}} \sum_{i=1}^N \Delta y_{i2} (y_{i3} - \theta y_{i2}) \\ &\xrightarrow{N \rightarrow \infty} \infty \\ \frac{1}{N} \sum_{i=1}^N y_{i2} \Delta y_{i2} &= \frac{1}{h_N(\theta_{0,N})\sqrt{N}} \frac{h_N(\theta_{0,N})}{\sqrt{N}} \sum_{i=1}^N y_{i2} \Delta y_{i2} \\ &\xrightarrow{N \rightarrow \infty} \infty \end{aligned} \quad (23)$$

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  $\theta$  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  $\theta$  by the Lev moment conditions is arbitrary for DGPs for which  $\theta_0$  is close to one and the variance of the initial observations is infinite when  $\theta_0$  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 (21), the limiting sequence  $\theta_{0,N}$  (17) is such that:

$$\begin{aligned} \text{DGP 1, 2, 5: } & (1 - \theta_{0,N})\sqrt{N} \xrightarrow{N \rightarrow \infty} 0 \quad \text{or} \quad \theta_{0,N} = 1 - \frac{e}{N^{\frac{1}{2}(1+\epsilon)}} \\ \text{DGP 3: } & (1 - \theta_{0,N}^2)N \xrightarrow{N \rightarrow \infty} 0 \quad \text{or} \quad \theta_{0,N} = 1 - \frac{e}{N^{1+\epsilon}} \\ \text{DGP 4: } & \frac{N}{g} \xrightarrow{N \rightarrow \infty, g \rightarrow \infty} 0, \end{aligned} \tag{24}$$

with  $e$  a constant and  $\epsilon$  some real number larger than zero. In case of DGP 4, (24) 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  $\theta$  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  $\theta$  when  $\theta_0$  is close to one and  $T = 3$ . To analyze the identification of  $\theta$  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 in (17) and (21). 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  $u_{it}$ ,  $i = 1, \dots, N$ ,  $t = 2, \dots, T$ ,  $T \geq 3$ , we can characterize the large sample behavior of the Dif, Lev, NL, AS and Sys sample moments and their derivatives for values  $\theta_0$  that drift to one according to (17) and (21) by:*

$$\begin{pmatrix} f_N^j(\theta) \\ q_N^j(\theta) \end{pmatrix} = \begin{pmatrix} A_f^j(\theta) \\ A_q^j(\theta) \end{pmatrix} \left[ \frac{1}{h_N(\theta_{0,N})\sqrt{N}}\psi + \iota \left( \lim_{N \rightarrow \infty} E((\theta_{0,N} - 1)u_{i1}^2) \right) \right] + \begin{pmatrix} \mu_f^j(\theta, \sigma^2) \\ \mu_q^j(\theta, \sigma^2) \end{pmatrix} + \frac{1}{\sqrt{N}} \begin{pmatrix} B_f^j(\theta) \\ B_q^j(\theta) \end{pmatrix} \psi_{uu}, \quad (25)$$

with  $j = \text{Dif, Lev, NL, AS, Sys}$ . Furthermore,  $\mu_f^j(\theta, \sigma^2)$  and  $\mu_q^j(\theta, \sigma^2)$  are constants,  $\sigma^2 = (\sigma_2^2 \dots \sigma_T^2)$ ,  $\psi$  and  $\psi_{uu}$  are mean zero finite variance normal random variables,  $\iota$  is a vector of ones and the dimensions of  $\iota$  and  $\psi$  are  $T - 1$  for Dif, AS and Sys and  $T - 2$  for Lev and NL.

The specifications of  $A_f^j(\theta)$ ,  $A_q^j(\theta)$ ,  $B_f^j(\theta)$ ,  $B_q^j(\theta)$ ,  $\mu_f^j(\theta, \sigma^2)$ ,  $\mu_q^j(\theta, \sigma^2)$ ,  $\psi$  and  $\psi_{uu}$  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  $\begin{pmatrix} A_f^j(\theta) \\ A_q^j(\theta) \end{pmatrix}$  since the latter components get multiplied by  $\frac{1}{h(\theta_{0,N})\sqrt{N}}\psi$ , which under (21) goes to infinity when the sample size increases. The only identifying information for  $\theta$  therefore results from that part of the sample moment which does not depend on  $\psi$ . Since  $\psi$  only affects the part of the sample moments spanned by  $A_f^j(\theta)$ , the sample moments are independent of  $\psi$  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

$$A_f^j(\theta)'_{\perp} f_N^j(\theta) \approx A_f^j(\theta)'_{\perp} \mu_f^j(\theta, \sigma^2) + \frac{1}{\sqrt{N}} A_f^j(\theta)'_{\perp} B_f^j(\theta) \psi_{cu}, \quad (26)$$

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 (25) in Theorem 1, the elements multiplied by  $A_f^j(\theta)$  have dropped out since  $A_f^j(\theta)'_{\perp} A_f^j(\theta) = 0$ . The right hand side of (26) now contains all remaining identifying elements of the original moment conditions. From expression (26), it is seen that identification results only when  $A_f^j(\theta)'_{\perp}$  is a well defined matrix and, furthermore,  $A_f^j(\theta)'_{\perp} \mu_f^j(\theta, \sigma^2)$  is non-zero.

We next briefly discuss what this implies for the different sets of moment conditions discussed previously.

**Dif and Lev conditions** When  $T = 3$  or  $4$ , the specifications of  $\begin{pmatrix} \mu_f^j(\theta, \sigma^2) \\ \mu_q^j(\theta, \sigma^2) \end{pmatrix}$  and  $\begin{pmatrix} A_f^j(\theta) \\ A_q^j(\theta) \end{pmatrix}$  for the Dif and Lev moment conditions, which are stated in the proof of Theorem 1 in Appendix A,<sup>10</sup> are:

$$\begin{aligned}
\mathbf{Dif:} \quad \mathbf{T} = \mathbf{3} \quad & \begin{pmatrix} \mu_f^{Dif}(\theta, \sigma^2) \\ \mu_q^{Dif}(\theta, \sigma^2) \end{pmatrix} = (0 \ 0)', \quad A_f^{Dif}(\theta) = (-\theta \ 1), \quad A_q^{Dif}(\theta) = (-1 \ 0). \\
\mathbf{T} = \mathbf{4} \quad & \begin{pmatrix} \mu_f^{Dif}(\theta, \sigma^2) \\ \mu_q^{Dif}(\theta, \sigma^2) \end{pmatrix} = (0 \dots 0)', \quad A_f^{Dif}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \end{pmatrix}, \quad A_q^{Dif}(\theta) = - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \\
\mathbf{Lev:} \quad \mathbf{T} = \mathbf{3} \quad & \begin{pmatrix} \mu_f^{Lev}(\theta, \sigma^2) \\ \mu_q^{Lev}(\theta, \sigma^2) \end{pmatrix} = \left( \sigma_2^2 \begin{pmatrix} 1 - \theta \\ 1 \end{pmatrix} \right), \quad A_f^{Lev}(\theta) = 1 - \theta, \quad A_q^{Lev}(\theta) = -1. \\
\mathbf{T} = \mathbf{4} \quad & \begin{pmatrix} \mu_f^{Lev}(\theta, \sigma^2) \\ \mu_q^{Lev}(\theta, \sigma^2) \end{pmatrix} = \left( \begin{pmatrix} 1 - \theta \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \end{pmatrix} \right), \quad A_f^{Lev}(\theta) = \begin{pmatrix} 1 - \theta & 0 \\ 0 & 1 - \theta \end{pmatrix}, \\
& A_q^{Lev}(\theta) = -I_2.
\end{aligned} \tag{27}$$

The expressions of  $A_f^{Dif}(\theta)$  and  $A_f^{Lev}(\theta)$  in (27) are all square or rectangular matrices. When  $T$  exceeds four, the expressions of  $A_f^{Lev}(\theta)$  remain square matrices<sup>11</sup> so the Lev moment conditions do not identify  $\theta$  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  $\theta$  for any value of  $T$ . Summarizing we have:

$$\begin{aligned}
\mathbf{Dif, T} = \mathbf{4:} \quad & A_f^{Dif}(\theta)_\perp \text{ does not exist.} \quad \text{No identification when } T = 4. \\
\mathbf{Dif, T} = \mathbf{3, T} > \mathbf{4:} \quad & A_f^{Dif}(\theta)'_\perp \mu_f^{Dif}(\theta, \sigma^2) = 0. \quad \text{No identification when } T = 3, T > 4. \\
\mathbf{Lev:} \quad & A_f^{Lev}(\theta)_\perp \text{ does not exist.} \quad \text{No identification when } T \geq 3.
\end{aligned} \tag{28}$$

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

$$\mathbf{NL:} \quad \begin{pmatrix} \mu_f^{NL}(\theta, \sigma^2) \\ \mu_q^{NL}(\theta, \sigma^2) \end{pmatrix} = \begin{pmatrix} (1-\theta)(\sigma_3^2 - \theta\sigma_2^2) \\ (2\theta-1)\sigma_2^2 - \sigma_3^2 \end{pmatrix}, \quad \begin{pmatrix} A_f^{NL}(\theta) \\ A_q^{NL}(\theta) \end{pmatrix} = \begin{pmatrix} \theta(\theta-1) & 1-\theta \\ 2\theta-1 & -1 \end{pmatrix}. \tag{29}$$

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  $\theta$ .

<sup>10</sup>The proofs in the Appendix A do not cover  $T = 3$  since it straightforwardly results from  $T = 4$ .

<sup>11</sup>We refer to the proof of Theorem 1 for the expressions of  $A_f^{Dif}(\theta)$  and  $A_f^{Lev}(\theta)$  when  $T = 5$ .

The expression of  $A_f^{NL}(\theta)$  for a larger number of time series observations<sup>12</sup> 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  $\theta$ .

**AS and Sys conditions** The expressions of  $\begin{pmatrix} \mu_f^j(\theta, \sigma^2) \\ \mu_q^j(\theta, \sigma^2) \end{pmatrix}$  and  $\begin{pmatrix} A_f^j(\theta) \\ A_q^j(\theta) \end{pmatrix}$  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:

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

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

$$\begin{aligned} \text{Sys: } \mathbf{T} = 4 \quad \mu_f^{Sys}(\theta, \sigma^2) &= (1-\theta) \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ \sigma_3^2 \end{pmatrix}, \quad A_f^{Sys}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \\ 1-\theta & 0 & 0 \\ 0 & 1-\theta & 0 \end{pmatrix}, \\ \mu_q^{Sys}(\theta, \sigma^2) &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ \sigma_3^2 \end{pmatrix}, \quad A_q^{Sys}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}. \end{aligned} \tag{30}$$

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  $\theta$  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 (30) 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

<sup>12</sup>The expression for  $T = 5$  is stated in the proof of Theorem 1 in Appendix A.

in the direction of  $A_f^j(\theta)$ , so that part cannot be used to identify  $\theta$ , the AS and Sys sample moments identify  $\theta$  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.

The preceding analysis leads to the following:

**Corollary 1 (Identification of  $\theta$ ).** *Under the assumptions of Theorem 1,  $\theta$  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,  $\theta$  is identified by the Sys moment conditions when there are more than three time periods. It also shows that  $\theta$  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 not need mean stationarity to hold. This shows that assuming mean stationarity for constructing additional moment conditions does not help to identify  $\theta$  when the convergence rate accords to (21), since the same identification results are obtained from moment conditions that do not assume mean stationarity.<sup>13</sup>

When the convergence rate accords with (19) the term involving  $\psi$  vanishes from the large sample behavior of the moment conditions in Theorem 1, and all of the original AS or Sys moments identify  $\theta$ . When the convergence rate accords with (21), however, Theorem 1 shows that only the part of the (AS or Sys) moment conditions in the direction of  $A_f^j(\theta)_\perp$  (26) now identifies  $\theta$ . This part of the original moment conditions is robust to the variance of the initial observations. 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. They can be specified as

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

where  $T$  indicates the number of time periods and  $G_{2,T}^j$  is such that  $G_{2,T}^{j'} \mu_f^j(\theta, \sigma^2) = 0$ . Furthermore,  $G_{f,T}^j(\theta)$  is the only part of  $A_f^j(\theta)_\perp$  that depends on  $\theta$ . The orthogonal complements are then such that the resulting, what we refer to as, robust moment conditions are quadratic in  $\theta$ :

$$g_{f,T}(\theta) = A_f(\theta)_\perp^{j'} f_N(\theta) = a\theta^2 + b\theta + d, \quad (32)$$

with for

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

---

<sup>13</sup>It can be shown that when mean stationarity does not hold both the Dif and NL, and consequently the AS, moment conditions identify  $\theta$  even when the convergence rate accords to (21).

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

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

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

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

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

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

and similar specifications of  $a$ ,  $b$  and  $d$  result for larger values of  $T$ . It is interesting to see that these robust moments only depend on differences of the data so the initial observations get differenced out. This explains why these moments are robust to the variance of the initial observations.

## 5 GMM estimators

We showed that the original AS and Sys moment conditions identify  $\theta$  when  $T$  exceeds three. Furthermore, we showed that the robust moments (32) contain the identifying elements of the original moment conditions when the panel data are highly persistent. This does, however, not imply that AS and Sys GMM estimators based on either original or robust moment conditions behave in the manner that we are used to when estimating identified parameters. We distinguish two different cases, i.e. the convergence rate accords with (19) or (21).



## 5.1 Highly persistent panel data

When the convergence rate accords with (21), only the part of the moment conditions in the direction of  $A_f^j(\theta)_\perp$ , see (26), identifies  $\theta$ . One and two step AS and Sys 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  $\theta$ . Usage of the first part results in an inconsistency so one and two step GMM estimators 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  $\theta$ .

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

$$\hat{\theta}_{1s}^{Sys} \xrightarrow{d} 1 + (\psi' A_q^{Sys}(1)' A_q^{Sys}(1) \psi)^{-1} \psi' A_q^{Sys}(1)' A_f^{Sys}(1) \psi, \quad (33)$$

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  $\theta$ . It furthermore 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  $\theta$  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. However, it can be shown that, under the conditions of Theorem 1, the limiting behavior of one- and two-step AS and Sys GMM estimators is similar to the non-standard results in *e.g.* Madsen (2003) or Kruiniger (2009).

Corollary 2 shows that the identification of  $\theta$  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. Conducting inference based on these estimators exploiting the usual Wald  $t$ -statistic is therefore hard when  $\theta_0$  is close to one. We show that similar results apply for the robust moments in (32). Note that the robust moments are defined purely in differences of the data, hence their validity does not depend on the mean stationarity assumption (6)-(7).<sup>14</sup> We therefore analyze the limiting behavior of the robust moments at the unit root also under deviations from mean stationarity. To obtain it, we first state the probability limits of the quantities  $a$ ,  $b$  and  $d$  in (32) when the true value of  $\theta$  is one.

---

<sup>14</sup>This suggests that the mean stationarity assumption is actually not helpful for identification of  $\theta$ . In our power analysis later on we indeed show that maximal attainable power for AS and Sys moment conditions coincides precisely in the case of mean stationarity.

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

$$\mathbf{T=4, Sys:} \quad a \xrightarrow{p} \begin{pmatrix} \omega + \sigma_2^2 \\ 0 \end{pmatrix}, b \xrightarrow{p} - \begin{pmatrix} 4\omega + \sigma_2^2 + \sigma_3^2 \\ \omega \end{pmatrix}, d \xrightarrow{p} \begin{pmatrix} 3\omega + \sigma_3^2 \\ \omega \end{pmatrix}.$$

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

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

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

**Proof.** see Appendix A. ■

The parameter  $\omega$  in Theorem 2 reflects the deviation from mean stationarity. Theorem 2 can be used to show that the Jacobian of the robust moment equation (32) is of full column rank whenever  $\omega \neq 0$  or  $\sigma_t^2 \neq \sigma^2$  for at least one value of  $t = 2, \dots, T$ . The standard asymptotic theory for GMM estimators based on the robust moments therefore applies for these cases.

Under mean stationarity and time series homoscedasticity results are, however, markedly different. Previously we referred to it as the worst case setting of the nuisance parameters. Using Theorem 2 we see that essentially only the first element of the components  $a$ ,  $b$  and  $d$  remains when  $\omega = 0$  and  $\sigma_t^2 = \sigma^2$ , since for values of  $T$  larger than four all non-zero elements of  $a$ ,  $b$  and  $d$  are identical. Therefore, for any number of time periods GMM estimation based on the robust moments (32) is then asymptotically equivalent to method of moments estimation based on the scalar sample moment condition:

$$g_N(\theta) = a\theta^2 + b\theta + d, \tag{34}$$

with

$$\begin{aligned} \text{Sys } a &= \frac{1}{N} \sum_{i=1}^N (\Delta y_{i2})^2, & b &= -\frac{1}{N} \sum_{i=1}^N (y_{i3} - y_{i1})^2, & d &= \frac{1}{N} \sum_{i=1}^N (y_{i4} - y_{i1}) \Delta y_{i3}, \\ \text{AS } a &= \frac{1}{N} \sum_{i=1}^N (y_{i3} - y_{i1}) \Delta y_{i2}, & b &= -\frac{1}{N} \sum_{i=1}^N (y_{i4} - y_{i1}) \Delta y_{i2} + (y_{i3} - y_{i1}) \Delta y_{i3}, \\ d &= \frac{1}{N} \sum_{i=1}^N (y_{i4} - y_{i1}) \Delta y_{i3}. \end{aligned}$$

We note that  $g_N(\theta)$  is a quadratic function with positive second order derivative  $2a$ . The equation  $g_N(\theta) = 0$  therefore has zero, one or two solutions, depending on the sign of the discriminant  $b^2 - 4ad$ . Lemma 1 states the asymptotic distribution of this discriminant when  $\theta_0 = 1$ ,  $\omega = 0$  and  $\sigma_t^2 = \sigma^2$ .

**Lemma 1.** *Under the conditions of Theorem 2 and when  $\omega = 0$  and  $\sigma_t^2 = \sigma^2$ ,  $t = 2, \dots, T$ , the limit behavior of the discriminant of the moment condition (34) when the true value of  $\theta$  is one is for both Sys and AS:*

1.  $\text{plim}_{N \rightarrow \infty} (b^2 - 4ad) = 0$ ,
2.  $\sqrt{N}(b^2 - 4ad) \xrightarrow{d} N(0, 32\sigma^4)$ .

**Proof.** see Appendix A. ■

The results of Lemma 1 imply that the limit probability of obtaining a positive discriminant, which is the probability that  $g_N(\theta) = 0$  has a real valued solution, is 50%. For these 50% of the realizations, we have:

$$\hat{\theta} = -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ad}}{2a}. \quad (35)$$

Since both the discriminant  $b^2 - 4ad$  and  $-b/(2a) - 1 = (-b - 2a)/(2a)$  are  $O_P(N^{-1/2})$ , it follows that  $\sqrt{b^2 - 4ad}/(2a) = O_P(N^{-1/4})$ , hence this is the dominating element of the limiting distribution of  $\hat{\theta} - 1$ . It implies that in the limit, in order to obtain an estimate of  $\theta$  which is less than one, we only choose the minus component of the solution in (35). Theorem 3 states the resulting limiting behavior of the GMM estimator.

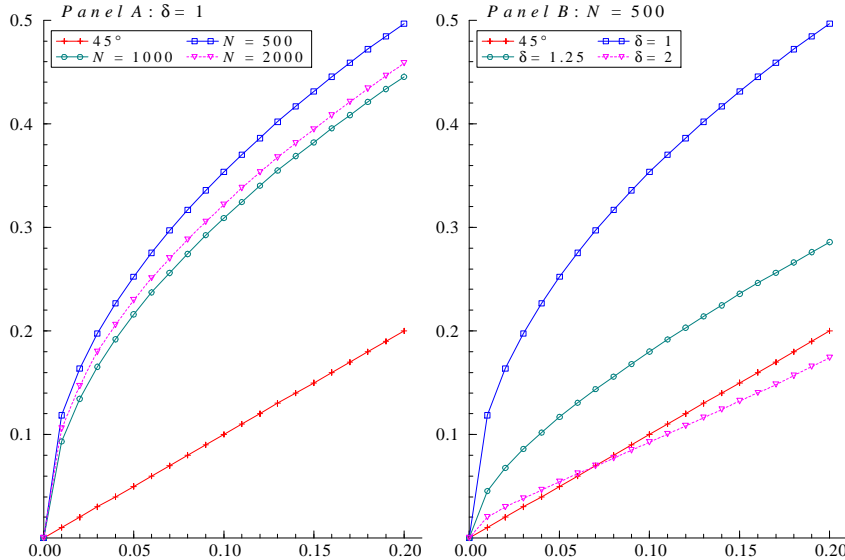
**Theorem 3.** *Under the conditions of Lemma 1 and when  $b^2 - 4ad > 0$ , the limit behavior of the Sys and AS GMM estimators based on the moment condition (34) when the true value of  $\theta$  is one is:*

1.  $\text{plim}_{N \rightarrow \infty} \hat{\theta} = 1$ ,
2.  $N^{1/4}(\hat{\theta} - 1) \xrightarrow{d} -\sqrt{|N(0, 2\sigma^2)|}$ .

**Proof.** see Appendix A. ■

Theorem 3 shows that, under mean stationarity and time series homoscedasticity, GMM estimators based on the AS and Sys robust moment conditions (34) are consistent, even when the true value of  $\theta$  is one. Interestingly they have identical asymptotic distributions, which again indicates that the assumption of mean stationarity is redundant in the worst case setting of the nuisance parameters. The asymptotic distribution is the square root of (the absolute value of) a normal random variable which is highly non-standard and convergence to it occurs at a low quartic root rate.<sup>15</sup> This slow rate results from the square root rate of the discriminant in (35), which itself has the standard  $\sqrt{N}$  convergence rate as shown in Lemma 1. Because of the quartic convergence rate and the non-normal asymptotic distribution, conventional Wald type inference is invalid. Moreover, it is unclear how to define the estimators in case of a negative value of the discriminant.<sup>16</sup>

Figure 1. P-value plots, Sys Wald t-test



Note: based on 10,000 replications from DGP 2 with  $T=4$ ,  $\sigma_c^2=1$  and  $\sigma^2=1$  in all experiments.

Panel A of Figure 1 illustrates the non-standard inference at the unit root in case of mean stationarity and time series homoscedasticity. It shows p-value plots of the Wald t-statistic testing  $H_0 : \theta = \theta_0$  using the Sys robust moment (34). The data have been generated according to DGP 2 with  $T = 4$ ,  $\theta_0 = 0.99$ ,  $\sigma_c^2 = 1$ ,  $\sigma^2 = 1$ . We show results for  $N = (500, 1000, 2000)$ . The size properties of the Wald t-statistic in Panel A are very poor and show

<sup>15</sup>Remarkably, a quartic root convergence rate and limiting distributions similar to those in Theorem 3 are also discussed in Ahn and Thomas (2006) and Kruiniger (2013, Theorem 4) for random effects maximum likelihood estimators.

<sup>16</sup>In case of a negative discriminant one can take for the estimator e.g. its minimum  $-b/2a$  or just one as in Ahn and Thomas (2006, Proposition 6) and Kruiniger (2013, Theorem 7).

severe over rejection at all significance levels. Due to the non-standard asymptotic behavior of the GMM estimator, see Theorem 3, the performance of the Wald test does not improve when  $N$  increases, as expected.

Panel B of Figure 1 shows the effect of mean-nonstationarity. We alter the initial condition (7) to  $y_{i1} = \delta_\mu \mu_i + u_{i1}$  to allow for deviations from mean stationarity. This parametrization is often used (see *e.g.* Hayakawa, 2009), and implies that  $\omega = (1 - \delta_\mu)^2 \sigma_c^2$  so mean stationarity corresponds to  $\delta_\mu = 1$ . We choose  $\delta_\mu = (1, 1.25, 2)$  while keeping  $N = 500$  and the same values for all other parameters. The pattern of size properties is now different. As  $\delta_\mu$  moves away from one, conventional asymptotic approximations become gradually more accurate and Wald type inference becomes valid.

## 5.2 Away from the unit root

When the convergence rate accords with (19), the element involving  $\psi$  vanishes from the large sample behavior of the sample moment conditions in Theorem 1. This case includes true values of  $\theta$  away from the unit root. Hence, standard first-order asymptotic theory applies leading to the usual normal limit distributions of one and two step GMM estimators. Usage of the original AS or Sys moment conditions instead of their robust counterpart (32) is then preferable since the AS and Sys moment conditions achieve the semi-parametric efficiency bound under standard first-order asymptotics which apply when the data are not persistent, see Ahn and Schmidt (1995) and Blundell and Bond (1998). Away from the unit root these moment conditions contain more information on the parameter of interest than their robust counterparts so AS and Sys GMM estimators have lower asymptotic variance than GMM estimators based on their robust moments.

To show the improved efficiency away from the unit root when using the original AS and Sys moments compared to their robust moments, Table 1 reports the means and standard deviations of two step AS and Sys estimators and GMM estimators based on the robust moments that result from a simulation experiment. The robust GMM estimator is computed along the lines explained below (35). Table 1 only reports results for moderate values of the autoregressive parameter to avoid identification issues. Table 1 clearly shows the efficiency gain from using the original AS and Sys moment conditions over their robust counterparts. It shows that especially the difference in standard deviation between the original Sys moments and their robust counterpart can be large.

Table 1: Finite sample properties of two-step GMM estimators

T	$\theta_0$	mean				standard deviation			
		original		robust		original		robust	
		AS	Sys	AS	Sys	AS	Sys	AS	Sys
4	0	-0.001	-0.002	-0.000	-0.001	0.042	0.039	0.044	0.044
4	0.25	0.250	0.249	0.250	0.250	0.056	0.048	0.059	0.059
4	0.50	0.508	0.501	0.516	0.510	0.092	0.061	0.124	0.110
5	0	-0.001	-0.002	0.000	0.000	0.031	0.028	0.031	0.031
5	0.25	0.249	0.248	0.251	0.251	0.037	0.033	0.037	0.037
5	0.50	0.499	0.498	0.507	0.503	0.050	0.039	0.061	0.056

Note: based on 1000 replications of DGP 2. N=500,  $\sigma_c^2=1$  and  $\sigma^2=1$  in all experiments.

### 5.3 Identification robust inference

We have shown that inference based on the robust moments (32) is difficult. The resulting estimators are essentially ‘worst case’ estimators since they only perform relatively well under worst case DGPs, *i.e.* when the true value of  $\theta$  is one. For DGPs with values of  $\theta$  less than one, estimators based on the original AS and Sys moment conditions outperform estimators based on the robust moments. Alongside this suboptimality also the large sample distributions can be highly non-standard, which is similar to the behavior of one-step and two-step GMM AS and Sys estimators (see Corollary 2 and Theorem 3). The estimators then have a quartic root convergence rate and a non-standard large sample distribution. The practical appeal of GMM estimators based on the robust moments is thus limited.

Instead of using the robust moments for estimation, we therefore use them to determine the optimal test procedure when the data is persistent and under worst case settings of the nuisance parameters. The robust moments contain all the identifying information for the parameter of interest in such instances. Since most test statistics are then size distorted (as we showed previously for Wald statistics), we use 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  $\theta_0$ .<sup>17</sup>

Theorem 1 shows that  $\theta$  is identified by the AS and Sys moment conditions so identification robust GMM statistics have discriminatory power. For settings of  $\theta_0$  and the nuisance parameters for which no identification issues exist, both the GMM-LM and KLM statistics

<sup>17</sup>The GMM-LM statistic is size correct in this case because under the null hypothesis no parameters are estimated.

are efficient and more powerful than the GMM-AR statistic. This standard notion of efficiency does, however, not apply to values of  $\theta_0$  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  $\theta$  under the worst case settings where its true value is one and the variance of the initial observations accords with (21). The implied low convergence rate is a result of the weak identification of  $\theta$ . There is a growing literature on optimal testing when identification is weak. The first paper in that literature is by Andrews *et. al.* (2006) on optimal testing in the linear instrumental variables regression model. Later papers on optimal testing in linear instrumental variables models with general covariance structure are by I. Andrews (2016) and Moreira and Moreira (2015). Our paper differs from these papers since the identification in the panel AR(1) model is not governed by a "first stage equation". It also depends on other nuisance parameters besides those present in the first stage equation, like, the initial observations and fixed effects. The optimality results obtained in these other papers do therefore not apply to our case. Many models estimated using GMM contain similar nuisance parameters that influence the identification so our paper contributions to optimal testing for those cases.

## 6 Robust moments and maximal attainable power in a worst case setting

We use the robust moments in (32) to determine the maximal attainable power of asymptotically size correct GMM testing procedures when the true value of  $\theta$  equals one and under the worst case setting of the nuisance parameters. In the usual local power analysis, we test  $H_0 : \theta = \theta^*$  against a local alternative, say,  $H_1 : \theta = \theta^* - \frac{e}{N^\xi}$ , for various values of the drifting parameter  $e$  with  $N^\xi$  the appropriate convergence rate and the data generating process corresponding with the model under the alternative hypothesis. Since we are interested in optimal power when  $\theta$  equals one, we focus on tests of  $H_0 : \theta = \theta^*$  against  $H_1 : \theta = \theta^* - \frac{e}{N^\xi} = 1$ , which implies that  $\theta^* = 1 + \frac{e}{N^\xi}$ . Hence, we use one point from the usual power curve, *i.e.* the rejection frequency of testing  $H_0 : \theta = 1 + \frac{e}{N^\xi}$  against  $H_1 : \theta = 1$ , and do so for various values of  $e$  to obtain an alternative power curve. This is then used to determine the optimal power under the worst case setting, which is characterized by the limiting sequence in (21).

When the true value of  $\theta$  equals one, the robust moment condition in (32) is the only part of the AS and Sys moment conditions that contains information on  $\theta$  and identifies it. We therefore use it to determine the maximal attainable power (MAP) of tests using the AS or SYS moment conditions of  $H_0 : \theta = 1 + \frac{e}{N^\xi}$  with 95% significance against the alternative (worst case) hypothesis  $H_1 : \theta = 1$  with nuisance parameters that are characterized by (21),

see Lehmann and Romano (2005):

$$\begin{aligned} MAP^j(\theta_0|\mathbf{H}_1) &= \lim_{N \rightarrow \infty} \max_{TS} \Pr [ts(\theta_0) > ac_{ts(\theta_0)}|\mathbf{H}_1] & j = AS, Sys \\ &= \lim_{N \rightarrow \infty} \max_{TS} \min_S \Pr [ts(\theta_0) > ac_{ts(\theta_0)}|\theta = 1] & j = AS, Sys \end{aligned} \quad (36)$$

where  $TS$  is the set of statistics testing  $\mathbf{H}_0$ ,  $ts(\theta_0)$  is an element of the set  $TS$ , *i.e.* a statistic that tests  $\mathbf{H}_0$ ,  $ac_{ts(\theta_0)}$  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 (36) as the maximal attainable power curve. It just focusses on the alternative hypothesis  $\mathbf{H}_1 : \theta = 1$  since the identification issues occur at this value of  $\theta$ . The standard optimality results 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 determine the rate  $\xi$  that we use in  $\mathbf{H}_0 : \theta = 1 + \frac{e}{N^\xi}$ . It is the slowest rate at which the hypothesized value of  $\theta$  under  $\mathbf{H}_0$  can drift away from one, whilst the true value of  $\theta$  equals one, such that the sample moment (32) converges to a random variable that is non-degenerate and remains finite with probability one. Theorem 4 states the convergence rate that we employ to obtain the maximal attainable power curve.

**Theorem 4.** *Under the conditions of Theorem 2, the local to unity drifting sequence for  $\theta$  under  $\mathbf{H}_0$  for the robust moments  $g_{f,T}(\theta)$  is such that:*

1. When  $\omega = 0$ ,  $\sigma_t^2 = \sigma^2$ ,  $t = 2, \dots, T : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ ,
  2. When  $\omega \neq 0$  or  $\sigma_t^2 \neq \sigma^2$ , for at least one value of  $t$ ,  $t = 2, \dots, T - 1 : \theta = 1 + \frac{e}{\sqrt{N}}$ ,
- with  $e < 0$  a finite constant.

**Proof.** see Appendix A. ■

Just like in Theorem 3 before, the quartic root convergence rate in Theorem 4 results since the robust moment equation (32) is quadratic in  $\theta$ . When we specify  $\theta$  as  $1 + \frac{e}{N^\xi}$  and  $\omega = 0$ ,  $\sigma_t^2 = \sigma^2$ ,  $t = 2, \dots, T$ , 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. Theorem 4 shows that mean stationarity,  $\omega = 0$ , and average variances which are constant over time,  $\sigma_t^2 = \sigma^2$ ,  $t = 2, \dots, T$ , lead to the largest asymptotic variance of estimators of  $\theta$ . Therefore, mean stationarity and homoscedasticity provide the worst case setting in which the power of tests is the lowest.

To construct the maximal attainable power curve, we first construct the large sample distribution of the GMM-AR statistic which tests  $\mathbf{H}_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  just using the robust moments in (32) whilst the true value of  $\theta$  is equal to one jointly with (21). The GMM-AR



statistic reads:

$$\text{GMM-AR}(e) = Ng_{f,T}(e)' \hat{V}_{gg}(e)^{-1} g_{f,T}(e), \quad (37)$$

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  $g_{f,T}(e)$ .

**Theorem 5.** *Under the conditions of Theorem 2 and when the true value of  $\theta$  is equal to one,  $\omega = 0$ ,  $\sigma_t^2 = \sigma^2$ ,  $t = 2, \dots, T$ , the large sample distribution of the GMM-AR statistic (37) for testing the hypothesis  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ , is characterized by*

$$\chi^2(\delta, p_{\max}), \quad (38)$$

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

$$B(N) = (\iota_3 \otimes I_{p_{\max}}) + \frac{e}{\sqrt[4]{N}} \left[ \left(2 + \frac{e}{\sqrt[4]{N}}\right) (e_{1,3} \otimes I_{p_{\max}}) + (e_{2,3} \otimes I_{p_{\max}}) \right], \quad (39)$$

$V_{abd}$  the covariance matrix of  $a$ ,  $b$  and  $d$ ,  $\iota_3$  a  $3 \times 1$  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  $3 \times 1$  dimensional unity vectors and  $\chi^2(\delta, p_{\max})$  a non-central  $\chi^2$  distribution with non-centrality parameter  $\delta$  and degrees of freedom parameter  $p_{\max}$ .

**Proof.** see Appendix A. ■

The expression of the large sample distribution in Theorem 5 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  $\theta$ , which is one, so they over identify  $\theta$ . More powerful statistics for testing a point null hypothesis on  $\theta$  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 (37) that leads to the largest value of the non-centrality parameter of the large sample distribution.

**Theorem 6.** *Under the conditions of Theorem 2 and when the true value of  $\theta$  is equal to one, the maximal attainable power curve for testing  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  is*

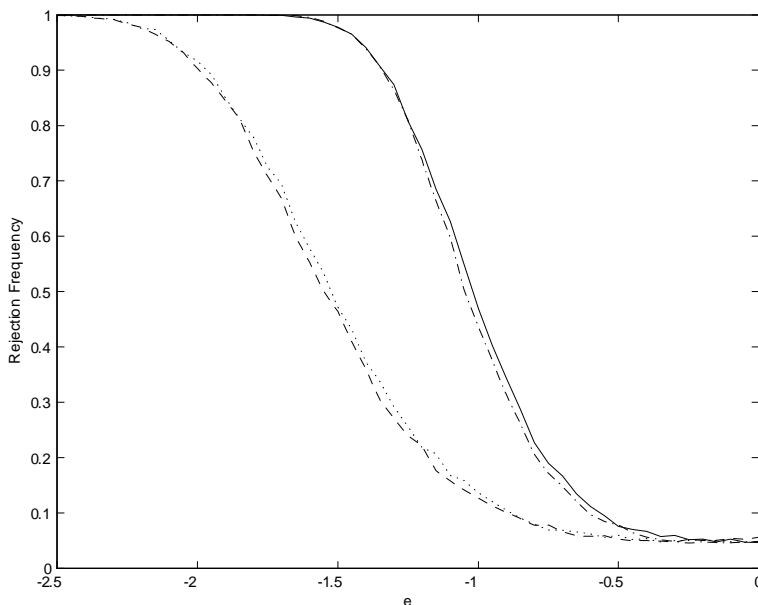
$$\chi^2(\delta, 1), \quad (40)$$

with  $\delta = e^4 \binom{\iota_p}{0}' (B(N)' V_{abd} B(N))^{-1} \binom{\iota_p}{0}$ ,  $\iota_p$  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 2 shows the maximal attainable power curves that result from the AS and Sys moment conditions when  $T = 4$  and 5. Figure 2 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 only the AS moment conditions. This novel result reveals that imposing mean stationarity is superfluous. Furthermore, Figure 2 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 2. Power envelope for testing  $H_0 : \theta = 1 + \frac{e}{\sqrt[3]{N}}$



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

## 6.1 Large individual fixed effect variance

Sofar we have focused on highly persistent panel data resulting from a large autoregressive parameter. However, the representation theorem for 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,  $\mu_i$ , 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 (21) 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 only 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  $\mu_q(\theta, \theta_0, \sigma^2)$ <sup>18</sup> accordingly. For example, the expressions of  $A_f(\theta, \theta_0)$  when  $T = 4$  for the AS and Sys moment conditions are:

$$\begin{aligned}
 \text{AS: } A_f^{AS}(\theta, \theta_0) &= \begin{pmatrix} \theta_0 - 1 - \theta & 1 & 0 \\ (\theta_0 - \theta)(\theta_0 - 1) & \theta_0 - 1 - \theta & 1 \\ \theta_0(\theta_0 - \theta)(\theta_0 - 1) & \theta_0(\theta_0 - 1 - \theta) & \theta_0 \\ (1 - \theta)(\theta_0 - \theta - 1) & 1 - \theta & 0 \end{pmatrix}, \\
 \text{Sys: } A_f^{Sys}(\theta, \theta_0) &= \begin{pmatrix} \theta_0 - 1 - \theta & 1 & 0 \\ (\theta_0 - \theta)(\theta_0 - 1) & \theta_0 - 1 - \theta & 1 \\ \theta_0(\theta_0 - \theta)(\theta_0 - 1) & \theta_0(\theta_0 - 1 - \theta) & \theta_0 \\ 1 - \theta & 0 & 0 \\ (1 - \theta)(\theta_0 - 1) & 1 - \theta & 0 \end{pmatrix}.
 \end{aligned} \tag{41}$$

The expressions in (41) are identical to those in (30) when  $\theta_0 = 1$ . Interestingly, the part of the orthogonal complement of  $A_f(\theta, \theta_0)$  which depends on  $\theta$ , *i.e.*  $G_{f,T}^j(\theta)$  in (31), remains unchanged:

$$\begin{aligned}
 \text{AS: } G_{f,T=4}^{AS}(\theta) &= \begin{pmatrix} -(1 - \theta) \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad G_{2,T=4}^{AS} = \begin{pmatrix} 0 \\ -\theta_0 \\ 1 \\ 0 \end{pmatrix}, \\
 \text{Sys: } G_{f,T=4}^{Sys}(\theta) &= \begin{pmatrix} -(1 - \theta) \\ 0 \\ 0 \\ -\theta \\ 1 \end{pmatrix}, \quad G_{2,T=4}^{Sys} = \begin{pmatrix} 0 \\ -\theta_0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.
 \end{aligned} \tag{42}$$

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. Theorems 3 and 4

<sup>18</sup>Instead of  $A_f(\theta)$ ,  $B_f(\theta)$ , ... we now use  $A_f(\theta, \theta_0)$ ,  $B_f(\theta, \theta_0)$ , ... since the results no longer apply to just the unit root value of  $\theta_0$ .

show 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 an 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 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 than those that occur because of the unit root value, we have only covered them briefly.

## 7 Power envelope and maximal attainable power curve

The results in Section 5 show that, although identification is achieved, the limiting behavior of estimators is not uniform since it depends on the data generating process at hand. Corresponding 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  $\theta$  is less than one, so  $\theta$  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<sup>19</sup>, 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 7.

**Theorem 7.** *Under the conditions from Theorem 2, the worst case large sample distributions, which apply under (21), mean stationarity (6)-(7) and  $\sigma_t^2 = \sigma^2$ ,  $t = 2, \dots, T$ , of the GMM-AR, GMM-LM and KLM statistics for testing the hypothesis  $H_0 : \theta = 1 + \frac{\epsilon}{\sqrt{N}}$  whilst*

---

<sup>19</sup>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 (37).

the true value of  $\theta$  equals one are characterized by

$$\begin{aligned} \text{GMM-AR}(e) &: \chi^2(\delta_{\text{GMM-AR}}, p_{\text{GMM-AR}}) \\ \text{KLM}(e) &: \chi^2(\delta_{\text{KLM}}, 1) \\ \text{GMM-LM}(e) &: \chi^2(\delta_{\text{GMM-LM}}, 1), \end{aligned} \quad (43)$$

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

$$\begin{aligned} \delta_{\text{GMM-AR}} &= (e\sigma)^4 \binom{\iota_p}{0}' (B(N)' V_{abd} B(N))^{-1} \binom{\iota_p}{0} \\ \delta_{\text{KLM}} &= \delta_{\text{GMM-AR}} \\ \delta_{\text{GMM-LM}} &= (e\sigma)^4 \binom{\iota_p}{0}' (B(N)' V_{abd} B(N))^{-\frac{1}{2}} P_{(B(N)' V_{abd} B(N))^{-\frac{1}{2}} (G_f(e)' A_q \psi)} \\ &= \begin{cases} (B(N)' V_{abd} B(N))^{-\frac{1}{2}} \binom{\iota_p}{0} & T > 4 \\ \delta_{\text{GMM-AR}} & T = 4, \end{cases} \end{aligned} \quad (44)$$

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

$$\lim_{N \rightarrow \infty} \text{var} \left( h_N(\theta_{0,N}) \begin{bmatrix} y_{1i} u_{i2} \\ \vdots \\ y_{1i} u_{iT} \end{bmatrix} \right). \quad (45)$$

**Proof.** see Appendix A. ■

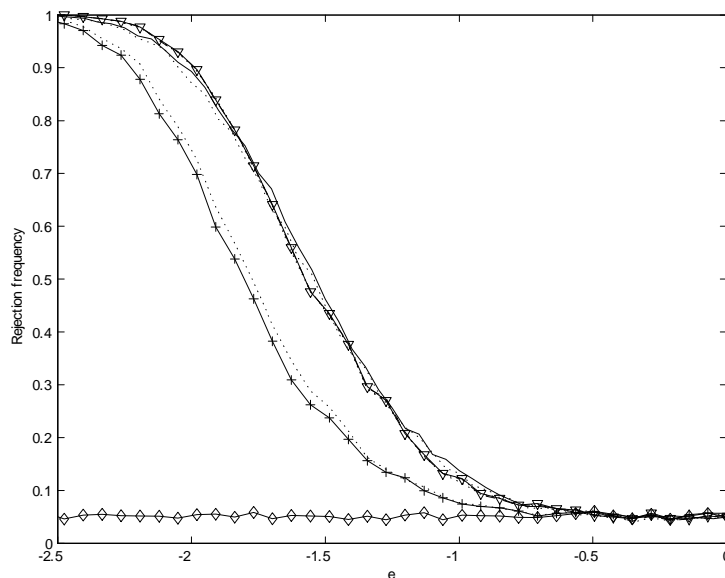
Theorem 7 shows that the lower power envelope of the KLM statistic coincides with the maximal attainable power curve as described in Theorem 6. For the GMM-LM statistic this only occurs when  $T = 4$ . It shows that the KLM statistic is, in a sense, optimal when  $\theta$  is equal to one. Since the KLM statistic is also efficient when  $\theta$  is less than one, it is efficient both when  $\theta$  is less than one or equal to one.

Figures 3 and 4 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. Worst case DGPs result from imposing mean stationarity (7) and time series homoscedasticity, *i.e.*  $\sigma_t^2 = \sigma^2$  for  $t = 1, \dots, T$ . Here we choose DGP 1 from Section 3 with a large value of  $\sigma_c^2$  (ten) compared to  $\sigma^2$  (one), which amplifies the variance of the initial conditions. The results of Theorems 6 and 7 follow from a quartic root convergence rate, and to provide a numerical assessment we fix  $N = 2000$ , a relatively large value. We next simulate for a wide range of values for  $\theta$ , which together with  $N$  provides a mapping to the constant  $e$  in Figures 3 and 4 (horizontal axis).

Since the Sys moment conditions do not identify  $\theta$  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

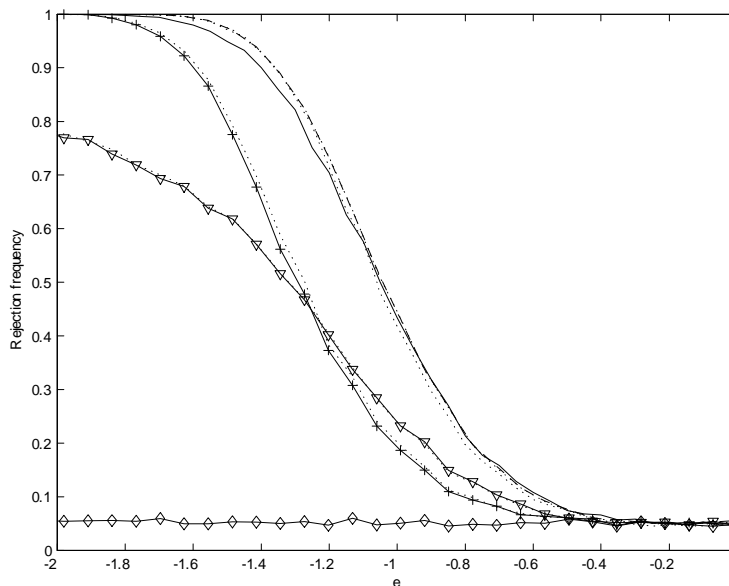
by themselves do not identify  $\theta$  when its true value is one, Figures 3 and 4 below also include the rejection frequencies that result from the GMM-AR statistic with Dif moment conditions. These rejection frequencies equal 5% for all values of  $\theta$  which shows that the Dif moment conditions do not identify  $\theta$  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.

Figure 3. 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).  $N=2000$ .

Figure 4. 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).  $N=2000$ .

Figures 3 and 4<sup>20</sup> provide a numerical proof of the main results from Theorems 6 and 7. Figure 3 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. The power envelopes of the GMM-AR statistic are below this power curve. Figure 3 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 7, the degrees of freedom parameter of the non-central  $\chi^2$  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 4 shows that, when  $T = 5$ , only 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 4 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 7 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  $\chi^2$  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.

## 8 Conclusions

We have analyzed GMM inference for dynamic panel data models involving highly persistent panel data. 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. The identification based on the AS and Sys moment conditions for the problematic cases of divergent initial observations then results from our novel robust moment conditions. They are combinations of either the AS or Sys moments and do not depend on the initial observations.

---

<sup>20</sup>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.



Despite the positive identification results for AS and Sys moment conditions, conventional Wald type inference based on GMM estimators is not valid since these estimators have non-standard limiting distributions near the unit root. We show that similar results hold for GMM estimators based on our novel robust moments. We therefore propose to use weak identification robust GMM inference which are size correct.

We use our novel robust moments to determine the maximal attainable power curves of size correct GMM tests based on the AS and Sys moment conditions under worst case settings. The robust moments then provide the only identifying information for the autoregressive parameter. We show that the maximal attainable power curves 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 condition.

Finally, we use the maximal attainable power curve to determine which size correct GMM statistic is optimal for testing when the autoregressive parameter is one and we have possibly divergent initial observations. 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 other less severe cases.

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 have only 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 specify the Dif sample moments and their derivatives as

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

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

Under (17) and (21), these expressions are approximately equal to:

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

Since

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

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

with  $\psi$  and  $\psi_{uu}$  normally distributed random variables, it is readily seen that

$$A_f^{Dif}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \end{pmatrix}, \quad B_q^{Dif}(\theta) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

$$A_q^{Dif}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad B_f^{Dif}(\theta) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\theta & 1 & 0 & 0 \end{pmatrix},$$

$$\mu_f^{Dif}(\theta, \sigma^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mu_q^{Dif}(\theta, \sigma^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$C_f^{Dif}(\theta) = A_f^{Dif}(\theta)\iota_3, \quad C_q^{Dif}(\theta) = A_q^{Dif}(\theta)\iota_3.$$

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 specify the Lev sample moments and their derivatives as

$$f_N^{Lev}(\theta) = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i3}\Delta y_{i2} - \theta y_{i2}\Delta y_{i2} \\ y_{i4}\Delta y_{i3} - \theta y_{i3}\Delta y_{i3} \end{pmatrix}$$

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

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

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

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

Under (17) and (21), these expressions are approximately equal to:

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

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

so this implies that

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

We can specify the NL sample moment and its derivative as

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

Under (17) and (21), these expressions are approximately equal to:

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

so this implies that:

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

Finally, regarding AS and Sys moment conditions we simply have

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

**T=5.** Using similar calculations we obtain:

$$\begin{aligned}
\psi &= \begin{pmatrix} \psi_{y_{i1}u_{i2}} \\ \psi_{y_{i1}u_{i3}} \\ \psi_{y_{i1}u_{i4}} \\ \psi_{y_{i1}u_{i5}} \end{pmatrix}, \\
A_f^{Dif}(\theta) &= \begin{pmatrix} -\theta & 1 & 0 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & -\theta & 1 & 0 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \\ 0 & 0 & -\theta & 1 \end{pmatrix}, \quad \mu_f^{Dif}(\theta, \sigma^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \\
A_f^{Lev}(\theta) &= \begin{pmatrix} 1-\theta & 0 & 0 & 0 \\ 0 & 1-\theta & 0 & 0 \\ 0 & 0 & 1-\theta & 0 \end{pmatrix}, \quad \mu_f^{Lev}(\theta, \sigma^2) = (1-\theta) \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \\ \sigma_4^2 \end{pmatrix}, \\
A_f^{NL}(\theta) &= \begin{pmatrix} \theta(\theta-1) & 1-\theta & 0 & 0 \\ 0 & \theta(\theta-1) & 1-\theta & 0 \end{pmatrix}, \quad \mu_f^{NL}(\theta, \sigma^2) = (1-\theta) \begin{pmatrix} \sigma_3^2 - \theta\sigma_2^2 \\ \sigma_4^2 - \theta\sigma_3^2 \end{pmatrix}.
\end{aligned}$$

**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  $\theta$ .

**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:

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

where  $T$  indicates the number of time periods and  $G_{2,T}^j$  is such that  $G_{2,T}^{j'} \mu_f^j(\theta, \sigma^2) = 0$ . 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 (30),  $G_{f,T=4}^j(\theta)$  and  $G_{2,T=4}^j$  for  $j = AS, Sys$  result as:

$$G_{f,T=4}^{AS}(\theta) = \begin{pmatrix} -(1-\theta) \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad G_{2,T=4}^{AS} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix},$$

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

From these expressions 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:

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

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

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

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

**Proof of Theorem 2.** Since

$$\begin{aligned}
\Delta y_{i2} &= c_i - (1 - \theta_0)y_{i1} + u_{i2} \\
\Delta y_{i3} &= \theta_0(c_i - (1 - \theta_0)y_{i1}) + (\theta_0 - 1)u_{i2} + u_{i3} \\
\Delta y_{i4} &= \theta_0^2(c_i - (1 - \theta_0)y_{i1}) + \theta_0(\theta_0 - 1)u_{i2} + (\theta_0 - 1)u_{i3} + u_{i4} \\
\Delta y_{i5} &= \theta_0^3(c_i - (1 - \theta_0)y_{i1}) + \theta_0^2(\theta_0 - 1)u_{i2} + \theta_0(\theta_0 - 1)u_{i3} + (\theta_0 - 1)u_{i4} + u_{i5} \\
y_{i3} - y_{i1} &= (1 + \theta_0)(c_i - (1 - \theta_0)y_{i1}) + \theta_0 u_{i2} + u_{i3} \\
y_{i4} - y_{i1} &= (1 + \theta_0 + \theta_0)(c_i - (1 - \theta_0)y_{i1}) + \theta_0 u_{i2} + \theta_0 u_{i3} + u_{i4} \\
y_{i4} - y_{i2} &= (\theta_0 + \theta_0)(c_i - (1 - \theta_0)y_{i1}) + (\theta_0 - 1)u_{i2} + \theta_0 u_{i3} + u_{i4} \\
y_{i5} - y_{i1} &= (1 + \theta_0 + \theta_0 + \theta_0)(c_i - (1 - \theta_0)y_{i1}) + \theta_0 u_{i2} + \theta_0 u_{i3} + \theta_0 u_{i4} + u_{i5} \\
y_{i5} - y_{i2} &= (\theta_0 + \theta_0 + \theta_0)(c_i - (1 - \theta_0)y_{i1}) + (\theta_0 - 1)u_{i2} + \theta_0 u_{i3} + \theta_0 u_{i4} + u_{i5}
\end{aligned}$$

it holds that for

**T=4, Sys:**

$$\begin{aligned}
a &\xrightarrow{p} \left( \frac{E((c_i - (1 - \theta_0)y_{i1})^2) + \sigma_2^2}{0} \right) \\
b &\xrightarrow{p} \left( \frac{-(1 + \theta_0)^2 E((c_i - (1 - \theta_0)y_{i1})^2) - \theta_0^2 \sigma_2^2 - \sigma_3^2}{-\theta_0^2 E((c_i - (1 - \theta_0)y_{i1})^2)} \right) \\
d &\xrightarrow{p} \left( \frac{\theta_0(1 + \theta_0 + \theta_0^2) E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0^2(\theta_0 - 1)\sigma_2^2 + \theta_0\sigma_3^2}{\theta_0^2 E((c_i - (1 - \theta_0)y_{i1})^2)} \right).
\end{aligned}$$



**T=4, AS:**

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

**T=5, Sys:**

$$\begin{aligned}
a &\xrightarrow{p} \begin{pmatrix} E((c_i - (1 - \theta_0)y_{i1})^2) + \sigma_2^2 \\ (1 + \theta_0)E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0(\theta_0 - 1)\sigma_2^2 + \sigma_3^2 \\ \theta_0^2E((c_i - (1 - \theta_0)y_{i1})^2) + (\theta_0 - 1)^2\sigma_2 + \sigma_3^2 \\ 0 \\ 0 \end{pmatrix} \\
b &\xrightarrow{p} - \begin{pmatrix} (1 + \theta_0)^2E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0^2\sigma_2^2 + \sigma_3^2 \\ (1 + \theta_0 + \theta_0^2)(\theta_0 + \theta_0^2)E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0^2(\theta_0^2 - 1)\sigma_2^2 + \theta_0^2\sigma_3^2 + \sigma_4^2 \\ (\theta_0 + \theta_0^2)^2(\theta_0 + \theta_0^2)E((c_i - (1 - \theta_0)y_{i1})^2) + (\theta_0^2 - 1)^2\sigma_2^2 + \theta_0^2\sigma_3^2 + \sigma_4^2 \\ \theta_0^2E((c_i - (1 - \theta_0)y_{i1})^2) \\ \theta_0^3E((c_i - (1 - \theta_0)y_{i1})^2) \end{pmatrix} \\
d &\xrightarrow{p} \begin{pmatrix} \theta_0(1 + \theta_0 + \theta_0^2)E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0^2(\theta_0 - 1)\sigma_2^2 + \theta_0\sigma_3^2 \\ \theta_0^2(1 + \theta_0 + \theta_0^2 + \theta_0^3)E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0^4(\theta_0 - 1)\sigma_2^2 + \theta_0^2(\theta_0 - 1)\sigma_3^2 + \theta_0\sigma_4^2 \\ \theta_0^2(\theta_0 + \theta_0^2 + \theta_0^3)E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0(\theta_0 - 1)(\theta_0^3 - 1)\sigma_2^2 + \theta_0^2(\theta_0 - 1)\sigma_3^2 + \theta_0\sigma_4^2 \\ \theta_0^3E((c_i - (1 - \theta_0)y_{i1})^2) \\ \theta_0^4E((c_i - (1 - \theta_0)y_{i1})^2) \end{pmatrix}.
\end{aligned}$$

**T=5, AS:**

$$\begin{aligned}
a &\xrightarrow{p} \begin{pmatrix} (1 + \theta_0)E((c_i - (1 - \theta_0)y_{i1})^2) + \sigma_2^2 \\ \theta_0(1 + \theta_0 + \theta_0^2)E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0^2(\theta_0 - 1)\sigma_2^2 + \theta_0\sigma_3^2 \\ \theta_0^2(1 + \theta_0)E((c_i - (1 - \theta_0)y_{i1})^2) + (\theta_0^2 - 1)(\theta_0 - 1)\sigma_2 + \theta_0\sigma_3^2 \\ 0 \\ 0 \end{pmatrix} \\
b &\xrightarrow{p} - \begin{pmatrix} -((1 + \theta_0)^2 + 1)E((c_i - (1 - \theta_0)y_{i1})^2) - \theta_0(2\theta_0 - 1)\sigma_2^2 - \sigma_3^2 \\ (\theta_0 + 2\theta_0^2(1 + \theta_0 + \theta_0^2))E((c_i - (1 - \theta_0)y_{i1})^2) + 2\theta_0^3(\theta_0 - 1)\sigma_2^2 + \theta_0(2\theta_0 - 1)\sigma_3^2 + \sigma_4^2 \\ \theta_0^2(1 + 2\theta_0(1 + \theta_0))E((c_i - (1 - \theta_0)y_{i1})^2) + (2\theta_0^3 - \theta_0 - 1)(\theta_0 - 1)\sigma_2^2 + \theta_0(2\theta_0 - 1)\sigma_3^2 + \sigma_4^2 \\ \theta_0^2E((c_i - (1 - \theta_0)y_{i1})^2) \\ \theta_0^3E((c_i - (1 - \theta_0)y_{i1})^2) \end{pmatrix}
\end{aligned}$$

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

**Proof of Lemma 1.** Theorem 2 implies for both AS and Sys moments that

$$\text{plim}_{N \rightarrow \infty} \begin{pmatrix} a \\ b \\ d \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \\ d_0 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix},$$

hence  $\text{plim}_{N \rightarrow \infty}(b^2 - 4ad) = 0$ . Furthermore, for Sys moments

$$\begin{aligned} \sqrt{N} \begin{pmatrix} a - a_0 \\ b - b_0 \\ d - d_0 \end{pmatrix} &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \begin{pmatrix} u_{i2}^2 - \sigma^2 \\ -(u_{i2}^2 + 2u_{i2}u_{i3} + u_{i3}^2) + 2\sigma^2 \\ (u_{i2} + u_{i3} + u_{i4})u_{i3} - \sigma^2 \end{pmatrix} \\ &\xrightarrow{d} N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \sigma^4 \begin{bmatrix} 2 & -2 & 0 \\ -2 & 8 & -4 \\ 0 & -4 & 4 \end{bmatrix} \right), \end{aligned}$$

and regarding AS we have

$$\begin{aligned} \sqrt{N} \begin{pmatrix} a - a_0 \\ b - b_0 \\ d - d_0 \end{pmatrix} &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \begin{pmatrix} (u_{i2} + u_{i3})u_{i2} - \sigma^2 \\ -(u_{i2}^2 + 2u_{i2}u_{i3} + u_{i3}^2) - u_{i2}u_{i4} + 2\sigma^2 \\ (u_{i2} + u_{i3} + u_{i4})u_{i3} - \sigma^2 \end{pmatrix} \\ &\xrightarrow{d} N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \sigma^4 \begin{bmatrix} 3 & -4 & 1 \\ -4 & 9 & -4 \\ 1 & -4 & 4 \end{bmatrix} \right). \end{aligned}$$

Using the delta method, this yields for both AS and Sys moments

$$\begin{aligned} \sqrt{N}(b^2 - 4ad) &= 2b_0\sqrt{N}(b - b_0) - 4d_0\sqrt{N}(a - a_0) - 4a_0\sqrt{N}(d - d_0) \\ &\xrightarrow{d} N(0, 32\sigma^4). \end{aligned}$$

**Proof of Theorem 3.** For those 50% of the realizations with a positive discriminant we have

$$\hat{\theta} = -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ad}}{2a},$$

so that

$$\text{plim}_{N \rightarrow \infty} \hat{\theta} = \text{plim}_{N \rightarrow \infty} \frac{-b}{2a} = 1.$$

Because both the discriminant  $b^2 - 4ad$  and  $-b/(2a) - 1 = (-b - 2a)/(2a)$  are  $O_P(N^{-1/2})$ , it follows that  $\sqrt{b^2 - 4ad}/(2a) = O_P(N^{-1/4})$ , which dominates the distribution of  $\hat{\theta} - 1$ . In summary, we find for realizations with a positive discriminant:

$$\begin{aligned} N^{1/4}(\hat{\theta} - 1) &= N^{-1/4}\sqrt{N} \left( -\frac{b}{2a} - 1 \right) - \frac{\sqrt{\sqrt{N}(b^2 - 4ad)}}{2a} \\ &= -\sqrt{\frac{\sqrt{N}(b^2 - 4ad)}{4a^2}} + O_P(N^{-1/4}). \end{aligned}$$

Substituting the limiting distribution derived under Lemma 1, we find for both AS and Sys:

$$N^{1/4}(\hat{\theta} - 1) \xrightarrow{d} -\sqrt{|N(0, 2\sigma^2)|}.$$

**Proof of Theorem 4.** The components  $a$ ,  $b$  and  $d$  in (32) are all sample averages so we can characterize their large sample behavior by

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

with  $\varepsilon_a$ ,  $\varepsilon_b$  and  $\varepsilon_d$  converging to mean zero random variables and the expressions for  $E(a)$ ,  $E(b)$  and  $E(d)$ , when the true value of  $\theta$  is one, are stated in Theorem 2. To determine the appropriate rate for the local to unity asymptotics regarding  $\theta$ , we insert

$$\theta = 1 + \frac{e}{N\xi}$$

in (32) and determine the appropriate value of  $\xi$  :

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

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

$$\begin{aligned} &\left(1 + \frac{e}{\sqrt[4]{N}}\right)^2 \left(E(a) + \frac{\varepsilon_a}{\sqrt{N}}\right) + \left(1 + \frac{e}{\sqrt[4]{N}}\right) \left(E(b) + \frac{\varepsilon_b}{\sqrt{N}}\right) + E(d) + \frac{\varepsilon_d}{\sqrt{N}} = \\ &E(a) + E(b) + E(d) + \frac{1}{\sqrt{N}}(\varepsilon_a + \varepsilon_b + \varepsilon_d + e^2 E(a)) + \\ &\frac{e}{\sqrt[4]{N}} \left(E(b) + 2E(a)\right) + \frac{e}{\sqrt{N}\sqrt[4]{N}}(\varepsilon_b + 2\varepsilon_a) + \frac{e^2 \varepsilon_a}{N} = \\ &\frac{1}{\sqrt{N}}(\varepsilon_a + \varepsilon_b + \varepsilon_d + e^2 E(a)) + \frac{e}{\sqrt{N}\sqrt[4]{N}}(\varepsilon_b + 2\varepsilon_a) + \frac{e^2 \varepsilon_a}{N} \end{aligned}$$

since  $E(a) + E(b) + E(d) = 0$  and  $E(b) = -2E(a)$  so the appropriate specification for  $\theta$  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  $\xi = \frac{1}{2}$  :

$$\begin{aligned} &\left(1 + \frac{e}{\sqrt{N}}\right)^2 \left(E(a) + \frac{\varepsilon_a}{\sqrt{N}}\right) + \left(1 + \frac{e}{\sqrt{N}}\right) \left(E(b) + \frac{\varepsilon_b}{\sqrt{N}}\right) + E(d) + \frac{\varepsilon_d}{\sqrt{N}} = \\ &E(a) + E(b) + E(d) + \frac{1}{\sqrt{N}}(\varepsilon_a + \varepsilon_b + \varepsilon_d + e(E(b) + 2E(a))) + \\ &\frac{e}{N}(2\varepsilon_a + \varepsilon_b + eE(a)) + \frac{e^2 \varepsilon_a}{N\sqrt{N}} = \\ &\frac{1}{\sqrt{N}}(\varepsilon_a + \varepsilon_b + \varepsilon_d - e(E(b) + 2E(a))) - \frac{e}{N}(2\varepsilon_a + \varepsilon_b - eE(a)) + \frac{e^2 \varepsilon_a}{N\sqrt{N}} \end{aligned}$$

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

**Proof of Theorem 5.** 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

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

Therefore, we have

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

and

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

with

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

and  $V_{abd}$  the covariance matrix of  $(a' : b' : d)'$ ,  $\iota_3$  a  $3 \times 1$  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  $3 \times 1$  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:

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

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

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

From the above, it then straightforwardly follows that

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

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 6.** 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  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  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

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

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

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

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

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

and the largest root then equals

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

so the maximal value of the non-centrality parameter is

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

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

**Proof of Theorem 7.** The specification of  $A_f^j(\theta)_\perp$  as equal to  $(G_{f,T}^j(\theta) \dot{=} G_{2,T}^j)$ , see (31), is such that  $(A_f^j(\theta) \dot{=} A_f^j(\theta)_\perp)$  is not invertible for the AS moment conditions both when  $T = 4$  and 5. The invertibility of  $(A_f^j(\theta) \dot{=} A_f^j(\theta)_\perp)$  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 specification of  $A_f^j(\theta)_\perp$ , we therefore use:

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

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

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

The specification of  $Q$  intends to economize on notation for the remaining of the proof. It implies that the same expressions can be used when the GMM statistics are based on either the Sys or AS sample moments.

**GMM-AR statistic** To construct the worst case limiting distribution of the GMM-AR statistic to test  $H_0 : \theta = 1 + \frac{e}{\sqrt[3]{N}}$  whilst the true value of  $\theta$  equals one, we first specify the GMM-AR statistic as

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

We now determine the limit behavior of the different components under the limit sequence in (21) and using the above alternative specification of  $A_f(e)_\perp$ . The large sample behavior of the different components of  $Q$  for the AS moment conditions are such that

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

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

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

Obviously for the Sys moment conditions  $\bar{Q}$  equals the identity matrix. 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

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

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

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

so combining,

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

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:

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

Under the worst case setting (21) also

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

so

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

Because  $A_f(e)'f_N(e)$  and  $A_f(e)'_\perp f_N(e)$  are uncorrelated under (21),

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

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:

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

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

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

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

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

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 altogether, the worst case large sample distribution of the GMM-AR statistic test  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  reads

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

**GMM-LM statistic** To obtain the large sample behavior of the GMM-LM statistic to test  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  when  $\theta_0$  is one and under the limiting sequence in (21), we determine the behavior of the different components of:

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

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 (21):

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

since under (21):

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



$A_f(e)'_{\perp} q_N(e)$  : 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:

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

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

$$\begin{aligned} \mathbf{Sys}: \quad \bar{e} &= e\sigma^2, \\ \mathbf{AS}: \quad &= e [2\sigma^2 - \lim_{\theta_0 \uparrow 1} E((\theta_0 - 1)u_{i1}^2)]. \end{aligned}$$

Combining:

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

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

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

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

$$\begin{aligned}
h_N(\theta_{0,N})Nq_N(e)'\hat{V}_{ff}(e)^{-1}f_N(e) &= h_N(\theta_{0,N})\sqrt{N} \left[ (h_N(\theta_{0,N})A_f(e) \dot{ : } A_f(e)_\perp)'q_N(e) \right]' \\
((h_N(\theta_{0,N})A_f(e) \dot{ : } A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \dot{ : } A_f(e)_\perp))^{-1}(h_N(\theta_{0,N})A_f(e) \dot{ : } A_f(e)_\perp)'\sqrt{N}f_N(e) &\approx \\
\left[ \left( \begin{array}{c} h_N(\theta_{0,N})A_f(e)'A_q(e) \\ \bar{Q}'(G_f(e)'A_q) \end{array} \right) \psi + h_N(\theta_{0,N})\sqrt{N} \left( \begin{array}{c} 0 \\ \bar{Q}'(-\frac{\varepsilon}{\sqrt{4N}}\sigma^2\iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq}) \\ \frac{1}{\sqrt{N}}\varepsilon_{bq} \end{array} \right) \right]' \\
\left( (A_f(e)'A_f(e))^{-1} \left[ \lim_{N \rightarrow \infty} \text{var}(h_N(\theta_{0,N})) \left( \begin{array}{c} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{array} \right) \right] \psi \right) &\xrightarrow{d} \\
\left( (\bar{Q}'B(N)'V_{abd}B(N)\bar{Q})^{-1} \bar{Q}' \left( e^2\sigma^2(\iota_p)_0 + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \right) & \\
\left( \begin{array}{c} G_f(e)'A_q(e)\psi \\ 0 \end{array} \right)' \bar{Q}' (\bar{Q}'B(N)'V_{abd}B(N)\bar{Q})^{-1} \bar{Q}' \left( e^2\sigma^2(\iota_p)_0 + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) &= \\
\left( \begin{array}{c} G_f(e)'A_q(e)\psi \\ 0 \end{array} \right)' (B(N)'V_{abd}B(N))^{-1} \left( e^2\sigma^2(\iota_p)_0 + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right), &
\end{aligned}$$

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 (21) of a smaller order of magnitude and therefore drop out.

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

$$\begin{aligned}
h_N(\theta_{0,N})^2Nq_N(e)'\hat{V}_{ff}(e)^{-1}q_N(e) &= h_N(\theta_{0,N})^2N \left[ (h_N(\theta_{0,N})A_f(e) \dot{ : } A_f(e)_\perp)'q_N(e) \right]' \\
((h_N(\theta_{0,N})A_f(e) \dot{ : } A_f(e)_\perp)'\hat{V}_{ff}(e)(h_N(\theta_{0,N})A_f(e) \dot{ : } A_f(e)_\perp))^{-1}(h_N(\theta_{0,N})A_f(e) \dot{ : } A_f(e)_\perp)'q_N(e) & \\
\xrightarrow{d} \left( \begin{array}{c} G_f(e)'A_q(e)\psi \\ 0 \end{array} \right)' \bar{Q} (\bar{Q}'B(N)'V_{abd}B(N)\bar{Q})^{-1} \bar{Q}' \left( \begin{array}{c} G_f(e)'A_q(e)\psi \\ 0 \end{array} \right) &= \\
\left( \begin{array}{c} G_f(e)'A_q(e)\psi \\ 0 \end{array} \right)' (B(N)'V_{abd}B(N))^{-1} \left( \begin{array}{c} G_f(e)'A_q(e)\psi \\ 0 \end{array} \right), &
\end{aligned}$$

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  $H_0 : \theta = 1 + \frac{\varepsilon}{\sqrt{4N}}$  under (21):

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

with

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

and independent from  $\psi$ , so

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

with  $\delta(\psi) = e^4 \sigma^4 \binom{\iota_p}{0}' (B(N)' V_{abd} B(N))^{-\frac{1}{2}} P_{(B(N)' V_{abd} B(N))^{-\frac{1}{2}} (G_f(e)' A_q(e) \psi)} (B(N)' V_{abd} B(N))^{-\frac{1}{2}} \binom{\iota_p}{0}$ .

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  $(G_f(e)' A_q \psi) = \binom{\iota_p}{0} G_f(e)' A_q \psi$  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  $H_0 : \theta = 1 + \frac{e}{\sqrt{N}}$  when  $\theta_0$  is equal to one and under the limiting sequence in (21), we first determine the behavior of

$$\begin{aligned} & (h_N(\theta_{0,N}) A_f(e) \dot{=} A_f(e)_\perp)' \hat{V}_{\theta_f}(e) (h_N(\theta_{0,N}) A_f(e) \dot{=} A_f(e)_\perp) \xrightarrow{p} \\ & \left( \begin{array}{c} A_f(e)' A_q \\ \bar{Q}' \binom{1}{h_N(\theta_{0,N})} G_f(e)' A_q \\ 0 \end{array} \right) \left[ \lim_{N \rightarrow \infty} \text{var}(h_N(\theta_{0,N}) \begin{pmatrix} y_{1i} u_{i2} \\ \vdots \\ y_{1i} u_{iT} \end{pmatrix}) \right] \begin{pmatrix} A_f(e)' A_f(e) \\ 0 \\ 0 \end{pmatrix}' + \\ & \begin{pmatrix} 0 & 0 \\ 0 & \bar{Q}' \binom{V_{aq,abd} B(N)}{V_{bq,abd} B(N)} \bar{Q} \end{pmatrix}, \end{aligned}$$

with  $V_{aq,abd}$ ,  $V_{bq,abd}$  the covariance between  $\varepsilon_{aq}$  and  $(\varepsilon'_a \dot{=} \varepsilon'_b \dot{=} \varepsilon'_d)'$  and  $\varepsilon_{bq}$  and  $(\varepsilon'_a \dot{=} \varepsilon'_b \dot{=} \varepsilon'_d)'$  respectively.

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

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

so

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

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

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

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

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

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

$$\begin{aligned}
\sqrt{N} \hat{D}_N(e)' \hat{V}_{ff}(e)^{-1} \hat{D}_N(e) &= \left[ \sqrt[4]{N}(h_N(\theta_{0,N})A_f(e) \dot{=} A_f(e)_\perp)' \hat{D}_N(e) \right]' \\
& ((h_N(\theta_{0,N})A_f(e) \dot{=} A_f(e)_\perp)' \hat{V}_{ff}(e) (h_N(\theta_{0,N})A_f(e) \dot{=} A_f(e)_\perp))^{-1} \left[ \sqrt[4]{N}(h_N(\theta_{0,N})A_f(e) \dot{=} A_f(e)_\perp)' \hat{D}_N(e) \right] \\
& \approx \left[ \binom{t_p}{0} \bar{e} + \frac{1}{\sqrt[4]{N}} \nu \right]' \bar{Q} \left( \bar{Q}' B(N)' V_{abd} B(N) \bar{Q} \right)^{-1} \bar{Q}' \left[ \binom{t_p}{0} \bar{e} + \frac{1}{\sqrt[4]{N}} \nu \right] \\
& \rightarrow \bar{e}^2 \binom{t_p}{0}' \bar{Q} \left( \bar{Q}' B(N)' V_{abd} B(N) \bar{Q} \right)^{-1} \bar{Q}' \binom{t_p}{0}
\end{aligned}$$

and

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

Upon combining everything, we obtain the limit behavior of the KLM statistic to test  $H_0$  :  $\theta = 1 + \frac{e}{\sqrt[4]{N}}$  under (21):

$$\text{KLM}(e) \xrightarrow{d} \eta' P (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-\frac{1}{2}} \bar{Q} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \bar{e} \eta$$

with

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

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

$$\text{KLM}(e) \xrightarrow{d} \chi^2(\delta_{KLM}, 1)$$

because

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

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):

$$E[f_i(\theta_0)] = 0, \quad i = 1, \dots, N, \quad (46)$$

which are a function of observed data and the unknown parameter vector  $\theta$ . The moment conditions are only satisfied at the true value of the  $p$ -dimensional vector  $\theta$ ,  $\theta_0$ , 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 (46) are estimated using the average sample moments,

$$f_N(\theta) = \frac{1}{N} \sum_{i=1}^N f_i(\theta). \quad (47)$$

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

$$q_N(\theta) = \frac{\partial}{\partial \theta'} f_N(\theta) = \frac{1}{N} \sum_{i=1}^N q_i(\theta), \quad (48)$$

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

The two step estimator results by minimizing the objective function:

$$Q(\theta, \theta^1) = N f_N(\theta)' \hat{V}_{ff}(\hat{\theta}^1)^{-1} f_N(\theta), \quad (49)$$

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

$$\hat{V}_{ff}(\theta) = \frac{1}{N} \sum_{i=1}^N (f_i(\theta) - f_N(\theta))(f_i(\theta) - f_N(\theta))'. \quad (50)$$

The two step estimator,  $\hat{\theta}_{2step}$ , uses the one step estimator  $\hat{\theta}_{1step}$  which equals the minimizer of (49) when we replace  $\hat{V}_{ff}(\theta)^{-1}$  by the identity matrix.

The expressions of the different statistics to test  $H_0 : \theta = \theta_0$  that we use read:

1. Two step Wald statistic:

$$W_{2step}(\theta_0) = N(\hat{\theta}_{2step} - \theta_0)' q_N(\hat{\theta}_{2step})' \hat{V}_{ff}(\hat{\theta}_{2step})^{-1} q_N(\hat{\theta}_{2step})(\hat{\theta}_{2step} - \theta_0). \quad (51)$$

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

$$LM(\theta_0) = N f_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} q_N(\theta_0) \left[ q_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} q_N(\theta_0) \right]^{-1} q_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} f_N(\theta_0). \quad (52)$$

3. The KLM statistic of Kleibergen (2005):

$$KLM(\theta_0) = N f_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} \hat{D}_N(\theta_0) \left[ \hat{D}_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} \hat{D}_N(\theta_0) \right]^{-1} \hat{D}_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} f_N(\theta_0), \quad (53)$$

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

$$\text{vec}(\hat{D}_N(\theta)) = \text{vec}(q_N(\theta)) - \hat{V}_{qf}(\theta) \hat{V}_{ff}(\theta)^{-1} f_N(\theta) \quad (54)$$

and

$$\hat{V}_{qf}(\theta) = \frac{1}{N} \sum_{i=1}^N \text{vec}[q_i(\theta) - q_N(\theta)](f_i(\theta) - f_N(\theta))'. \quad (55)$$

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

$$GMM-AR(\theta) = N f_N(\theta)' \hat{V}_{ff}(\theta)^{-1} f_N(\theta) = Q(\theta, \theta). \quad (56)$$

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

$$\begin{aligned} f_i^{Dif}(\theta) &= Z_i^{Dif} \varphi_i^{Dif}(\theta) \\ q_i^{Dif}(\theta) &= -Z_i^{Dif} \Delta y_{-1,i}, \end{aligned} \quad (57)$$

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

$$Z_i^{Dif} = \begin{pmatrix} y_{i1} & 0 \dots 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 \dots 0 & \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT-2} \end{pmatrix} \end{pmatrix} : \frac{1}{2}(T-1)(T-2) \times (T-2). \quad (58)$$

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

$$\begin{aligned} f_i^{Lev}(\theta) &= Z_i^{Lev} \varphi_i^{Lev}(\theta) \\ q_i^{Lev}(\theta) &= Z_i^{Lev} y_{-1,i}, \end{aligned} \quad (59)$$

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

$$Z_i^{Lev} = \begin{pmatrix} \Delta y_{i2} & 0 \dots 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 \dots 0 & \Delta y_{iT-1} \end{pmatrix} : (T-2) \times (T-2). \quad (60)$$

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

$$\begin{aligned} f_i^{NL}(\theta) &= Z_i^{NL} \varphi_i^{NL}(\theta) \\ q_i^{NL}(\theta) &= Z_i^{NL} \frac{\partial}{\partial \theta} \varphi_i^{NL}(\theta), \end{aligned} \quad (61)$$

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

$$Z_i^{NL} = \begin{pmatrix} 1 & 0 \dots 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 \dots 0 & 1 \end{pmatrix} : (T-3) \times (T-3). \quad (62)$$

The specification of the moment functions for the AS moment conditions results by stacking the moment conditions in (57) and (61) 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 (57) and (59) so  $k_{Sys}$  equals  $\frac{1}{2}(T+1)(T-2)$ .

**Acknowledgements:** 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<sup>2</sup> meeting in Maastricht, Groningen, Leuven, Monash, the 19-th International conference on panel data in London, Oxford, the Tinbergen Institute in Amsterdam, Toulouse and UCL for helpful comments and stimulating discussion.

## 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.
- Andrews, I. (2015). Conditional Linear Combination Tests for Weakly Identified Models. *Working Paper, Harvard University*.
- 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 effects. *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*.
- Moreira, H. and M.J. Moreira (2013). Contributions to the Theory of Optimal Tests. *Working paper, Getulio Vargas Foundation*.
- 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.