



## UvA-DARE (Digital Academic Repository)

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

Bun, M.J.G.; Kleibergen, F.

**Publication date**

2013

**Document Version**

Submitted manuscript

[Link to publication](#)

**Citation for published version (APA):**

Bun, M. J. G., & Kleibergen, F. (2013). *Identification and inference in moments based analysis of linear dynamic panel data models*. (UvA-Econometrics Discussion Paper; No. 2013/07). University of Amsterdam.

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Discussion Paper: 2013/07

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

Maurice J.G. Bun and Frank Kleibergen

[www.ase.uva.nl/uva-econometrics](http://www.ase.uva.nl/uva-econometrics)

**Amsterdam School of Economics**

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

UvA  UNIVERSITEIT VAN AMSTERDAM



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

Maurice Bun\*      Frank Kleibergen†

June 20, 2013

## Abstract

We show that Dif(ference), see Arellano and Bond (1991), Lev(el), see Arellano and Bover (1995) and Blundell and Bond (1998), or the N(on-)L(inear) moment conditions of Ahn and Schmidt (1995) do not identify the parameters of a first-order autoregressive panel data model when the autoregressive parameter is equal to one. Combinations of the Dif and Lev, resulting in Sys(tem), moment conditions and the Dif and NL, resulting in A(hn-)S(chmidt), moment conditions identify the parameters when there are four or more time periods. The behavior of one step and two step GMM estimators, however, remains non-standard. We therefore use size correct GMM statistics, like, the GMM-AR, GMM-LM or KLM statistic, to conduct inference. We compare their worst case large sample distributions with the power envelope to determine the optimal statistic. The power envelope involves a quartic root convergence rate which further indicates the non-standard identification issues. The worst case large sample distribution of the KLM statistic coincides with the power envelope whilst the one of the GMM-LM statistic only does so when there are four time periods. It shows that the KLM statistic is efficient both when the autoregressive parameter is one or less than one. The power envelopes for the AS and Sys moment conditions are identical so assuming mean stationarity does not help for identification.

---

\*Department of Economics & Econometrics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands. Email: M.J.G.Bun@uva.nl. Homepage: <http://www1.feb.uva.nl/pp/mjgbun/>. The research of the author has been funded by the NWO Vernieuwingsimpuls research grant ‘Causal Inference with Panel Data’.

†Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States. Email: Frank.Kleibergen@brown.edu. Homepage: [http://www.econ.brown.edu/fac/Frank\\_Kleibergen](http://www.econ.brown.edu/fac/Frank_Kleibergen).

# 1 Introduction

The moment conditions for linear dynamic panel data models 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). All these moment conditions involve first differences of the series to remove the fixed effects. This first difference operation removes the information in the series at the unit root value of the autoregressive parameter. The Dif moment conditions do therefore not identify the autoregressive parameter when its true value is one. This has led to the development of the Lev and NL moment conditions which are considered to identify the autoregressive parameter when its true value is equal to one.

The moment conditions are semi-parametric with respect to the initial observations, fixed effects and variances of the disturbances. We show that these affect the identification of the autoregressive parameter and that none of the previous moment conditions identifies the autoregressive parameter when its true value is one for all specifications of these nuisance parameters. To exhaust all the information on the autoregressive parameter, the Dif, Lev and NL moment conditions are combined. 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. In a combination of all three sets of moments conditions, the NL moment conditions are redundant. The Sys moment conditions exhaust all information on the autoregressive parameter that is present under mean stationarity. The AS moment conditions exhaust all information whilst not assuming mean stationarity to hold, see Ahn and Schmidt (1995). We show that from four time periods onwards both of these combined sets of moment conditions identify the autoregressive parameter for all specifications of the nuisance parameters. Hence, the autoregressive parameter is identified by the AS and Sys moment conditions when there are at least four time periods.

Despite that the AS and Sys moment conditions identify the autoregressive parameter from four time periods onwards, the large sample distributions of one step and two step generalized method of moments (GMM) estimators remain non-standard when the true value of the autoregressive parameter is (close to) one. Only a specific part

of the AS and Sys sample moments identifies the autoregressive parameter whilst the remaining part of these sample moments diverges and is therefore not helpful for identification. Since one and two step estimators result from the full unweighted sample moment, they are combinations of the identifying and diverging parts of the sample moments. Hence, they are inconsistent and have non-standard large sample distributions. This explains their large biases and the size distortions of their corresponding t-statistics when the series are persistent, see *e.g.* Bond and Windmeijer (2005), Bond *et. al.* (2005), Kruiniger (2009) and Bun and Windmeijer (2010).

Instead of using the full sample moment, we can also estimate the autoregressive parameter just from that part of the sample moment that identifies it. The resulting estimator, however, only outperforms one and two step GMM estimators when the true value of the autoregressive parameter is one and under certain (worst case) specifications of the nuisance parameters. The identifying part of the sample moment is also a quadratic polynomial of the autoregressive parameter with a discriminant whose expected value is equal to zero for the worst case specifications of the nuisance parameters. It leads to a quartic root convergence rate and a non-standard large sample distribution under these worst case specifications.

Since estimation using the identifying part of the sample moments is complicated, we use it instead to establish the optimal test procedure amongst GMM statistics that remain size correct for values of the autoregressive parameter close to one. This provides an extension of Andrews *et. al.* (2009) from the linear instrumental variables regression model with one included endogenous parameter towards the panel autoregressive model of order one. The size correct GMM statistics that we use are the GMM-A(nderson-)R(ubin) statistic of Anderson and Rubin (1949) and Stock and Wright (2000), the GMM-L(agrang-)M(ultiplier) statistic of Newey and West (1987) and the KLM statistic of Kleibergen (2005). The GMM-LM and KLM statistics are both Lagrange multiplier or score statistics and are therefore efficient when the true value of the autoregressive parameter is less than one. To determine if one of them is efficient when the true value of the autoregressive parameter is equal to one, we construct the power envelope. The power envelope shows the maximal rejection frequency under the worst case specifications of the nuisance parameters. Under these worst case specifications, the only part of the sample moments that contains information on the autoregressive parameter is the identifying part of the sample moments. The power envelope therefore results from it. The power envelope involves a quartic root convergence rate which further reflects the non-standard manner in which the autoregressive parameter is identified by the moment conditions. We compare the power envelope

with the large sample distributions of the GMM-AR, GMM-LM and KLM statistics that result under the worst case specifications of the nuisance parameters. They show that, for these worst case specifications, the GMM-AR statistic leads to rejection frequencies below the power envelope, the rejection frequencies of the GMM-LM statistic are on the power envelope when there are four time periods and below it when there are more time periods and the rejection frequencies that result from the KLM statistic are on the power envelope for all number of time periods. It shows that the KLM statistic is efficient when the autoregressive parameter equals one. Hence, since it is also efficient for other values of the autoregressive parameter, it is efficient in general.

The power envelopes that result for the AS and Sys moment conditions coincide. It shows that assuming mean stationarity to construct the moment conditions does not help to identify the autoregressive parameter. This results since the worst case specifications of the nuisance parameters accord with mean stationarity.

The paper is organized as follows. Section 2 introduces the linear dynamic panel data model and the different moment conditions used to identify its parameters. In Section 3 we show that the Dif and Lev moment conditions and the Sys moment conditions with three time periods do not identify the autoregressive parameter. The limiting distributions of several one and two step GMM estimators and corresponding Wald statistics are then non-standard. In Section 4 we use a representation theorem, akin to the cointegration representation theorem, see Engle and Granger (1987) and Johansen (1991), to show that the AS and Sys moment conditions identify the parameters when there are more than three time periods. In Section 5, we construct the power envelope. In Section 6 we determine which one, if any, of the GMM-AR, GMM-LM and KLM statistics lead to rejection frequencies that lie on the power envelope under worst case data generating processes. The final section concludes. Proofs of theorems and definitions of test statistics are provided in Appendices A and B respectively.

## 2 Moment conditions

We analyze the 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. 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:

$$\begin{aligned} E[u_{it}u_{it-j}] &= 0, & j = 1, \dots, t-1; t = 1, \dots, T, \\ E[u_{it}c_i] &= 0, & t = 1, \dots, T. \end{aligned} \quad (2)$$

Alongside the moment assumptions in (2), we do not impose any additional assumptions on the variances of the disturbances  $u_{it}$  and fixed effects  $c_i$  except for that they are finite. Under these assumptions, the moments of the  $T^2$  interactions of  $\Delta y_{it}$  and  $y_{it}$ :

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

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, like, for example, homoscedasticity or initial condition assumptions, to do so, 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>1</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), result 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. Ahn and Schmidt (1995) show that these (combined) 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. In the next Section we analyze the (non-) robustness with respect to initial conditions in more detail, and show that identification by Lev moment conditions is arbitrary in case of highly persistent panel data. The argument extends to the NL moment conditions and Sys moment conditions with  $T = 3$  so identification is then problematic as well.

---

<sup>1</sup>The Lev moment conditions in (5) are the  $T - 2$  moments who are not redundant when applied in system GMM. In principle we can base inference also on the complete set of  $\frac{1}{2}(T - 1)(T - 2)$  sample moments from the levels model only, i.e. extend the Lev moment conditions in (5) with 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 similar identification issues result.



### 3 Initial observations and identification when $T = 3$

The Dif, Lev and NL moment conditions that we use to identify  $\theta$  are semi-parametric with respect to the fixed effects, variances and initial observations so they identify  $\theta$  for a variety of different specifications of them. These specifications, however, still influence the identification of  $\theta$  for persistent values of it, *i.e.* values that are close to one. To exemplify this, we first consider the simplest setting which has  $T$  equal to three.<sup>2</sup>

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

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

with Jacobians:

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

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 (10) is equal to zero when  $\theta_0$  is equal to one.<sup>3</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

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

so the Lev moment conditions seem 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 and, despite that  $y_{i1}$  and  $u_{i2}$  are uncorrelated, we then do not know the value of  $E(y_{i1} u_{i2})$  which is an element of the Jacobian in (11). 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 disturbance of the initial observation under mean stationarity (6)-(7).

---

<sup>2</sup>We note that the NL (and hence AS) moment conditions are not defined for  $T = 3$ , hence we only analyze Dif, Lev and Sys moment conditions in this Section but our results extend likewise to the NL moment conditions for  $T = 4$  and larger.

<sup>3</sup>Exceptions are when mean stationarity (6)-(7) does not hold or, for example, in case of covariance stationarity so  $\text{var}(u_{i1}) = \frac{\sigma^2}{1-\theta_0^2}$ .

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

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

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

**Lev moment condition** 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:

$$\begin{aligned} \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N \Delta y_{i2}(y_{i3} - \theta y_{i2}) &= (1 - \theta) \left\{ \frac{1}{N} \sum_{i=1}^N u_{i2}^2 + \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N u_{i2} y_{i1} + \right. \\ &\quad \left. \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N (1 - \theta_0) u_{i1} y_{i1} \right\} \\ \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N y_{i2} \Delta y_{i2} &= \frac{1}{N} \sum_{i=1}^N u_{i2}^2 + \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N u_{i2} y_{i1} + \\ &\quad \lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N (1 - \theta_0) u_{i1} y_{i1}. \end{aligned} \quad (13)$$

We left out all elements in (13) that do not affect the large sample behavior when  $\theta_0$  goes to one. Since  $u_{i2}$  and  $y_{i1}$  are uncorrelated, for some function  $h(\theta_0)$  it holds that

$$\lim_{\theta_0 \uparrow 1} h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2} y_{i1} \xrightarrow{d} \psi_2, \quad (14)$$

with  $\psi_2$  a normal random variable with mean zero and variance  $\sigma_2^2 = \text{var}(u_{i2})$  so  $h(\theta_0)^{-2} = \text{var}(y_{i1})$ , which explains why  $\frac{1}{N} \sum_{i=1}^N u_{i2} y_{i1}$  appears in (13).

Similarly, under the moment conditions in (2) and mean stationarity (6)-(7),

$$\lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N (1 - \theta_0) u_{i1} y_{i1} = \lim_{\theta_0 \uparrow 1} E((1 - \theta_0) u_{i1}^2), \quad (15)$$

so, for example, if  $\text{var}(u_{i1}) = \frac{\sigma^2}{1 - \theta_0^2}$ ,<sup>4</sup> then

$$\lim_{\theta_0 \uparrow 1} \frac{1}{N} \sum_{i=1}^N (1 - \theta_0) u_{i1} y_{i1} = \frac{\sigma^2}{2}, \quad (16)$$

with  $\sigma^2$  a non-zero constant.

The rate at which the sample size goes to infinity relative to the convergence of  $h(\theta_0)$  to zero, or put differently the rate at which the variance of the initial observations

---

<sup>4</sup>This is the special case of covariance stationarity to which we do not confine ourselves.

goes to infinity, determines the behavior of the Jacobian of the Lev moment condition. For example, when

$$h(\theta_0)\sqrt{N} \xrightarrow{N \rightarrow \infty, \theta_0 \uparrow 1} \infty, \quad (17)$$

it holds that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N y_{i2} \Delta y_{i2} &= \frac{1}{N} \sum_{i=1}^N u_{i2}^2 + \frac{1}{h(\theta_0)\sqrt{N}} \left[ h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2} y_{i1} \right] + \\ &\xrightarrow[p]{N \rightarrow \infty, \theta_0 \uparrow 1} \sigma_2^2 + \lim_{\theta_0 \uparrow 1} E((1 - \theta_0) u_{i1}^2), \end{aligned} \quad (18)$$

while when

$$h(\theta_0)\sqrt{N} \xrightarrow{N \rightarrow \infty, \theta_0 \uparrow 1} 0, \quad (19)$$

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

$$\begin{aligned} h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta y_{i2} (y_{i3} - \theta y_{i2}) &= (1 - \theta) \left\{ h(\theta_0)\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N u_{i2}^2 \right] + \right. \\ &\quad \left. h(\theta_0)\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N (1 - \theta_0) u_{i1} y_{i1} \right] + \right. \\ &\quad \left. \left[ h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2} y_{i1} \right] \right\} \\ &\xrightarrow[d]{N \rightarrow \infty, \theta_0 \uparrow 1} (1 - \theta) \psi_2 \\ h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N y_{i2} \Delta y_{i2} &= h(\theta_0)\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N u_{i2}^2 \right] + \\ &\quad h(\theta_0)\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N (1 - \theta_0) u_{i1} y_{i1} \right] + \\ &\quad h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i2} y_{i1} \\ &\xrightarrow[d]{N \rightarrow \infty, \theta_0 \uparrow 1} \psi_2. \end{aligned} \quad (20)$$

It shows that when (19) holds that  $\theta$  is not identified when  $\theta_0$  is equal to one, since  $\psi_2$  is a mean zero random variable.

Since any assumption about the convergence rates of the sample size and the variance of the initial observations is arbitrary, also the identification of  $\theta$  by the Lev moment conditions is arbitrary for DGPs for which  $\theta_0$  is close to one and  $h(\theta_0)$  is equal to zero 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.**  $\sigma_c^2 = \text{var}(c_i)$ ,  $h(\theta_0) = (1 - \theta_0)/\sigma_c$ .

**DGP 2.**  $\sigma_c^2 = \text{var}(c_i)$ ,  $\sigma_1^2 = \frac{\sigma_c^2}{1 - \theta_0^2}$ ,  $h(\theta_0) = (1 - \theta_0)/\sigma_c$ .

**DGP 3.**  $\sigma_\mu^2 = \text{var}(\mu_i)$ ,  $\sigma_1^2 = \frac{\sigma^2}{1-\theta_0^2}$ ,  $h(\theta_0) = \left(\sqrt{1-\theta_0^2}\right) / \sigma_1$ .

**DGP 4.**  $\sigma_\mu^2 = \text{var}(\mu_i)$ ,  $\sigma_1^2 = \sigma^2 \frac{1-\theta_0^{2(g+1)}}{1-\theta_0^2}$ ,  $\lim_{\theta_0 \uparrow 1} h(\theta_0) = \frac{1}{\sigma_1} g^{-\frac{1}{2}}$ .

**DGP 5.**  $\sigma_c^2 = \text{var}(c_i)$ ,  $\sigma_1^2 = \sigma^2 \frac{1-\theta_0^{2(g+1)}}{1-\theta_0^2}$ ,  $\lim_{\theta_0 \uparrow 1} h(\theta_0) = (1-\theta_0) / \sigma_c$ .

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.

The convergence rate of  $\theta_0$  that results from (19) for DGPs 1-5 is such that:

$$\begin{aligned} \text{DGP 1, 2, 5 : } & (1-\theta_0)\sqrt{N} \xrightarrow[N \rightarrow \infty, \theta_0 \uparrow 1]{} 0 \quad \text{or} \quad \theta_0 = 1 - \frac{e}{N^{\frac{1}{2}(1+\epsilon)}} \\ \text{DGP 3 : } & (1-\theta_0^2)N \xrightarrow[N \rightarrow \infty, \theta_0 \uparrow 1]{} 0 \quad \text{or} \quad \theta_0 = 1 - \frac{e}{N^{1+\epsilon}} \\ \text{DGP 4 : } & \frac{N}{g} \xrightarrow[N \rightarrow \infty, g \rightarrow \infty]{} 0, \end{aligned} \tag{21}$$

with  $e$  a constant and  $\epsilon$  some real number larger than zero. In case of DGP 4, (21) 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. For example, the identification condition might hold but it can lead to large size distortions of test statistics as in case of weak instruments, see *e.g.* Staiger and Stock (1997).

**Dif moment condition** Under mean stationarity (6)-(7), the Dif moment conditions do in general not identify  $\theta$  when  $\theta_0$  is equal to one. Exceptions are DGPs for which the expectation in (15) is non-zero, like , for example, covariance stationarity. When the convergence of the sample size and the variance of the initial observations is, however, in line with (19), the large sample behavior for values of  $\theta_0$  close to one of the Dif

sample moment and its derivative are such that:

$$\begin{aligned} h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N y_{i1} (\Delta y_{i3} - \theta \Delta y_{i2}) &\xrightarrow[N \rightarrow \infty, \theta_0 \uparrow 1]{d} \psi_3 - \theta \psi_2 \\ -h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N y_{i1} \Delta y_{i2} &\xrightarrow[N \rightarrow \infty, \theta_0 \uparrow 1]{d} -\psi_2 \end{aligned} \quad (22)$$

with

$$\lim_{\theta_0 \uparrow 1} h(\theta_0) \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i3} y_{i1} \xrightarrow{d} \psi_3, \quad (23)$$

and  $\psi_3$  a normal random variable with mean zero and variance  $\sigma_3^2 = \text{var}(u_{i3})$ . Since  $\psi_2$  has mean zero, (22) shows that the Dif moment conditions do not identify  $\theta$  when  $\theta_0 = 1$  for all DGPs that accord with mean stationarity even those for which (15) is non-zero.

The above implies that  $\theta$  is not identified by the Dif and Lev moment conditions when  $\theta_0$  is equal to one so it is of interest to analyze if this extends to the Sys moment conditions which are a combination of the Dif and Lev moment conditions.

**Sys moment conditions** The Sys sample moments, reflected by  $f_N(\theta)$ , and their derivative, reflected by  $q_N(\theta)$ , when  $T = 3$  read:

$$f_N(\theta) = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} (\Delta y_{i3} - \theta \Delta y_{i2}) \\ \Delta y_{i2} (y_{i3} - \theta y_{i2}) \end{pmatrix}, \quad q_N(\theta) = -\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta y_{i2} \\ \Delta y_{i2} y_{i2} \end{pmatrix}. \quad (24)$$

In large samples and when  $\theta_0$  converges to one according to (19), the behavior of the Sys sample moments and its derivative is characterized by

$$\lim_{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \sqrt{N} h(\theta_0) \begin{pmatrix} f_N(\theta) \\ q_N(\theta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} -\theta & 1 \\ 1 - \theta & 0 \\ -1 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \psi_2 \\ \psi_3 \end{pmatrix}. \quad (25)$$

The result in (25) is implied by the expressions in (20) and (22). The large sample behavior of the Sys moment conditions in (25) does, since the means of  $\psi_2$  and  $\psi_3$  are equal to zero, not identify  $\theta$ . This shows that also for the Sys moment conditions the identification of  $\theta$  is arbitrary for unit values of  $\theta_0$  when  $T = 3$  since it depends on an high level assumption that concerns the convergence rates of the variance of the initial observations and the sample size.

The non-identification of  $\theta$  by its moment conditions for specific convergence sequences, that concern the variance of the initial observations, implies that the limit

behavior of estimators is non-standard. We state this limit behavior for the one and two step estimators that result for the Dif, Lev and Sys moment conditions when  $T = 3$  in Theorem 1. The two step estimator that results from the Sys moment conditions is computed using the usual Eicker-White covariance matrix estimator evaluated at the estimate from the first step, see White (1980). Since the number of Lev and Dif moment conditions equals the number of elements of  $\theta$  when  $T = 3$ , the GMM estimators based on these moment conditions do not depend on the covariance matrix estimator.

**Theorem 1.** *Under Assumption 1, the conditions in (2), mean stationarity (6)-(7), finite eighth moments of  $c_i$  and  $u_{it}$ ,  $i = 1, \dots, N$ ,  $t = 2, \dots, T$  and when (19) holds, the large sample behavior of the one and two step GMM estimators that result from the Dif, Lev and Sys moment conditions when  $T = 3$  read:*

$$\begin{aligned}
\hat{\theta}_{Dif} &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \frac{\psi_3}{\psi_2} = 1 + \frac{\psi_3 - \psi_2}{\psi_2} \\
h(\theta_0)^{-1}(\hat{\theta}_{Lev} - 1) &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \\
\hat{\theta}_{Sys,1step} &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} 1 + \frac{\psi_3 - \psi_2}{2\psi_2} \\
\hat{\theta}_{Sys,2step} &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} 1 - \frac{\psi_3^2 - \psi_3 \psi_2}{2\psi_2^2} \frac{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}
\end{aligned} \tag{26}$$

with

$$\begin{aligned}
h(\theta_0)^2 \left[ \frac{1}{N} \sum_{i=1}^N \left( y_{i1} \begin{pmatrix} \Delta y_{i2} \\ \Delta y_{i3} \end{pmatrix} - \begin{pmatrix} \overline{y_{i1} \Delta y_{i2}} \\ \overline{y_{i1} \Delta y_{i3}} \end{pmatrix} \right) \left( y_{i1} \begin{pmatrix} \Delta y_{i2} \\ \Delta y_{i3} \end{pmatrix} - \begin{pmatrix} \overline{y_{i1} \Delta y_{i2}} \\ \overline{y_{i1} \Delta y_{i3}} \end{pmatrix} \right) \right] \\
\qquad \qquad \qquad \xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} V_{y_1 \Delta y, y_1 \Delta y},
\end{aligned} \tag{27}$$

and  $V_{y_1 \Delta y, y_1 \Delta y} = \text{var}(y_{i1} u_{i2}, y_{i1} u_{i3})$ ,  $\overline{y_{i1} \Delta y_{i2}} = \frac{1}{N} \sum_{i=1}^N y_{i1} \Delta y_{i2}$ ,  $\overline{y_{i1} \Delta y_{i3}} = \frac{1}{N} \sum_{i=1}^N y_{i1} \Delta y_{i3}$ .

**Proof.** see Appendix A. ■

Theorem 1 shows that all GMM estimators have large sample distributions with non-standard convergence rates when the convergence accords with (19). The Dif and one and two step Sys estimators are all inconsistent under the convergence sequence in

(19) while the Lev estimator is consistent but with an unusual convergence rate since, under (19),  $h(\theta_0)^{-1}$  goes to infinity faster than  $\sqrt{N}$ . The large sample distributions of the estimators in Theorem 1 are all non-standard which implies that the large sample distributions of the Wald and/or  $t$ -statistics, whose definitions are stated in Appendix B, associated with them are non-standard as well.

The distribution of  $\hat{\theta}_{Dif}$  in Theorem 1 is identical to the distribution in Kruiniger (2009) and Madsen (2003). In Kruiniger (2009) also the large sample distributions of  $\hat{\theta}_{Lev}$  and  $\hat{\theta}_{Sys,2step}$  are constructed albeit using a different DGP for the initial observations. The qualitative conclusions from Kruiniger (2009) that  $\hat{\theta}_{Lev}$  is consistent but with a non-standard large sample distribution and that  $\hat{\theta}_{Sys,2step}$  is inconsistent and converges to a random variable result as well from Theorem 1. These non-standard results for two-step GMM coefficient estimators carry over to their corresponding Wald or  $t$ -statistics as stated in Theorem 2.

**Theorem 2.** *Under Assumption 1, the conditions in (2), mean stationarity (6)-(7), finite eighth moments of  $c_i$  and  $u_{it}$ ,  $i = 1, \dots, N$ ,  $t = 2, \dots, T$  and when (19) holds, the large sample distributions of the Wald statistics associated with the two step GMM estimators that result from the Dif, Lev and Sys moment conditions when  $T = 3$  read:*

$$\begin{aligned}
W_{Dif}(\theta) &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \psi_2^2 \frac{(\psi_3 - \theta \psi_2)^2}{\begin{pmatrix} \psi_3 \\ -\psi_2 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} \psi_3 \\ -\psi_2 \end{pmatrix}} \\
W_{Lev}(\theta) &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \left( \psi_{cu,1} + \psi_{cu,3} + (1 - \theta) \frac{\psi_2}{h(\theta_0)} \right)^2 \\
&\quad \left[ \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right]^2 V_{y_1 \Delta y, y_1 \Delta y, 11} + V_{cu,11} + V_{cu,33} \right]^{-1} \\
W_{Sys,2step}(\theta) &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \psi_2^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left( \frac{2\psi_2^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}{(\psi_3^2 - \psi_3 \psi_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}} \right)^2 \\
&\quad \left( 1 - \theta - \frac{\psi_3^2 - \psi_3 \psi_2}{2\psi_2^2} \frac{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}} \right)^2.
\end{aligned} \tag{28}$$

**Proof.** see Appendix A. ■

Unlike the large sample distributions of the Wald statistics in Theorem 2, the large sample distribution of the GMM-LM statistic proposed by Newey and West (1987), which is defined in Appendix B, remains standard  $\chi^2$  when the true value of  $\theta$  gets close to one. Because the moment conditions do not identify  $\theta$  when the true value of  $\theta$  gets close to one according to (19), it remains standard  $\chi^2$  for all tested values of  $\theta$  as stated in Theorem 3.

**Theorem 3.** *Under Assumption 1, the conditions in (2), mean stationarity (6)-(7), finite fourth moments of  $c_i$  and  $u_{it}$ ,  $i = 1, \dots, N$ ,  $t = 2, \dots, T$  and when (19) holds, the large sample distributions of the GMM-LM statistics of Newey and West (1987) testing  $H_0 : \theta = \theta$  that result from the Dif, Lev and Sys moment conditions when  $T = 3$  read:*

$$\begin{aligned} \text{GMM-LM}_{Dif}(\theta) &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{\frac{\rightarrow}{d}} \chi^2(1) \\ \text{GMM-LM}_{Lev}(\theta) &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{\frac{\rightarrow}{d}} \chi^2(1) \\ \text{GMM-LM}_{Sys}(\theta) &\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{\frac{\rightarrow}{d}} \chi^2(1). \end{aligned} \tag{29}$$

**Proof.** see Appendix A. ■

The large sample distributions of the GMM-LM statistic in Theorem 3 properly reflect that  $\theta$  is not identified by the moment conditions when its true value is equal to one and the convergence is according to (19). Accordingly, the large sample distribution of the GMM-LM statistic is the same for all values of  $\theta$  since it is not-identified.

Alongside the Wald and GMM-LM statistics, we also use the identification robust KLM statistic proposed in Kleibergen (2002,2005) and the GMM Anderson-Rubin (GMM-AR) statistic, see Anderson and Rubin (1949) and Stock and Wright (2000). The expressions of these statistics are stated in Appendix B. When  $T$  equals 3, these statistics are both identical to the GMM-LM statistic when we use either the Dif or Lev moment conditions since  $\theta$  is exactly identified by the moment conditions. When we use the Sys moment conditions, so  $\theta$  is over-identified by the moment conditions, these statistics differ from the GMM-LM statistic. The large sample distribution of the KLM and AR statistics do not alter when the tested parameter becomes non-identified so the large sample distributions of them are standard  $\chi^2$  for all values of  $\theta$  under (19) albeit with different degrees of freedom. We further discuss these statistics in the next Sections.



**Simulation experiment** To illustrate the identification issues when  $T = 3$ , we generate observations from DGP 3 with  $T = 3$ ,  $N = 500$ ,  $\theta_0 = 0.95$  and  $\sigma_c^2 = 1$ ,  $\sigma_t^2 = 1$ ,  $t = 1, \dots, T$ . We use them to compute the distributions of the different estimators and test statistics stated in Theorems 1 and 2. These are shown in Panel 1 and Figure 2. Figure 1.1 in Panel 1 shows the distribution functions of the Dif, Lev, Sys one step and Sys two step estimators. Figure 1.2 shows the distributions functions of the t-statistics that result from the Dif, Lev and Sys two step estimators.

The distributions functions in Figure 1.1 show the inconsistency/non-normality of the Dif, Sys one step and Sys two step estimators. The distribution function of the Lev estimator hints its consistency since it lies close to the true value of 0.95 as stated in Theorem 1.

Panel 1. Distribution functions of estimators and t-statistics in DGP 3 with  $\theta = 0.95$ ,  $T = 3$ ,  $\sigma_c^2 = \sigma_t^2 = 1$ ,  $t = 1, \dots, 3$ ;  $N = 500$ . Dif (solid), Lev (dashed), Sys two step (dash-dot) and Sys one step (dotted).

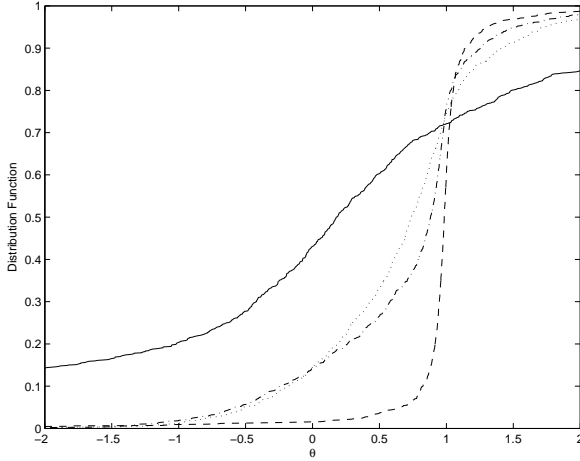


Figure 1.1. Estimators.

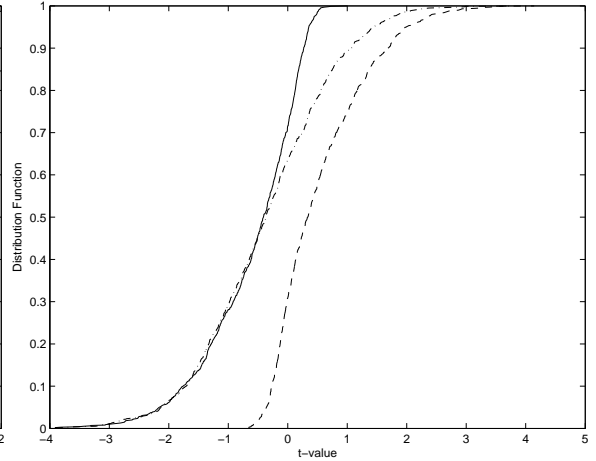
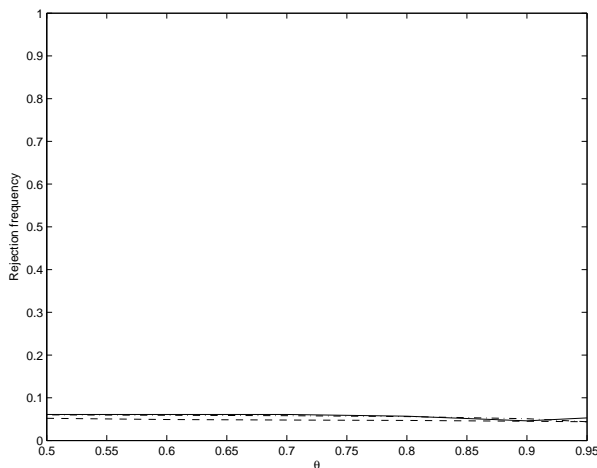


Figure 1.2. t-statistics.

The distributions functions in Figure 1.2 show that the Dif and Lev t-statistics are not normally distributed. The distribution function of the Sys two step t-statistic is, despite that the Sys two step estimator is not normally distributed, close to a normal distribution. This results from the expression of the two step Wald statistic in Theorem 2. If we use  $\theta = 0.95$ , the expressions in the large brackets roughly cancel out against each other which leaves us with the elements at the front of the expression. It results in a  $\chi^2(1)$  distribution of the Wald statistic so the t-statistic has a standard normal distribution. This results is, however, specific to  $T = 3$ .

Figure 2 uses the same DGP as used for Panel 1 but with a true value of  $\theta$  equal to one. It reports the rejection frequencies of 95% significance tests using the GMM-LM statistic with Dif, Lev and Sys moment conditions for testing for different values of  $\theta$  while the true value of  $\theta$  is equal to one. Figure 2 shows that the rejection frequency of all three GMM-LM statistics is equal to the size of the test for all values of  $\theta$  when the true value of  $\theta$  is equal to one which is in line with Theorem 3.

Figure 2. Rejection frequencies of 95% significance tests using GMM-LM statistic with Dif (solid), Lev (dashed) and Sys (dash-dot) moments, for different values of  $\theta$  while the true value is one using DGP 3,  $T = 3$ ,  $\sigma_c^2 = \sigma_t^2 = 1$ ,  $N = 500$ .



## 4 Identification when $T > 3$

We just showed that neither the Dif, Lev or Sys moment conditions identify  $\theta$  when  $\theta_0$  is close to one and  $T = 3$ . Since the NL, and hence AS, moment conditions are not defined for  $T = 3$ ,  $\theta$  is not identified by either one of the moment conditions when  $\theta_0 = 1$  and  $T = 3$ . When  $T = 4$  and under mean stationarity (6)-(7), the Jacobian of the NL moment condition is such that

$$\begin{aligned}
 \text{NL:} \quad & -\frac{\partial}{\partial \theta} E[(y_{it} - \theta y_{it-1})(\Delta y_{it-1} - \theta \Delta y_{it-2})] = \\
 & E[y_{it-1}(\Delta y_{it-1} - \theta \Delta y_{it-2}) + (y_{it} - \theta y_{it-1})\Delta y_{it-2}] = \\
 & E[u_{it-1}^2 + u_{it-2}^2 - 2\theta u_{it-2}^2] \neq 0 \quad \text{when } \theta_0 = 1,
 \end{aligned} \tag{30}$$

which again gives the impression that the NL moment conditions identify  $\theta$  irrespective of the value of  $\theta_0$ , see Ahn and Schmidt (1995) and Arellano and Alvarez (2004).<sup>5</sup>

<sup>5</sup>When  $\theta = 1$  and  $E(u_{it-1}^2) = E(u_{it-2}^2)$ , the Jacobian (30) is equal to zero which makes Arellano and Alvarez (2004) refer to this case as first order underidentification.

The same argument as for the Lev moment conditions applies, however, here as well since  $E[y_{ij}\Delta y_{it}]$ ,  $j, t = 1, \dots, T$ , is not defined under the convergence sequence in (19). To analyze the identification of  $\theta$  when  $T$  exceeds three, we therefore start out with a representation theorem. It states the behavior of the sample moments and their derivatives for the different moment conditions. When the mean stationarity conditions (6)-(7) do not hold, the Dif and NL moment conditions identify  $\theta$  at  $\theta_0 = 1$  and the Lev, and hence the Sys, moment conditions are violated. Hence, the representation theorem states the behavior of the different sample moments and their derivatives when the mean stationarity conditions (6)-(7) apply.

**Theorem 4 (Representation Theorem).** *Under Assumption 1, the conditions in (2), mean stationarity (6)-(7), finite fourth moments of  $c_i$  and  $u_{it}$ ,  $i = 1, \dots, N$ ,  $t = 2, \dots, T$ ,  $T > 3$ , we can characterize the large sample behavior of the Dif, Lev, NL, AS and Sys sample moments and their derivatives for values of  $\theta_0$  close to one by*

$$\begin{pmatrix} f_N^j(\theta) \\ q_N^j(\theta) \end{pmatrix} \approx \begin{pmatrix} A_f^j(\theta) \\ A_q^j(\theta) \end{pmatrix} \left[ \frac{1}{h(\theta_0)\sqrt{N}}\psi + \iota (\lim_{\theta_0 \uparrow 1} E((\theta_0 - 1)u_{i1}^2)) \right] + \begin{pmatrix} \mu_f^j(\theta, \sigma^2) \\ \mu_q^j(\theta, \sigma^2) \end{pmatrix} + \frac{1}{\sqrt{N}} \begin{pmatrix} B_f^j(\theta) \\ B_q^j(\theta) \end{pmatrix} \psi_{cu}, \quad (31)$$

with  $j = Dif, Lev, NL, AS, Sys$ . Furthermore,  $\mu_f^j(\theta, \sigma^2)$  and  $\mu_q^j(\theta, \sigma^2)$  are constants,  $\psi$  and  $\psi_{cu}$  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  $\psi_{cu}$  for values of  $T$  equal to 4 and 5 are stated in Appendix A.

**Proof.** see Appendix A. ■

The representation theorem in Theorem 4 is reminiscent of the cointegration representation theorem, see *e.g.* Engle and Granger (1987) and Johansen (1991). Identical to that representation theorem, Theorem 4 shows that the behavior of the moment series changes over different directions. To show what this implies for the different moment conditions discussed previously, we briefly discuss their resulting large sample properties.

**Dif and Lev conditions** When  $T = 4$ , 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 4 in

Appendix A, are:

$$\begin{aligned}
\text{Dif: } & \begin{pmatrix} \mu_f^{Dif}(\theta, \sigma^2) \\ \mu_q^{Dif}(\theta, \sigma^2) \end{pmatrix} = (0 \dots 0)', \quad A_f^{Dif}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \end{pmatrix}, \quad A_q^{Dif}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \\
\text{Lev: } & \begin{pmatrix} \mu_f^{Lev}(\theta, \sigma^2) \\ \mu_q^{Lev}(\theta, \sigma^2) \end{pmatrix} = \left( \begin{pmatrix} 1 - \theta \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \sigma_2^2 \\ \sigma_3^2 \end{pmatrix} \right), \quad A_f^{Lev}(\theta) = \begin{pmatrix} 1 - \theta & 0 \\ 0 & 1 - \theta \end{pmatrix}, \\
& A_q^{Lev}(\theta) = I_2.
\end{aligned} \tag{32}$$

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  $\frac{1}{h(\theta_0)\sqrt{N}}\psi$ , which under (19) goes to infinity when the sample size increases. The only identifying information for  $\theta$  therefore results from that part of the sample moment which is independent of  $\psi$ . Since  $\psi$  only affects the part of the sample moments spanned by  $A_f^j(\theta)$ , they are independent of  $\psi$  in the direction of the orthogonal complement of  $A_f^j(\theta)$ . The expressions of  $A_f^{Dif}(\theta)$  and  $A_f^{Lev}(\theta)$  in (32) are both square matrices so their orthogonal complements do not exist. It implies that the Dif and Lev moment conditions do not identify  $\theta$  when  $T = 4$ .

When  $T$  exceeds 4, the expressions of  $A_f^{Lev}(\theta)$  remain square matrices<sup>6</sup> so the Lev moment conditions also do not identify  $\theta$  for larger numbers of time series observations. The expression of  $A_f^{Dif}(\theta)$  is no longer a square matrix when  $T$  exceeds 4 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  $\theta$  when  $T$  exceeds 4. Summarizing we have:

$$\begin{aligned}
\text{Dif, } T = 4: & \quad A_f^{Dif}(\theta)_\perp \text{ does not exist.} \quad \text{No identification when } T = 4. \\
\text{Dif, } T > 4: & \quad A_f^{Dif}(\theta)'_\perp \mu_f^{Dif}(\theta, \sigma^2) = 0. \quad \text{No identification when } T > 4. \\
\text{Lev: } & \quad A_f^{Lev}(\theta)_\perp \text{ does not exist.} \quad \text{No identification when } T \geq 4.
\end{aligned} \tag{33}$$

**NL condition** For the NL moment 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  $T = 4$  read

$$\text{NL: } \begin{pmatrix} \mu_f^{NL}(\theta, \sigma^2) \\ \mu_q^{NL}(\theta, \sigma^2) \end{pmatrix} = \begin{pmatrix} (1-\theta)(\sigma_3^2 - \theta\sigma_2^2) \\ 2(\theta-1)\sigma_2^2 - \sigma_3^2 \end{pmatrix}, \quad \begin{pmatrix} A_f^{NL}(\theta) \\ A_q^{NL}(\theta) \end{pmatrix} = \begin{pmatrix} \theta(\theta-1) & 1-\theta \\ 2(\theta-1)+1 & -1 \end{pmatrix}. \tag{34}$$

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 also the NL moment condition

---

<sup>6</sup>We refer to the proof of Theorem 4 for the expressions of  $A_f^{Dif}(\theta)$  and  $A_f^{Lev}(\theta)$  when  $T = 5$ .

does not identify  $\theta$ .

The expression of  $A_f^{NL}(\theta)$  for larger number of time series observations<sup>7</sup> are also such that their ranks exceed 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  $(\mu_f^j(\theta, \sigma^2))$  and  $(A_f^j(\theta))$  when  $T = 4$  for the AS and Sys moment conditions result from stacking those of the Dif and NL and Dif and Lev moment conditions resp.:

$$\begin{aligned}
\text{AS: } \mu_f^{AS}(\theta, \sigma^2) &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ (1-\theta)(\sigma_3^2 - \theta\sigma_2^2) \end{pmatrix}, \quad A_f^{AS}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \\ \theta(\theta-1) & 1-\theta & 0 \end{pmatrix}, \\
\mu_q^{AS}(\theta, \sigma^2) &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2(\theta-1)\sigma_2^2 - \sigma_3^2 \end{pmatrix}, \quad A_q^{Dif}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 2(\theta-1)+1 & -1 & 0 \end{pmatrix}. \\
\text{Sys: } \mu_f^{Sys}(\theta, \sigma^2) &= (1-\theta) \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ \sigma_3^2 \end{pmatrix}, \quad A_f^{Sys}(\theta) = \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \\ 1-\theta & 0 & 0 \\ 0 & 1-\theta & 0 \end{pmatrix}, \\
\mu_q^{Sys}(\theta, \sigma^2) &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ \sigma_3^2 \end{pmatrix}, \quad A_q^{Sys}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}.
\end{aligned} \tag{35}$$

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,  $\mu_f^j(\theta, \sigma^2)$  is non-zero. When we therefore pre-multiply the sample moments by the orthogonal complement of  $A_f^j(\theta)$ , we obtain

$$A_f^j(\theta)_\perp' f_N^j(\theta) \approx A_f^j(\theta)_\perp' \mu_f^j(\theta, \sigma^2) + \frac{1}{\sqrt{N}} A_f^j(\theta)_\perp' B_f^j(\theta) \psi_{cu}, \quad j = AS, Sys, \tag{36}$$

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

---

<sup>7</sup>The expression for  $T = 5$  is stated in the proof of Theorem 4 in Appendix A.

with the expression in Theorem 4 (31), the elements multiplied by  $A_f^j(\theta)$  have both dropped out since  $A_f^j(\theta)'_{\perp} A_f^j(\theta) = 0$ .

The specification of  $\mu_f^j(\theta, \sigma^2)$  for the AS and Sys moment conditions in (35) is such that  $A_f^j(\theta)'_{\perp} \mu_f^j(\theta, \sigma^2) \neq 0$ . It implies that although the AS and Sys sample moments diverge in the direction of  $A_f^j(\theta)$ , so that part cannot be used to identify  $\theta$ , 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 4 in Appendix A show that this argument extends to all values of  $T$  larger than three.

**Corollary 1 (Identification of  $\theta$ ).** *Under the assumptions of Theorem 4,  $\theta$  is identified by the AS and Sys moment conditions when  $T$  exceeds 3 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 larger numbers of time series observations. Hence,  $\theta$  is identified by the Sys moment conditions when there are more than three time periods. When there are three time periods, the number of elements of  $f_N^{Sys}(\theta)$  and  $\psi$  are the same (see the previous Section) so  $A_f^{Sys}(\theta)$  is a square matrix and its orthogonal complement is not defined.

Corollary 1 also shows that  $\theta$  is identified by the AS moment conditions. We used mean stationarity to construct the large sample behavior in Theorem 4 and Corollary 1. Unlike the Sys moment conditions, the AS moment conditions do, however, not need mean stationarity to hold. It shows that assuming mean stationarity for constructing moment conditions does not help to identify  $\theta$  when  $\theta_0 = 1$  since the same identification results are obtained from moment conditions that do not assume mean stationarity. When mean stationarity does not hold both the Dif and NL, and consequently the AS, moment conditions identify  $\theta$  when  $\theta_0 = 1$ .

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

1. The convergence rate accords with (17) so  $\psi$  vanishes from the large sample behavior of the moment conditions in Theorem 4. It leads to standard behavior of one and two step GMM estimators.

2. The convergence rate accords with (19). Only the part of the moment conditions in the direction of  $A_f^j(\theta)_{\perp}$  (36) now identifies  $\theta$ . One step and two step GMM estimators,

however, use both the part of the sample moment that lies in the direction of  $A_f^j(\theta)$ , which diverges, and the part which lies in the direction of its orthogonal complement, which identifies  $\theta$ . Usage of the first part results in an inconsistency so one and two step GMM estimators are inconsistent and have non-standard limiting distributions. To exemplify this, Corollary 2 states the limiting distribution of the one step estimator based on the Sys moment conditions which results in a straightforward manner from Theorem 4 since the Sys moment conditions are linear in  $\theta$ .

**Corollary 2.** *Under the assumptions of Theorem 4, the limiting behavior of the one step estimator for the Sys moment conditions is characterized by*

$$\hat{\theta}_{1s}^{Sys} \xrightarrow{d} 1 - (\psi' A_q^{Sys}(1)' A_q^{Sys}(1) \psi)^{-1} \psi' A_q^{Sys}(1)' A_f^{Sys}(1) \psi, \quad (37)$$

*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 also shows that the limiting distribution of the one step estimator is non-standard. Similar results hold for the one step GMM estimator based on the AS moment conditions and the two step GMM estimator based on either the AS or Sys moment conditions. These are more involved to obtain since the AS moment conditions are a quadratic function of  $\theta$  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.

Corollary 2 shows that the identification of  $\theta$  by the AS and Sys moment conditions when  $T$  exceeds three does not lead to standard behavior of one and two step GMM estimators. Conducting inference based on these estimators is therefore hard when  $\theta_0$  is close to one. The GMM-LM, KLM and GMM-AR statistics are, however, size correct for such values of  $\theta_0$ . Since  $\theta$  is identified by the AS and Sys moment conditions, they also have discriminatory power. For smaller values of  $\theta_0$ , for which no identification issues exist, both the GMM-LM and KLM statistics are efficient and more powerful than the GMM-AR statistic. This standard notion of efficiency does, however, not apply to values of  $\theta_0$  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 construct the power envelope that results from the AS and Sys moment conditions for testing values of  $\theta$  while its true value is close to one. To construct the power envelope, we use the identifying part of the sample moments (36).

## 5 Moment conditions robust to initial observations

The orthogonal complements with respect to  $A_f^j(\theta)$  of the AS and Sys sample moments identify  $\theta$  when  $T$  is larger than three. Expressions of the orthogonal complements of  $A_f^j(\theta)$  for  $T = 4$  and 5 for the AS and Sys moment conditions are stated in Appendix A. They are such that the rotated AS and Sys moment conditions are quadratic in  $\theta$  :

$$g_{f,T}(\theta) = A_f(\theta)_{\perp}^j f_N(\theta) = a\theta^2 + b\theta + d, \quad (38)$$

with for

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

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

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

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

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

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

Under the limiting sequence in (19), the moment condition in (38) is the only part of the AS and Sys moment conditions that identifies  $\theta$ . The limiting sequence in (19)



therefore characterizes the worst case for identifying  $\theta$ . Since the moment condition (38) is the only part that identifies  $\theta$  in this worst case setting, the power envelope for testing hypotheses on  $\theta$ , which reflects the largest rejection frequency under the worst possible data generating processes, results from it. Hence, we use  $g_{f,T}(\theta)$  to construct the power envelope for testing hypotheses on  $\theta$  while its true value is equal to one. In order to do so, we first determine the slowest rate at which the hypothesized value of  $\theta$  can drift away from one, while the true value of  $\theta$  equals one, such that the sample moment (38) converges to a random variable that is non-degenerate and remains finite with probability one. To determine this rate, we first state the probability limits of  $a$ ,  $b$  and  $d$  when the true value of  $\theta$  is one.

**Theorem 5.** *Under Assumption 1, 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  $g_T(\theta)$  when  $\theta_0$  is equal to one is characterized by:*

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

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

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

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

**Proof.** see Appendix A. ■

We use the probability limits in Theorem 5 to determine the convergence rate for our local to unity asymptotics that we employ to construct the power envelope.

**Theorem 6.** *Under the conditions of Theorem 5, the drifting sequence for  $\theta$  for the robust moments  $g_T(\theta)$  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  $e$  a finite constant.*

**Proof.** see Appendix A. ■

The quartic root convergence rate in Theorem 6 results since the moment equation (38) is quadratic in  $\theta$ . When we specify  $\theta$  as  $1 + \frac{e}{\sqrt[4]{N}}$  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 makes all these components of the same order of magnitude.

Instead of using the moment equation in (38) for testing hypotheses on  $\theta$ , it can also be used to estimate  $\theta$ . The estimator that results from it, is, however, a worst case estimator since it only does relatively well under worst case DGPs. For DGPs with values of  $\theta$  less than one, estimators based on the other moment conditions outperform it. Alongside this suboptimality also its large sample distribution is, identical to the other estimators, non-standard. This holds since the expected value of the discriminant of the quadratic equation (38) is equal to zero under the worst case DGPs, *i.e.* case 1 in Theorem 6. The estimator has then a quartic root convergence rate and a non-standard large sample distribution. For DGPs covered by case 2 in Theorem 6 and when the true value of  $\theta$  is less than one, the large sample distribution of the estimator is normal with a square root convergence rate.

Theorem 6 shows that mean stationarity (6)-(7), under which  $\omega = 0$ , and mean variances that are constant over time lead to the slowest converge rate for  $\theta$ . It implies that jointly with (19) this setting provides the worst case data generating process under  $\theta_0 = 1$ . We therefore use it to establish the power envelope.

To construct the power envelope, we first obtain an approximation of the finite sample distribution of the GMM-AR statistic which tests  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  just using the moments in (38) while the true value of  $\theta$  is equal to one:

$$\text{GMM-AR}(e) = N g_T(e)' \hat{V}_{gg}(e)^{-1} g_T(e), \quad (39)$$

with  $g_T(e)$  the moments in (38) 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  $g_T(e)$ .

**Theorem 7.** *Under the conditions of Theorem 5 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 (39) for testing the hypothesis  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ , is characterized by

$$\chi^2(\delta, p_{\max}), \quad (40)$$

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

$$B(N) = (\iota_3 \otimes I_{p_{\max}}) + \frac{e}{\sqrt[4]{N}} \left[ \left(2 + \frac{e}{\sqrt[4]{N}}\right) (e_{1,3} \otimes I_{p_{\max}}) + \frac{e}{\sqrt[4]{N}} (e_{2,3} \otimes I_{p_{\max}}) \right], \quad (41)$$

$V_{abd}$  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 7 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.

The moment conditions in  $g_T(e)$  over-identify  $\theta$ . We therefore obtain the power envelope from the weighted average of the individual moments in  $g_T(e)$  that leads to the largest rejection frequency when testing  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$ .

**Theorem 8.** Under the conditions of Theorem 5 and when the true value of  $\theta$  is equal to one, the power envelope for testing  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  is

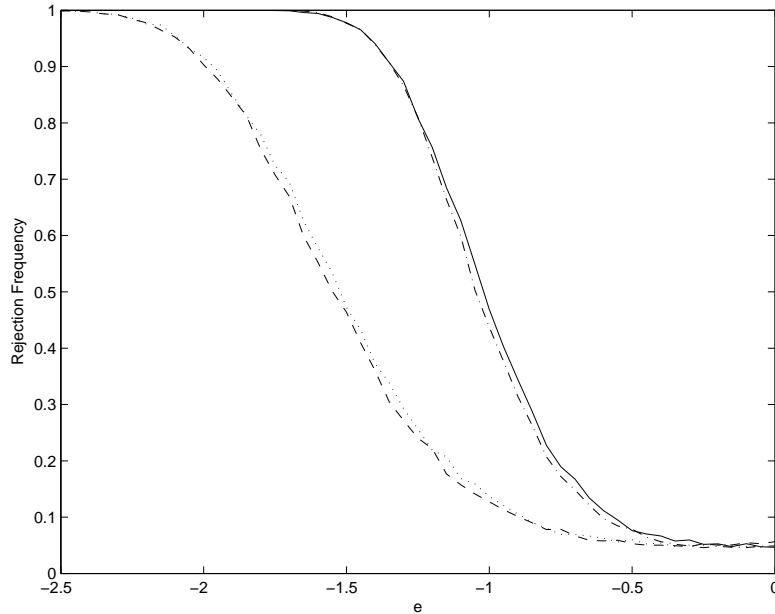
$$\chi^2((e\sigma)^4 (\iota_p)' (B(N)' V_{abd} B(N))^{-1} (\iota_p), 1), \quad (42)$$

with  $\iota_p$  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 3 shows the power envelopes that result from the AS and Sys moment conditions when  $T = 4$  and 5. Figure 3 shows that the power envelopes that result from the AS and Sys moment conditions are identical which is surprising. Figure 3 also shows that the power envelopes that result for a larger number of time series observations dominate those that result for smaller number of time series observations.

Figure 3. Power envelope for testing  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  at the 95% significance level and the true value of  $\theta$  is one:  $T = 4$ ,  $N = 500$  : Sys (dashed), AS (dotted);  $T = 5$  : Sys (solid), AS (dash-dotted).



## 6 Which statistic has rejection frequencies on the power envelope

Theorems 1, 4 and the convergence rate of the power envelope stated in Theorem 6 imply that the limiting behavior of estimators is not uniform since the limiting behavior of estimators depends on the data generating process at hand. Wald statistics based on them 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. Since these statistics are size correct, the recommended statistic to use 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 determine which one, if any, of the GMM-AR<sup>8</sup>, GMM-LM and KLM statistics leads to rejection frequencies that lie on the power envelope under the worst case data generating process which accords with (19), mean stationarity and  $\sigma_t^2 = \sigma^2$ ,  $t = 2, \dots, T$ .

**Theorem 9.** *Under the conditions from Theorem 5, the worst case large sample distributions, which apply under (19), 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  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  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,

$$\begin{aligned} \delta_{\text{GMM-AR}} &= (e\sigma)^4 \binom{t_p}{0}' (B(N)' V_{abd} B(N))^{-1} \binom{t_p}{0} \\ \delta_{\text{KLM}} &= \delta_{\text{GMM-AR}} \\ \delta_{\text{GMM-LM}} &= (e\sigma)^4 \binom{t_p}{0}' (B(N)' V_{abd} B(N))^{-\frac{1}{2}} P_{(B(N)' V_{abd} B(N))^{-\frac{1}{2}} (G_f(e)' A_q \psi)} \\ &= \delta_{\text{GMM-AR}} \quad \begin{array}{l} T = 5, \dots \\ T = 4, \end{array} \end{aligned} \quad (44)$$

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

$$\lim_{\theta_0 \uparrow 1, \frac{h(\theta_0)}{p} \sqrt{N} \rightarrow 0} \text{var} \left( h(\theta_0) \begin{pmatrix} y_{1i} u_{i2} \\ \vdots \\ y_{1i} u_{iT} \end{pmatrix} \right). \quad (45)$$

**Proof.** see Appendix A. 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  $(G_f(e)' A_q \psi) = \binom{t_p}{0} G_f(e)' A_q \psi$  and  $G_f(e)' A_q \psi$  cancels out of the expression of the non-centrality parameter. ■

---

<sup>8</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 (39).

Panel 4. Rejection frequencies under worst case scenario and power envelope.  
 Sys moment conditions: KLM statistic (dashed), GMM-AR (solid with plusses),  
 GMM-LM (dash-dotted), power envelope (solid). AS moment conditions (dotted lines).  
 Dif moment conditions: GMM-AR (solid with diamonds).  $N = 2000$ .

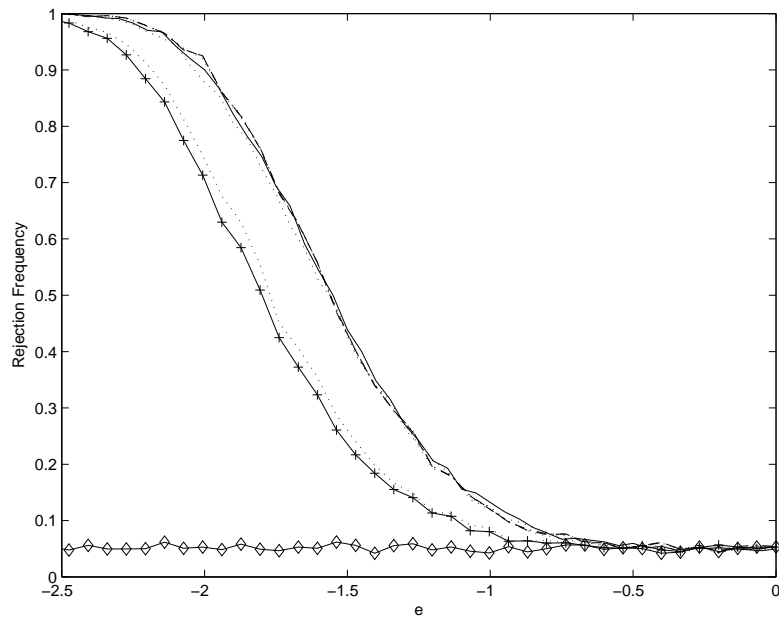


Figure 4.1.  $T = 4$

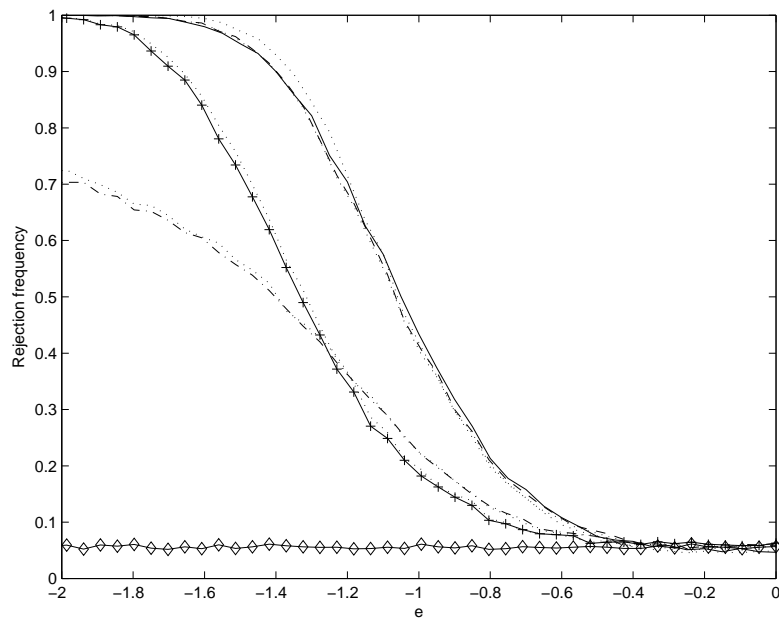


Figure 4.2.  $T = 5$

Theorem 9 shows that the KLM statistic attains the power envelope under the worst case DGPs and the GMM-LM statistic does so only 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.

Panel 4 shows the rejection frequencies 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. The worst case DGP that we use results from DGP 1 in Section 3 with a large value of  $\sigma_c^2$  (ten) compared to  $\sigma_t^2$ ,  $t = 1, \dots, T$  (one). Since the AS and 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, the figures in Panel 4 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 non-linear part of the AS moment conditions or instead of the GMM-AR statistic use the GMM-LM or KLM statistic.

Figure 4.1 in Panel 4 shows that, when  $T = 4$ , the rejection frequencies of the KLM and GMM-LM statistics are on the power envelope when we use the Sys or AS moment conditions as stated in Theorem 9. The rejection frequencies of the GMM-AR statistic are below the power envelope. Figure 4.1 also shows that the rejection frequencies of the GMM-AR statistic which uses the AS moment conditions are slightly above those that result from the GMM-AR statistic that uses the Sys moment conditions. This results since, as stated in Theorem 9, 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 4.2 shows that, when  $T = 5$ , the rejection frequencies that result from using the KLM statistic with either the AS or Sys moment conditions are on the power envelope. Figure 4.2 also shows that the rejection frequencies which result from the GMM-LM and GMM-AR statistics are below the power envelope which is in line with Theorem 9 since the simplification of the worst case large sample distribution of the GMM-LM statistic only applies to  $T = 4$ . The rejection frequencies that result from using either the AS or Sys moment conditions are the same for the KLM and

GMM-LM statistics while for those for 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 using the AS moment conditions compared to the Sys moment conditions while they have identical non-centrality parameters.

## 7 Conclusions

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

Despite the identification from four time periods onwards, GMM estimators based on the AS and Sys moment conditions behave in a non-standard manner. Inference based on these estimators using, for example, Wald statistics is then difficult. We therefore use size correct GMM statistics, like, the GMM-AR, GMM-LM or KLM statistic, to conduct inference. To recommend which one to use, we compare their large sample distributions under worst case DGPs with a value of the autoregressive parameter equal to one with the power envelope. The resulting rejection frequencies of the GMM-AR statistic for these worst case DGPs are below the power envelope whilst the rejection frequencies of the GMM-LM statistic are on the power envelope when there are four time periods and below it for more time periods. The rejection frequencies of the KLM statistic are on the power envelope for all number of time periods. This makes it our recommended statistic since it is also efficient for smaller values of the autoregressive parameter.

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

For expository purposes, we 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.** When  $T = 3$ , we can specify the large sample behavior of the Sys moment conditions by

$$\begin{aligned} f_N^{Sys}(\theta) &= \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1}(\Delta y_{i3} - \theta \Delta y_{i2}) \\ \Delta y_{i2}(y_{i3} - \theta y_{i2}) \end{pmatrix} \\ &\approx \mu^{Sys}(\sigma^2) + A_f^{Sys}(\theta) \left( \frac{1}{h(\theta_0)\sqrt{N}} \psi + \iota_2 E(\lim_{\theta_0 \uparrow 1} (1 - \theta_0) u_{i1}^2) \right) + \frac{1}{\sqrt{N}} B_f^{Sys}(\theta) \psi_{cu}, \end{aligned}$$

with

$$\begin{aligned} \mu^{Sys}(\sigma^2) &= \begin{pmatrix} 0 \\ \sigma_2^2 \end{pmatrix}, \quad A_f^{Sys}(\theta) = \begin{pmatrix} -\theta & 1 \\ 1 - \theta & 0 \end{pmatrix}, \quad B_f^{Sys}(\theta) = \begin{pmatrix} 0 & 0 & 0 \\ 1 - \theta & 1 - \theta & 1 \end{pmatrix} \\ \frac{h(\theta_0)}{\sqrt{N}} \sum_{i=1}^N \begin{pmatrix} y_{i1} u_{i2} \\ y_{i1} u_{i3} \end{pmatrix} &\xrightarrow{d} \psi = \begin{pmatrix} \psi_2 \\ \psi_3 \end{pmatrix}, \quad \frac{1}{\sqrt{N}} \sum_{i=1}^N \begin{pmatrix} c_i u_{i2} \\ u_{i2}^2 - \sigma_2^2 \\ u_{i2} u_{i3} \end{pmatrix} \xrightarrow{d} \psi_{cu} = \begin{pmatrix} \psi_{cu,1} \\ \psi_{cu,2} \\ \psi_{cu,3} \end{pmatrix} \end{aligned}$$

and  $\psi$  and  $\psi_{cu}$  are normally distributed random variables. Under the limit behavior in (19), we can then characterize the large sample behavior of the Dif, Lev and one step Sys estimators by

$$\begin{aligned} \hat{\theta}_{Dif} &= \frac{\frac{1}{N} \sum_{i=1}^N y_{i1} \Delta y_{i3}}{\frac{1}{N} \sum_{i=1}^N y_{i1} \Delta y_{i2}} \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \frac{\psi_3}{\psi_2} = 1 + \frac{\psi_3 - \psi_2}{\psi_2} \\ \hat{\theta}_{Lev} &= \frac{\frac{1}{N} \sum_{i=1}^N y_{i3} \Delta y_{i2}}{\frac{1}{N} \sum_{i=1}^N y_{i2} \Delta y_{i2}} = \theta_0 + \frac{\frac{1}{N} \sum_{i=1}^N (c_i + u_{i3}) \Delta y_{i2}}{\frac{1}{N} \sum_{i=1}^N y_{i2} \Delta y_{i2}} \\ h(\theta_0)(\hat{\theta}_{Lev} - \theta_0) &\xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \end{aligned}$$

and

$$\begin{aligned} \hat{\theta}_{Sys,1step} &= \frac{\left( \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta y_{i2} \\ y_{i2} \Delta y_{i2} \end{pmatrix} \right)' \left( \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta y_{i3} \\ y_{i3} \Delta y_{i2} \end{pmatrix} \right)}{\left( \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta y_{i2} \\ y_{i2} \Delta y_{i2} \end{pmatrix} \right)' \left( \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta y_{i2} \\ y_{i2} \Delta y_{i2} \end{pmatrix} \right)} \\ &\xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \frac{\psi_2^2 + \psi_3 \psi_2}{2\psi_2^2} = 1 + \frac{\psi_3 - \psi_2}{2\psi_2} \end{aligned}$$

For the two step Sys estimator, we first need to characterize the behavior of the Eiker-White covariance estimator  $\hat{V}_{ff}(\theta)$ :

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

which reads

$$h(\theta_0)^2 \hat{V}_{ff}(\theta) \xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} A_f^{Sys}(\theta) V_{y_1 \Delta y, y_1 \Delta y} A_f^{Sys}(\theta)'$$

with

$$h(\theta_0)^2 \left[ \frac{1}{N} \sum_{i=1}^N \left( y_{i1} \begin{pmatrix} \Delta y_{i2} \\ \Delta y_{i3} \end{pmatrix} - \begin{pmatrix} \overline{y_{i1} \Delta y_{i2}} \\ \overline{y_{i1} \Delta y_{i3}} \end{pmatrix} \right) \left( y_{i1} \begin{pmatrix} \Delta y_{i2} \\ \Delta y_{i3} \end{pmatrix} - \begin{pmatrix} \overline{y_{i1} \Delta y_{i2}} \\ \overline{y_{i1} \Delta y_{i3}} \end{pmatrix} \right)' \right] \xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} V_{y_1 \Delta y, y_1 \Delta y},$$

and where  $\overline{y_{i1} \Delta y_{i2}} = \frac{1}{N} \sum_{i=1}^N y_{i1} \Delta y_{i2}$ ,  $\overline{y_{i1} \Delta y_{i3}} = \frac{1}{N} \sum_{i=1}^N y_{i1} \Delta y_{i3}$ . The limit behavior of the two step Sys estimator then results as

$$\begin{aligned} & \hat{\theta}_{Sys, 2Step} - \theta_0 \\ &= \left[ \left( \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta y_{i2} \\ y_{i2} \Delta y_{i2} \end{pmatrix} \right)' \hat{V}_{ff}(\hat{\theta}_{Sys, 1step})^{-1} \left( \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta y_{i2} \\ y_{i2} \Delta y_{i2} \end{pmatrix} \right) \right]^{-1} \\ & \quad \left[ \left( \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta y_{i2} \\ y_{i2} \Delta y_{i2} \end{pmatrix} \right)' \hat{V}_{ff}(\hat{\theta}_{Sys, 1step})^{-1} \left( \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i1} \Delta u_{i3} \\ (c_i + u_{i3}) \Delta y_{i2} \end{pmatrix} \right) \right] \\ & \xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \left[ \psi_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}' A_f^{Sys}(\hat{\theta}_{Sys, 1step})^{-1} V_{y_1 \Delta y, y_1 \Delta y}^{-1} A_f^{Sys}(\hat{\theta}_{Sys, 1step})^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \psi_2 \right]^{-1} \\ & \quad \psi_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}' A_f^{Sys}(\hat{\theta}_{Sys, 1step})^{-1} V_{y_1 \Delta y, y_1 \Delta y}^{-1} A_f^{Sys}(\hat{\theta}_{Sys, 1step})^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \psi_3 \\ &= (1 - \hat{\theta}_{Sys, 1step}) \frac{\psi_3}{\psi_2} \left[ \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= -\frac{\psi_3^2 - \psi_3 \psi_2}{2\psi_2^2} \frac{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}} \end{aligned}$$

where we used that  $A_f^{Sys}(\theta)^{-1} = \frac{1}{1-\theta} \begin{pmatrix} 0 \\ 1-\theta \\ \theta \end{pmatrix}$  so  $A_f^{Sys}(\theta)^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{1-\theta} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  and  $A_f^{Sys}(\theta)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ .

**Proof of Theorem 2.** The two step Wald statistic is defined as:

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

For the Dif estimator, we can characterize the limit behavior of  $Nq_N(\hat{\theta}_{2s})'\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}q_N(\hat{\theta}_{2s})$  under (19) by

$$\begin{aligned} Nq_N(\hat{\theta}_{2s})'\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}q_N(\hat{\theta}_{2s}) &\xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \psi_2 \left[ \begin{pmatrix} -\hat{\theta}_{2s} \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} -\hat{\theta}_{2s} \\ 1 \end{pmatrix} \right]^{-1} \psi_2 \\ &= \psi_2^4 \left[ \begin{pmatrix} -\psi_3 \\ \psi_2 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} -\psi_3 \\ \psi_2 \end{pmatrix} \right]^{-1} \end{aligned}$$

so

$$\begin{aligned} W_{Dif}(\theta) &\xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \left( \frac{\psi_3 - \theta \psi_2}{\psi_2} \right)^2 \frac{\psi_2^4}{\begin{pmatrix} \psi_2 \\ -\psi_3 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} \psi_2 \\ -\psi_3 \end{pmatrix}} \\ &= (\psi_3 - \theta \psi_2)^2 \frac{\psi_2^2}{\begin{pmatrix} \psi_2 \\ -\psi_3 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} \psi_2 \\ -\psi_3 \end{pmatrix}}. \end{aligned}$$

For the Lev estimator, we use that  $\hat{\theta}_{Lev} \approx 1 + h(\theta_0) \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right]$  so  $\hat{\theta}_{Lev} - \theta \approx 1 - \theta + h(\theta_0) \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right]$  and

$$\begin{aligned} \Delta y_{i2}(y_{i3} - \hat{\theta}_{Lev} y_{i2}) &\approx \Delta y_{i2} \left( (1 - \theta - h(\theta_0) \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right]) y_{i2} + c_i + u_{i3} \right) \\ &= \Delta y_{i2} \left( h(\theta_0) \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right] (y_{i1} + c_i + u_{i2}) + c_i + u_{i3} \right) \end{aligned}$$

which implies that

$$\hat{V}_{ff}(\hat{\theta}_{Lev}) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right]^2 V_{y_1 \Delta y, y_1 \Delta y, 11} + V_{cu,11} + V_{cu,33},$$

with  $V_{cu,11}$  and  $V_{cu,33}$  the first and third diagonal elements of the covariance matrix of  $(c_i \Delta y_{i2} : u_{i2}^2 - \sigma_2^2 : u_{i2} u_{i3})'$  and  $V_{y_1 \Delta y, y_1 \Delta y, 11}$  the first diagonal element of  $V_{y_1 \Delta y, y_1 \Delta y}$ . The large sample behavior of the Wald statistic using the Lev moment conditions then results as

$$\begin{aligned} W_{Lev}(\theta) &\xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \left( 1 - \theta + h(\theta_0) \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right] \right)^2 \left( \frac{\psi_2}{h(\theta_0)} \right)^2 \left[ \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right]^2 V_{y_1 \Delta y, y_1 \Delta y, 11} + V_{cu,11} + V_{cu,33} \right]^{-1} \\ &= \left( \psi_{cu,1} + \psi_{cu,3} + (1 - \theta) \frac{\psi_2}{h(\theta_0)} \right)^2 \left[ \left[ \frac{\psi_{cu,1} + \psi_{cu,3}}{\psi_2} \right]^2 V_{y_1 \Delta y, y_1 \Delta y, 11} + V_{cu,11} + V_{cu,33} \right]^{-1}. \end{aligned}$$

The large sample behavior of  $\hat{\theta}_{Sys, 2step} - \theta$ :

$$\begin{aligned} \hat{\theta}_{Sys, 2step} - \theta &= \hat{\theta}_{Sys, 2step} - 1 + 1 - \theta \\ &\xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} 1 - \theta - \frac{\psi_3^2 - \psi_3 \psi_2}{2\psi_2^2} \frac{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}} \end{aligned}$$

so since

$$q_N(\hat{\theta}_{Sys,2step})' \hat{V}_{ff}(\hat{\theta}_{Sys,2step})^{-1} q_N(\hat{\theta}_{Sys,2step}) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \frac{\psi_2^2}{(1-\hat{\theta}_{2s})^2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

we can characterize the limit behavior of the two step Sys Wald statistic by:

$$\begin{aligned} & W_{Sys,2step}(\theta) \\ &= (\hat{\theta}_{Sys,2step} - \theta)' \left[ q_N(\hat{\theta}_{Sys,2step})' \hat{V}_{ff}(\hat{\theta}_{Sys,2step})^{-1} q_N(\hat{\theta}_{Sys,2step}) \right] (\hat{\theta}_{Sys,2step} - \theta) \\ & \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \psi_2^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left( \frac{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}{(\psi_3^2 - \psi_3 \psi_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}} \right)^2 \\ & \quad \left( \frac{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}} \right)^2 \cdot \end{aligned}$$

**Proof of Theorem 3.** The large sample behavior of the GMM-LM statistic of Newey and West (1987) testing  $H_0 : \theta = \theta$ :

$$\begin{aligned} \text{GMM-LM}(\theta) &= N f_N(\theta)' \hat{V}_{ff}(\theta)^{-1} q_N(\theta) \left[ q_N(\theta)' \hat{V}_{ff}(\theta)^{-1} q_N(\theta) \right]^{-1} \\ & \quad q_N(\theta)' \hat{V}_{ff}(\theta)^{-1} f_N(\theta), \end{aligned}$$

while the true value  $\theta_0$  converges to one according to (19) results from its different components. For the Dif and Lev moment conditions, it reads:

$$\begin{aligned} & \text{GMM-LM}_{Dif}(\theta) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \\ & \psi' A_f(\theta)'_1 [A_f(\theta)_1 V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)'_1]^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}' \psi \left[ \psi' \begin{pmatrix} 1 \\ 0 \end{pmatrix}' [A_f(\theta)_1 V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)'_1]^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \psi \right]^{-1} \\ & [A_f(\theta)_1 V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)'_1]^{-1} A_f(\theta)_1 \psi \\ & = \psi' A_f(\theta)'_1 [A_f(\theta)_1 V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)'_1]^{-1} A_f(\theta)_1 \psi \sim \chi^2(1), \\ & \text{GMM-LM}_{Lev}(\theta) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \\ & \psi' A_f(\theta)'_2 [A_f(\theta)_2 V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)'_2]^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}' \psi \left[ \psi' \begin{pmatrix} 0 \\ 1 \end{pmatrix}' [A_f(\theta)_2 V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)'_2]^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \psi \right]^{-1} \\ & [A_f(\theta)_2 V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)'_2]^{-1} A_f(\theta)_2 \psi \\ & = \psi' A_f(\theta)'_2 [A_f(\theta)_2 V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)'_2]^{-1} A_f(\theta)_2 \psi \sim \chi^2(1), \end{aligned}$$

with  $A_f(\theta)_1$  and  $A_f(\theta)_2$  the first and second row of  $A_f(\theta) = \begin{pmatrix} -\theta & 1 \\ 1-\theta & 0 \end{pmatrix}$  and  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$  proportional to the first and second row of  $A_q(\theta) = -\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ . For the Sys moment conditions, the large sample behavior of the LM statistic results as:

$$\begin{aligned} & \text{GMM-LM}_{Sys}(\theta) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \\ & \psi' A_f(\theta)' [A_f(\theta) V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)']^{-1} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \psi \left[ \psi' \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}' [A_f(\theta) V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)']^{-1} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \psi \right]^{-1} \\ & \quad [A_f(\theta) V_{y_1 \Delta y, y_1 \Delta y} A_f(\theta)']^{-1} A_f(\theta) \psi \\ & \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \psi' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left[ \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}' V_{y_1 \Delta y, y_1 \Delta y}^{-1} \psi \\ & \sim \chi^2(1), \end{aligned}$$

where we used that  $A_f(\theta)^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{1-\theta} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ .

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

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

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

For values of  $\theta_0$  close to one these expressions are approximately equal to:

$$\begin{aligned}
f_N^{Dif}(\theta) &\approx \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} -\theta & 1 & 0 \\ 0 & -\theta & 1 \\ 0 & -\theta & 1 \end{pmatrix} \left[ \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_0 \uparrow 1} (1 - \theta_0) u_{i1}^2) \right] + \\
&\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} 0 \\ 0 \\ (c_i + u_{i2})(u_{i4} - \theta u_{i3}) \end{pmatrix}, \\
q_N^{Dif}(\theta) &\approx -\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \left[ \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_0 \uparrow 1} (1 - \theta_0) u_{i1}^2) \right] \\
&-\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} 0 \\ 0 \\ (c_i + u_{i2})u_{i3} \end{pmatrix}.
\end{aligned}$$

Because

$$\begin{aligned}
&\frac{h(\theta_0)}{\sqrt{N}} \sum_{i=1}^N \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_{y_{i1}u_{i2}} \\ \psi_{y_{i1}u_{i3}} \\ \psi_{y_{i1}u_{i4}} \end{pmatrix} = \psi, \\
&\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} c_i u_{i2} \\ c_i u_{i3} \\ c_i u_{i4} \\ u_{i2}^2 \\ u_{i2} u_{i3} \\ u_{i2} u_{i4} \\ u_{i3}^2 \\ u_{i3} u_{i4} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ 0 \\ 0 \\ \sigma_3^2 \\ 0 \end{pmatrix} \right] \xrightarrow{d} \begin{pmatrix} \psi_{c_i u_{i2}} \\ \psi_{c_i u_{i3}} \\ \psi_{c_i u_{i4}} \\ \psi_{u_{i2} u_{i2}} \\ \psi_{u_{i2} u_{i3}} \\ \psi_{u_{i2} u_{i4}} \\ \psi_{u_{i3} u_{i3}} \\ \psi_{u_{i3} u_{i4}} \end{pmatrix} = \psi_{cu},
\end{aligned}$$

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

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

We can write the Lev sample moments and their derivatives as

$$\begin{aligned} f_N^{Lev}(\theta) &= \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i3}\Delta y_{i2} - \theta y_{i2}\Delta y_{i2} \\ y_{i4}\Delta y_{i3} - \theta y_{i3}\Delta y_{i3} \end{pmatrix} \\ &= \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} \theta_0(\theta_0 - \theta) & 0 & 0 \\ \theta_0^2(\theta_0 - 1)(\theta_0 - \theta) & \theta_0^2(\theta_0 - \theta) & 0 \end{pmatrix} \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} + (\theta_0 - 1)y_{i1}u_{i1} \begin{pmatrix} \theta_0(\theta_0 - \theta) \\ \theta_0^3(\theta_0 - \theta) \end{pmatrix} + \\ &\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} (1 + (\theta_0 - \theta))c_i\Delta y_{i2} + (\theta_0 - \theta)u_{i2}\Delta y_{i2} + u_{i3}\Delta y_{i2} \\ (1 + (\theta_0 - \theta)(1 + \theta_0))c_i\Delta y_{i3} + (\theta_0 - \theta)u_{i3}\Delta y_{i3} + (\theta_0 - \theta)\theta_0 u_{i2}\Delta y_{i3} + u_{i4}\Delta y_{i3} \end{pmatrix}, \\ q_N^{Lev}(\theta) &= -\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i2}\Delta y_{i2} \\ y_{i3}\Delta y_{i3} \end{pmatrix} \\ &= -\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} \theta_0 & 0 & 0 \\ \theta_0^2(\theta_0 - 1) & \theta_0^2 & 0 \end{pmatrix} \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - (\theta_0 - 1)y_{i1}u_{i1} \begin{pmatrix} \theta_0 \\ \theta_0^3 \end{pmatrix} \\ &\quad - \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} c_i\Delta y_{i2} + u_{i2}\Delta y_{i2} \\ (1 + \theta_0)c_i\Delta y_{i3} + u_{i3}\Delta y_{i3} + \theta_0 u_{i2}\Delta y_{i3} \end{pmatrix}. \end{aligned}$$

For values of  $\theta_0$  close to one these expressions are approximately equal to:

$$\begin{aligned} f_N^{Lev}(\theta) &= \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} 1 - \theta & 0 & 0 \\ 0 & 1 - \theta & 0 \end{pmatrix} \left[ \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_0 \uparrow 1} (1 - \theta_0)u_{i1}^2) \right] + \\ &\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} (2 - \theta)c_i u_{i2} + (1 - \theta)u_{i2}^2 + u_{i2}u_{i3} \\ (1 + 2(1 - \theta))c_i u_{i3} + (1 - \theta)u_{i3}^2 + (1 - \theta)u_{i2}u_{i3} + u_{i3}u_{i4} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned}
q_N^{Lev}(\theta) &= -\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} y_{i2} \Delta y_{i2} \\ y_{i3} \Delta y_{i3} \end{pmatrix} \\
&= -\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \left[ \begin{pmatrix} y_{i1} u_{i2} \\ y_{i1} u_{i3} \\ y_{i1} u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_0 \uparrow 1} (1 - \theta_0) u_{i1}^2) \right] \\
&\quad - \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} c_i u_{i2} + u_{i2}^2 \\ 2c_i u_{i3} + u_{i3}^2 + u_{i2} u_{i3} \end{pmatrix},
\end{aligned}$$

so this implies that

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

We can write the NL sample moment and its derivative as

$$\begin{aligned}
f_N^{NL}(\theta) &= \frac{1}{N} \sum_{i=1}^N ((y_{i4} - \theta y_{i3} - c_i) (\Delta y_{i3} - \theta \Delta y_{i2})) \\
&= \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} \theta_0^2 (\theta - \theta_0) (\theta - \theta_0 + 1) & (\theta_0 - \theta) \theta_0^2 & 0 \end{pmatrix} \begin{pmatrix} y_{i1} u_{i2} \\ y_{i1} u_{i3} \\ y_{i1} u_{i4} \end{pmatrix} + (\theta_0 - 1) y_{i1} u_{i1} (\theta_0^2 (\theta_0 - \theta)^2) \\
&\quad + \frac{1}{N} \sum_{i=1}^N ((\theta_0 - \theta) (1 + \theta_0) c_i + \theta_0 (\theta_0 - \theta) u_{i2} + (\theta_0 - \theta) u_{i3} + u_{i4}) (\Delta y_{i3} - \theta \Delta y_{i2}), \\
q_N^{NL}(\theta) &= -\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} (2(\theta - \theta_0) + 1) \theta_0^2 & -\theta_0^2 & 0 \end{pmatrix} \begin{pmatrix} y_{i1} u_{i2} \\ y_{i1} u_{i3} \\ y_{i1} u_{i4} \end{pmatrix} + (\theta_0 - 1) y_{i1} u_{i1} 2\theta_0^2 (\theta - \theta_0) \\
&\quad + \frac{1}{N} \sum_{i=1}^N [(2\theta - \theta_0) ((1 + \theta_0) c_i + \theta_0 u_{i2} + u_{i3}) - u_{i4}] \Delta y_{i2} \\
&\quad - \frac{1}{N} \sum_{i=1}^N [(1 + \theta_0) c_i + \theta_0 u_{i2} + u_{i3}] \Delta y_{i3}.
\end{aligned}$$



For values of  $\theta_0$  close to one these expressions are approximately equal to:

$$f_N^{NL}(\theta) = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} \theta(\theta-1) & 1-\theta & 0 \end{pmatrix} \left[ \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_0 \uparrow 1} (1-\theta_0)u_{i1}^2) \right] + \frac{1}{N} \sum_{i=1}^N (2(1-\theta)c_i + (1-\theta)u_{i2} + (1-\theta)u_{i3} + u_{i4})(u_{i3} - \theta u_{i2}),$$

$$q_N^{NL}(\theta) = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} (2(\theta-1)+1) & -1 & 0 \end{pmatrix} \left[ \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} - \iota_3 E(\lim_{\theta_0 \uparrow 1} (1-\theta_0)u_{i1}^2) \right] + \frac{1}{N} \sum_{i=1}^N [(2\theta-1)(2c_i + u_{i2} + u_{i3}) - u_{i4}]u_{i2} - \frac{1}{N} \sum_{i=1}^N [2c_i + u_{i2} + u_{i3}]u_{i3},$$

so this implies that:

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

Finally, regarding AS and Sys moment conditions we simply have

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

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

**T=5.** Using similar calculations we can write:

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

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

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

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

**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

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

where  $T$  indicates the number of time periods and  $G_{2,T}^j$  is such that  $G_{2,T}^{j'} \mu_f^j(\theta, \sigma^2) = 0$ . 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 (35),  $G_{f,T=4}^j(\theta)$  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:

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

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

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

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

The above specification of  $A_f^j(\theta)_{\perp}$  as equal to  $(G_{f,T}^j(\theta) : G_{2,T}^j)$  is such that  $(A_f^j(\theta) : A_f^j(\theta)_{\perp})$  is not invertible for the AS moment conditions both when  $T = 4$  and 5. The invertibility of  $(A_f^j(\theta) : A_f^j(\theta)_{\perp})$  is not needed for the construction of the power envelope. 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 then use a specification of  $A_f^j(\theta)_{\perp}$  :

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

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

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

**Proof of Theorem 5.** Since

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

it holds that for

**T=4, Sys:**

$$\begin{aligned}
a &\xrightarrow[p]{} \left( \frac{E((c_i - (1 - \theta_0)y_{i1})^2) + \sigma_2^2}{0} \right) \\
b &\xrightarrow[p]{} \left( \frac{-(1 + \theta_0)^2 E((c_i - (1 - \theta_0)y_{i1})^2) - \theta_0^2 \sigma_2^2 - \sigma_3^2}{\theta_0^2 E((c_i - (1 - \theta_0)y_{i1})^2)} \right) \\
d &\xrightarrow[p]{} \left( \frac{\theta_0(1 + \theta_0 + \theta_0^2) E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0^2(\theta_0 - 1)\sigma_2^2 + \theta_0\sigma_3^2}{\theta_0^3 E((c_i - (1 - \theta_0)y_{i1})^2)} \right).
\end{aligned}$$

**T=4, AS:**

$$\begin{aligned}
a &\xrightarrow[p]{} \left( \frac{(1 + \theta_0) E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0 \sigma_2^2}{0} \right) \\
b &\xrightarrow[p]{} \left( \frac{-((1 + \theta_0)^2 + 1) E((c_i - (1 - \theta_0)y_{i1})^2) - \theta_0(2\theta_0 - 1)\sigma_2^2 - \sigma_3^2}{\theta_0^2 E((c_i - (1 - \theta_0)y_{i1})^2)} \right) \\
d &\xrightarrow[p]{} \left( \frac{\theta_0(1 + \theta_0 + \theta_0^2) E((c_i - (1 - \theta_0)y_{i1})^2) + \theta_0^2(\theta_0 - 1)\sigma_2^2 + \theta_0\sigma_3^2}{\theta_0^3 E((c_i - (1 - \theta_0)y_{i1})^2)} \right).
\end{aligned}$$

**T=5, Sys:**

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

**T=5, AS:**

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

**Proof of Theorem 6.** The components  $a$ ,  $b$  and  $d$  of (38) 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 5. To determine the appropriate rate for the local to unity asymptotics regarding  $\theta$ , we insert

$$\theta = 1 + \frac{e}{N^\mu}$$

in the expression of (38) and determine the appropriate value of  $\mu$  :

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

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

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

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  $t$ ,  $t$ , and  $\mu = \frac{1}{2}$  :

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

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

**Proof of Theorem 7.** When  $\omega = 0$  and  $\sigma_t^2 = \sigma^2$ ,  $g_T(e)$  is characterized by

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

so

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

and

$$\sqrt{N}g_T(e) \simeq N(e^2 E(a), B(N)'V_{abd}B(N)),$$

with

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

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.

Since  $\frac{e}{\sqrt[4]{N}}$  converges to zero,  $B(N)$  converges to  $(\iota_3 \otimes I_{p_{\max}})$ . For most sample sizes,  $\frac{e}{\sqrt[4]{N}}$  is, however, non-negligible and it is important to incorporate the terms multiplied by it in order to obtain an accurate approximation of the distribution of the GMM-AR statistic in finite samples:

$$\text{GMM-AR}(e) \sim \chi^2(\delta, p_{\max}),$$

with  $\delta = e^4 E(a)' [B(N)'V_{abd}B(N)]^{-1} E(a)$  and  $\chi^2(\delta, p_{\max})$  a non-central  $\chi^2$  distribution with non-centrality parameter  $\delta$  and degrees of freedom parameter  $p$ .

**Proof of Theorem 8.** When we instead of the full vector  $g_T(e)$  use a linear combination of it, say  $w'g_T(e)$  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  $w'g_T(e)$  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

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

so the maximal value of the non-centrality parameter is

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

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

**Proof of Theorem 9. GMM-AR statistic** To construct the worst case limiting distribution of the GMM-AR statistic to test  $H_0 : \theta = 1 + \frac{e}{\sqrt[4]{N}}$  while the true value of  $\theta = 1$ , we first specify the GMM-AR statistic as

$$\begin{aligned} \text{GMM-AR}(e) &= N f_N(e)' \hat{V}_{ff}(e)^{-1} f_N(e) \\ &= \left[ \sqrt{N} (h(\theta_0) A_f(e) \vdash A_f(e)_\perp)' f_N(e) \right]' \\ &\quad \left[ (h(\theta_0) A_f(e) \vdash A_f(e)_\perp)' \hat{V}_{ff}(e) (h(\theta_0) A_f(e) \vdash A_f(e)_\perp) \right]^{-1} \\ &\quad \left[ \sqrt{N} (h(\theta_0) A_f(e) \vdash A_f(e)_\perp)' f_N(e) \right]. \end{aligned}$$

We now determine the limit behavior of the different components under the limit sequence in (19). The specification of  $A_f(e)_\perp$  that we use is such that

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

with  $Q$  equal to the identity matrix for the Sys moment conditions and equal to the specifications stated previously in this Appendix for the AS moment conditions. The large sample behavior of the different components of  $Q$  for the AS moment conditions are such that

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

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

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

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

$$\sqrt{N}A_f(e)'_\perp f_N(e) = Q' \left( \sqrt{N}g_T(e) \right),$$

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

$$\sqrt{N}A_f(e)'_\perp f_N(e) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \bar{Q}' \left[ e^2 \sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right].$$

The limit behavior of  $\sqrt{N}h(\theta_0)A_f(e)'f_N(e)$  accords with

$$\sqrt{N}h(\theta_0)A_f(e)'f_N(e) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} A_f(e)'A_f(e)\psi,$$

so combining,

$$\left[ \sqrt{N}(h(\theta_0)A_f(e) : A_f(e)_\perp)'f_N(e) \right] \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \begin{bmatrix} A_f(e)'A_f(e)\psi \\ \bar{Q}' \left( e^2 \sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \end{bmatrix}.$$

Under mean stationarity,  $\omega = 0$  and  $g_T(e)$  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:

$$h(\theta_0)A_f(e)'\hat{V}_{ff}(e)A_f(e)_\perp \xrightarrow[\substack{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0 \\ p}]{} 0.$$

Under the worst case setting (19) also

$$\begin{aligned} h(\theta_0)^2 A_f(e)'\hat{V}_{ff}(e)A_f(e) &\xrightarrow[\substack{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0 \\ p}]{} A_f(e)'A_f(e) \left[ \lim_{\substack{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0 \\ p}} \text{var}(h(\theta_0) \begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{pmatrix}) \right] \\ A_f(e)'_\perp \hat{V}_{ff}(e)A_f(e)_\perp &\xrightarrow[\substack{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0 \\ p}]{} \bar{Q}'B(N)'V_{abd}B(N)\bar{Q} \end{aligned}$$

so

$$\begin{pmatrix} (h(\theta_0)A_f(e) : A_f(e)_\perp)'\hat{V}_{ff}(e)(h(\theta_0)A_f(e) : A_f(e)_\perp) & \xrightarrow[\substack{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0 \\ p}]{} \\ \left( A_f(e)'A_f(e) \left[ \lim_{\substack{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0 \\ p}} \text{var}(h(\theta_0) \begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{pmatrix}) \right] \right. & A_f(e)'A_f(e) & 0 \\ 0 & \bar{Q}'B(N)'V_{abd}B(N)\bar{Q} \end{pmatrix}.$$

Because  $A_f(e)'f_N(e)$  and  $A_f(e)'_{\perp}f_N(e)$  are uncorrelated under (19),

$$\left[ (h(\theta_0)A_f(e) : A_f(e)_{\perp})' \hat{V}_{ff}(e) (h(\theta_0)A_f(e) : A_f(e)_{\perp}) \right]$$

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:

$$\begin{aligned} \text{i. } & \left[ \sqrt{N}h(\theta_0)A_f(e)'f_N(e) \right]' \left[ h(\theta_0)^2 A_f(e)' \hat{V}_{ff}(e) A_f(e) \right]^{-1} \left[ \sqrt{N}h(\theta_0)A_f(e)'f_N(e) \right] \\ & \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \psi' \left[ \lim_{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \text{var}(h(\theta_0) \begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{pmatrix}) \right]^{-1} \psi \sim \chi^2(p_{GMM-AR} - p_{\max}) \\ \text{ii. } & \left( \sqrt{N}A_f(e)'_{\perp}f_N(e) \right)' \left[ A_f(e)'_{\perp} \hat{V}_{ff}(e) A_f(e)_{\perp} \right]^{-1} \left( \sqrt{N}A_f(e)'_{\perp}f_N(e) \right) \\ & \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \chi^2(\delta, p_{\max}) \end{aligned}$$

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

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

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

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

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  $H_0 : \theta = 1 + \frac{e}{\sqrt[3]{N}}$  reads

$$\text{GMM-AR}(e) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \chi^2(\delta, p_{GMM-AR}).$$

**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 (19), we determine the behavior of the different components of:

$$(h(\theta_0)A_f(e) : A_f(e)_\perp)'q_N(e)$$

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

$h(\theta_0)'A_f(e)'q_N(e)$  : Under the worst case DGPs characterized by (19):

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

since under (19):

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

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

$$A_f(e)'_\perp q_N(e) = \bar{Q} \begin{pmatrix} G_{f,T}(e)'q_N(e) \\ G'_{2,T}q_N(e) \end{pmatrix} \approx \bar{Q}' \begin{pmatrix} \frac{1}{h(\theta_0)\sqrt{N}}G_f(e)'A_q(e)\psi - \frac{e}{\sqrt[4]{N}}\sigma^2\iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq} \\ \frac{1}{\sqrt{N}}\varepsilon_{bq} \end{pmatrix},$$

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:

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

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

$$A_f(e)'_\perp q_N(e) \approx \bar{Q}' \begin{pmatrix} \frac{1}{h(\theta_0)\sqrt{N}}G_f(e)'A_q(e)\psi - \frac{\bar{e}}{\sqrt[4]{N}}\iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq} \\ \frac{1}{\sqrt{N}}\varepsilon_{bq} \end{pmatrix},$$

with

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

Combining:

$$\begin{aligned}(h(\theta_0)A_f(e) : A_f(e)_\perp)'q_N(e) &= (h(\theta_0)A_f(e) : (G_{f,T}(e) : G_{2,T})Q)'q_N(e) \\ &\approx \begin{pmatrix} \frac{1}{\sqrt{N}}A_f(e)'A_q \\ \bar{Q}'\left(\frac{1}{h(\theta_0)\sqrt{N}}G_f(e)'A_q\right)_0 \end{pmatrix} \psi + \begin{pmatrix} 0 \\ \bar{Q}'\left(-\frac{\bar{e}}{\sqrt{N}}\iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq}\right) \\ \frac{1}{\sqrt{N}}\varepsilon_{bq} \end{pmatrix}.\end{aligned}$$

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

$$\begin{aligned}\sqrt{N}((h(\theta_0)A_f(e) : A_f(e)_\perp)'\hat{V}_{ff}(e)(h(\theta_0)A_f(e) : A_f(e)_\perp))^{-1}(h(\theta_0)A_f(e) : A_f(e)_\perp)'f_N(e) \\ \xrightarrow[\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0]{d} \begin{pmatrix} (A_f(e)'A_f(e))^{-1} \left[ \lim_{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \text{var}(h(\theta_0) \begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{pmatrix}) \right]^{-1} \psi \\ (\bar{Q}'B(N)'V_{abd}B(N)\bar{Q})^{-1} \bar{Q}' \left( e^2\sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \end{pmatrix},\end{aligned}$$

we now construct the limit behavior of  $h(\theta_0)Nq_N(e)'\hat{V}_{ff}(e)^{-1}f_N(e)$  :

$$\begin{aligned}h(\theta_0)Nq_N(e)'\hat{V}_{ff}(e)^{-1}f_N(e) &= h(\theta_0)\sqrt{N} \left[ (h(\theta_0)A_f(e) : A_f(e)_\perp)'q_N(e) \right]' \\ ((h(\theta_0)A_f(e) : A_f(e)_\perp)'\hat{V}_{ff}(e)(h(\theta_0)A_f(e) : A_f(e)_\perp))^{-1}(h(\theta_0)A_f(e) : A_f(e)_\perp)'\sqrt{N}f_N(e) &\approx \\ \left[ \begin{pmatrix} h(\theta_0)A_f(e)'A_q(e) \\ \bar{Q}'\left(G_f(e)'A_q\right)_0 \end{pmatrix} \psi + h(\theta_0)\sqrt{N} \begin{pmatrix} 0 \\ \bar{Q}'\left(-\frac{\bar{e}}{\sqrt{N}}\sigma^2\iota_p + \frac{1}{\sqrt{N}}\varepsilon_{aq}\right) \\ \frac{1}{\sqrt{N}}\varepsilon_{bq} \end{pmatrix} \right]' & \\ \left( \begin{pmatrix} (A_f(e)'A_f(e))^{-1} \left[ \lim_{\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0} \text{var}(h(\theta_0) \begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT} \end{pmatrix}) \right]^{-1} \psi \\ (\bar{Q}'B(N)'V_{abd}B(N)\bar{Q})^{-1} \bar{Q}' \left( e^2\sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \end{pmatrix} \xrightarrow[\theta_0 \uparrow 1, h(\theta_0)\sqrt{N} \rightarrow 0]{d} & \\ \left( \begin{pmatrix} G_f(e)'A_q(e)\psi \\ 0 \end{pmatrix}' \bar{Q}' (\bar{Q}'B(N)'V_{abd}B(N)\bar{Q})^{-1} \bar{Q}' \left( e^2\sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) = & \\ \left( \begin{pmatrix} G_f(e)'A_q(e)\psi \\ 0 \end{pmatrix}' (B(N)'V_{abd}B(N))^{-1} \left( e^2\sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right), &\end{aligned}$$

where  $\bar{Q}$  drops out for the same reason as discussed for the GMM-AR statistic. The elements multiplied by  $h(\theta_0)$  or  $h(\theta_0)\sqrt{N}$  are under (19) of a smaller order of magnitude and therefore drop out.

The limit behavior of  $h(\theta_0)^2 N q_N(e)' \hat{V}_{ff}(e)^{-1} q_N(e)$  results in a similar manner:

$$\begin{aligned} h(\theta_0)^2 N q_N(e)' \hat{V}_{ff}(e)^{-1} q_N(e) &= h(\theta_0)^2 N \left[ (h(\theta_0) A_f(e) : A_f(e)_\perp)' q_N(e) \right]' \\ &((h(\theta_0) A_f(e) : A_f(e)_\perp)' \hat{V}_{ff}(e) (h(\theta_0) A_f(e) : A_f(e)_\perp))^{-1} (h(\theta_0) A_f(e) : A_f(e)_\perp)' q_N(e) \\ \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)_d \sqrt{N} \rightarrow 0} &\begin{pmatrix} G_f(e)' A_q(e) \psi \\ 0 \end{pmatrix}' \bar{Q} (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-1} \bar{Q}' \begin{pmatrix} G_f(e)' A_q(e) \psi \\ 0 \end{pmatrix} = \\ &\begin{pmatrix} G_f(e)' A_q(e) \psi \\ 0 \end{pmatrix}' (B(N)' V_{abd} B(N))^{-1} \begin{pmatrix} G_f(e)' A_q(e) \psi \\ 0 \end{pmatrix}, \end{aligned}$$

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  $H_0 : \theta = 1 + \frac{\epsilon}{\sqrt[4]{N}}$  under (19):

$$\text{GMM-LM}(e) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)_d \sqrt{N} \rightarrow 0} \eta' P_{(B(N)' V_{abd} B(N))^{-\frac{1}{2}} (G_f(e)' A_q(e) \psi)_0} \eta$$

with

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

and independent from  $\psi$ , so

$$\text{GMM-LM}(e) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)_d \sqrt{N} \rightarrow 0} \chi^2(\delta(\psi), 1),$$

with  $\delta(\psi) = e^4 \sigma^4 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' (B(N)' V_{abd} B(N))^{-\frac{1}{2}} P_{(B(N)' V_{abd} B(N))^{-\frac{1}{2}} (G_f(e)' A_q(e) \psi)_0} (B(N)' V_{abd} B(N))^{-\frac{1}{2}} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}$ .

**KLM statistic.** To obtain the large sample distribution of the KLM statistic to test  $H_0 : \theta = 1 + \frac{\epsilon}{\sqrt[4]{N}}$  when  $\theta_0$  is equal to one and under the limiting sequence in (19), we first determine the behavior of

$$\begin{aligned} &(h(\theta_0) A_f(e) : A_f(e)_\perp)' \hat{V}_{\theta f}(e) (h(\theta_0) A_f(e) : A_f(e)_\perp) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)_d \sqrt{N} \rightarrow 0} \\ &\begin{pmatrix} A_f(e)' A_q \\ \bar{Q}' \begin{pmatrix} \frac{1}{h(\theta_0)} G_f(e)' A_q \\ 0 \end{pmatrix} \end{pmatrix} \left[ \lim_{\theta_0 \uparrow 1, h(\theta_0)_p \sqrt{N} \rightarrow 0} \text{var}(h(\theta_0) \begin{pmatrix} y_{1i} u_{i2} \\ \vdots \\ y_{1i} u_{iT} \end{pmatrix}) \right] \begin{pmatrix} A_f(e)' A_f(e) \\ 0 \\ 0 \end{pmatrix}' + \\ &\begin{pmatrix} 0 & 0 \\ 0 & \bar{Q}' \begin{pmatrix} V_{aq,abd} B(N) \\ V_{bq,abd} B(N) \end{pmatrix} \bar{Q} \end{pmatrix}, \end{aligned}$$

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

Combining with the limit behavior of  $\sqrt{N}((h(\theta_0)A_f(e) : A_f(e)_\perp)' \hat{V}_{ff}(e)(h(\theta_0)A_f(e) : A_f(e)_\perp))^{-1}(h(\theta_0)A_f(e) : A_f(e)_\perp)' f_N(e) :$

$$\sqrt{N}((h(\theta_0)A_f(e) : A_f(e)_\perp)' \hat{V}_{ff}(e)(h(\theta_0)A_f(e) : A_f(e)_\perp))^{-1}(h(\theta_0)A_f(e) : A_f(e)_\perp)' f_N(e)$$

$$\xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \left( \begin{array}{c} (A_f(e)' A_f(e))^{-1} \left[ \lim_{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \text{var}(h(\theta_0) \begin{pmatrix} y_{1i} u_{i2} \\ \vdots \\ y_{1i} u_{iT} \end{pmatrix}) \right]^{-1} \psi \\ (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-1} \bar{Q}' \left( e^2 \sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \end{array} \right)$$

so

$$\begin{aligned} & \sqrt{N}(h(\theta_0)A_f(e) : A_f(e)_\perp)' \hat{V}_{\theta f}(e) \hat{V}_{ff}(e)^{-1} f_N(e) = \\ & \sqrt{N}(h(\theta_0)A_f(e) : A_f(e)_\perp)' \hat{V}_{\theta f}(e)(h(\theta_0)A_f(e) : A_f(e)_\perp) \\ & ((h(\theta_0)A_f(e) : A_f(e)_\perp)' \hat{V}_{ff}(e)(h(\theta_0)A_f(e) : A_f(e)_\perp))^{-1}(h(\theta_0)A_f(e) : A_f(e)_\perp)' f_N(e) \\ & \xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \left( \begin{array}{c} A_f(e)' A_q \\ \bar{Q}' \binom{1}{h(\theta_0)} G_f(e)' A_q \end{array} \right) \psi + \left( \begin{array}{c} 0 \\ \bar{Q}' \binom{V_{aq,abd} B(N)}{V_{bq,abd} B(N)} \bar{Q} \end{array} \right) \\ & (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-1} \bar{Q}' \left( e^2 \sigma^2 \binom{\iota_p}{0} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right). \end{aligned}$$

Upon combining with the limit behavior of  $\sqrt{N}(h(\theta_0)A_f(e) : A_f(e)_\perp)' q_N(e)$ , the convergence behavior of  $\sqrt[4]{N}(h(\theta_0)A_f(e) : A_f(e)_\perp)' \hat{D}_N(e)$  then results as

$$\begin{aligned} \sqrt[4]{N}(h(\theta_0)A_f(e) : A_f(e)_\perp)' \hat{D}_N(e) &= \frac{1}{\sqrt[4]{N}} \left\{ \left[ \sqrt{N}(h(\theta_0)A_f(e) : A_f(e)_\perp)' q_N(e) - \right. \right. \\ & \left. \left. \sqrt{N}(h(\theta_0)A_f(e) : A_f(e)_\perp)' \hat{V}_{\theta f}(e) \hat{V}_{ff}(e)^{-1} f_N(e) \right] \approx \left( \begin{array}{c} 0 \\ \bar{Q}' \left( \binom{\iota_p}{0} \bar{e} + \frac{1}{\sqrt[4]{N}} \nu \right) \right) \right. \\ & \left. \xrightarrow[\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0]{d} \left( \begin{array}{c} 0 \\ \bar{Q}' \binom{\iota_p}{0} \end{array} \right) \bar{e}, \right. \end{aligned}$$

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

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

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

$$\begin{aligned} \sqrt{N} \hat{D}_N(e)' \hat{V}_{ff}(e)^{-1} \hat{D}_N(e) &= \left[ \sqrt[4]{N} (h(\theta_0) A_f(e) : A_f(e)_\perp)' \hat{D}_N(e) \right]' \\ &((h(\theta_0) A_f(e) : A_f(e)_\perp)' \hat{V}_{ff}(e) (h(\theta_0) A_f(e) : A_f(e)_\perp))^{-1} \left[ \sqrt[4]{N} (h(\theta_0) A_f(e) : A_f(e)_\perp)' \hat{D}_N(e) \right] \\ &\approx \left[ \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \bar{e} + \frac{1}{\sqrt[4]{N}} \nu \right]' \bar{Q} (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-1} \bar{Q}' \left[ \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \bar{e} + \frac{1}{\sqrt[4]{N}} \nu \right] \\ &\xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \bar{e}^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \bar{Q} (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-1} \bar{Q}' \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} N^{\frac{3}{4}} \hat{D}_N(e)' \hat{V}_{ff}(e)^{-1} f_N(e) &= \left[ \sqrt[4]{N} (h(\theta_0) A_f(e) : A_f(e)_\perp)' \hat{D}_N(e) \right]' \\ &((h(\theta_0) A_f(e) : A_f(e)_\perp)' \hat{V}_{ff}(e) (h(\theta_0) A_f(e) : A_f(e)_\perp))^{-1} \sqrt{N} \left[ (h(\theta_0) A_f(e) : A_f(e)_\perp)' f_N(e) \right] \\ &\approx \left[ (\bar{Q} B(N)' V_{abd} B(N) \bar{Q})^{-\frac{1}{2}} \bar{Q}' \left[ \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \bar{e} + \frac{1}{\sqrt[4]{N}} \nu \right] \right]' \\ &(\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-\frac{1}{2}} \bar{Q}' \left( e^2 \sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \\ &\xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix}' \bar{Q} (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-1} \bar{Q}' \left( e^2 \sigma^2 \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} + B(N)' \begin{pmatrix} \varepsilon_a \\ \varepsilon_b \\ \varepsilon_d \end{pmatrix} \right) \end{aligned}$$

Upon combining everything, we obtain the limit behavior of the KLM statistic to test  $H_0 : \theta = 1 + \frac{\varepsilon}{\sqrt[4]{N}}$  under (19):

$$\text{KLM}(e) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0) \sqrt{N} \rightarrow 0} \eta' P (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-\frac{1}{2}} \bar{Q} \begin{pmatrix} \iota_p \\ 0 \end{pmatrix} \bar{e} \eta$$

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

$$\text{KLM}(e) \xrightarrow{\theta_0 \uparrow 1, h(\theta_0)/\sqrt{N} \rightarrow 0} \chi^2(\delta_{KLM}, 1)$$

because

$$\begin{aligned} \delta_{KLM} &= (e\sigma)^4 (\iota_0^p)' \bar{Q} (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-\frac{1}{2}} P_{(\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-\frac{1}{2}} \bar{Q} (\iota_0^p)} \\ &\quad (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-\frac{1}{2}} \bar{Q} (\iota_0^p) \\ &= (e\sigma)^4 (\iota_0^p)' \bar{Q} (\bar{Q}' B(N)' V_{abd} B(N) \bar{Q})^{-1} \bar{Q} (\iota_0^p) \\ &= (e\sigma)^4 (\iota_0^p)' (B(N)' V_{abd} B(N))^{-1} (\iota_0^p), \end{aligned}$$

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  $\theta$ . 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,  $\hat{\theta}_{2s}$ , uses the one step estimator  $\hat{\theta}^1$  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:

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

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

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

3. The KLM statistic of Kleibergen (2005):

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

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

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

and

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

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

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

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

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

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

$$\begin{aligned} f_i^{NL}(\theta) &= Z_i^{NL} \varphi_i^{NL}(\theta) & i = 1, \dots, N, \\ q_i^{NL}(\theta) &= Z_i^{NL} \frac{\partial}{\partial \theta} \varphi_i^{NL}(\theta) & i = 1, \dots, N, \end{aligned} \quad (61)$$

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  $k_{NL}$  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  $\frac{1}{2}(T + 1)(T - 2)$ .

## References

- Ahn, S.C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* **68**, 5–27.
- Alvarez, J. and M. Arellano (2004). Robust Likelihood Estimation of Dynamic Panel Data Models. CEMFI, Working Paper.
- Anderson, T.W. and C. Hsiao (1981). Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association* **76**, 598–606.

- Anderson, T.W. and H. Rubin (1949). Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics* **21**, 570–582.
- Andrews, D.W.K., M.J. Moreira and J.H. Stock (2006). Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica* **74**, 715–752.
- Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* **68**, 29–51.
- Arellano, M. and S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* **58**, 277–297.
- Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* **87**, 115–143.
- Bond, S. and F. Windmeijer (2005). Reliable Inference for GMM Estimators? Finite Sample Properties of Alternative Test Procedures in Linear Panel Data Models. *Econometric Reviews* **24**, 1–37.
- Bond, S., C. Nauges and F. Windmeijer (2005). Unit Roots: Identification and Testing in Micro Panels. *CEMMAP Working Paper CWP07/05*.
- Engle, R.F. and C.W.J. Granger (1987). Co-integration and error correction : Representation, estimation and testing. *Econometrica* **55**, 251–276.
- Hansen, L.P. (1982). Large Sample Properties of Generalized Method Moments Estimators. *Econometrica* **50**, 1029–1054.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551–1580.
- Kleibergen, F. (2002). Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica* **70**, 1781–1803.
- 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.
- Madsen, E. (2003). GMM Estimators and Unit Root Tests in the AR(1) Panel Data Model. *Working Paper, Centre for Applied Micro Econometrics, University of Copenhagen*.

- Newey, W.K. and K.D. West (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review* **28**, 777–787.
- Nickell, S.J. (1981). Biases in dynamic models with fixed effects. *Econometrica* **49**, 1417–1426.
- Staiger, D. and J.H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica* **65**, 557–586.
- Stock, J.H. and J.H. Wright (2000). GMM with Weak Identification. *Econometrica* **68**, 1055–1096.
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica* **48**, 817–838.