UvA ECONOMETRICS

Discussion Paper: 2011/04

Identification in linear dynamic panel data models

Maurice Bun and Frank Kleibergen

www.feb.uva.nl/ke/UvA-Econometrics

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

UvA UNIVERSITEIT VAN AMSTERDAM

Identification in linear dynamic panel data models

Maurice Bun[*] Frank Kleibergen[†]
University of Amsterdam Brown University

Version of September 1-st 2011

Abstract

Neither the Dif(ference) moment conditions, see Arellano and Bond (1991), nor the Lev(el) moment conditions, see Arellano and Bover (1995) and Blundell and Bond (1998), identify the parameters of linear dynamic panel data models for all data generating processes for the initial observations that accord with them when the data is persistent. The combined Dif-Lev (Sys) moment conditions do not always identify the parameters either when there are three time series observations but do so for larger numbers of time series observations. Thus the Sys moment conditions always identify the parameters when there are more than three time series observations. To determine the optimal GMM procedure for analyzing the parameters in linear dynamic panel data models, we construct the power envelope and find that the KLM statistic from Kleibergen (2005) maximizes the rejection frequency under the worst case alternative hypothesis whilst always being size correct under the null hypothesis.

[*]Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: M.J.G.Bun@uva.nl. Homepage: http://www1.feb.uva.nl/pp/mjgbun/ .

[†]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.

1 Introduction

Many empirical studies employ dynamic panel data methods. These are typically generalized method of moments (GMM) based and use the moment conditions that result from either Arellano and Bond (1991), to which we refer as Dif moment conditions, or Arellano and Bover (1995), to which we refer as Lev moment conditions, or a combination of these two sets of moment conditions to which we refer as Sys moment conditions, see also Anderson and Hsiao (1981) and Blundell and Bond (1998). The Dif moment conditions do not identify the parameters of dynamic panel data models when the data is persistent which has led to the development of the Lev moment conditions which are supposed to identify the parameters when the data is persistent. GMM based procedures based on the Lev moment conditions, however, often still lead to large biases of estimators and size distortions of test statistics when the data is persistent, see *e.g.* Bond and Windmeijer (2005) and Bond *et. al.* (2005).

We show that the Lev moment conditions do not identify the parameters of dynamic panel data models when the data is persistent for a range of different data generating processes for the initial observations. This explains the unsatisfactory performance of estimators and test statistics in such instances. Thus both the Dif and Lev moment conditions do not identify the parameters of the dynamic panel data model when the data is persistent because identification has to hold for all data generating processes that accord with the moment conditions. The Sys moment conditions are a combination of the Dif and Lev moment conditions so we expect that they do not identify the parameters either. This, however, only holds when there are three time series observations. We show that the Sys moment conditions identify the parameters of the dynamic panel data model when there are more than three time series observations for all data generating processes for the initial observations that accord with the moment conditions. The non-identification of the parameters from the Lev moment conditions when the data is persistent results from the divergence of the initial observations and therefore also of products of it with other variables. When there are three time series observations, the number of divergent components equals the number of Sys moment conditions so we cannot identify the parameters. When there are more than three moment conditions, the number of such divergent components is less than the number of Sys moment conditions so we can identify the parameters from those parts of the Sys moment conditions that do not depend on the divergent components.

To alleviate the biases and size distortions of estimators and test statistics alternative approximations for the large sample distributions of them in case of persistent data alongside corrections of the estimators and test statistics themselves have been proposed. Kruiniger (2009) and Madsen (2003) construct such alternative approximations of the large sample distributions using an approach similar to the weak instrument asymptotics employed in Staiger and Stock (1997). To show the non-identification of the parameters, we use a similar setting with a joint limit sequence where the variance of the initial observations goes to infinity jointy with the sample size. The resulting expressions for the large sample distributions of the estimators are similar to those in Kruiniger (2009) except for when there are more than three time series observations and we use the Sys moment conditions. We show that the parameters are then identified which is left unmentioned in Kruiniger (2009).

We often use identification robust GMM statistics, like, for example, the GMM extension of the Anderson-Rubin (AR) statistic, see Anderson and Rubin (1949) and Stock and Wright (2000), and the KLM statistic, see Kleibergen (2002,2005), to analyze the identification issues with the different moment conditions. The large sample distributions of these statistics are not affected by the identification issues that result from the initial observations. Hence, we can use them to analyze such issues in a clear manner since the results that we obtain from them are not blurred by size distortions etc. that result from the identification problems. The correction of the Wald statistic proposed in Windmeijer (2004) is another statistic that overcomes size distortions. It is closely related to the KLM statistic from Kleibergen (2005). Since it uses the two step estimator and does not incorporate all components of the KLM statistic, the large sample distribution of Windmeijer's correction of the Wald statistic is not fully robust to identification failure whilst the large sample distribution of the KLM statistic is.

We only analyze GMM based procedures. Alongside GMM based procedures, likelihood based procedures have been proposed to analyze dynamic panel data, see *e.g.* Lancaster (2002), Moreira (2009) and Hsiao *et. al.* (2002). None of these likelihood based procedures does, however, identify the parameters when the data is persistent which we show can be achieved through GMM based procedures when there are more than three time series observations.

The paper is organized as follows. The second section introduces the linear dynamic panel data model and the different moment conditions that are used to identify its parameters. In the third second, we show the non-identification of the parameters that results from the Dif and Lev moment conditions and the Sys moment conditions with three time series observations. In the fourth section, we show that the Sys moment conditions always identify the parameters when there are more than three time series observations. We also construct the large sample distributions of the estimators and statistics for persistent data when there are more than three time series observations. In the fifth section we construct the power envelope for the statistics when the data is persistent for different numbers of time series observations. The sixth section concludes.

We use the following notation throughout the paper: $P_A = A(A'A)^{-1}A'$ is a projection on the columns of the full rank A $k \times k$ dimensional matrix A and $M_A = I_k - P_A$ is a projection on the space orthogonal to A. Convergence in probability is denoted by "$\underset{p}{\rightarrow}$" and convergence in distribution by "$\underset{d}{\rightarrow}$".

2 Moment conditions

We analyze the dynamic panel data model

$$\begin{array}{rll} y_{it} = & c_i + \theta y_{it-1} + u_{it}, & i = 1, \dots, N,\ t = 2, \dots, T, \\ y_{i1} = & \mu_i + u_{i1} & i = 1, \dots, N, \end{array} \tag{1}$$

with $c_i = \mu_i(1-\theta)$, 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 θ by means of least squares leads to a biased estimator in samples with a finite value of T, see *e.g.* Nickel (1981). We therefore estimate it using GMM. We obtain the GMM moment conditions from the conditional moment assumption:

$$E[u_{it}|u_{it-1}, \dots, u_{i1}, c_i] = 0, \qquad t = 1, \dots, T,\ i = 1, \dots, N, \tag{2}$$

which implies that for every i :

$$\begin{array}{rll} E[u_{it}u_{it-j}] = & 0, & j = 1, \dots, t-1;\ t = 1, \dots, T, \\ E[u_{it}c_i] = & 0, & t = 1, \dots, T. \end{array} \tag{3}$$

Alongside the conditional moment assumption 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 Δy_{it} and y_{it} :

$$E[\Delta y_{it} y_{ij}] \qquad j = 1, \dots, T,\ t = 2, \dots, T \tag{4}$$

can be used to construct functions which identify the parameter of interest θ. The covariance between c_i and y_{it-1} leads to inconsistency of the least squares estimator of θ when T is finite and N gets large so the moments in (4) do not contain any product of c_i and y_{it-1}, since Δy_{it} does not depend on c_i, as to avoid the origin of this inconsistency. We do not use products of Δy_{it} to identify θ either 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 (4), are commonly used to identify θ :

1. Difference (Dif) moment conditions:

$$E[y_{ij}(\Delta y_{it} - \theta \Delta y_{it-1})] = 0 \qquad j = 1, \ldots, t-2;\ t = 3, \ldots, T, \tag{5}$$

   as proposed by *e.g.* Anderson and Hsiao (1981) and Arellano and Bond (1991).

2. Level (Lev) moment conditions:

$$E[\Delta y_{it-1}(y_{it} - \theta y_{it-1})] = 0 \qquad t = 3, \ldots, T, \tag{6}$$

   as proposed by Arellano and Bover (1995), see also Blundell and Bond (1998).

These moments can be used separately or jointly to identify θ. If we use the moment conditions in (5) and (6) jointly, we refer to them as system (Sys) moment conditions, see Arellano and Bover (1995) and Blundell and Bond (1998).

Without any additional homoscedasticity or initial observation assumptions, the sample analogs of the moments in (4),

$$\tfrac{1}{N} \sum_{i=1}^{N} \Delta y_{it} y_{ij} \qquad j = 1, \ldots, T,\ t = 2, \ldots, T, \tag{7}$$

provide the sufficient statistics for θ. Under our assumptions, the Sys moment conditions extract all available information on θ from these sufficient statistics.

The Dif moment conditions do not identify θ when its true value is equal to one while the Lev moment conditions are supposed to do, see Arellano and Bover (1995) and Blundell and Bond (1998). It has therefore become customary to use the Sys moment conditions so θ is identified throughout by the moment conditions. The identification results in Blundell and Bond (1998) are, however, silent about their sensitivity with respect to the initial observations.

In Lancaster (2002) and Moreira (2009), the observations are analyzed in deviation from the initial observations which preserves their autoregressive structure and sets the initial observations to zero. The transformed observations which are analyzed using likelihood based procedures, however, no longer satisfy the Lev moment conditions. Likelihood based procedures are also used in Hsiao *et. al.* (2002) but identical to the likelihood based procedures proposed in Lancaster (2002) and Moreira (2009), they do not apply to unit values of θ.

3 Initial observations and identification

The Dif and Lev moment conditions that we use to identify θ are semi-parametric with respect to the fixed effects and initial observations so they identify θ for a variety of different specifications of them. These specifications, however, still influence the identification of θ 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.

3.1 Identification when $T = 3$.

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

$$\begin{array}{rll} \text{Dif:} & E[y_{i1}(\Delta y_{i3} - \theta \Delta y_{i2})] & = 0 \\ \text{Lev:} & E[\Delta y_{i2}(y_{i3} - \theta y_{i2})] & = 0 \end{array} \tag{8}$$

with Jacobians:

$$\begin{array}{rl} \text{Dif:} & -E[y_{i1}\Delta y_{i2}] \\ \text{Lev:} & -E[y_{i2}\Delta y_{i2})]. \end{array} \tag{9}$$

The Jacobians in (9) show that for many data generating processes (DGPs) for the initial observations y_{i1}, the Dif moment condition does not identify θ when its true value, θ_0, is equal to one since $E[y_{i1}\Delta y_{i2}]$ is then equal to zero[1]. The Jacobian of the Lev moment condition is such that

$$\begin{aligned} E(y_{i2}\Delta y_{i2}) = & \; E((c_i + \theta_0 y_{i1} + u_{i2})u_{i2}) \\ = & \; E(u_{i2}^2) \\ = & \; \sigma_2^2 \neq 0, \end{aligned} \tag{10}$$

so the Lev moment conditions seem to identify θ irrespective of the value of θ_0, see Arellano and Bover (1995) and Blundell and Bond (1997). There is a caveat though since for many data generating processes y_{i1} is not defined when θ_0 is equal to one and, despite that y_{i1} and u_{i2} are uncorrelated, we then do not know the value of $E(y_{i1}u_{i2})$

[1]An example of a DGP for y_{i1} for which the Dif Jacobian condition does hold at $\theta_0 = 1$ is $y_{i1} \sim N(\mu, \frac{\sigma^2}{1-\theta_0^2})$ (covariance stationarity) so $E(y_{i1}\Delta y_{i2}) = \frac{\sigma^2}{2}$.

which is used in the construction of the Jacobian in (10). To ascertain the identification of θ by the Lev moment conditions when θ_0 is equal to one, we therefore consider a joint limit process where both θ_0 converges to one and the sample size goes to infinity. In order to do so, we first make an assumption about the mean of the initial observations.

Assumption 1. *When θ_0 goes to one, the initial observations are such that*

$$\lim_{\theta_0 \to 1}(1-\theta_0)y_{i1} = c_i. \tag{11}$$

Assumption 1 implies that the constant term c_i in the autoregressive model in (1) is associated with the mean of the initial observations μ_i and that

$$\lim_{\theta_0 \to 1}(1-\theta_0)u_{i1} = 0. \tag{12}$$

It allows both for c_i or μ_i fixed for different values of θ_0. If Assumption 1 does not hold, the Lev moment conditions are not satisfied since

$$\Delta y_{i2} = c_i + (\theta_0 - 1)y_{i1} + u_{i2} \tag{13}$$

so if Assumption 1 does not hold, Δy_{i2} depends on c_i when θ_0 is equal to one and the Lev moment condition does not hold.

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 θ_0 converges to one (we rule out explosive values of θ_0) for which we just list their elements that matter for the large sample behavior for some DGPs for the initial observations:

$$\begin{aligned}
\lim_{\theta_0 \uparrow 1} \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Delta y_{i2}(y_{i3} - \theta y_{i2}) = \; & (1-\theta)\Big\{\tfrac{1}{N}\textstyle\sum_{i=1}^{N} u_{i2}^2 + \lim_{\theta_0 \uparrow 1}\tfrac{1}{N}\sum_{i=1}^{N} u_{i2}y_{i1} + \\
& \lim_{\theta_0 \uparrow 1}\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(1-\theta_0)u_{i1}y_{i1}\Big\} \\
\lim_{\theta_0 \uparrow 1} \tfrac{1}{N}\textstyle\sum_{i=1}^{N} y_{i2}\Delta y_{i2} = \; & \tfrac{1}{N}\textstyle\sum_{i=1}^{N} u_{i2}^2 + \lim_{\theta_0 \uparrow 1}\tfrac{1}{N}\sum_{i=1}^{N} u_{i2}y_{i1} + \\
& \lim_{\theta_0 \uparrow 1}\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(1-\theta_0)u_{i1}y_{i1}.
\end{aligned} \tag{14}$$

Since u_{i2} and y_{i1} are uncorrelated, for some function $h(\theta_0)$ it holds that

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

with ψ_2 a normal random variable with mean zero and variance σ_2^2 =var(u_{i2}) so $h(\theta_0)^{-2}$ =var(y_{i1}), which explains why $\frac{1}{N}\sum_{i=1}^{N} u_{i2}y_{i1}$ appears in (14). Similarly, if var$(u_{i1}) = \frac{\sigma^2}{1-\theta_0^2}$ then

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

with σ^2 a non-zero constant.[2]

The speed with which the sample size goes to infinity compared to the convergence of $h(\theta_0)$ to zero determines the behavior of the Jacobian of the Lev moment condition. For example, when

$$h(\theta_0)\sqrt{N} \underset{N\to\infty,\ \theta_0\uparrow 1}{\to} \infty, \tag{17}$$

it holds that

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N} y_{i2}\Delta y_{i2} \underset{N\to\infty,\ \theta_0\uparrow 1}{\underset{p}{\to}} \sigma_2^2 + \lim_{\theta_0\uparrow 1} E((1-\theta_0)u_{i1}^2), \tag{18}$$

while when

$$h(\theta_0)\sqrt{N} \underset{N\to\infty,\ \theta_0\uparrow 1}{\to} 0, \tag{19}$$

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

$$\begin{array}{rcl} h(\theta_0)\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Delta y_{i2}(y_{i3}-\theta y_{i2}) & \underset{N\to\infty,\ \theta_0\uparrow 1}{\underset{d}{\to}} & (1-\theta)\psi_2 \\ h(\theta_0)\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i2}\Delta y_{i2} & \underset{N\to\infty,\ \theta_0\uparrow 1}{\underset{d}{\to}} & \psi_2. \end{array} \tag{20}$$

This shows that θ is not identified if θ_0 is equal to one, since ψ_2 has mean zero, and the convergence of the sample size and θ_0 accords with (19). Since any assumption about the convergence rates of the sample size and θ_0 is arbitrary, also the identification of θ by the Lev moment conditions is arbitrary for DGPs for which θ_0 is close to one and $h(\theta_0)$ is equal to zero when θ_0 equals one. Some of the most plausible DGPs for the initial observations belong to this category. A few examples of these, and their specifications of $h(\theta_0)$, are:

DGP 1. $y_{i1} \sim N(\frac{c_i}{1-\theta_0},\ \sigma_1^2)$, $\sigma_c^2 = \text{var}(c_i)$, $h(\theta_0) = (1-\theta_0)/\sigma_c$.

DGP 2. $y_{i1} \sim N(\frac{c_i}{1-\theta_0},\ \frac{\sigma_1^2}{1-\theta_0^2})$, $\sigma_c^2 = \text{var}(c_i)$, $h(\theta_0) = (1-\theta_0)/\sigma_c$.

DGP 3. $y_{i1} \sim N(\mu_i,\ \frac{\sigma_1^2}{1-\theta_0^2})$, $\sigma_\mu^2 = \text{var}(\mu_i)$, $c_i = \mu_i(1-\theta_0)$, $h(\theta_0) = \left(\sqrt{1-\theta_0^2}\right)/\sigma_1$.

DGP 4. $y_{i1} \sim N(\mu_i,\ \sigma_1^2\frac{1-\theta_0^{2(g+1)}}{1-\theta_0^2})$, $\sigma_\mu^2 = \text{var}(\mu_i)$, $c_i = \mu_i(1-\theta_0)$, $\lim_{\theta_0\uparrow 1} h(\theta_0) = \frac{1}{\sigma_1}g^{-\frac{1}{2}}$.

DGP 5. $y_{i1} \sim N(\frac{c_i}{1-\theta_0},\ \sigma_1^2\frac{1-\theta_0^{2(g+1)}}{1-\theta_0^2})$, $\sigma_c^2 = \text{var}(c_i)$, $\lim_{\theta_0\uparrow 1} h(\theta_0) = (1-\theta_0)/\sigma_c$.

[2] We note that this is just the special case of covariance stationarity to which we do not confine ourselves.

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 1 and 3 are also used by Blundell and Bond (1997) while Arellano and Bover (1995) use DGP 3.

The conditions on $h(\theta)$ that result from (19) for DGPs 1-5 are:

$$\begin{array}{lll} \text{DGP } 1,\ 2,\ 5: & (1-\theta_0)\sqrt{N} \underset{N\to\infty,\ \theta_0\uparrow 1}{\to} 0 & \text{or} \quad \theta_0 = 1 - \frac{d}{N^{\frac{1}{2}(1+\epsilon)}} \\ \text{DGP } 3: & (1-\theta_0^2)N \underset{N\to\infty,\ \theta_0\uparrow 1}{\to} 0 & \text{or} \quad \theta_0 = 1 - \frac{d}{N^{1+\epsilon}} \\ \text{DGP } 4: & \frac{N}{g} \underset{N\to\infty,\ g\to\infty}{\to} 0, & \end{array} \tag{21}$$

with d a constant and ϵ some real number larger than zero, so the process should have been running longer than the sample size in case of DGP 4. Kruiniger (2009) uses the above specification for DGP 3 with $\epsilon = 0$ and DGP 4 with N/g going 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 in order to obtain results that apply generally. While the (non-) identification conditions for identifying θ that result from the above data generating processes might be (in)plausible, it is the arbitrariness of them which is problematic for practical purposes. 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 Staiger and Stock (1997).

Dif moment condition Because the Dif moment condition does not identify θ when θ_0 is equal to one, except, for example, for the covariance stationary DGP for the initial observations mentioned previously, the above identification issues that concern the Lev moment conditions do not in general alter the identification of θ when we use the Dif moment condition. When (19) holds, the large sample behavior for values of θ_0 close to one for the Dif sample moments and its derivative are such that:

$$\begin{array}{rcl} h(\theta_0)\frac{1}{\sqrt{N}}\sum_{i=1}^N y_{i1}(\Delta y_{i3} - \theta\Delta y_{i2}) & \underset{\substack{d \\ N\to\infty,\ \theta_0\uparrow 1}}{\to} & \psi_3 - \theta\psi_2 \\ -h(\theta_0)\frac{1}{\sqrt{N}}\sum_{i=1}^N y_{i1}\Delta y_{i2} & \underset{\substack{d \\ N\to\infty,\ \theta_0\uparrow 1}}{\to} & -\psi_2 \end{array} \tag{22}$$

where

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

with ψ_3 a normal random variable with mean zero and variance σ_3^2 =var(u_{i3}). Since ψ_2 has mean zero, θ remains unidentified when θ_0 is equal to one. This also shows that although a covariance stationary DGP for the initial observations, like DGP 3, seems to satisfy the Jacobian identification condition, it does not identify θ at $\theta_0 = 1$ when the convergence of the sample size and θ_0 is in accordance with (19).

The above implies that θ is not necessarily identified by the Dif and Lev moment conditions when θ_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, which we reflect by $f_N(\theta)$, and their derivative, which we reflect by $q_N(\theta)$, 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 θ_0 converges to one according to (19), the behavior of the Sys sample moments and their derivative is characterized by

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

The large sample behavior of the Sys moment conditions in (25) does, since the means of ψ_2 and ψ_3 are equal to zero, not identify θ. This shows that also for the Sys moment conditions the identification of θ is arbitrary for unit values of θ_0 when $T = 3$ since it depends on an high level assumption that concerns the convergence rates of θ_0 and the sample size.

The non-identification of θ by its moment conditions for specific convergence sequences concerning 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 θ 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 (3), finite fourth moments of c_i and u_{it}, $i = 1, \ldots, N$, $t = 2, \ldots, T$ 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:*

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

with

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

and $\overline{y_{i1}\Delta y_{i2}} = \frac{1}{N} \sum_{i=1}^N y_{i1} \Delta y_{i2}$, $\overline{y_{i1}\Delta y_{i3}} = \frac{1}{N} \sum_{i=1}^N y_{i1} \Delta y_{i3}$.

Proof. see Appendix A. ■

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

Theorem 2. *Under Assumption 1, the conditions in (3), finite fourth moments of* c_i *and* u_{it}, $i = 1, \ldots, N$, $t = 2, \ldots, T$ *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:*

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

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), whose definition is stated in Appendix B, remains standard χ^2 when the true value of θ gets close to one. Because the moment conditions do not identify θ when the true value of θ gets close to one according to (19), it remains standard χ^2 for all tested values of θ as stated in Theorem 3.

Theorem 3. *Under Assumption 1, the conditions in (3), finite fourth moments of* c_i *and* u_{it}, $i = 1, \ldots, N$, $t = 2, \ldots, T$ *and when (19) holds, the large sample distributions of the GMM-LM statistics of Newey and West (1987) that result from the Dif, Lev and*

Sys moment conditions when $T = 3$ read:

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

Proof. see Appendix A. ■

The large sample distributions of the GMM-LM statistic in Theorem 3 properly reflect that θ 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 θ 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), to analyze the identification of θ by the different moment conditions when its true value gets close to one. Their definitions 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 θ is then exactly identified by the moment conditions. When we use the Sys moment conditions, so θ is over-identified by the moment conditions, these statistics differ from the GMM-LM statistic. The large sample distributions of the KLM and GMM-AR statistics do not alter when the tested parameter becomes non-identified so the large sample distributions of them remain standard χ^2 for all values of θ under (19) albeit with different degrees of freedom. We further discuss the large sample distributions of these statistics in Theorem 6 in the next section.

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

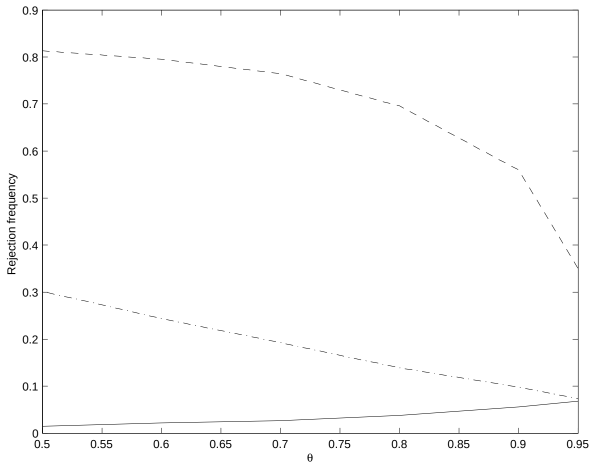

Figure 1.1. Wald statistics.

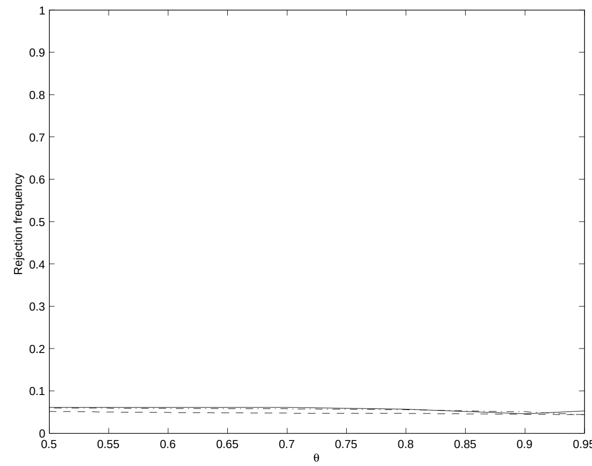

Figure 1.2. GMM-LM statistics.

Identification in simulated data To illustrate the identification issues with the different moment conditions when $T = 3$, we generate observations from DGP 1 with $T = 3$, $N = 500$ and $\sigma_c^2 = 1$, $\sigma_t^2 = 1$, $t = 1, \ldots, T$. We use them to compute the large sample distributions of the Wald and GMM-LM statistics stated in Theorems 2 and 3. These are shown in Figures 1.1 and 1.2 in Panel 1. Figure 1.1 contains the simulated distributions of the Wald statistic while Figure 1.2 contains the simulated distributions of the GMM-LM statistic.

Because of the specification of the data generating process, the ratios involved in the expressions of the large sample distributions of the Wald statistics that use the Dif and Sys moment conditions simplify considerably at $\theta = 1$. For the Wald statistic using the Dif moment condition, the numerator cancels out with the denominator and the remaining constant is also such that it scales ψ_2^2 to a $\chi^2(1)$ distributed random variable. This explains why this Wald statistic is size correct, so the rejection frequency is around 5%, when $\theta = 1$ as shown in Figure 1.1.

For the Wald statistic using the Sys moment conditions when we test $\theta = 1$, the squared terms in the brackets cancel out against each other so the large sample distribution only consists of the two elements at the front of the expression in Theorem

2. These are again such that the second element scales out the variance of the first so we are left with a standard $\chi^2(1)$ distributed random variable. This explains why the Wald statistic with the Sys moment conditions is size correct as well as shown in Figure 1.1.

The large sample distribution of the Wald statistic using the Lev moment condition in Theorem 2 clearly depends on nuisance parameters through $V_{y_1\Delta y,y_1\Delta y,11}$. When $V_{y_1\Delta y,y_1\Delta y,11}$ equals zero, which basically means that the convergence is according to (17) instead of (19), it is size correct while it is size distorted for non-zero values of $V_{y_1\Delta y,y_1\Delta y,11}$ as shown in Figure 1.1.

Although the Wald statistics using the Dif and Sys moment conditions are size correct, their rejection frequencies for other values of θ are non-standard because of the non-standard large sample distributions. Figure 1.1 therefore shows that the rejection frequency of the Wald statistic using the Dif moment condition declines when we get further away from θ while the rejection frequency of the Wald statistic using the Sys moment conditions increases. Figure 1.1 also shows that the rejection frequency of the Wald statistic using the Lev moment conditions is an increasing function of the distance towards one.

Figure 1.2 confirms the findings from Theorem 3 and shows that the rejection frequency of all three GMM-LM statistics is equal to the size of the test for all values of θ when the true value of θ is equal to one.

3.2 Identification from Lev and Dif moment conditions when $T > 3$.

The identification issues discussed before extend to the Lev and Dif moment conditions for larger numbers of time series observations. The derivatives of the Dif and Lev sample moments read:

$$\begin{array}{lll} \text{Dif:} & -\frac{1}{N}\sum_{i=1}^{N} y_{ij}\Delta y_{it-1} & j=1,\ldots,t-2;\ t=3,\ldots,T \\ \text{Lev:} & -\frac{1}{N}\sum_{i=1}^{N} y_{it-1}\Delta y_{it-1} & t=3,\ldots,T, \end{array} \tag{30}$$

and their convergence behavior under (19) is characterized by

$$h(\theta_0)\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{ij}\Delta y_{it} \underset{N\to\infty,\ \theta_0\uparrow 1}{\vec{d}} \psi_t \qquad j=1,\ldots,t-1;\ t=2,\ldots,T-1, \tag{31}$$

with ψ_t, $t=2,\ldots,T-1$, independently distributed normal random variables with mean zero and variance $\sigma_t^2 = \text{var}(u_{it})$. This shows that the Dif and Lev moment conditions do not identify θ for larger number of time series observations under the convergence scheme in (30). Hence, the identification of θ using the Dif and Lev moments conditions is arbitrary when θ_0 is close to one and $h(\theta_0)$ is equal to zero when θ_0 equals one.

Identification in simulated data The consequences of the non-identification of θ by the Dif and Lev moment conditions when the true value of θ is equal to one for values of T larger than three, has similar consequences for the large sample distributions of estimators and test statistics as when T equals three. We therefore do not derive these consequences analytically but just illustrate them using a simulation experiment similar to the one used to construct the figures in Panel 1.

To illustrate the identification issues with the different moment conditions when T exceeds 3, we generate observations from DGP 1 with $T = 4$ and 5, $N = 500$ and $\sigma_c^2 = 1$, $\sigma_t^2 = 1$, $t = 1, \ldots, T$. Panel 2 shows the rejection frequencies of testing for different values of θ when its true value is equal to one using the Wald and GMM-LM statistics for values of T equal to 4, Figures 2.1 and 2.2, and 5, Figures 2.1 and 2.2. Figures 2.1 and 2.3 that contain the rejection frequencies of the Wald statistics show that its non-standard large sample distributions are now such that it is size distorted also when we just use the Dif moment conditions. Besides that the rejection frequencies also decrease when we move further away from one when we use the Dif moment conditions which is similar to Figure 1.1. Figures 2.2 and 2.4 are identical to Figure 1.2 and show that the large sample distributions of the GMM-LM statistic are flat which we expect for statistics that test parameters that are not identified since the rejection frequencies are almost the same and equal to the size for all tested values of the parameter.

Panel 2. Rejection frequencies of 95% significance tests for different values of θ while the true value is one using DGP 1, $T = 4$ and 5, $\sigma_c^2 = 10$, $N = 500$ and Dif (solid) and Lev (dashed) moments.

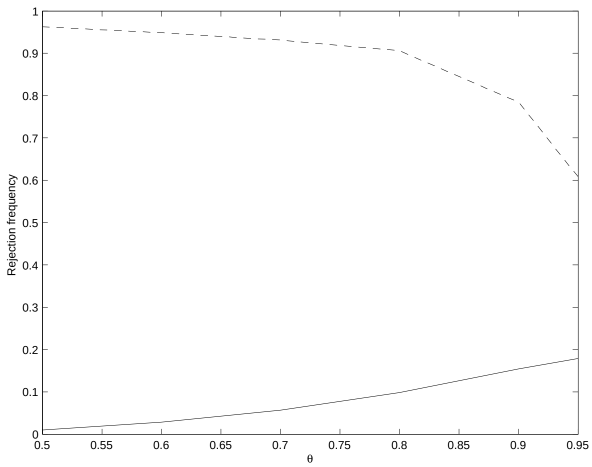

Figure 2.1. Wald statistics, $T = 4$.

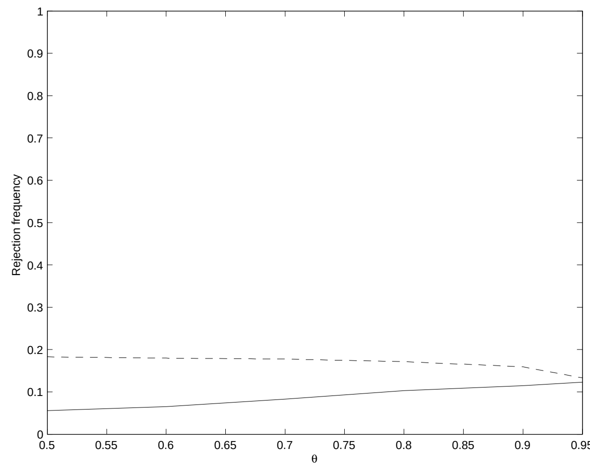

Figure 2.2. GMM-LM statistics, $T = 4$.

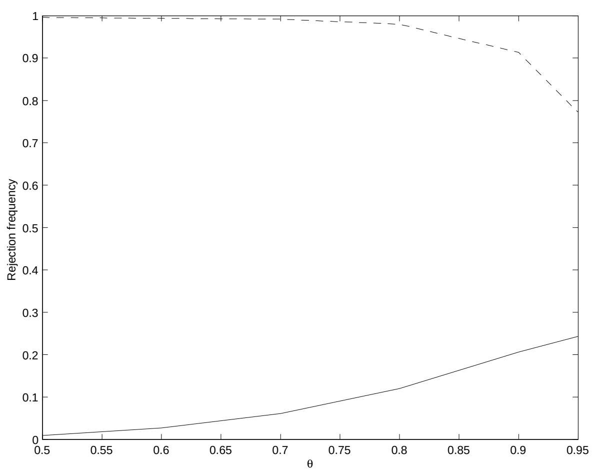

Figure 2.3. Wald statistics, $T = 5$.

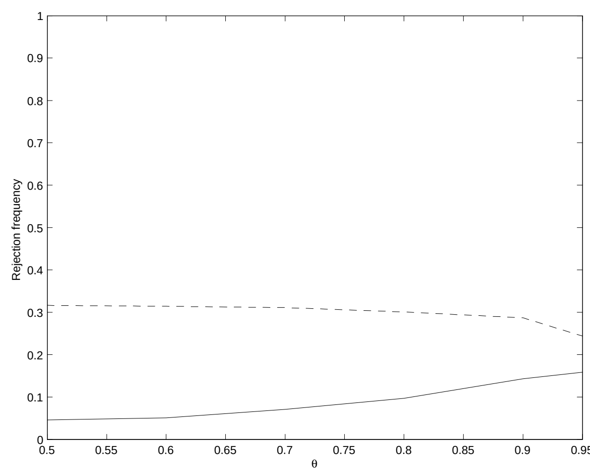

Figure 2.3. GMM-LM statistics, $T = 5$.

4 Identification using Sys moments

We just showed that also for larger numbers of time periods neither the Dif nor the Lev moment conditions identify θ for all data generating processes for which the initial observations satisfy these moment conditions. We therefore expect the same to hold for the Sys moment conditions for such number of time periods which is, as we showed previously, the case when the number of time periods is equal to three. To analyze the identification of θ using the Sys moment conditions, we begin with a representation theorem.

Theorem 4 (Representation Theorem). *Under Assumption 1, the conditions in (3), finite fourth moments of c_i and u_{it}, $i = 1, \ldots, N$, $t = 2, \ldots, T$, we can characterize the large sample behavior of the Sys sample moments and their derivatives for values of θ_0 close to one by*

$$\begin{array}{c}\begin{pmatrix} f_N(\theta) \\ q_N(\theta) \end{pmatrix} \approx \left[\begin{pmatrix} 1-\theta \\ 1 \end{pmatrix} \otimes I_k \right] \mu(\sigma^2) + \frac{1}{h(\theta_0)\sqrt{N}} \begin{pmatrix} A_f(\theta) \\ A_q \end{pmatrix} \psi + \\ \begin{pmatrix} A_f(\theta) \\ A_q \end{pmatrix} (\iota_2 \otimes \iota_{T-2}) \left(\lim_{\theta_0 \uparrow 1} E((\theta_0 - 1)u_{i1}^2)\right) + \frac{1}{\sqrt{N}} \begin{pmatrix} B_f(\theta) \\ B_q \end{pmatrix} \psi_{cu},\end{array} \tag{32}$$

with ι_j a $j \times 1$ dimensional vector of ones, $\mu(\sigma^2)$ a constant and ψ and ψ_{cu} are mean zero normal random variables and for:

T=3 *we have two moment conditions and the dimension of ψ is two.*

T=4 *we have five moment conditions and the dimension of ψ is three.*

T=5 *we have nine moment conditions and the dimension of ψ is four.*

General T $\frac{1}{2}(T+1)(T-2)$ *moment condtions and the dimension of ψ is $T-1$.*

The exact specification of $A_f(\theta)$, A_q, $B_f(\theta)$, B_q, $\mu(\sigma^2)$, ψ and ψ_{cu} for values of T equal to 3, 4 and 5 is given in Appendix A.

Proof. see Appendix A. ■

Theorem 4 shows that

$$\sqrt{N} h(\theta_0) \begin{pmatrix} f_N(\theta) \\ q_N(\theta) \end{pmatrix} \underset{\theta_0 \uparrow 1,\ h(\theta_0)\sqrt{N} \to 0}{\underset{d}{\to}} \begin{pmatrix} A_f(\theta) \\ A_q \end{pmatrix} \psi, \tag{33}$$

which seems to indicate that θ is not identified by the Sys moment conditions for any number of time periods when the limit sequence in (19) holds since ψ is a mean zero normal random variable. The expression in (33) is qualitatively the same as the one in (25), which applies to $T = 3$, except for that the number of components of ψ can be less than that of either $f_N(\theta)$ or $q_N(\theta)$ while these numbers are equal in (25). If we therefore pre-multiply the Sys moment conditions in (32) by the orthogonal complement of $A_f(\theta)$, we obtain

$$A_f(\theta)'_\perp f_N(\theta) \approx (1-\theta)A_f(\theta)'_\perp \mu(\sigma^2) + \tfrac{1}{\sqrt{N}} A_f(\theta)'_\perp B_f(\theta)\psi_{cu}, \tag{34}$$

with $A_f(\theta)_\perp$ a $\frac{1}{2}(T+1)(T-2) \times \left(\frac{1}{2}(T-1)(T-2)-1\right)$ dimensional matrix which is such that $A_f(\theta)'_\perp A_f(\theta) \equiv 0$, $A_f(\theta)'_\perp A_f(\theta)_\perp \equiv I_{\left(\frac{1}{2}(T-1)(T-2)-1\right)}$. The rotated Sys moment conditions in (34) do not suffer from the degeneracies caused by the divergence of $\frac{1}{h(\theta_0)\sqrt{N}}\psi$. It implies that θ is identified by the Sys moment conditions regardless of the process generating the initial conditions when there are more than three time periods. When there are three time periods, the orthogonal complement of $A_f(\theta)$ is not defined since $A_f(\theta)$ is a square matrix so θ is not necessarily identified. From three time periods onwards we therefore expect estimators of θ to be consistent regardless of the process that generated the initial observations as long as it satisfies the Sys moment conditions.

Theorem 5. *Under Assumption 1, the conditions in (3), finite fourth moments of c_i and u_{it}, $i = 1, \ldots, N$, $t = 2, \ldots, T$, with T larger than three, the large sample behavior of the one step, two step and continuous updating estimators under the convergence sequence in (19) reads:*

1. *One step estimator:*

$$\hat{\theta}_{1s} \underset{d}{\rightarrow} 1 - (\psi' A_q' A_q \psi)^{-1} \psi' A_q A_f(1)\psi, \tag{35}$$

which is inconsistent since $A_f(1)$ does not equal zero.

2. *Two step estimator:*

$$\begin{aligned}
&h(\theta_0)^{-1}(\hat{\theta}_{2s} - 1) \underset{d}{\rightarrow} \\
&\left(\psi' A_q' A_f(\hat{\theta}_{1s})_\perp \left[A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V_{uu,uu.y_1\Delta y} B_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s})_\perp\right]^{-1} A_f(\hat{\theta}_{1s})'_\perp A_q \psi\right)^{-1} \\
&\psi' A_q' A_f(\hat{\theta}_{1s})_\perp \left[A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V_{uu,uu.y_1\Delta y} B_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s})_\perp\right]^{-1} A_f(\hat{\theta}_{1s})'_\perp \\
&\left[(1-\hat{\theta}_{1s})\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\hat{\theta}_{1s}) \left(\psi_{cu} - V'_{y_1\Delta y,uu} V^{-1}_{y_1\Delta y, y_1\Delta y}\psi\right)\right],
\end{aligned} \tag{36}$$

which shows that the two step estimator is consistent but with a non-standard convergence rate, $h(\theta_0)$, and a non-standard large sample distribution.

3. *Continuous updating estimator (CUE), see Hansen et al (1996):*

$$\sqrt{N}(\hat{\theta}_{CUE}-1) \underset{d}{\rightarrow} \left[\mu(\sigma^2)'A_f(1)_\perp \left[A_f(1)'_\perp B_f(1)V_{uu,uu.y_1\Delta y}B_f(1)'A_f(1)_\perp\right]^{-1} A_f(1)'_\perp \mu(\sigma^2)\right]^{-1} \mu(\sigma^2)'A_f(1)_\perp \left[A_f(1)'_\perp B_f(1)V_{uu,uu.y_1\Delta y}B_f(1)'A_f(1)_\perp\right]^{-1} A_f(1)'_\perp B_f(1)\left[\psi_{cu} - V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y}\psi\right], \tag{37}$$

so the CUE is consistent with a standard convergence rate and large sample distribution.

where

$$\begin{aligned} V_{uu,uu} &= var(c_iu_{i2}, c_iu_{i3}, c_iu_{i4}, u_{i2}^2, u_{i2}u_{i3}, u_{i2}u_{i4}, u_{i3}^2, u_{i3}u_{i4}) \\ V_{y_1\Delta y,y_1\Delta y} &= var(y_{i1}u_{i2}, y_{i1}u_{i3}, y_{i1}u_{i4}) \\ V_{uu,y_1\Delta y} &= cov((c_iu_{i2}, c_iu_{i3}, c_iu_{i4}, u_{i2}^2, u_{i2}u_{i3}, u_{i2}u_{i4}, u_{i3}^2, u_{i3}u_{i4}), \\ &\qquad (y_{i1}u_{i2}, y_{i1}u_{i3}, y_{i1}u_{i4})) \\ V_{uu,uu.y_1\Delta y} &= V_{uu,uu} - V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y}V_{y_1\Delta y,uu}. \end{aligned} \tag{38}$$

Proof. see Appendix A. ∎

Theorem 5 states the large sample distributions of the one step, two step and continuous updating estimators. Because of the divergence of the Sys sample moments in the direction of $A_f(\theta)$, the one step estimator is inconsistent and converges to a random variable. The two step estimator uses the one step estimator as input which explains why the two step estimator is consistent but with a non-standard convergence rate and large sample distribution. When we further iterate to obtain the continuous updating estimator, Theorem 5 shows that we do obtain a standard $\sqrt{N}$ convergence rate and a normal large sample distribution. Because only the large sample distribution of the fully iterated CUE is normal, we, however, expect that the finite sample distribution of the CUE is considerably different from a normal distribution.

In Kruiniger (2009), the large sample distribution of the two step estimator for values of θ_0 close to one and when T exceeds 3 is constructed as well. Kruiniger (2009) shows that the two step estimator is inconsistent. This probably results because it is not mentioned in Kruiniger (2009) that the number of divergent components in the Sys moment conditions is less than the number of elements of the moment conditions so we can identify θ using that part of the moment conditions that does not depend on the divergent components. This is exactly what happens for the estimators since all their components result from the orthogonal complement of $A_f(\theta)$, $A_f(\theta)_\perp$, which is further reflected by the expressions of the large sample distributions in Theorem 5.

Panel 3. Rejection frequencies of 95% tests for different values of θ while the true value value is one using the two step and CUE Wald t-statistics with Sys moment conditions and varying numbers of T, DGP 1, $\sigma_c^2 = 1,\ \sigma_t^2 = 1,\ t = 1, \ldots, T,\ N = 500:\ T = 3$ (solid), 4 (dashed), 5 (dash-dotted), 6 (dotted), 7 (solid with plusses).

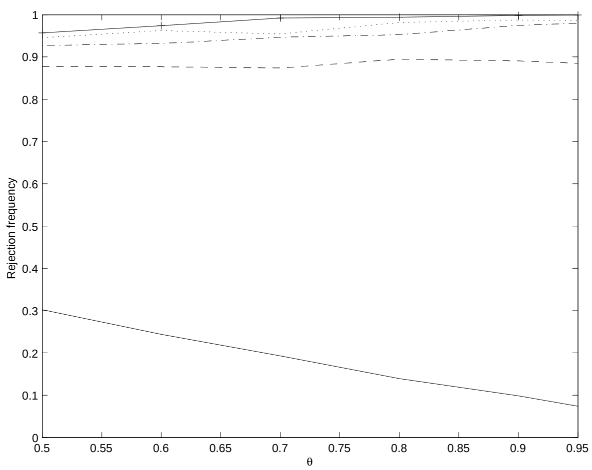

Figure 3.1. Two step Wald statistic

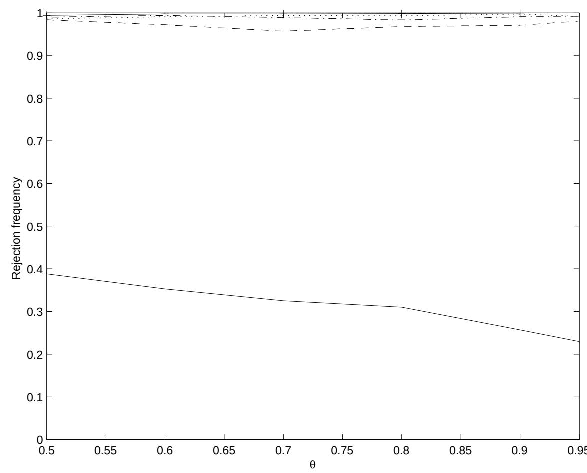

Figure 3.2. CUE Wald statistic

To assess the convergence speed of the finite sample distributions of the two step estimator and CUE to their (non-) standard limiting distributions, we compute the rejection frequencies of testing for different values of θ while the data is generated using DGP 1 with a value of θ_0 close to one. Panel 3 shows the rejection frequencies of the Wald statistics for both estimators. Both Figure 3.1, which contains the rejection frequencies that result for the two step Wald statistic, and 3.2, which contains the rejection frequencies that result for the CUE Wald statistic, show that the Wald statistics are severely size distorted. This shows that the normality of the large sample distribution of the CUE in Theorem 5 is often a bad approximation of its small sample distribution when the true value of θ is close to one. For the two step estimator, the size distortions are as expected since its large sample distribution is not normal.

Since the finite sample distributions of the two step and CUE Wald statistics are badly approximated by their large sample distributions, we construct the large sample distributions of the GMM-AR, GMM-LM and KLM statistics, whose definitions are stated in Appendix B, in order to assess whether they provide better approximations of their finite sample distributions. We also use these statistics, in the next section, to compute the power envelope.

Theorem 6. *Under Assumption 1, the conditions in (3), finite fourth moments of c_i and u_{it}, $i = 1, \ldots, N$, $t = 2, \ldots, T$ and under the convergence sequence (19), the large sample distributions of the GMM-AR, KLM and GMM-LM statistics for values of T larger than three are such that:*

1. *The GMM-AR statistic, defined in Appendix B, has a non-central χ^2 distribution with as many degrees of freedom as moment conditions and non-centrality parameter*

$$N(1-\theta)^2\mu(\sigma^2)'A_f(\theta)_\perp(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-1}A_f(\theta)'_\perp\mu(\sigma^2). \tag{39}$$

2. *The KLM statistic, defined in Appendix B, has a non-central $\chi^2(1)$ distribution with non-centrality parameter*

$$\begin{array}{c} N(1-\theta)^2\mu(\sigma^2)'A_f(\theta)_\perp(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-\frac{1}{2}} \\ P_{g(\theta)}(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-\frac{1}{2}}A_f(\theta)_\perp)'\mu(\sigma^2) \end{array} \tag{40}$$

where $g(\theta)$ is such that

$$\begin{array}{rl} g(\theta) = & (A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-\frac{1}{2}} \\ & A_f(\theta)'_\perp \Big\{ -[I_{\frac{1}{2}(T+1)(T-2)} - (1-\theta)A_q(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)']\mu(\sigma^2) + \\ & [A_qV_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} + B_qV'_{y_1\Delta y,uu} \vdots B_qV_{uu,uu}B_f(\theta)'A_f(\theta)_\perp] \\ & \begin{pmatrix} V_{y_1\Delta y,y_1\Delta y} & V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\ & \begin{pmatrix} 0 \\ (1-\theta)A_f(\theta)'_\perp\mu(\sigma^2) \end{pmatrix} \Big\}. \end{array} \tag{41}$$

3. *The GMM-LM statistic, defined in Appendix B, has a non-central $\chi^2(1)$ distribution with non-centrality parameter*

$$\begin{array}{c} N(1-\theta)^2\mu(\sigma^2)'A_f(\theta)_\perp(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-\frac{1}{2}}P_{h(\theta)} \\ (A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-\frac{1}{2}}A_f(\theta)'_\perp\mu(\sigma^2) \end{array} \tag{42}$$

where $h(\theta)$ is such that

$$h(\theta) = (A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-\frac{1}{2}}A_f(\theta)'_\perp A_q\psi \tag{43}$$

so $h(\theta)$ is a random variable independent of the normal random variable that constitutes the quadratic form that makes up the statistic.

Proof. see Appendix A. ■

The large sample distributions of the GMM-AR and KLM statistics are as expected which also applies to their non-centrality parameters. The non-centrality parameter of the large sample distribution of the GMM-LM statistic is rather unusual since it depends on ψ.

5 Power Envelope

We just showed that the identification of θ gradually improves when the number of time periods increases and the true value of θ is close to one. To determine the strength of identification under all applicable data generating processes, *i.e.* those that satisfy the Sys moment conditions, we construct the power envelope when the true value of θ is equal to one. Usually the power envelope is obtained from the likelihood ratio statistic testing point null against point alternative hypothezes which results from the Neyman-Pearson lemma, see *e.g.* Andrews *et. al.* (2005). Because of the incidental parameter problem that results from the fixed effects c_i, we can, however, not use the likelihood ratio statistic to construct the power envelope since the maximum likelihood estimator is inconsistent. We therefore obtain the power envelope using the least favorable alternative. The least favorable data generating process satisfies the Sys moment conditions and leads to the smallest rejection frequency of $\mathrm{H}_0 : \theta = \theta_0$ for values of θ_0 smaller than one when the true value of θ is equal to one. It results from those data generating processes that have the maximum rate for $h(\theta)$ while still satisfying the Sys moment conditions. This maximum rate results from Assumption 1 so, since

$$y_{i1} = \mu_i + u_{i1}, \tag{44}$$

it has to hold that $\mu_i = \frac{c_i}{1-\theta_0}$ and the variance of u_{i1} is at most of order $(1-\theta_0)^{-2+\varepsilon}$, for some $\varepsilon > 0$. Hence, the maximal rate for $h(\theta_0)$ is proportional to $1-\theta_0$. We also note that the Lev moment conditions do not hold when $\mu_i = \frac{c_i}{(1-\theta_0)^{1+\varepsilon}}$, for some $\varepsilon > 0$. The maximal rate for $h(\theta_0)$ is therefore attained for DGPs 1, 2 and 5.

We construct the power envelope for the identification robust GMM-AR, KLM and GMM-LM statistics.

Theorem 7. *Under Assumption 1, the conditions in (3), finite fourth moments of c_i and u_{it}, $i = 1, \ldots, N$, $t = 2, \ldots, T$, the least favorable data generating process for discriminating between values of θ less than one from a value of θ equal to one while* *the true value is equal to one is such that the non-centrality parameter in the limiting distribution in Theorem 6 has a value of the covariance matrix* $V_{uu,uu.y_1\Delta y}$ *equal to*

$$V_{uu,uu.y_1\Delta y} = \begin{pmatrix} 0 & 0 \\ 0 & \text{diag}(E(u_{i2}^2 - \sigma_2^2)^2,\ \sigma_2^2\sigma_3^2,\ \sigma_2^2\sigma_4^2,\ E(u_{i3}^2 - \sigma_3^2)^2,\ \sigma_3^2\sigma_4^2)) \end{pmatrix}. \quad (45)$$

Proof. see Appendix A. ■

Theorem 7 shows that the large sample distributions under the least favorable alternative do not depend on the variance of the fixed effects, σ_c^2. They only depend on the variance of the disturbances at the different time points and the fourth order moments of the disturbances. We use Theorem 6 to construct the power envelopes for the different statistics. To compute them we use values of σ_t^2 that are constant over time and disturbances that result from a normal distribution so the excess kurtosis is equal to three.

Panels 4-8 contain the power envelopes that result for the GMM-AR, GMM-LM and KLM statistics. All these power envelopes are computed by simulation and for none of them is there any size distortion at the true value. This shows that the large sample distributions of the statistics stated in Theorem 6 provide good approximations of their small sample distributions.

Figures 4.1-4.5 in Panel 4 contain the power envelopes for the GMM-AR statistic where each figure shows the power envelopes that hold for a varying number of cross sectional observations and the same number of time series observations. Figures 5.1-5.3 in Panel 5 contain the power envelopes for the same number of cross sectional observations and varying numbers of time series observations.

Figure 4.1, which is for three time series periods, shows that the Sys moment conditions do not identify θ when $T = 3$ since the power envelope which results from the least favorable alternative (from the perspective of the tested hypothesis) is flat at 5%. This accords with our discussion in Section 3.1. Figures 4.2-4.5 show that the identification improves when either the number of time series observations or the number of cross sectional observations increases. The former is further highlighted in Figures 5.1-5.3 in Panel 5. These figures clearly show that θ is not identified in case of three time series observations and becomes identified as the number of time series observations increases.

Panel 6 contains the power envelopes for the identification robust KLM statistic. It again shows that Sys moment conditions do not identify θ when the true value is close to one and the number of time periods is equal to three. When we increase the number of time periods, θ gradually becomes better identified. For the same number of time series and cross sectional observations, the power envelope of the KLM statistic lies above the power envelope that results for the GMM-AR statistic. This results because of the smaller degree of freedom parameter of the limiting distribution of the KLM statistic compared to that of the GMM-AR statistic.

Panel 7 contains the power envelopes of the GMM-LM and the two step Wald t-statistics. The large sample distributions of both of these statistics are sensitive to the identification of the parameters. Their power envelopes are therefore non-standard. First for the GMM-LM statistic, we see that it is size correct but the power envelope partially goes down when we have more time series observations. Second for the two step Wald statistic, we see that it gets more and more size distorted for larger number of time series observations. This results from the non-standard limiting distribution of the two step estimator stated in Theorem 5 and is shown in Figure 3.1 as well.

Panel 8 contains the power envelopes of the GMM-AR, GMM-LM and KLM statistics for different numbers of time periods and $N = 500$. The power envelope of all these statistics are flat at 5% when $T = 3$ which shows the non-identification of θ by the Sys moment conditions for values of θ_0 close to one. For larger number of time periods, the power envelope of the KLM statistics lies above the power envelope of the other statistic so the KLM statistic is the statistic that maximizes the rejection frequency under the worst possible scenario.

6 Conclusions

We analyze the identification of the parameters by the moment conditions in linear dynamic panel data models. We show that the Dif and Lev moment conditions for general numbers of time series observations and the Sys moment conditions for three time series observations do not always identify the parameters of the panel data model. When there are more than three time series observations, the Sys moment conditions identify the parameters of the panel data model for every data generating process for the initial observations that accords with the moment conditions. The non-identification of the parameters results from the divergence of the initial observations for some, common, data generating processes. This divergence leads to non-standard behavior of estimators or slow convergence to their limiting distributions even in case of the Sys moment conditions with more than three time periods. Statistics based on these estimators, like Wald statistics, therefore often behave badly. The performance of the identification robust GMM statistics, like the GMM-AR and KLM statistic, is, however, excellent. We use these statistics to compute the power envelope which shows that the KLM statistic is the statistic that minimizes the rejection frequency for the worst case data generating process with persistent data.

We analyzed the simplest panel autoregressive model in order to study the identification issues in an isolated manner. In future work, we plan to extend the analysis by incorporating more lags of the dependent variable and additional explanatory variables.

Panel 4. Power envelopes of 95% significance tests of different values of θ while the true value is one using the GMM-AR statistic with Sys moment conditions and varying numbers of T and N, $\sigma_t^2 = 1,\ t = 1, \ldots, T$. $N = 250$ (dashed), 500 (solid), 1000 (dash-dotted).

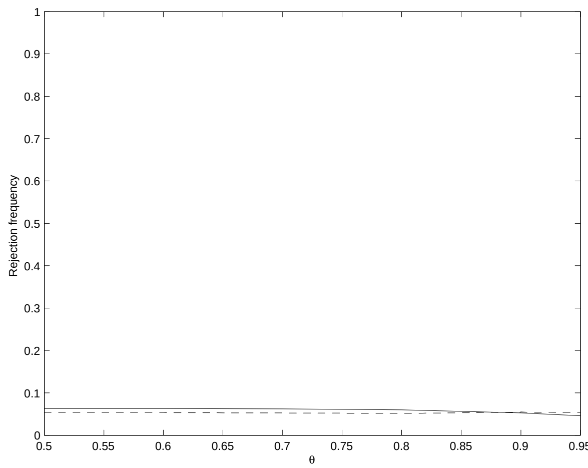

Figure 4.1. $T = 3$

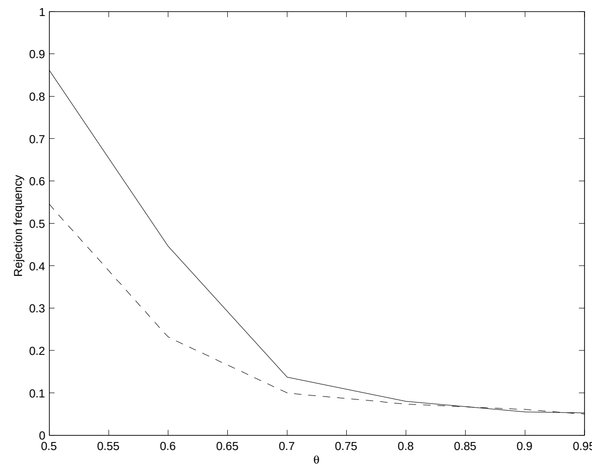

Figure 4.2. $T = 4$

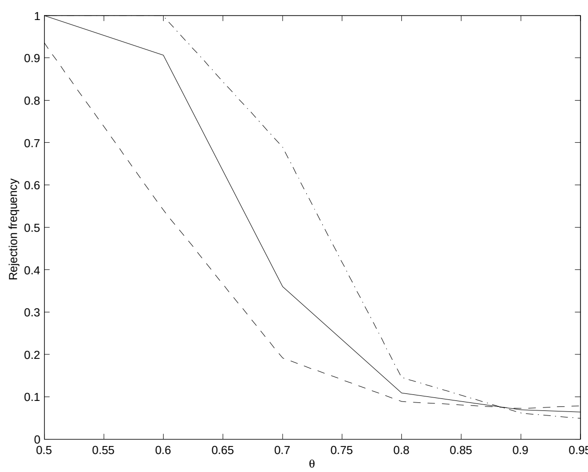

Figure 4.3. $T = 5$

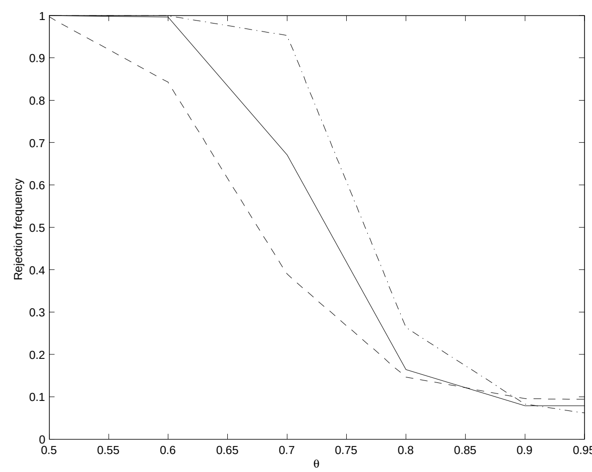

Figure 4.4. $T = 6$

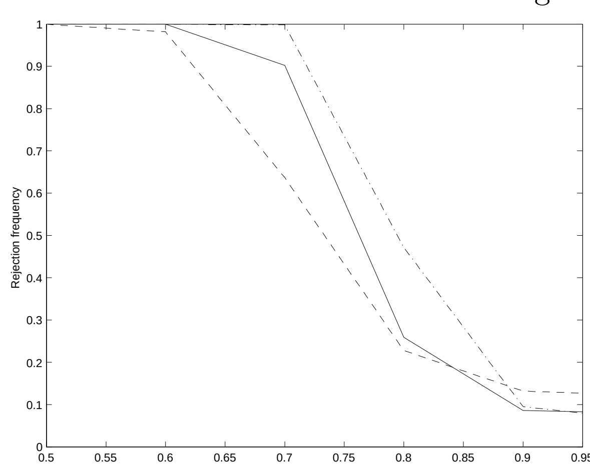

Figure 4.5. $T = 7$

Panel 5. Power envelopes of 95% tests for different values of θ while the true value is one using the GMM-AR statistic with Sys moment conditions and varying numbers of T and N, $\sigma_t^2 = 1$, $t = 1, \ldots, T$. $T = 3$ (solid), 4 (dashed), 5 (dash-dotted), 6 (dotted), 7 (solid with plusses).

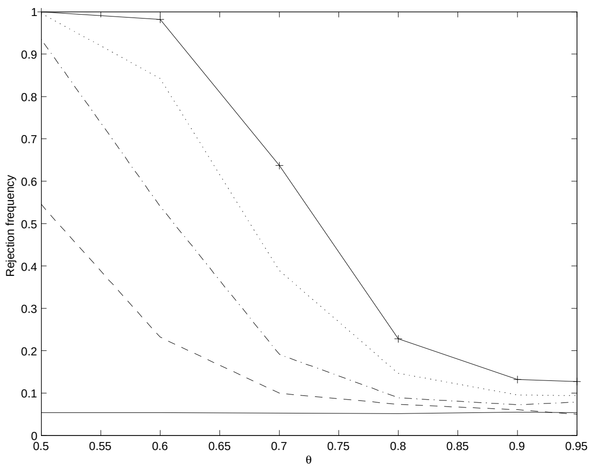

Figure 5.1. $N = 250$

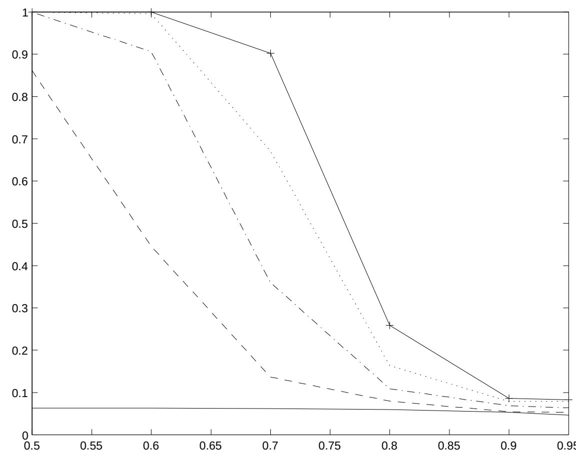

Figure 5.2. $N = 500$

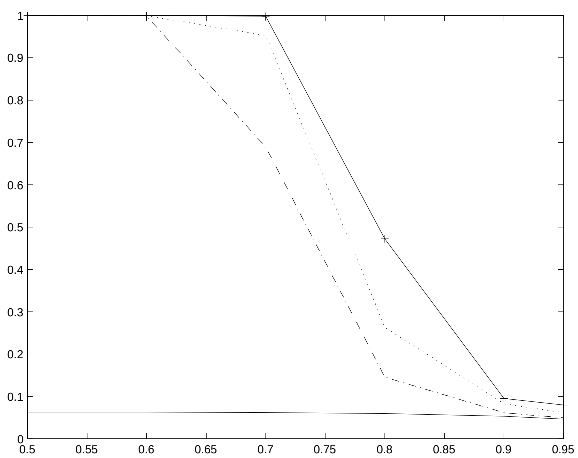

Figure 5.3. $N = 1000$

Panel 6. Power envelopes of 95% tests for different values of θ while the true value value is one using the KLM statistic with Sys moment conditions and varying numbers of T and N, $\sigma_t^2 = 1$, $t = 1, \ldots, T$. $T = 3$ (solid), 4 (dashed), 5 (dash-dotted), 6 (dotted), 7 (solid with plusses).

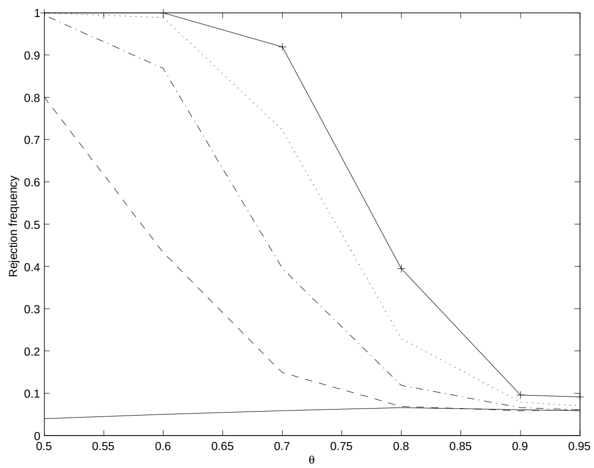

Figure 6.1. $N = 250$

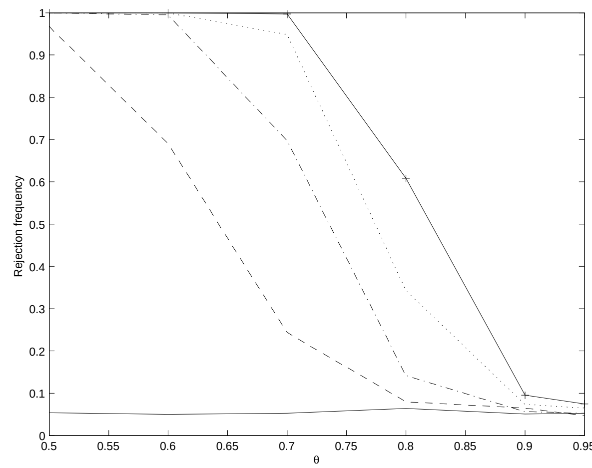

Figure 6.2. $N = 500$

Panel 7. Power envelopes of 95% tests for different values of θ while the true value value is one using the GMM-LM and 2 step Wald t-statistics with Sys moment conditions and varying numbers of T and N, $\sigma_t^2 = 1$, $t = 1, \ldots, T$. $T = 3$ (solid), 4 (dashed), 5 (dash-dotted), 6 (dotted), 7 (solid with plusses).

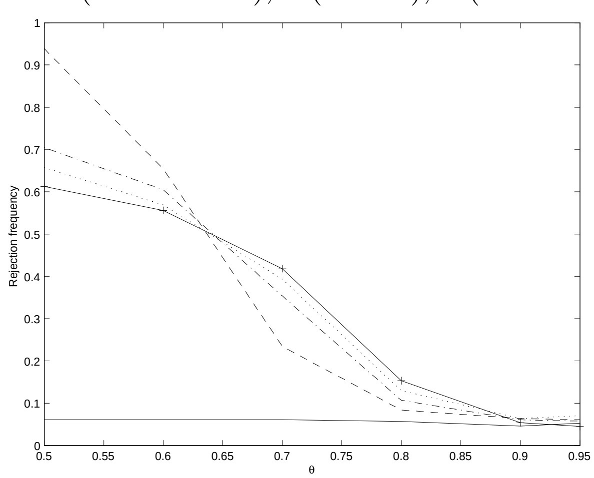

Figure 7.1. $N = 500$, GMM-LM

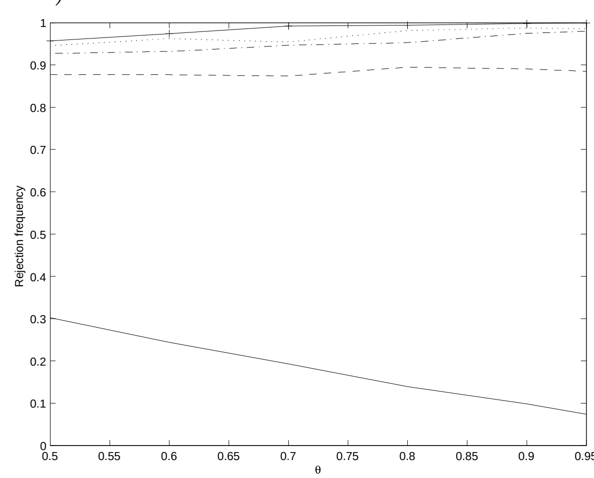

Figure 7.2. $N = 500$, two step Wald t-statistic

Panel 8. Power envelopes of 95% significance tests of different values of θ while the true value is one using the GMM-AR (dash-dot), KLM (dashed) and GMM-LM (solid) statistics, $N = 500$

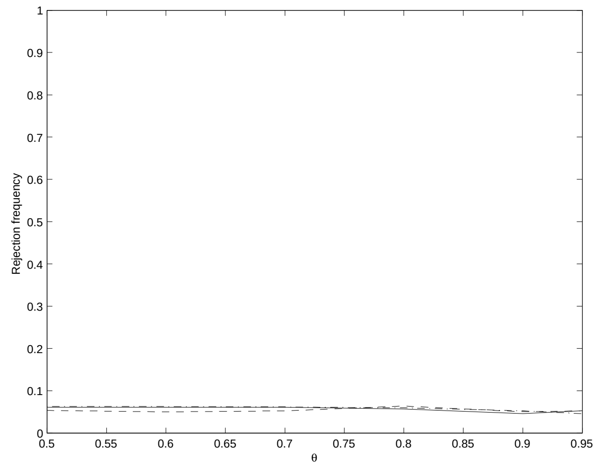

Figure 8.1. $T = 3$

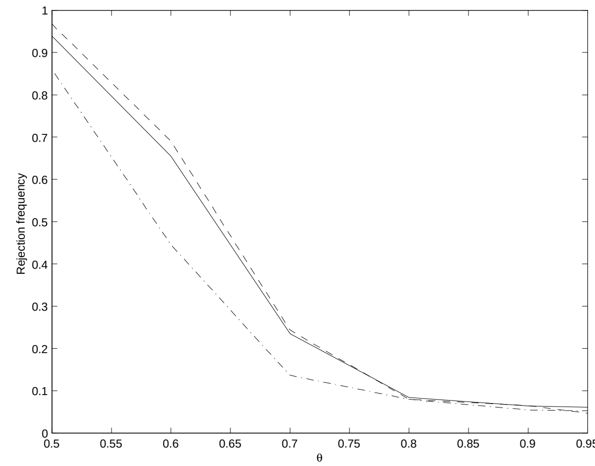

Figure 8.2. $T = 4$

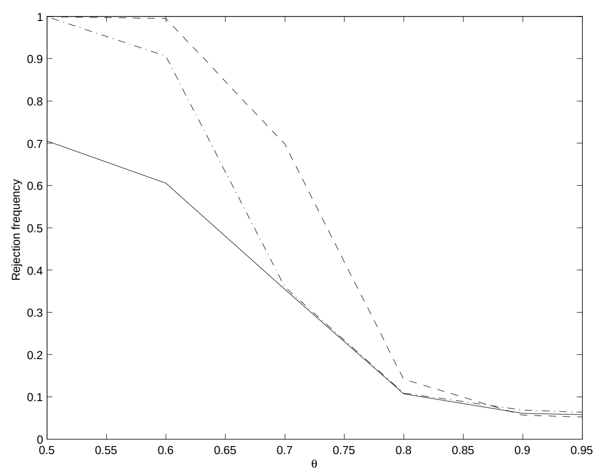

Figure 8.3. $T = 5$

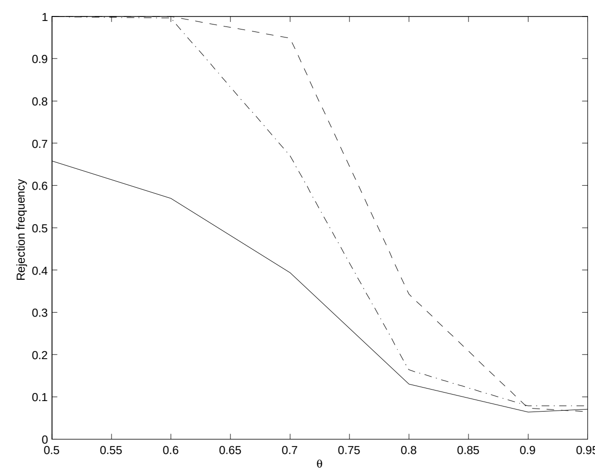

Figure 8.4. $T = 6$

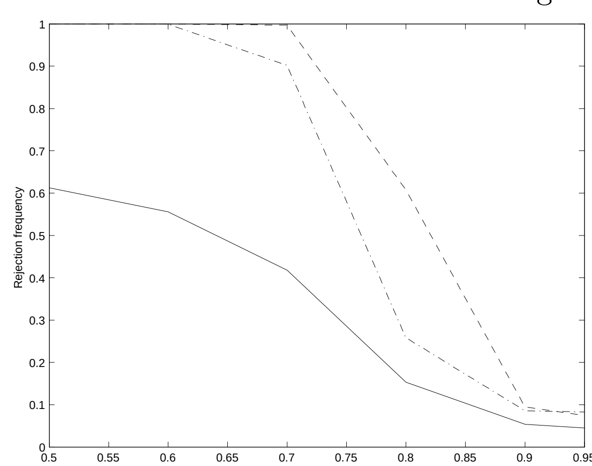

Figure 8.5. $T = 7$

Appendix A. Proofs

Proof of Theorem 1. When $T = 3$, we can specify the large sample behavior of the Sys moment conditions by

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

with

$$\mu(\sigma^2) = \begin{pmatrix} 0 \\ \sigma_2^2 \end{pmatrix}, \ A_f(\theta) = \begin{pmatrix} -\theta & 1 \\ 1-\theta & 0 \end{pmatrix}, \ B_f(\theta) = \begin{pmatrix} 0 & 0 & 0 \\ 1-\theta & 1-\theta & 1 \end{pmatrix}$$

$$\tfrac{h(\theta_0)}{\sqrt{N}}\sum_{i=1}^N \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \end{pmatrix} \underset{d}{\rightarrow} \psi = \begin{pmatrix} \psi_2 \\ \psi_3 \end{pmatrix}, \ \tfrac{1}{\sqrt{N}}\sum_{i=1}^N \begin{pmatrix} c_i u_{i2} \\ u_{i2}^2 - \sigma_2^2 \\ u_{i2}u_{i3} \end{pmatrix} \underset{d}{\rightarrow} \psi_{cu} = \begin{pmatrix} \psi_{cu,1} \\ \psi_{cu,2} \\ \psi_{cu,3} \end{pmatrix}$$

and ψ and ψ_{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

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

and

$$\hat{\theta}_{Sys,1step} = \frac{\left(\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} y_{i1}\Delta y_{i2} \\ y_{i2}\Delta y_{i2} \end{pmatrix}\right)'\left(\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} y_{i1}\Delta y_{i3} \\ y_{i3}\Delta y_{i2} \end{pmatrix}\right)}{\left(\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} y_{i1}\Delta y_{i2}) \\ y_{i2}\Delta y_{i2} \end{pmatrix}\right)'\left(\frac{1}{N}\sum_{i=1}^N \begin{pmatrix} y_{i1}\Delta y_{i2}) \\ y_{i2}\Delta y_{i2} \end{pmatrix}\right)}$$

$$\underset{\substack{\theta_0\uparrow 1,\ h(\theta_0)\sqrt{N}\rightarrow 0 \\ d}}{\longrightarrow} \frac{\psi_2^2 + \psi_3\psi_2}{2\psi_2^2} = 1 + \frac{\psi_3 - \psi_2}{2\psi_2}$$

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

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

which reads

$$h(\theta_0)^2\hat{V}_{ff}(\theta) \underset{\theta_0\uparrow 1,\ h(\theta_0)\sqrt{N}\to 0}{\underset{d}{\longrightarrow}} A_f(\theta)V_{y_1\Delta y,y_1\Delta y}A_f(\theta)'$$

with

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

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

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

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

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

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

so

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

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

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

which implies that

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

with $V_{cu,11}$ and $V_{cu,33}$ the first and third diagonal elements of the covariance matrix of $(c_i\Delta y_{i2}\ \vdots\ u_{i2}^2-\sigma_2^2\ \vdots\ u_{i2}u_{i3})'$ 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

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

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

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

so since

$$q_N(\hat{\theta}_{Sys,2step})'\hat{V}_{ff}(\hat{\theta}_{Sys,2step})^{-1}q_N(\hat{\theta}_{Sys,2step}) \underset{\underset{d}{\theta_0\uparrow 1,\ h(\theta_0)\sqrt{N}\to 0}}{\to} \frac{\psi_2^2}{(1-\hat{\theta}_{2s})^2}\binom{1}{1}'V_{y_1\Delta y,y_1\Delta y}^{-1}\binom{1}{1},$$

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

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

Proof of Theorem 3. The large sample behavior of the GMM-LM statistic of Newey and West (1987):

$$\begin{aligned}\text{GMM-LM}(\theta_0) = \quad & Nf_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}q_N(\theta_0)\left[q_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}q_N(\theta_0)\right]^{-1} \\ & q_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}f_N(\theta_0),\end{aligned}$$

results from its different components. For the Dif and Lev moment conditions, it reads:

$$\begin{aligned} & \text{GMM-LM}_{Dif}(\theta_0) \underset{\underset{d}{\theta_0\uparrow 1,\ h(\theta_0)\sqrt{N}\to 0}}{\to} \\ & \psi'A_f(\theta)_1'\left[A_f(\theta)_1V_{y_1\Delta y,y_1\Delta y}A_f(\theta)_1'\right]^{-1}\binom{1}{0}'\psi\left[\psi'\binom{1}{0}'\left[A_f(\theta)_1V_{y_1\Delta y,y_1\Delta y}A_f(\theta)_1'\right]^{-1}\binom{1}{1}\psi\right]^{-1} \\ & \left[A_f(\theta)_1V_{y_1\Delta y,y_1\Delta y}A_f(\theta)_1'\right]^{-1}A_f(\theta)_1\psi \\ & =\psi'A_f(\theta)_1'\left[A_f(\theta)_1V_{y_1\Delta y,y_1\Delta y}A_f(\theta)_1'\right]^{-1}A_f(\theta)_1\psi\sim\chi^2(1), \\ & \text{GMM-LM}_{Lev}(\theta_0) \underset{\underset{d}{\theta_0\uparrow 1,\ h(\theta_0)\sqrt{N}\to 0}}{\to} \\ & \psi'A_f(\theta)_2'\left[A_f(\theta)_2V_{y_1\Delta y,y_1\Delta y}A_f(\theta)_2'\right]^{-1}\binom{0}{1}'\psi\left[\psi'\binom{0}{1}\left[A_f(\theta)_2V_{y_1\Delta y,y_1\Delta y}A_f(\theta)_2'\right]^{-1}\binom{0}{1}'\psi\right]^{-1} \\ & \left[A_f(\theta)_2V_{y_1\Delta y,y_1\Delta y}A_f(\theta)_2'\right]^{-1}A_f(\theta)_2\psi \\ & =\psi'A_f(\theta)_2'\left[A_f(\theta)_2V_{y_1\Delta y,y_1\Delta y}A_f(\theta)_2'\right]^{-1}A_f(\theta)_2\psi\sim\chi^2(1),\end{aligned}$$

with $A_f(\theta)_1$ and $A_f(\theta)_2$ the first and second row of $A_f(\theta)$. For the Sys moment conditions, the large sample behavior of the LM statistic results as:

$$
\begin{aligned}
&\text{GMM-LM}_{Sys}(\theta_0) \underset{\substack{\theta_0\uparrow 1,\ h(\theta_0)\sqrt{N}\to 0\\ d}}{\longrightarrow} \\
&\psi' A_f(\theta)' \left[A_f(\theta) V_{y_1\Delta y, y_1\Delta y} A_f(\theta)'\right]^{-1} \left(\begin{smallmatrix}1&0\\1&0\end{smallmatrix}\right) \psi \left[\psi' \left(\begin{smallmatrix}1&0\\1&0\end{smallmatrix}\right)' \left[A_f(\theta) V_{y_1\Delta y, y_1\Delta y} A_f(\theta)'\right]^{-1} \left(\begin{smallmatrix}1&0\\1&0\end{smallmatrix}\right) \psi\right]^{-1} \\
&\qquad \left[A_f(\theta) V_{y_1\Delta y, y_1\Delta y} A_f(\theta)'\right]^{-1} A_f(\theta)\psi \\
&\underset{\substack{\theta_0\uparrow 1,\ h(\theta_0)\sqrt{N}\to 0\\ d}}{\longrightarrow} \psi' V^{-1}_{y_1\Delta y, y_1\Delta y} \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) \left[\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)' V^{-1}_{y_1\Delta y, y_1\Delta y} \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)\right]^{-1} \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)' V^{-1}_{y_1\Delta y, y_1\Delta y} \psi \\
&\sim \chi^2(1),
\end{aligned}
$$

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

Proof of Theorem 4. T=3. The specification of $A_f(\theta)$, A_q, $B_f(\theta)$, B_q, $\mu(\sigma^2)$, ψ and ψ_{cu} results from the proof of Theorem 1:

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

with

$$
\mu(\sigma^2) = \begin{pmatrix} 0 \\ \sigma_2^2 \end{pmatrix},\ A_f(\theta) = \begin{pmatrix} -\theta & 1 \\ 1-\theta & 0 \end{pmatrix},\ B_f(\theta) = \begin{pmatrix} 0 & 0 & 0 \\ 1-\theta & 1-\theta & 1 \end{pmatrix}
$$

$$
A_q = \begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix},\ B_q = \begin{pmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \end{pmatrix}
$$

$$
\tfrac{h(\theta_0)}{\sqrt{N}} \textstyle\sum_{i=1}^N \begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \end{pmatrix} \underset{d}{\to} \psi = \begin{pmatrix} \psi_2 \\ \psi_3 \end{pmatrix},\ \tfrac{1}{\sqrt{N}} \textstyle\sum_{i=1}^N \begin{pmatrix} c_i u_{i2} \\ u_{i2}^2 - \sigma_2^2 \\ u_{i2}u_{i3} \end{pmatrix} \underset{d}{\to} \psi_{cu} = \begin{pmatrix} \psi_{cu,1} \\ \psi_{cu,2} \\ \psi_{cu,3} \end{pmatrix}
$$

T=4. We can specify the Sys sample moments and their derivative as

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

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

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

$$\frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} 0 \\ 0 \\ (c_i+u_{i2})\Delta y_{i3} \\ c_i\Delta y_{i2}+u_{i2}\Delta u_{i2} \\ (1+\theta_0)c_i\Delta y_{i3}+u_{i3}\Delta y_{i3}+\theta_0 u_{i2}\Delta y_{i3} \end{pmatrix},$$

so

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

$$B_f(\theta) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\theta & 1 & 0 & -\theta & 1 & 0 & 0 \\ 2-\theta & 0 & 0 & 1-\theta & 1 & 0 & 0 & 0 \\ 0 & 1+2(1-\theta) & 0 & 0 & 1-\theta & 0 & 1-\theta & 1 \end{pmatrix},\ A_q = \begin{pmatrix} 1&0&0 \\ 0&1&0 \\ 0&1&0 \\ 1&0&0 \\ 0&1&0 \end{pmatrix}$$

and

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

$$\frac{h(\theta_0)}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}u_{i2}\\ y_{i1}u_{i3}\\ y_{i1}u_{i4}\end{pmatrix}\underset{d}{\rightarrow}\begin{pmatrix}\psi_{y_{i1}u_{i2}}\\ \psi_{y_{i1}u_{i3}}\\ \psi_{y_{i1}u_{i4}}\end{pmatrix}=\psi.$$

T=5. We can specify the Sys sample moments and their derivative as

$$f_N(\theta)\approx$$

$$\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} -\theta & 1 & 0 & 0\\ 0 & -\theta & 1 & 0\\ 0 & -\theta\theta_0 & \theta_0 & 0\\ 0 & 0 & -\theta & 1\\ 0 & 0 & -\theta\theta_0 & \theta_0\\ 0 & 0 & -\theta\theta_0^2 & \theta_0^2\\ (\theta_0-\theta)\theta_0 & 0 & 0 & 0\\ 0 & (\theta_0-\theta)\theta_0^2 & 0 & 0\\ 0 & 0 & (\theta_0-\theta)\theta_0^3 & 0\end{pmatrix}\left[\begin{pmatrix} y_{i1}u_{i2}\\ y_{i1}u_{i3}\\ y_{i1}u_{i4}\\ y_{i1}u_{i5}\end{pmatrix}+\iota_4 E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2\right]+$$

$$\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} 0\\ 0\\ (c_i+u_{i2})(\Delta y_{i4}-\theta\Delta y_{i3})\\ 0\\ (c_i+u_{i2})(\Delta y_{i5}-\theta\Delta y_{i4})\\ (c_i(1+\theta_0)+\theta_0u_{i2}+u_{i3})(\Delta y_{i5}-\theta\Delta y_{i4})\\ (1+(\theta_0-\theta))c_i\Delta y_{i2}+(\theta_0-\theta)u_{i2}\Delta y_{i2}+u_{i3}\Delta y_{i2}\\ (1+(\theta_0-\theta)(1+\theta_0))c_i\Delta y_{i3}+(\theta_0-\theta)u_{i3}\Delta y_{i3}+(\theta_0-\theta)\theta_0u_{i2}\Delta y_{i3}+u_{i4}\Delta y_{i3}\\ (1+(\theta_0-\theta)(1+\theta_0+\theta_0^2))c_i\Delta y_{i4}+(\theta_0-\theta)\theta_0^2u_{i2}\Delta y_{i4}+\\ +(\theta_0-\theta)\theta_0u_{i3}\Delta y_{i4}+(\theta_0-\theta)u_{i4}\Delta y_{i4}+u_{i5}\Delta y_{i4}\end{pmatrix}$$

$$q_N(\theta) \approx -\frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \theta_0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \theta_0 & 0 \\ 0 & 0 & \theta_0^2 & 0 \\ \theta_0 & 0 & 0 & 0 \\ 0 & \theta_0^2 & 0 & 0 \\ 0 & 0 & \theta_0^3 & 0 \end{pmatrix} \left[\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \\ y_{i1}u_{i5} \end{pmatrix} + \iota_4 E(\lim_{\theta_0 \uparrow 1}(1-\theta_0)u_{i1}^2 \right] -$$

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

so

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

$$B_f(\theta) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\theta & 1 & 0 & 0 & -\theta & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\theta & 1 & 0 & 0 & -\theta & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2\theta & 2 & 0 & 0 & -\theta & 1 & 0 & -\theta & 1 & 0 & 0 \\ 2-\theta & 0 & 0 & 0 & 1-\theta & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3-2\theta & 0 & 0 & 0 & 1-\theta & 0 & 0 & 1-\theta & 1 & 0 & 0 & 0 \\ 0 & 0 & 4-3\theta & 0 & 0 & 0 & 1-\theta & 0 & 0 & 1-\theta & 0 & 1-\theta & 1 \end{pmatrix}$$

$$B_q = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & -1 & 0 \end{pmatrix}$$

and

$$\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} c_i u_{i2} \\ c_i u_{i3} \\ c_i u_{i4} \\ c_i u_{i5} \\ u_{i2}^2 \\ u_{i2}u_{i3} \\ u_{i2}u_{i4} \\ u_{i2}u_{i5} \\ u_{i3}^2 \\ u_{i3}u_{i4} \\ u_{i3}u_{i4} \\ u_{i4}^2 \\ u_{i4}u_{i5} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \sigma_2^2 \\ 0 \\ 0 \\ 0 \\ \sigma_3^2 \\ 0 \\ 0 \\ \sigma_4^2 \\ 0 \end{pmatrix}\right] \underset{d}{\rightarrow} \begin{pmatrix} \psi_{c_i u_{i2}} \\ \psi_{c_i u_{i3}} \\ \psi_{c_i u_{i4}} \\ \psi_{c_i u_{i5}} \\ \psi_{u_{i2}u_{i2}} \\ \psi_{u_{i2}u_{i3}} \\ \psi_{u_{i2}u_{i4}} \\ \psi_{u_{i2}u_{i5}} \\ \psi_{u_{i3}u_{i3}} \\ \psi_{u_{i3}u_{i4}} \\ \psi_{u_{i3}u_{i5}} \\ \psi_{u_{i4}u_{i4}} \\ \psi_{u_{i4}u_{i5}} \end{pmatrix} = \psi_{cu}$$

$$\frac{h(\theta_0)}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} y_{i1}u_{i2} \\ y_{i1}u_{i3} \\ y_{i1}u_{i4} \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_{y_{i1}u_{i2}} \\ \psi_{y_{i1}u_{i3}} \\ \psi_{y_{i1}u_{i4}} \end{pmatrix} = \psi.$$

Proof of Theorem 5. We show the consistency of the two step estimator and CUE in two different manners. First, we show the consistency of the two step estimator using its closed form expression that involves the one step estimator. Second, for the consistency of the CUE, we use the convergence of the objective function and then invoke the results from Newey and McFadden (1994). Before we derive either one of these two results, we analyze the large sample behavior of the covariance estimator

$$\hat{V}_{ff}(\theta) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N}(f_i(\theta) - \bar{f})(f_i(\theta) - \bar{f})',$$

its inverse and some functions of it. To obtain the large sample behavior of the covariance matrix estimator, it is convenient to specify it as

$$(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'\hat{V}_{ff}(\theta)(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp),$$

with $A_f(\theta)_\perp$ a $\frac{1}{2}(T+1)(T-2) \times \left(\frac{1}{2}(T-1)(T-2)-1\right)$ dimensional matrix which is the orthogonal complement of $A_f(\theta)$, so $A_f(\theta)_\perp' A_f(\theta) \equiv 0$ and $A_f(\theta)_\perp' A_f(\theta)_\perp \equiv$

$I_{\left(\frac{1}{2}(T-1)(T-2)-1\right)}$. For values of θ_0 close to one, we decompose $\hat{V}_{ff}(\theta)$ as follows:

$$(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'\hat{V}_{ff}(\theta)(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp) =$$
$$\begin{pmatrix} \frac{1}{h(\theta_0)^2}\hat{V}_{y_1\Delta y, y_1\Delta y} & \frac{1}{h(\theta_0)}\hat{V}_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ \frac{1}{h(\theta_0)}A_f(\theta)'_\perp B_f(\theta)\hat{V}'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)\hat{V}_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix} \underset{p}{\rightarrow}$$
$$\begin{pmatrix} \frac{1}{h(\theta_0)^2}V_{y_1\Delta y, y_1\Delta y} & \frac{1}{h(\theta_0)}V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ \frac{1}{h(\theta_0)}A_f(\theta)'_\perp B_f(\theta)V'_{c\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}.$$

where

$$\hat{V}_{y_1\Delta y, y_1\Delta y} = \frac{h(\theta_0)^2}{N}\sum_{i=1}^N \left[\begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT}\end{pmatrix} - \begin{pmatrix} \overline{y_{1i}u_{i2}} \\ \vdots \\ \overline{y_{1i}u_{iT}}\end{pmatrix}\right]\left[\begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT}\end{pmatrix} - \begin{pmatrix} \overline{y_{1i}u_{i2}} \\ \vdots \\ \overline{y_{1i}u_{iT}}\end{pmatrix}\right]'$$

$$\hat{V}_{y_1\Delta y, uu} = \frac{h(\theta_0)}{N}\sum_{i=1}^N \left[\begin{pmatrix} y_{1i}u_{i2} \\ \vdots \\ y_{1i}u_{iT}\end{pmatrix} - \begin{pmatrix} \overline{y_{1i}u_{i2}} \\ \vdots \\ \overline{y_{1i}u_{iT}}\end{pmatrix}\right]\left[\begin{pmatrix} c_iu_{i2} \\ \vdots \\ c_iu_{iT} \\ u_{i2}^2 \\ u_{i2}u_{i3} \\ \vdots \\ u_{i2}u_{iT} \\ \vdots \\ u_{iT-1}u_{iT}\end{pmatrix} - \begin{pmatrix} \overline{c_iu_{i2}} \\ \vdots \\ \overline{c_iu_{iT}} \\ \overline{u_{i2}u_{i2}} \\ \overline{u_{i2}u_{i3}} \\ \vdots \\ \overline{u_{i2}u_{iT}} \\ \vdots \\ \overline{u_{iT-1}u_{iT}}\end{pmatrix}\right]'$$

$$\hat{V}_{uu,uu} = \frac{1}{N}\sum_{i=1}^N \left[\begin{pmatrix} c_iu_{i2} \\ \vdots \\ c_iu_{iT} \\ u_{i2}^2 \\ u_{i2}u_{i3} \\ \vdots \\ u_{i2}u_{iT} \\ \vdots \\ u_{iT-1}u_{iT}\end{pmatrix} - \begin{pmatrix} \overline{c_iu_{i2}} \\ \vdots \\ \overline{c_iu_{iT}} \\ \overline{u_{i2}u_{i2}} \\ \overline{u_{i2}u_{i3}} \\ \vdots \\ \overline{u_{i2}u_{iT}} \\ \vdots \\ \overline{u_{iT-1}u_{iT}}\end{pmatrix}\right]\left[\begin{pmatrix} c_iu_{i2} \\ \vdots \\ c_iu_{iT} \\ u_{i2}^2 \\ u_{i2}u_{i3} \\ \vdots \\ u_{i2}u_{iT} \\ \vdots \\ u_{iT-1}u_{iT}\end{pmatrix} - \begin{pmatrix} \overline{c_iu_{i2}} \\ \vdots \\ \overline{c_iu_{iT}} \\ \overline{u_{i2}u_{i2}} \\ \overline{u_{i2}u_{i3}} \\ \vdots \\ \overline{u_{i2}u_{iT}} \\ \vdots \\ \overline{u_{iT-1}u_{iT}}\end{pmatrix}\right]'$$

with $\overline{ab} = \frac{1}{N}\sum_{i=1}^{N} a_i b_i$. Using the properties of the partioned inverse, we can show that the inverse of the above covariance matrix estimator behaves like

$$\left[(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'\hat{V}_{ff}(\theta)(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)\right]^{-1} =$$

$$\begin{pmatrix} h(\theta_0)^2\hat{V}^{-1}_{y_1\Delta y,y_1\Delta y.uu} \\ -h(\theta_0)\left[A_f(\theta)'_\perp B_f(\theta)\hat{V}_{uu,uu}B_f(\theta)'A_f(\theta)_\perp\right]^{-1} A_f(\theta)'_\perp B_f(\theta)\hat{V}'_{y_1\Delta y,uu}\hat{V}^{-1}_{y_1\Delta y,y_1\Delta y.uu} \\ -h(\theta_0)\hat{V}^{-1}_{y_1\Delta y,y_1\Delta y.uu}\hat{V}_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp\left[A_f(\theta)'_\perp B_f(\theta)\hat{V}_{uu,uu}B_f(\theta)'A_f(\theta)_\perp\right]^{-1} \\ \left[A_f(\theta)'_\perp B_f(\theta)\hat{V}_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp\right]^{-1} \end{pmatrix}$$

$$\underset{p}{\rightarrow} \begin{pmatrix} h(\theta_0)^2 V^{-1}_{y_1\Delta y,y_1\Delta y.uu} \\ -h(\theta_0)\left[A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp\right]^{-1} A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y.uu} \\ -h(\theta_0)V^{-1}_{y_1\Delta y,y_1\Delta y.uu}V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp\left[A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp\right]^{-1} \\ \left[A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp - A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y}V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp\right]^{-1} \end{pmatrix}$$

with $\hat{V}_{y_1\Delta y,y_1\Delta y.uu} = \hat{V}_{y_1\Delta y,c\Delta y} - \hat{V}_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp(A_f(\theta)'_\perp B_f(\theta)\hat{V}_{uu,uu}B_f(\theta)'A_f(\theta)_\perp)^{-1}A_f(\theta)'_\perp B_f(\theta)\hat{V}'_{y_1}$ $\hat{V}_{uu,uu.y_1\Delta y} = \hat{V}_{uu,uu} - \hat{V}'_{y_1\Delta y,uu}\hat{V}^{-1}_{y_1\Delta y,y_1\Delta y}\hat{V}_{y_1\Delta y,uu}$.

We also decompose the Sys sample moments:

$$(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'f_N(\theta) = \begin{pmatrix} i \\ ii \end{pmatrix}$$

$$\begin{aligned} i = &\ \tfrac{1}{h(\theta_0)\sqrt{N}}\psi + \iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2) + (A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \\ ii = &\ A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \end{aligned}$$

which implies that

$$\left[(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'\hat{V}_{ff}(\theta)(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)\right]^{-1}$$
$$(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'f_N(\theta) \underset{d}{\rightarrow} \begin{pmatrix} \text{i} \\ \text{ii} \end{pmatrix}$$

with

$$\begin{aligned} \text{i} = &\ h(\theta_0)V^{-1}_{y_1\Delta y,y_1\Delta y.uu}\Big\{(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\Big[h(\theta_0)\Big[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}}B_f(\theta)\psi_{cu} + \\ &\ A_f(\theta)\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)\Big] + \tfrac{1}{\sqrt{N}}A_f(\theta)\psi\Big] - V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp(A_f(\theta)'_\perp B_f(\theta) \\ &\ V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp)^{-1}A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right]\Big\} \\ = &\ h(\theta_0)V^{-1}_{y_1\Delta y,y_1\Delta y.uu}\Big\{\tfrac{1}{\sqrt{N}}\psi + h(\theta_0)(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'[(1-\theta)\mu(\sigma^2) + \\ &\ A_f(\theta)\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)]\Big\} - h(\theta_0)V^{-1}_{y_1\Delta y,y_1\Delta y.uu}V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ &\ (A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp)^{-1}A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \end{aligned}$$

$$\text{ii} = \left[A_f(\theta)'_\perp B_f(\theta) V_{uu,uu.y_1\Delta y} B_f(\theta)' A_f(\theta)_\perp\right]^{-1} \left\{-A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y,uu} V^{-1}_{y_1\Delta y,y_1\Delta y} \right.$$
$$(A_f(\theta)'A_f(\theta))^{-1} A_f(\theta)' \left[h(\theta_0)\left((1-\theta)\mu(\sigma^2) + A_f(\theta)\iota_{T-2} E(\lim_{\theta_0 \uparrow 1}(1-\theta_0)u_{i1}^2)\right) + \right.$$
$$\left.\left.\tfrac{1}{\sqrt{N}} A_f(\theta)\psi + h(\theta_0)\tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu}\right] + A_f(\theta)'_\perp \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu}\right]\right\}.$$

We use the above results to obtain the large sample distribution of the two step estimator. The two step estimator uses the one step estimator

$$\hat{\theta}_{1s} = -(\psi' A_q' A_q \psi)^{-1} \psi' A_q A_{f,1} \psi = 1 - (\psi' A_q' A_q \psi)^{-1} \psi' A_q A_f(1)\psi,$$

since $A_{f,1} = A_f(1) - A_q$, which is inconsistent as $A_f(1)$ does not go to zero and where we used that the large sample behavior of $q_N(\theta)$ is governed by $\frac{1}{h(\theta)} A_q \psi$, see Theorem 4. The large sample expression of the two step estimator then reads

$$\begin{aligned}
\hat{\theta}_{2s} = {} & -\left(\tfrac{1}{h(\theta_0)}\psi' A_q' \hat{V}_{ff}(\hat{\theta}_{1s})^{-1} A_q \tfrac{1}{h(\theta_0)}\psi\right)^{-1} \tfrac{1}{h(\theta_0)}\psi' A_q \hat{V}_{ff}(\hat{\theta}_{1s})^{-1} \tfrac{1}{h(\theta_0)} A_{f,1}\psi \\
= {} & 1 - \left(\tfrac{1}{h(\theta_0)}\psi' A_q' \hat{V}_{ff}(\hat{\theta}_{1s})^{-1} A_q \psi \tfrac{1}{h(\theta_0)}\right)^{-1} \tfrac{1}{h(\theta_0)}\psi' A_q \hat{V}_{ff}(\hat{\theta}_{1s})^{-1} A_f(1)\psi \tfrac{1}{h(\theta_0)}, \\
= {} & 1 - \Big(\tfrac{1}{h(\theta_0)^2}\psi' A_q' A_f(\hat{\theta}_{1s})_\perp \left[A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V_{uu,uu.y_1\Delta y} B_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s})_\perp\right]^{-1} A_f(\hat{\theta}_{1s})'_\perp A_q \psi\Big)^{-1} \\
& \tfrac{1}{h(\theta_0)}\psi' A_q' A_f(\hat{\theta}_{1s})_\perp \left[A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V_{uu,uu.y_1\Delta y} B_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s})_\perp\right]^{-1} \\
& \Big[A_f(\hat{\theta}_{1s})'_\perp \left[(1-\hat{\theta}_{1s})\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\hat{\theta}_{1s})\psi_{cu}\right] - \\
& A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V'_{y_1\Delta y,uu} V^{-1}_{y_1\Delta y,y_1\Delta y} (A_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s}))^{-1} A_f(\hat{\theta}_{1s})' \tfrac{1}{\sqrt{N}} A_f(\hat{\theta}_{1s})\psi\Big] \\
= {} & 1 - h(\theta_0)\Big(\psi' A_q' A_f(\hat{\theta}_{1s})_\perp \left[A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V_{uu,uu.y_1\Delta y} B_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s})_\perp\right]^{-1} A_f(\hat{\theta}_{1s})'_\perp A_q \psi\Big)^{-1} \\
& \psi' A_q' A_f(\hat{\theta}_{1s})_\perp \left[A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V_{uu,uu.y_1\Delta y} B_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s})_\perp\right]^{-1} \\
& \Big[A_f(\hat{\theta}_{1s})'_\perp \left[(1-\hat{\theta}_{1s})\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\hat{\theta}_{1s})\psi_{cu}\right] - \\
& A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V'_{y_1\Delta y,uu} V^{-1}_{y_1\Delta y,y_1\Delta y} (A_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s}))^{-1} A_f(\hat{\theta}_{1s})' \tfrac{1}{\sqrt{N}} A_f(\hat{\theta}_{1s})\psi\Big],
\end{aligned}$$

which shows that the large sample distribution of the two step estimator for values of θ_0 close to one is characterized by

$$\begin{aligned}
& h(\theta_0)^{-1}(\hat{\theta}_{2s} - 1) \underset{d}{\rightarrow} \\
& \Big(\psi' A_q' A_f(\hat{\theta}_{1s})_\perp \left[A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V_{uu,uu.y_1\Delta y} B_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s})_\perp\right]^{-1} A_f(\hat{\theta}_{1s})'_\perp A_q \psi\Big)^{-1} \\
& \psi' A_q' A_f(\hat{\theta}_{1s})_\perp \left[A_f(\hat{\theta}_{1s})'_\perp B_f(\hat{\theta}_{1s}) V_{uu,uu.y_1\Delta y} B_f(\hat{\theta}_{1s})' A_f(\hat{\theta}_{1s})_\perp\right]^{-1} A_f(\hat{\theta}_{1s})'_\perp \\
& \left[(1-\hat{\theta}_{1s})\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\hat{\theta}_{1s})\left(\psi_{cu} - V'_{y_1\Delta y,uu} V^{-1}_{y_1\Delta y,y_1\Delta y}\psi\right)\right],
\end{aligned}$$

so the two step estimator is consistent but with a non-standard convergence rate, $h(\theta_0)$, and a non-standard large sample distribution.

To show the consistency of the CUE, we construct the large sample behavior of the objective function for the CUE, the GMM-AR statistic divided by N :

$$Q_{CUE}(\theta) = \frac{1}{N} GMM\text{-}AR(\theta) = f_N(\theta)' \hat{V}_{ff}(\theta)^{-1} f_N(\theta).$$

As stated in Theorem 6, the large sample behavior of GMM-AR statistic for values of θ_0 close to one is characterized by

$$GMM\text{-}AR(\theta) \underset{d}{\rightarrow} \begin{pmatrix} \psi \\ A_f(\theta)'_\perp \left[\sqrt{N}(1-\theta)\mu(\sigma^2) + B_f(\theta)\psi_{cu}\right] \end{pmatrix}' \begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y, uu} B_f(\theta)' A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y, uu} & A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)' A_f(\theta)_\perp \end{pmatrix}^{-1} \begin{pmatrix} \psi \\ A_f(\theta)'_\perp \left[\sqrt{N}(1-\theta)\mu(\sigma^2) + B_f(\theta)\psi_{cu}\right] \end{pmatrix}.$$

When we divide by N, we obtain the large sample behavior of the CUE objective function

$$Q_{CUE}(\theta) \underset{d}{\rightarrow} \begin{pmatrix} \frac{1}{\sqrt{N}}\psi \\ A_f(\theta)'_\perp \left[(1-\theta)\mu(\sigma^2) + B_f(\theta)\frac{1}{\sqrt{N}}\psi_{cu}\right] \end{pmatrix}' \begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y, uu} B_f(\theta)' A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y, uu} & A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)' A_f(\theta)_\perp \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{N}}\psi \\ A_f(\theta)'_\perp \left[(1-\theta)\mu(\sigma^2) + B_f(\theta)\frac{1}{\sqrt{N}}\psi_{cu}\right] \end{pmatrix}.$$

Alternatively, this can be considered as the large sample behavior of the GMM objective function that results from the sample moment

$$(h(\theta_0) A_f(\theta)(A_f(\theta)' A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)' f_N(\theta).$$

The expected value of this sample moment for large samples equals

$$\begin{pmatrix} 0 \\ A_f(\theta)'_\perp (1-\theta)\mu(\sigma^2) \end{pmatrix}$$

so the quadratic form with respect to the covariance matrix of this rescaled sample moment, is clearly minimized when θ equals 1. Since the sample objective function uniformly converges to the population objective function, the CUE is consistent by Lemma 2.3 in Newey and McFadden (1994).

To obtain the convergence rate and large sample distribution of the CUE, we first need the large sample distribution of $\hat{D}_N(\theta)$ for which we construct the expression for $\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_N(\theta)$, for which we start out with $\hat{V}_{\theta f}(\theta)(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)$:

$$\begin{array}{l}
\hat{V}_{\theta f}(\theta)(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp) = \\
\quad \Big\{ \frac{1}{(h(\theta_0))^2} A_q \hat{V}_{y_1\Delta y, y_1\Delta y} + \frac{1}{h(\theta_0)}(A_q \hat{V}_{y_1\Delta y, uu} B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} + \\
\quad B_q \hat{V}'_{y_1\Delta y, uu}) + B_q \hat{V}_{uu,uu} B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots \\
\quad \frac{1}{h(\theta_0)} A_q \hat{V}_{y_1\Delta y, uu} B_f(\theta)'A_f(\theta)_\perp + B_q \hat{V}_{uu,uu} B_f(\theta)'A_f(\theta)_\perp \Big\} \\
\underset{p}{\rightarrow} \left(A_q \vdots 0\right) \begin{pmatrix} \frac{1}{h(\theta_0)^2} V_{y_1\Delta y, y_1\Delta y} & \frac{1}{h(\theta_0)} V_{y_1\Delta y, uu} B_f(\theta)'A_f(\theta)_\perp \\ \frac{1}{h(\theta_0)} A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y, uu} & A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)'A_f(\theta)_\perp \end{pmatrix} + \\
\quad \frac{1}{h(\theta_0)} \left(A_q V_{y_1\Delta y, uu} B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} + B_q V'_{y_1\Delta y, uu} \vdots 0 \right) + \\
\quad B_q V_{uu,uu} B_f(\theta)'(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)
\end{array}$$

We post-multiply it with our limit previous expression for

$$\left[(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)' \hat{V}_{ff}(\theta)(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp) \right]^{-1}$$

$$(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)' f_N(\theta),$$

to obtain the limit expression of $\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_N(\theta)$:

$$
\begin{aligned}
&-\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_N(\theta) \\
&\underset{d}{\rightarrow}\left[\left(A_q \vdots 0\right)\begin{pmatrix} \frac{1}{h(\theta_0)^2}V_{y_1\Delta y,y_1\Delta y} & \frac{1}{h(\theta_0)}V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ \frac{1}{h(\theta_0)}A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}+\right. \\
&\frac{1}{h(\theta_0)}\left(A_qV_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1}+B_qV'_{y_1\Delta y,uu}\vdots 0\right)+ \\
&B_qV_{uu,uu}B_f(\theta)'(A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1}\vdots A_f(\theta)_\perp)] \\
&\begin{pmatrix} \frac{1}{h(\theta_0)^2}V_{y_1\Delta y,y_1\Delta y} & \frac{1}{h(\theta_0)}V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ \frac{1}{h(\theta_0)}A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\
&\begin{pmatrix} \frac{1}{h(\theta_0)\sqrt{N}}\psi+(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\left[(1-\theta)\mu(\sigma^2)+\frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}+A_f(\theta)\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)\right] \\ A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2)+\frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \end{pmatrix} \\
&=\frac{1}{h(\theta_0)\sqrt{N}}A_q\psi+A_q(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\left[(1-\theta)\mu(\sigma^2)+A_f(\theta)\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)+\right. \\
&\left.\frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right]\left[\left(A_qV_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1}+B_qV'_{y_1\Delta y,uu}\vdots 0\right)+\right. \\
&B_qV_{uu,uu}B_f(\theta)'(h(\theta_0)A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1}\vdots A_f(\theta)_\perp)] \\
&\begin{pmatrix} V_{y_1\Delta y,y_1\Delta y} & V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\
&\begin{pmatrix} \frac{1}{\sqrt{N}}\psi+h(\theta_0)(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\left[(1-\theta)\mu(\sigma^2)+\frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}+A_f(\theta)\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)\right] \\ A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2)+\frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \end{pmatrix}
\end{aligned}
$$

so the expression for $\hat{D}_N(\theta)=q_N(\theta)-\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_N(\theta)$ becomes:

$$
\begin{aligned}
\hat{D}_N(\theta)= &\ -\mu(\sigma^2)-\frac{1}{h(\theta_0)\sqrt{N}}A_q\psi-\frac{1}{\sqrt{N}}B_q\psi_{cu}-\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_N(\theta) \\
\underset{d}{\rightarrow} &\ -[I_5-(1-\theta)A_q(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)']\mu(\sigma^2)+ \\
&[A_qV_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1}+B_qV'_{y_1\Delta y,uu}\vdots B_qV_{uu,uu}B_f(\theta)'A_f(\theta)_\perp] \\
&\begin{pmatrix} V_{y_1\Delta y,y_1\Delta y} & V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\
&\begin{pmatrix} \frac{1}{\sqrt{N}}\psi \\ A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2)+\frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \end{pmatrix}.
\end{aligned}
$$

so

$$\begin{array}{l}(h(\theta_0)A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'\hat{D}_N(\theta) = \\ \underset{d}{\rightarrow} \quad -A_f(\theta)'_\perp[I_5 - (1-\theta)A_q(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)']\mu(\sigma^2) + \\ \quad A_f(\theta)'_\perp[A_qV_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} + B_qV'_{y_1\Delta y,uu} \vdots B_qV_{uu,uu}B_f(\theta)'A_f(\theta)_\perp] \\ \quad \begin{pmatrix} V_{y_1\Delta y,y_1\Delta y} & V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\ \quad \begin{pmatrix} \frac{1}{\sqrt{N}}\psi \\ A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2) + \frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \end{pmatrix}.\end{array}$$

The CUE results from the first order condition:

$$\hat{D}_N(\hat{\theta}_{CUE})'\hat{V}_{ff}(\hat{\theta}_{CUE})^{-1}f_N(\hat{\theta}_{CUE}) = 0.$$

The derivative of $\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta)$ is such that

$$\tfrac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta) = \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta),$$

see Kleibergen and Mavroeidis (2010), which we use to obtain a Taylor expansion of $\hat{V}_{ff}(\hat{\theta}_{CUE})^{-\frac{1}{2}}f_N(\hat{\theta}_{CUE})$ around one:

$$\hat{V}_{ff}(\hat{\theta}_{CUE})^{-\frac{1}{2}}f_N(\hat{\theta}_{CUE}) \approx \hat{V}_{ff}(1)^{-\frac{1}{2}}f_N(1) + \hat{V}_{ff}(1)^{-\frac{1}{2}}\hat{D}_N(1)(\hat{\theta}_{CUE} - 1),$$

which we insert into the first order condition to obtain:

$$\hat{\theta}_{CUE} \approx 1 + \left(\hat{D}_N(\hat{\theta}_{CUE})'\hat{V}_{ff}(\hat{\theta}_{CUE})^{-\frac{1}{2}}\hat{V}_{ff}(1)^{-\frac{1}{2}}\hat{D}_N(1)\right)^{-1}\hat{D}_N(\hat{\theta}_{CUE})'\hat{V}_{ff}(\hat{\theta}_{CUE})^{-\frac{1}{2}}\hat{V}_{ff}(1)^{-\frac{1}{2}}f_N(1),$$

because $\hat{\theta}_{CUE}$ is consistent, we equate it to the true value of 1 to obtain its large sample distribution and convergence speed:

$$\begin{array}{rcl}\hat{\theta}_{CUE} & \approx & 1 + \left(\hat{D}_N(\hat{\theta}_{CUE})'\hat{V}_{ff}(\hat{\theta}_{CUE})^{-\frac{1}{2}}\hat{V}_{ff}(1)^{-\frac{1}{2}}\hat{D}_N(1)\right)^{-1}\hat{D}_N(\hat{\theta}_{CUE})'\hat{V}_{ff}(\hat{\theta}_{CUE})^{-\frac{1}{2}}\hat{V}_{ff}(1)^{-\frac{1}{2}}f_N(1) \\ & \approx & 1 + \left(\hat{D}_N(1)'\hat{V}_{ff}(1)^{-1}\hat{D}_N(1)\right)^{-1}\hat{D}_N(1)'\hat{V}_{ff}(1)^{-1}\hat{V}_{ff}(1)^{-\frac{1}{2}}f_N(1) \\ & \approx & 1 + \left[\mu(\sigma^2)'A_f(1)_\perp\left[A_f(1)'_\perp B_f(1)V_{uu,uu.y_1\Delta y}B_f(1)'A_f(1)_\perp\right]^{-1}A_f(1)'_\perp\mu(\sigma^2)\right]^{-1} \\ & & \mu(\sigma^2)'A_f(1)_\perp\left[A_f(1)'_\perp B_f(1)V_{uu,uu.y_1\Delta y}B_f(1)'A_f(1)_\perp\right]^{-1} \\ & & \left[A_f(1)'_\perp\frac{1}{\sqrt{N}}B_f(1)\psi_{cu} - A_f(1)'_\perp B_f(1)V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y}\frac{1}{\sqrt{N}}\psi\right]\end{array}$$

so

$$\begin{array}{rcl}\sqrt{N}(\hat{\theta}_{CUE} - 1) & \underset{d}{\rightarrow} & \left[\mu(\sigma^2)'A_f(1)_\perp\left[A_f(1)'_\perp B_f(1)V_{uu,uu.y_1\Delta y}B_f(1)'A_f(1)_\perp\right]^{-1}A_f(1)'_\perp\mu(\sigma^2)\right]^{-1} \\ & & \mu(\sigma^2)'A_f(1)_\perp\left[A_f(1)'_\perp B_f(1)V_{uu,uu.y_1\Delta y}B_f(1)'A_f(1)_\perp\right]^{-1} \\ & & A_f(1)'_\perp B_f(1)\left[\psi_{cu} - V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y}\psi\right],\end{array}$$

which shows that the CUE is consistent with a standard convergence rate and a normal large sample distribution.

Proof of Theorem 6. GMM-AR statistic. Using the results from the proof of Theorem 5 on the large sample distribution of the two step estimator, it holds that the GMM-AR statistic equals

$$\begin{aligned}\text{GMM-AR}(\theta) = & \; N f_N(\theta)' \hat{V}_{ff}(\theta)^{-1} f_N(\theta) \\ = & \; N \times i' \times \text{i} + N \times ii' \times \text{ii}.\end{aligned}$$

The two components of the GMM-AR statistic behave according to:

$$\begin{aligned}
& N \times i' \times \text{i} = \\
& N \left[\tfrac{1}{h(\theta_0)\sqrt{N}} \psi + (A_f(\theta)' A_f(\theta))^{-1} A_f(\theta)' \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu} + A_f(\theta)\iota_{T-2} E(\lim_{\theta_0 \uparrow 1}(1-\theta_0)u_{i1}^2) \right]\right] \\
& \left[h(\theta_0) V^{-1}_{y_1\Delta y, y_1\Delta y.uu} \left\{ (A_f(\theta)' A_f(\theta))^{-1} A_f(\theta)' \left[\tfrac{1}{\sqrt{N}} A_f(\theta)\psi + h(\theta_0) \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu} + \right.\right.\right.\right. \\
& A_f(\theta)\iota_{T-2} E(\lim_{\theta_0 \uparrow 1}(1-\theta_0)u_{i1}^2)]] - V_{y_1\Delta y, uu} B_f(\theta)' A_f(\theta)_\perp (A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)' A_f(\theta)_\perp)^{-1} A_f(\theta)'_\perp \\
& \left.\left.\left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu} \right] \right\}\right] \\
& = \left[\psi + h(\theta_0)\sqrt{N} (A_f(\theta)' A_f(\theta))^{-1} A_f(\theta)' \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu} + \right.\right. \\
& A_f(\theta)\iota_{T-2} E(\lim_{\theta_0 \uparrow 1}(1-\theta_0)u_{i1}^2)]]' \, V^{-1}_{y_1\Delta y, y_1\Delta y.uu} \left[\psi + h(\theta_0)\sqrt{N} (A_f(\theta)' A_f(\theta))^{-1} A_f(\theta)' \right. \\
& \left.\left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu} + A_f(\theta)\iota_{T-2} E(\lim_{\theta_0 \uparrow 1}(1-\theta_0)u_{i1}^2) \right] \right] - \\
& \left[\psi + h(\theta_0)\sqrt{N} (A_f(\theta)' A_f(\theta))^{-1} A_f(\theta)' \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu} + A_f(\theta)\iota_{T-2} E(\lim_{\theta_0 \uparrow 1}(1-\theta_0)u_{i1}^2) \right]\right]' \\
& V^{-1}_{y_1\Delta y, y_1\Delta y.uu} V_{y_1\Delta y, uu} B_f(\theta)' A_f(\theta)_\perp (A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)' A_f(\theta)_\perp)^{-1} A_f(\theta)'_\perp \left[\sqrt{N}(1-\theta)\mu(\sigma^2) + B_f(\theta \right.
\end{aligned}$$

where we used that $h(\theta_0)\sqrt{N} \to 0$ when $\theta_0 \to 1$, and

$$\begin{aligned}
& N \times ii' \times \text{ii} = \\
& N \left[A_f(\theta)'_\perp \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu} \right]\right]' \\
& [A_f(\theta)'_\perp B_f(\theta) V_{uu,uu,y_1\Delta y} B_f(\theta)' A_f(\theta)_\perp]^{-1} \left\{ -A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y, uu} V^{-1}_{y_1\Delta y, y_1\Delta y} (A_f(\theta)' A_f(\theta))^{-1} A_f(\theta)' \right. \\
& \left[h(\theta_0)(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} A_f(\theta)\psi + h(\theta_0) A_f(\theta)\iota_{T-2} E(\lim_{\theta_0 \uparrow 1}(1-\theta_0)u_{i1}^2) + \right. \\
& \left.\left. \tfrac{1}{\sqrt{N}} h(\theta_0) B_f(\theta)\psi_{cu} \right] + A_f(\theta)'_\perp \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu} \right] \right\}
\end{aligned}$$

$$\begin{aligned}
= \; & \left[\sqrt{N}(1-\theta)\mu(\sigma^2) + B_f(\theta)\psi_{cu} \right]' A_f(\theta)_\perp \left[A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)' A_f(\theta)_\perp \right]^{-1} \\
& A_f(\theta)'_\perp \left[\sqrt{N}(1-\theta)\mu(\sigma^2) + B_f(\theta)\psi_{cu} \right] - \\
& \left[\sqrt{N}(1-\theta)\mu(\sigma^2) + B_f(\theta)\psi_{cu} \right]' A_f(\theta)_\perp \left[A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)' A_f(\theta)_\perp \right]^{-1} \\
& A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y, uu} V^{-1}_{y_1\Delta y, y_1\Delta y.uu} (A_f(\theta)' A_f(\theta))^{-1} A_f(\theta)' \left[\sqrt{N} h(\theta_0)(1-\theta)\mu(\sigma^2) + \right. \\
& \left. A_f(\theta)\psi + h(\theta_0) B_f(\theta)\psi_{cu} \right]
\end{aligned}$$

since $h(\theta_0)\sqrt{N} \to 0$ when $\theta_0 \to 1$. Hence, the behavior of GMM-AR(θ) is characterized by

$$
\begin{aligned}
&\text{GMM-AR}(\theta) \\
&= Nf_N(\theta)'V_{ff}(\theta)^{-1}f_N(\theta) \\
&= N \times i' \times \text{i} + N \times ii' \times \text{ii} \\
&\underset{d}{\to} \begin{pmatrix} \left[\psi + h(\theta_0)\sqrt{N}(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\left[(1-\theta)\mu(\sigma^2) + \frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu} + A_f(\theta)\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)\right.\right. \\ \sqrt{N}A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2) + \frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \end{pmatrix} \\
&\begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y, uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y, uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\
&\begin{pmatrix} \left[\psi + h(\theta_0)\sqrt{N}(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\left[(1-\theta)\mu(\sigma^2) + \frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu} + A_f(\theta)\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2\right.\right. \\ \sqrt{N}A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2) + \frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right] \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
= &\begin{pmatrix} \psi \\ A_f(\theta)'_\perp\left[\sqrt{N}(1-\theta)\mu(\sigma^2) + B_f(\theta)\psi_{cu}\right] \end{pmatrix}' \\
&\begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y, uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y, uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\
&\begin{pmatrix} \psi \\ A_f(\theta)'_\perp\left[\sqrt{N}(1-\theta)\mu(\sigma^2) + B_f(\theta)\psi_{cu}\right] \end{pmatrix}
\end{aligned}
$$

which shows that the GMM-AR statistic has a $\chi^2(\frac{1}{2}(T+1)(T-2))$ distribution in large samples when $\theta = \theta_0 = 1$ and a non-central χ^2 with $\frac{1}{2}(T+1)(T-2)$ degrees of freedom and non-centrality parameter

$$
\begin{aligned}
&\begin{pmatrix} 0 \\ \sqrt{N}(1-\theta)A_f(\theta)'_\perp\mu(\sigma^2) \end{pmatrix}' \begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y, uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y, uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\
&\begin{pmatrix} 0 \\ \sqrt{N}(1-\theta)A_f(\theta)'_\perp\mu(\sigma^2) \end{pmatrix} \\
&= N(1-\theta)^2\mu(\sigma^2)'A_f(\theta)_\perp(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-1}A_f(\theta)'_\perp\mu(\sigma^2)
\end{aligned}
$$

for values of θ unequal to one. The non-centrality parameter shows that the power of the test increases when θ decreases and N increases as confirmed by our simulation experiments.

KLM statistic. To obtain the expression of the KLM statistic, we use the expression for $\hat{D}_N(\theta)$ from the proof of Theorem 5 on the large sample distribution of the CUE:

$$\hat{D}_N(\theta) = -\begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma^2_{u_2} \\ \sigma^2_{u_3} \end{pmatrix} - \tfrac{1}{h(\theta_0)\sqrt{N}} A_q\psi - \tfrac{1}{\sqrt{N}} B_q \psi_{cu} - \hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1} f_N(\theta)$$

$$\underset{d}{\rightarrow} -[I_5 - (1-\theta)A_q(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)']\mu(\sigma^2) + \\ [A_q V_{y_1\Delta y,uu} B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} + B_q V'_{y_1\Delta y,uu} \vdots B_q V_{uu,uu} B_f(\theta)'A_f(\theta)_\perp] \\ \begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y,uu} B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\ \begin{pmatrix} \tfrac{1}{\sqrt{N}}\psi \\ A_f(\theta)'_\perp \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu}\right] \end{pmatrix}.$$

so

$$(h(\theta_0)A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'\hat{D}_N(\theta) = \\ \underset{d}{\rightarrow} -A_f(\theta)'_\perp[I_5 - (1-\theta)A_q(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)']\mu(\sigma^2) + \\ A_f(\theta)'_\perp[A_q V_{y_1\Delta y,uu} B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} + B_q V'_{y_1\Delta y,uu} \vdots B_q V_{uu,uu} B_f(\theta)'A_f(\theta)_\perp] \\ \begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y,uu} B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1} \\ \begin{pmatrix} \tfrac{1}{\sqrt{N}}\psi \\ A_f(\theta)'_\perp \left[(1-\theta)\mu(\sigma^2) + \tfrac{1}{\sqrt{N}} B_f(\theta)\psi_{cu}\right] \end{pmatrix}$$

The important element of $\hat{D}_N(\theta)$ is that it does not have a $\frac{1}{h(\theta_0)\sqrt{N}}\psi$ component and that its random components are uncorrelated with those of $f_N(\theta)$. The large sample distribution of the KLM statistic is therefore a non-central $\chi^2(1)$ with non-centrality parameter:

$$N(1-\theta)^2\mu(\sigma^2)'(0 \vdots A_f(\theta)_\perp) \begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y,uu} B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-\frac{1}{2}\prime} \\ P_{g(\theta)} \begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y,uu} B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta) V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta) V_{uu,uu} B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-\frac{1}{2}} (0 \vdots A_f(\theta)_\perp)'\mu(\sigma^2)$$

where

$$g(\theta) = \begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y, uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y, uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-\frac{1}{2}}$$
$$(h(\theta_0)A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'$$
$$\{-[I_5 - (1-\theta)A_q(A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)']\mu(\sigma^2)+$$
$$[A_qV_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} + B_qV'_{y_1\Delta y,uu} \vdots B_qV_{uu,uu}B_f(\theta)'A_f(\theta)_\perp]$$
$$\begin{pmatrix} V_{y_1\Delta y, y_1\Delta y} & V_{y_1\Delta y, uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y, uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp \end{pmatrix}^{-1}$$
$$\left.\begin{pmatrix} 0 \\ (1-\theta)A_f(\theta)'_\perp\mu(\sigma^2) \end{pmatrix}\right\}.$$

GMM-LM statistic: To determine the large sample behavior of the GMM-LM statistic for values of θ close to one, we first determine the behavior of $(h(\theta_0)A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'q_N(\theta)$:

$$\begin{aligned} &(h(\theta_0)A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)'q_N(\theta) \\ &= (h(\theta_0)A_f(\theta)(A_f(\theta)'A_f(\theta))^{-1} \vdots A_f(\theta)_\perp)' \\ &\quad \left[\mu(\sigma^2) + A_q\left(\tfrac{1}{h(\theta_0)\sqrt{N}}\psi + \iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)\right) + \tfrac{1}{\sqrt{N}}B_q\psi_{cu}\right] \\ &= \begin{pmatrix} (A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\frac{1}{\sqrt{N}}A_q\psi \\ A_f(\theta)'_\perp\left[\mu(\sigma^2) + A_q\left(\frac{1}{h(\theta_0)\sqrt{N}}\psi + \iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)\right) + \frac{1}{\sqrt{N}}B_q\psi_{cu}\right] \end{pmatrix}. \end{aligned}$$

so the behavior of $f_N(\theta)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta)$ is characterized by

$$
\begin{array}{l}
f_N(\theta)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta) \\
= \left(\begin{array}{c} \frac{1}{\sqrt{N}}\psi \\ A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2)+\frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right]\end{array}\right)' \\
\left(\begin{array}{cc} V_{y_1\Delta y,y_1\Delta y} & V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp\end{array}\right)^{-1} \\
\left(\begin{array}{c} (A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\frac{1}{\sqrt{N}}A_q\psi \\ A_f(\theta)'_\perp\left[\mu(\sigma^2)+A_q\left(\frac{1}{h(\theta_0)\sqrt{N}}\psi+\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)\right)+\frac{1}{\sqrt{N}}B_q\psi_{cu}\right]\end{array}\right) \\
= \left(\begin{array}{c} \frac{1}{\sqrt{N}}\psi \\ A_f(\theta)'_\perp\left[(1-\theta)\mu(\sigma^2)+\frac{1}{\sqrt{N}}B_f(\theta)\psi_{cu}\right]\end{array}\right)' \\
\left(\begin{array}{cc} V_{y_1\Delta y,y_1\Delta y} & V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp\end{array}\right)^{-1} \\
\left(\begin{array}{c} 0 \\ \frac{1}{h(\theta_0)\sqrt{N}}A_f(\theta)'_\perp A_q\psi\end{array}\right) \\
= \frac{1}{h(\theta_0)}\left[(1-\theta)\mu(\sigma^2)+\frac{1}{\sqrt{N}}B_f(\theta)\left(\psi_{cu}-V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y}\psi\right)\right]'A_f(\theta)_\perp \\
(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-1}A_f(\theta)'_\perp A_q\psi.
\end{array}
$$

While $q_N(\theta)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta)$ is characterized by

$$
\begin{array}{l}
q_N(\theta)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta) \\
= \left(\begin{array}{c} (A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\frac{1}{\sqrt{N}}A_q\psi \\ A_f(\theta)'_\perp\left[\mu(\sigma^2)+A_q\left(\frac{1}{h(\theta_0)\sqrt{N}}\psi+\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)\right)+\frac{1}{\sqrt{N}}B_q\psi_{cu}\right]\end{array}\right)' \\
\left(\begin{array}{cc} V_{y_1\Delta y,y_1\Delta y} & V_{y_1\Delta y,uu}B_f(\theta)'A_f(\theta)_\perp \\ A_f(\theta)'_\perp B_f(\theta)V'_{y_1\Delta y,uu} & A_f(\theta)'_\perp B_f(\theta)V_{uu,uu}B_f(\theta)'A_f(\theta)_\perp\end{array}\right)^{-1} \\
\left(\begin{array}{c} (A_f(\theta)'A_f(\theta))^{-1}A_f(\theta)'\frac{1}{\sqrt{N}}A_q\psi \\ A_f(\theta)'_\perp\left[\mu(\sigma^2)+A_q\left(\frac{1}{h(\theta_0)\sqrt{N}}\psi+\iota_{T-2}E(\lim_{\theta_0\uparrow 1}(1-\theta_0)u_{i1}^2)\right)+\frac{1}{\sqrt{N}}B_q\psi_{cu}\right]\end{array}\right) \\
= \frac{1}{h(\theta_0)^2N}\psi'A_q'(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-1}A_q\psi
\end{array}
$$

Combining these results, we obtain the large sample distribution of the GMM-LM statistic:

$$
\begin{array}{l}
\text{GMM-LM}(\theta) = Nf_N(\theta)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta)(q_N(\theta)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta))^{-1}q_N(\theta)'\hat{V}_{ff}(\theta)^{-1}f_N(\theta) \\
= \left[(1-\theta)\sqrt{N}\mu(\sigma^2)+B_f(\theta)\left(\psi_{cu}-V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y}\psi\right)\right]'A_f(\theta)_\perp \\
(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-\frac{1}{2}}P_{h(\theta)}(A_f(\theta)'_\perp B_f(\theta)V_{uu,uu.y_1\Delta y}B_f(\theta)'A_f(\theta)_\perp)^{-\frac{1}{2}} \\
A_f(\theta)'_\perp\left[(1-\theta)\sqrt{N}\mu(\sigma^2)+B_f(\theta)\left(\psi_{cu}-V'_{y_1\Delta y,uu}V^{-1}_{y_1\Delta y,y_1\Delta y}\psi\right)\right]
\end{array}
$$

with

$$h(\theta) = \quad (A_f(\theta)'_\perp B_f(\theta) V_{uu,uu.y_1\Delta y} B_f(\theta)' A_f(\theta)_\perp)^{-\frac{1}{2}} A_f(\theta)'_\perp A_q \psi$$

which is a non-central $\chi^2(1)$ since $\psi_{cu} - V'_{y_1\Delta y,uu} V^{-1}_{y_1\Delta y,y_1\Delta y}\psi$ is independent of ψ with (an independently distributed random) non-centrality parameter

$$\begin{array}{c} N(1-\theta)^2 \mu(\sigma^2)' A_f(\theta)_\perp (A_f(\theta)'_\perp B_f(\theta) V_{uu,uu.y_1\Delta y} B_f(\theta)' A_f(\theta)_\perp)^{-\frac{1}{2}} P_{h(\theta)} \\ (A_f(\theta)'_\perp B_f(\theta) V_{uu,uu.y_1\Delta y} B_f(\theta)' A_f(\theta)_\perp)^{-\frac{1}{2}} A_f(\theta)'_\perp \mu(\sigma^2). \end{array}$$

Proof of Theorem 7. To obtain the power envelopes, we need the different elements of the covariance matrix $V_{uu,uu.y_1\Delta y} = V_{uu,uu} - V'_{y_1\Delta y,uu} V^{-1}_{y_1\Delta y,y_1\Delta y} V_{y_1\Delta y,uu}$, which is involved in the non-centrality parameter of the non-central χ^2 distributions of the GMM-AR, GMM-LM and KLM statistics in Theorem 7 under DGP 1 when θ_0 goes to one:

$$\begin{array}{rcl} V_{uu,uu} &=& \left(\begin{array}{cc} \sigma_c^2 \mathrm{diag}(\sigma_2^2,\ \sigma_3^2,\ \sigma_4^2) & 0 \\ 0 & \mathrm{diag}(E(u_{i2}^2 - \sigma_2^2)^2,\ \sigma_2^2\sigma_3^2,\ \sigma_2^2\sigma_4^2,\ E(u_{i3}^2 - \sigma_3^2)^2,\ \sigma_3^2\sigma_4^2)) \end{array} \right) \\ V_{y_1\Delta y,y_1\Delta y} &=& \lim_{\theta_0\uparrow 1} \frac{\sigma_c^2}{(1-\theta_0)^2} \mathrm{diag}(\sigma_2^2,\ \sigma_3^2,\ \sigma_4^2) + \sigma_1^2 \mathrm{diag}(\sigma_2^2,\ \sigma_3^2,\ \sigma_4^2) \\ V^{-1}_{y_1\Delta y,y_1\Delta y} &=& (\lim_{\theta_0\uparrow 1}(1-\theta_0)^2)\, \mathrm{diag}(\sigma_2^{-2},\ \sigma_3^{-2},\ \sigma_4^{-2}) - (\lim_{\theta_0\uparrow 1}(1-\theta_0)^4)\, \mathrm{diag}(\sigma_2^{-2},\ \sigma_3^{-2},\ \sigma_4^{-2}) \\ && \left[(\lim_{\theta_0\uparrow 1}(1-\theta_0)^2)\, \mathrm{diag}(\sigma_2^{-2},\ \sigma_3^{-2},\ \sigma_4^{-2}) + \sigma_1^2 \mathrm{diag}(\sigma_2^{-2},\ \sigma_3^{-2},\ \sigma_4^{-2}) \right]^{-1} \\ && \mathrm{diag}(\sigma_2^{-2},\ \sigma_3^{-2},\ \sigma_4^{-2}) \\ V_{y_1\Delta y,uu} &=& \left(\begin{array}{c} \lim_{\theta_0\uparrow 1} \frac{\sigma_c^2}{1-\theta_0} \mathrm{diag}(\sigma_2^2,\ \sigma_3^2,\ \sigma_4^2) \\ 0 \end{array} \right). \end{array}$$

Combining these we obtain the limit of the covariance matrix involved in the non-central χ^2 distributions when θ_0 is equal to one.

$$\lim_{\theta_0\uparrow 1} V_{uu,uu.y_1\Delta y} = \left(\begin{array}{cc} 0 & 0 \\ 0 & \mathrm{diag}(E(u_{i2}^2 - \sigma_2^2)^2,\ \sigma_2^2\sigma_3^2,\ \sigma_2^2\sigma_4^2,\ E(u_{i3}^2 - \sigma_3^2)^2,\ \sigma_3^2\sigma_4^2)) \end{array} \right).$$

Appendix B. Definitions

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

$$E[f_i(\theta_0)] = 0, \qquad i = 1, \ldots, N, \tag{46}$$

which are a function of observed data and the unknown parameter θ. The moment conditions are only satisfied at the true value of the p-dimensional vector θ, θ_0, and k is at least as large as p. The population moments in (46) are estimated using the average sample moments,

$$f_N(\theta) = \tfrac{1}{N}\textstyle\sum_{i=1}^N f_i(\theta). \tag{47}$$

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

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

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

For the Dif moment conditions in (5), k equals $\frac{1}{2}(T-2)(T-1)$ and the specifications of $f_i(\theta)$ and $q_i(\theta)$ read

$$\begin{array}{rll} f_i(\theta) = & X_i\varphi_i(\theta) & i = 1, \ldots, N, \\ q_i(\theta) = & -X_i\Delta y_{-1,i} & i = 1, \ldots, N, \end{array} \tag{49}$$

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

$$X_i = \begin{pmatrix} y_{i1} & 0\ldots0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\ldots0 & \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT-2} \end{pmatrix} \end{pmatrix} : \tfrac{1}{2}(T-1)(T-2) \times (T-2), \tag{50}$$

while for the Lev moment conditions in (6), k equals $T-2$ while the moment functions can be specified as

$$\begin{array}{rll} f_i(\theta) = & W_i\eta_i(\theta) & i = 1, \ldots, N, \\ q_i(\theta) = & W_i y_{-1,i} & i = 1, \ldots, N, \end{array} \tag{51}$$

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

$$W_i = \begin{pmatrix} \Delta y_{i2} & 0\ldots0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\ldots0 & \Delta y_{iT-1} \end{pmatrix} : (T-2) \times (T-2). \tag{52}$$

The specification of the moment functions for the Sys moment conditions results by stacking the moment conditions in (49) and (51) so k equals $\frac{1}{2}(T+1)(T-2)$. While we could extend the Lev moment conditions with additional interactions of Δy_{it-j} and $y_{it}-\theta y_{it-1}$, for $j=2,\ldots,t-2$, like the extension of the moment conditions used by Anderson and Hsiao (1981) towards those used by Arellano and Bond (1991), we cannot incorporate these into the Sys moment conditions since some of the Sys moment functions then result as a linear combination of the others.[3] We therefore do not extend the Lev moment functions with these interactions.

We use five different GMM statistics: the two step and CUE Wald statistics, the GMM-LM statistic of Newey and West (1987), the KLM statistic of Kleibergen (2002,2005) and the GMM extension of the Anderson-Rubin statistic, see Anderson and Rubin (1949) and Stock and Wright (2000). The two step estimator and the CUE result by minimizing the objective function:

$$Q(\theta,\theta^1)=Nf_N(\theta)'\hat{V}_{ff}(\theta^1)^{-1}f_N(\theta), \tag{53}$$

with $\hat{V}_{ff}(\theta^1)$ the Eicker-White covariance matrix estimator:

$$\hat{V}_{ff}(\theta^1)=\ \frac{1}{N}\sum_{i=1}^N(f_i(\theta^1)-f_N(\theta^1))(f_i(\theta^1)-f_N(\theta^1))'. \tag{54}$$

The CUE of Hansen *et. al.* (1996), $\hat{\theta}_{CUE}$, results by replacing θ^1 by θ in (53) while the two step estimator, $\hat{\theta}_{2s}$, uses a value of θ^1 equal to the minimizer of (53) when we replace $\hat{V}_{ff}(\theta^1)^{-1}$ by the identity matrix.[4] The expressions of the different statistics to test $\mathrm{H}_0:\theta=\theta_0$ that we use read:

1. Two step Wald statistic:

$$W_{2s}(\theta_0)=N(\hat{\theta}_{2s}-\theta_0)'q_N(\hat{\theta}_{2s})'\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}q_N(\hat{\theta}_{2s})(\hat{\theta}_{2s}-\theta_0). \tag{55}$$

2. CUE Wald statistic:

$$W_{CUE}(\theta_0)=N(\hat{\theta}_{CUE}-\theta_0)'\hat{D}_N(\hat{\theta}_{CUE})'\hat{V}_{ff}(\hat{\theta}_{CUE})^{-1}\hat{D}_N(\hat{\theta}_{CUE})(\hat{\theta}_{CUE}-\theta_0), \tag{56}$$

[3]To show this consider that $T=4$. Two of the (three) Dif sample moments are then $\frac{1}{N}\sum_{i=1}^N y_{i1}(\Delta y_{i4}-\theta\Delta y_{i3})$ and $\frac{1}{N}\sum_{i=1}^N y_{i2}(\Delta y_{i4}-\theta\Delta y_{i3})$ while two of the (three) Lev sample moments are $\frac{1}{N}\sum_{i=1}^N \Delta y_{i2}(y_{i3}-\theta y_{i2})$ and $\frac{1}{N}\sum_{i=1}^N \Delta y_{2i}(y_{i4}-\theta y_{i3})$. If we substract the first of these Dif sample moments from the second and add the first of the Lev sample moments to it, the second Lev sample moment results.

[4]For the two step Wald statistic with the Dif moment conditions, we use a value of θ^1 that results from minimizing (53) using the covariance matrix estimator that results under homoscedasticity.

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

$$\text{vec}(\hat{D}_N(\theta)) = \text{vec}(q_N(\theta)) - \hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_N(\theta) \tag{57}$$

and

$$\hat{V}_{\theta f}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \text{vec}[q_i(\theta) - q_N(\theta)](f_i(\theta) - f_N(\theta))'. \tag{58}$$

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

$$\begin{aligned} LM(\theta_0) = \; & N f_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}q_N(\theta_0)\left[q_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}q_N(\theta_0)\right]^{-1} \\ & q_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}f_N(\theta_0). \end{aligned} \tag{59}$$

4. The KLM statistic of Kleibergen (2005):

$$\begin{aligned} KLM(\theta_0) = \; & N f_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_N(\theta_0)\left[\hat{D}_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_N(\theta_0)\right]^{-1} \\ & \hat{D}_N(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}f_N(\theta_0). \end{aligned} \tag{60}$$

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

$$GMM\text{-}AR(\theta) = N f_N(\theta)'\hat{V}_{ff}(\theta)^{-1}f_N(\theta) = Q(\theta, \theta). \tag{61}$$

All of the above statistics are general GMM statistics. The GMM-AR statistic is the only statistic whose large sample distributions is not $\chi^2(p)$ under H_0. The large sample distribution of the GMM-AR statistic is a $\chi^2(k)$ distribution under H_0. The last two of the above statistics are so-called weak instrument robust statistics so their (conditional) large sample distributions remain free from nuisance parameters when the Jacobian identification condition fails. We can therefore use them to analyze the influence of the fixed effects and initial observations on the identification of θ in a clear manner since the results are not distorted by the change of the large sample distribution of the statistic caused by the identification failure.

References

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*.

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

Hansen, L.P., J. Heaton and A. Yaron (1996). Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics* **14**, 262–280.

Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (2002). Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models covering Short Time Periods. *Journal of Econometrics* **109**, 107–150.

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.

Lancaster, T. (2002). Orthogonal Parameters and Panel Data. *Review of Economic Studies* **69**, 647–666.

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

Moreira, M.J. A maximum likelihood method for the incidental parameter problem. *Annals of Statistics* **37**, 3660–3696.

Newey, W.K. and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden (Ed.), *Handbook of Econometrics, Volume 4*, Chapter 36, pp. 2113–2148. Elsevier Science B.V.

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.

Windmeijer, F. (2004). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* **126**, 25–51.