

Inference in Second-Order Identified Models

Prosper Dovonon

Concordia University¹

Alastair R. Hall

University of Manchester²

and

Frank Kleibergen

University of Amsterdam³

January 9, 2017

¹Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada. E-mail: prosper.dovonon@concordia.ca.

²Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK. E-mail: alastair.hall@manchester.ac.uk.

³Department of Quantitative Economics, Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1001 NJ Amsterdam, NL; email f.r.kleibergen@uva.nl

Abstract

First-order asymptotic analyses of the Generalized Method of Moments (GMM) estimator and its associated statistics are based on the assumption that the population moment condition identifies the parameter vector both globally and locally at first order. In linear models, global and first-order local identification are equivalent but in nonlinear models they are not. In certain econometric models of interest, parameters are globally identified but only identified locally at second order. In these scenarios the standard GMM inference techniques based on first-order asymptotics are invalid, see Dovonon and Renault (2013) and Dovonon and Hall (2016). In this paper, we explore how to perform inference in moment condition models that only identify the parameters locally to second order. For inference about the parameters, we consider inference based on conventional Wald and LM statistics, and also the Generalized Anderson Rubin (GAR) statistic (Anderson and Rubin, 1949; Dufour, 1997; Staiger and Stock, 1997; Stock and Wright, 2000) and the KLM statistic (Kleibergen, 2002, 2005). Both the GAR and KLM statistics have been proposed as methods of inference in the presence of weak identification and are known to be “identification robust” in the sense that their limiting distribution is the same under first-order and weak identification. For inference about the model specification, we consider the identification-robust J statistic (Kleibergen, 2005) and the GAR statistic. In each case, we derive the limiting distribution of statistics under both null and local alternative hypotheses. We show that under their respective null hypotheses the GAR, KLM and J statistics have the same limiting distribution as would apply under first-order or weak identification, thus showing their identification robustness extends to second-order identification. We explore the power properties in detail in two empirically relevant models with second-order identification. In the panel autoregressive (AR) model of order one, our analysis indicates that the Wald test of whether the AR parameter is one has superior power to the corresponding GAR test which, in turn, dominates the KLM and LM tests. For the conditionally heteroskedastic factor model, we compare Kleibergen’s (2005) J and the GAR statistics to Hansen’s (1982) overidentifying restrictions test (previously analyzed in this context by Dovonon and Renault, 2013) and find the power ranking depends on the sample size. Collectively, our results suggest that tests with meaningful power can be conducted in second-order identified models.

Keywords: Generalized Method of Moments estimation, First-order identification failure, Identification-robust inference

1 Introduction

Generalized Method of Moments (GMM) is a popular method for estimating the parameters of econometric models based on the information in population moment conditions. In his seminal article introducing GMM, Hansen (1982) proves the consistency of the estimator and provides a framework for inference based on first-order asymptotic statistical arguments. This original framework includes confidence intervals for the parameters and the overidentifying restrictions statistic that can be used to test the model specification, and it has been subsequently extended to a wide variety of inference procedures, similarly based on first-order asymptotic arguments. However, the statistical arguments that justify these inference techniques are predicated on certain regularity conditions among which are the assumptions that the population moment condition is valid and identifies the parameters both globally and also locally at first order.

Over the last 25 years, there has been a growing awareness that this first-order asymptotic theory may provide a poor approximation to the finite sample behaviour of GMM-based statistics in finite samples. Attention has focussed primarily on cases where the assumed identification conditions fail or are close to failure. To derive alternative approximations to the behaviour of GMM-based statistics under this scenario, Staiger and Stock (1997) introduced the concept of weak identification. Within this framework, parameters are globally and first-order locally identified in finite samples but the information provided by the population moment declines (at a prescribed rate) as the sample size increases resulting in the parameters being globally unidentified in the limit. Under weak identification, the large sample properties of the conventional GMM-based statistics are different from those derived in Hansen's (1982) analysis, see Staiger and Stock (1997) and Stock and Wright (2000). Furthermore, once the possibility of weak identification is admitted, the conventional approach to constructing confidence intervals based on GMM estimators - "estimator plus/minus a multiple of the standard error" - is invalid, see Dufour (1997). This has led to a focus on inferences based on so-called "identification robust" statistics whose distribution is invariant to the quality of the identification. Leading examples of such statistics are the generalized Anderson-Rubin (GAR) statistic (Anderson and Rubin, 1949; Dufour, 1997; Staiger and Stock, 1997; Stock and Wright, 2000), the KLM statistic (Kleibergen, 2002, 2005), the J statistic (Kleibergen, 2005), and the conditional likelihood ratio statistic (Moreira, 2003; Kleibergen, 2005). In each case, inferences are performed by inverting the statistic in question to calculate parameter values consistent with the null hypothesis at the chosen level of confidence/significance.

However, weak identification and its variants are not the only way in which first order local identification can fail.¹ In linear models, first-order local and global identification are the same, but in nonlinear models, they are not: identification can fail at first order locally but hold at a higher order. In this paper, we focus on the case where parameters are globally identified, identification fails locally at first order but holds at second order. This pattern of identification has been shown to arise in a number of situations in statistics and econometrics such as: ML for skew-normal distributions, Azzalini (2005); ML for binary response models based on skew-normal distributions, Stingo, Stanghellini, and Capobianco (2011); ML for missing not at random (MNAR) models, Jansen and *et al* (2006); GMM estimation of conditionally heteroskedastic factor models, Dovonon and Renault (2009, 2013); GMM estimation of panel data models using second moments, Madsen (2009), Bun and Kleibergen (2016); ML estimation of panel data models, Kruiniger (2014).

Within this second-order identification framework, GMM estimators are consistent but the limiting distribution of statistics based on the estimator is both different from its first-order asymptotic

¹For a recent review of methods for inference under weak identification and its extensions, see Hall (2015).

counterpart and also sensitive to the nature of the first-order identification failure. Local identification relates to the behaviour of the population moment condition as the parameter moves away from the true value. First order identification can fail in some or all directions, and the large sample behaviour of GMM-based statistics is sensitive to the number of directions in which local identification is at second order and not first order. For the case where first order identification only fails in one direction, the limiting distribution of the GMM estimator has been characterized by Dovonon and Hall (2016), extending earlier results by Sargan (1983) and Rotnitzky, Cox, Bottai, and Robins (2000) for estimators obtained respectively by IV in a nonlinear in parameters model and Maximum Likelihood. Dovonon and Renault (2009, 2013) derive the limiting distribution of the overidentifying restrictions statistic for an arbitrary number of directions in which local identification is at second and not first order.

In this paper, we study the power of commonly used test procedures when the parameter of interest is only locally second order identified. We analyze tests on the value of the parameter itself and the specification of the moment function. To conduct tests on the parameter of interest, we employ the traditional Wald and Lagrange multiplier (LM) statistics as well as the identification robust GAR and KLM statistics. For tests on the specification of the moment function, we use the GAR statistic and Kleibergen's (2005) J statistic (hereafter denoted as the K-J statistic). For each type of test, we define the appropriate local alternatives and derive the limiting distributions of all tests under both null and local alternatives. We also illustrate the power properties of the tests in two empirically relevant models: the panel autoregressive model of order one and the conditionally heteroskedastic factor model. For the panel data model, it is well known that the autoregressive parameter is plagued by identification issues if the autoregressive parameter is one. Bun and Kleibergen (2016) construct a specific moment equation which second order identifies the autoregressive parameter at this value. For the conditionally heteroskedastic factor model, Dovonon and Renault (2013) establish that the parameters are second-order identified by a moment condition used as a basis for testing for a common factor structure. Because of the second order identification, GMM estimators have a quartic root convergence rate and so we observe a very slow convergence of the finite sample distributions of the tests towards their limiting distributions under local alternatives. We therefore focus on the finite sample distributions of the tests for varying numbers of observations. For the panel autoregressive model, the Wald statistic has a surprising amount of discriminatory power and dominates the other tests, although the GAR statistic exhibits comparable power in large samples. The powers of the KLM and LM statistics are much less than that of the GAR statistic which is explained by the second-order identification. Because of it, the parameter of interest is not well identified and it is known that the GAR statistic compares favorably to the KLM statistic in such settings in terms of power. For the conditionally heteroskedastic factor model, we compare the power properties of K-J and the GAR tests with those of Hansen's (1982) overidentifying restrictions test, previously analyzed in this context by Dovonon and Renault (2013). Our results indicate that the power ranking is sensitive to the sample size: in small to moderate sample sizes the K-J test dominates the other two, which have comparable power; but in large sample sizes this ranking is reversed.

The paper is organized as follows. In the second section, we set up notation, introduce the concept of second order identification and the two running examples of the panel autoregressive model and the conditionally heteroskedastic factor model. In the third section, we introduce the different test statistics and their limiting distributions under the null hypothesis. In the fourth section, we discuss these distributions under appropriate local alternatives. The fifth section explores the finite sample power properties of the tests. Finally the sixth section concludes. All proofs are relegated

to a mathematical appendix.

2 Second-order identification: definition and examples

Suppose it is desired to estimate a parameter vector $\theta_0 \in \Theta \subset \mathbb{R}^p$ that indexes an econometric model. This model may explain behaviour of individual economic agents in a population and so be estimated from a random sample from that population or the model may explain the behaviour of economic variables over time and be estimated from time series data. Second-order identification can arise in either case, as demonstrated by our two examples below, and our results apply equally in both scenarios. However, certain definitions are different in the two cases. For ease of presentation, we first describe GMM estimations for the case where the data are obtained from a random sample, and then briefly note how those definitions need to be adapted for time series in footnote 3 below.

To this end, let X denote a random vector with probability distribution P and sample space \mathcal{X} modeling the variables in the econometric model. We consider the case where this model implies the following population moment condition:

$$E[f(X, \theta_0)] = 0, \quad (1)$$

where $f : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^k$ is twice continuously differentiable in θ almost everywhere and $k \geq p$. Associated with this population moment condition is a matrix $G(\theta_0)$ known as the Jacobian and defined via: $G(\theta) = E[q(X, \theta)]$, $q(\theta) = \partial f(X, \theta) / \partial \theta' |_{\theta = \bar{\theta}}$. Let $\{x_i, i = 1, \dots, N\}$ be a random sample of observations for X , and define the sample moment function to be $\bar{f}_N(\theta) = \frac{1}{N} \sum_{i=1}^N f_i(\theta)$ where $f_i(\theta) \equiv f(x_i, \theta)$.

Following Hansen (1982), we define a GMM estimator of θ_0 based on (1) as:

$$\hat{\theta}(W_N) = \arg \min_{\theta \in \Theta} N \bar{f}_N(\theta)' W_N \bar{f}_N(\theta), \quad (2)$$

where W_N is $k \times k$ weighting matrix that converges in probability to W , a symmetric positive definite matrix W . As emphasized by the notation, the GMM estimator depends on the choice of weighting matrix. Hansen (1982) shows that the optimal choice of weighting matrix is one that satisfies $W = \{V_{ff}(\theta_0)\}^{-1}$ where $V_{ff}(\theta_0) = Var[f(X, \theta_0)]$, assumed nonsingular throughout. This optimal choice is implemented via a two-step procedure in which a first-step GMM estimation is used to obtain a preliminary - "first-step GMM" - estimator, $\hat{\theta}_{1,s} = \hat{\theta}(W_N)$, based on a sub-optimal choice of W_N . This first-step GMM estimator is used to construct a consistent estimator of $Var[f(X, \theta_0)]$, the inverse of which is used as weighting matrix on a second-step estimation. Defining

$$\hat{V}_{ff}(\theta) = \frac{1}{N} \sum_{i=1}^N [f_i(\theta) - \bar{f}_N(\theta)] [f_i(\theta) - \bar{f}_N(\theta)]'$$

and

$$Q(\theta, \bar{\theta}) = N \bar{f}_N(\theta)' \hat{V}_{ff}(\bar{\theta})^{-1} \bar{f}_N(\theta),$$

the two-step GMM estimator is:

$$\hat{\theta}_N = \arg \min_{\theta \in \Theta} Q(\theta, \hat{\theta}_{1,s}). \quad (3)$$

Within this framework, two statistics are naturally of interest: $\hat{\theta}_N$ and the overidentifying restrictions test statistic $Q(\hat{\theta}_N, \hat{\theta}_{1,s})$. The former is the basis for inference about θ_0 and the latter

can be used to assess if the data are consistent with (1) being true in the population, often thought of as a test of the model specification.² Hansen (1982) establishes the limiting properties of both these statistics under a set of regularity conditions.³ Specifically, he shows that $\hat{\theta}_N$ is consistent for θ_0 ,

$$N^{1/2}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0, V_\theta), \quad (4)$$

where $V_\theta = \{G(\theta_0)'V_{ff}(\theta_0)^{-1}G(\theta_0)\}^{-1}$, and

$$Q(\hat{\theta}, \hat{\theta}_{1,s}) \xrightarrow{d} \chi_{k-p}^2. \quad (5)$$

For our purposes here, it suffices to highlight three of these regularity conditions. To this end, it is useful to condense our notation and write $m(\theta) = E[f(X, \theta)]$. The aforementioned three conditions are then:

- (i) $m(\theta_0) = 0$ so that the estimation is based on valid information;
- (ii) $m(\bar{\theta}) \neq 0$ for all $\bar{\theta} \neq \theta_0$ so that θ_0 is globally identified;
- (iii) $\text{rank}\{G(\theta_0)\} = p$ so that θ_0 is first-order locally identified.⁴

Of these three, the consistency of the GMM estimator only requires (i) and (ii) to hold; but the distributional results in (4) and (5) require all three conditions to hold.

As noted in the introduction, first-order local identification is not a necessary condition for global identification in nonlinear models. In this paper we focus on the case where first-order local identification fails but the parameters are identified at second order. To formally introduce this scenario, we let

$$H_s(\bar{\theta}) = E \left[\frac{\partial^2 f_s(X, \theta)}{\partial \theta \partial \theta'} \Big|_{\theta = \bar{\theta}} \right], \quad s = 1, 2, \dots, k$$

where $f_s(X, \theta)$ is the s -th element of $f(X, \theta)$. The following assumption defines the identification configuration maintained throughout our analysis.

Assumption 1. (a) $\forall \theta \in \Theta, \quad m(\theta) = 0 \Leftrightarrow \theta = \theta_0$; (b) For all u in the range of $G(\theta_0)'$ and all v in the null space of $G(\theta_0)$,

$$\left(G(\theta_0)u + (v' H_s(\theta_0)v)_{1 \leq s \leq k} = 0 \right) \Rightarrow (u = v = 0).$$

Assumption 1(a) combines conditions (i) and (ii) above, and provides the necessary and sufficient identification condition for consistent estimation of θ_0 . Assumption 1(b) is the second-order local identification condition introduced by Dovonon and Renault (2009). This is a sufficient condition for local identification that extends the standard first-order local identification (property (iii) above). If $\text{rank}\{G(\theta_0)\} = p$, then the null space of $G(\theta_0)$ is the null vector and Assumption 1(b) holds

²Although some caution needs to be exercised in interpreting the outcome of this test, see Newey (1985) and Hall (2005)[Section 5.1].

³If the model involves (stationary ergodic) time series then X is replaced by X_t in (1) with t denoting the time index, and replacing i in the definitions above. In this case the optimal choice of weighting matrix is $V_{ff} = \lim_{T \rightarrow \infty} \text{Var} \left[N^{-1/2} \sum_{t=1}^N f(X_t) \right]$ and $\hat{V}_{ff}(\theta)$ by a member of the class of Heteroskedasticity Autocorrelation Covariance (HAC) estimators, for example see Andrews (1991).

⁴Sometimes referred to as the rank condition for identification.

trivially. If $G(\theta_0)$ is rank deficient, this assumption ensures that the direction of the parameter that belongs to the range of $G(\theta_0)'$ is identified by the first order approximation of the moment function whereas the direction in the null space of the Jacobian is identified by the second-order approximation. In the extreme case where $G(\theta_0) = 0$, the whole parameter vector is identified by the second-order terms in the expansion of the moment function.

Dovonon and Renault (2009) establish that the components of the GMM estimator in the direction of the range of $G(\theta_0)'$ have the standard rate of convergence (\sqrt{N}) while the components in the direction of the null space of $G(\theta_0)$ have a non-standard rate of convergence ($N^{1/4}$) and those rates are sharp. It is thus evident that the distributional result in (4) does not apply if local identification holds at second but not first order, and Dovonon and Renault (2013) show that (5) is similarly invalid. We return to this issue in the next section. To conclude this section, we consider two examples where first-order identification fails but Assumption 1 holds.

2.1 Panel data example

Consider the first-order linear dynamic panel data model

$$y_{i,t} = c_i + \theta_0 y_{i,t-1} + u_{it} \quad i = 1, \dots, N, \quad t = 2, \dots, T, \quad (6)$$

where c_i denotes the (unobserved) fixed effect, T equals the number of time periods and N equals the number of cross section observations. The assumptions commonly used to identify the parameters of this model are that the error terms are independently distributed from each other and the fixed effect so that

$$\begin{aligned} E[u_{i,t} u_{i,s}] &= 0, & s \neq t; \quad t = 2, \dots, T, \\ E[u_{i,t} c_i] &= 0, & t = 2, \dots, T, \\ E[u_{i,t} y_{i,1}] &= 0, & t = 2, \dots, T. \end{aligned} \quad (7)$$

Based on these assumptions, different moment functions have been proposed to identify the autoregressive parameter of which the most commonly used are, perhaps, those proposed by Anderson and Hsiao (1981), Arellano and Bond (1991), Ahn and Schmidt (1995) and Blundell and Bond (1998). All these moment conditions have difficulty identifying the autoregressive parameter when its true value is close to one and the variance of the initial observations and/or fixed effects becomes large, see Bun and Kleibergen (2016). Bun and Kleibergen (2016) show that a non-linear combination of these moment conditions does, however, identify the autoregressive parameter in such settings. This non-linear combination leads to so-called robust moments that do not depend on the initial observations and fixed effects. Bun and Kleibergen (2016) show that for $T = 4$ the specification of the sample moment function associated with these robust moments is:

$$\bar{f}_N(\theta) = a\theta^2 + b\theta + d, \quad (8)$$

where

$$a = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} (\Delta y_{i,2})^2 \\ 0 \end{pmatrix}, \quad b = -\frac{1}{N} \sum_{i=1}^N \begin{pmatrix} (y_{i,3} - y_{i,1})^2 \\ \Delta y_{i,2} \Delta y_{i,3} \end{pmatrix}, \quad d = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} (y_{i,4} - y_{i,1}) \Delta y_{i,3} \\ \Delta y_{i,2} \Delta y_{i,4} \end{pmatrix}.$$

Under the assumptions above, the expectation of these terms is given by:

$$\begin{aligned}
E[a] &= \begin{pmatrix} E[(c_i - (1 - \theta_0)y_{i,1})^2] + \sigma_2^2 \\ 0 \end{pmatrix}, \\
E[b] &= \begin{pmatrix} (1 + \theta_0)^2 E[(c_i - (1 - \theta_0)y_{i,1})^2] - \theta_0^2 \sigma_2^2 - \sigma_3^2 \\ -\theta_0^2 E[(c_i - (1 - \theta_0)y_{i,1})^2] \end{pmatrix}, \\
E[d] &= \begin{pmatrix} \theta_0(1 + \theta_0 + \theta_0^2) E[(c_i - (1 - \theta_0)y_{i,1})^2] + \theta_0^2(\theta_0 - 1)\sigma_2^2 + \theta_0\sigma_3^2 \\ \theta_0^2 E[(c_i - (1 - \theta_0)y_{i,1})^2] \end{pmatrix}.
\end{aligned} \tag{9}$$

with $\sigma_t^2 = E[u_{it}^2]$. If we assume mean-stationarity⁵ - so that $E[(c_i - (1 - \theta)y_{i1})^2] = 0$ - and the errors are homoskedastic - $\sigma_t^2 = \sigma^2$ - then these expected values simplify to

$$E[a] = \sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad E[b] = -\sigma^2 \begin{pmatrix} \theta_0^2 + 1 \\ 0 \end{pmatrix}, \quad E[d] = \sigma^2 \begin{pmatrix} \theta_0^2(\theta_0 - 1) + \theta_0 \\ 0 \end{pmatrix}. \tag{10}$$

From (8) and (10), it follows that if $\theta_0 = 1$ then:

$$m(\theta_0) = 0_{2 \times 1}, \quad G(\theta_0) = 0_{2 \times 1}, \quad H_1(\theta_0) = 2\sigma^2, \quad H_2(\theta_0) = 0, \tag{11}$$

where we have emphasized the dimensions of the null vectors for clarity. It can be seen from (11) that if $\theta_0 = 1$ then this model is not first-order locally identified but satisfies Assumption 1 and so is second-order locally identified. In our subsequent analysis of this model, we focus on the inference about whether or not $\theta_0 = 1$.

2.2 Conditionally heteroskedastic factor models

Conditionally heteroskedastic factor (CHF) models are widely used to study the volatility of financial asset returns.⁶ Within this approach, the volatility of a vector of assets is assumed to derive from two sources: a latent common factor that exhibits conditional variation and an idiosyncratic component that is conditionally homoskedastic. In practice, the number of latent factors is assumed to be smaller than the number of assets and thus the CHF model provides a relatively parsimonious way of capturing the conditional variances and covariances of the assets.

Before basing inferences on the model, it is important to assess whether the sample covariance structure is consistent with this type of specification. Engle and Kozicki (1993) propose a general methodology for testing for common features in economic time series based on the GMM overidentifying restrictions test, and propose using it to test the validity of the CHF model. However, they base their decision rule on standard first-order asymptotic behaviour of the overidentifying restrictions test. Dovonon and Renault (2013) show that this theory is invalid in this case because the moment condition in question only identifies the parameters locally to second order.

To elaborate, consider the following CHF model for the $p \times 1$ vector of asset returns Y_{t+1} :

$$E[Y_{t+1} | \mathfrak{F}_t] = 0, \tag{12}$$

$$Var[Y_{t+1} | \mathfrak{F}_t] = \Lambda D_t \Lambda' + \Omega, \tag{13}$$

⁵See Blundell and Bond (1998).

⁶The approach is introduced in Diebold and Nerlove (1989); see also *inter alia* Engle, Ng, and Rothschild (1990), Fiorentini, Sentana, and Shephard (2004) and Doz and Renault (2006).

where D_t is a $L \times L$ diagonal matrix with ℓ^{th} diagonal element equal to $\sigma_{\ell,t}^2$ for $\ell = 1, 2, \dots, L$, Λ is a $p \times L$ matrix, and Ω is a $p \times p$ symmetric positive semi-definite matrix. The stochastic processes $\{Y_t\}_{t \geq 0}$ and $\left\{\sigma_{\ell,t}^2\right\}_{1 \leq \ell \leq L, t \geq 0}$ are adapted with respect to the increasing filtration $\{\mathfrak{F}_t\}_{t \geq 0}$. It is assumed that $rank(\Lambda) = L$ and $Var[\sigma_{\ell,t}^2] > 0$ for all $\ell = 1, 2, \dots, L$. If $L < p$ then the factors can be viewed as “common features” in the sense that there are fewer sources of conditional variation than the number of assets.

Engle and Kozicki’s (1993) test for common features can be motivated as follows. If $L < p$ then there exists $\theta_0 \neq 0$ such that $E[(\theta_0' Y_{t+1})^2 | \mathfrak{F}_t] = \mu$, for some constant μ , and so for any $k \times 1$ vector $z_t \in \mathfrak{F}_t$, with $k > p$, θ_0 satisfies

$$m(\theta_0) = 0 \tag{14}$$

where $m(\theta) = E[f_t(\theta)]$,

$$f_t(\theta) = z_t\{(\theta' Y_{t+1})^2 - c(\theta)\}, \tag{15}$$

and $c(\theta) = E[(\theta' Y_{t+1})^2]$. Clearly (14) only identifies θ up to some normalizing constant, and so in practice some normalization needs to be adopted. However for our purposes here, we can sidestep this issue.⁷ The population moment condition in (14) can be used as a basis for estimation of θ_0 , and the existence of the common feature can be tested by testing whether (14) holds using the overidentifying restrictions statistic.

However, the population moment condition in (14) does not locally identify θ_0 at first order. Dovonon and Renault (2013) show that

$$G(\theta) = 2E[(z_t - E[z_t])\theta_0' (\Lambda D_t \Lambda' + \Omega)], \tag{16}$$

and that under the assumptions above,

$$E[(\theta_0' Y_{t+1})^2 | \mathfrak{F}_t] = \mu \quad \Leftrightarrow \quad \theta_0' \Lambda = 0. \tag{17}$$

Therefore, $G(\theta_0)$ is the null matrix by construction under the null hypothesis of the test. However, θ_0 is second-order locally identified under plausible conditions because

$$H_s(\theta) = \Lambda' C_s \Lambda, \tag{18}$$

where C_s is the $L \times L$ diagonal matrix with ℓ^{th} main diagonal element equal to $Cov[z_{s,t}, \sigma_{\ell,t}^2]$. Dovonon and Renault (2013) argue this rank condition can be ensured by picking a sufficiently broad group of instruments z_t such that at least one instrument is correlated with every possible linear combination of the volatilities $\left\{\sigma_{\ell,t}^2\right\}$.⁸

Finally, we emphasize that in this model, the value of θ_0 is not of primary interest: the key issue is whether $m(\theta_0) = 0$.

3 Test statistics and their limiting distributions under their null hypotheses

In this section, we consider methods for testing two types of hypotheses in models that satisfy Assumption 1. In the first type, the null hypothesis takes the form: $H_0 : \theta_0 = \theta_*$. Notice that

⁷See Dovonon and Renault (2013) for further discussion and also Section 5.2 for an example.

⁸Specifically, they assume $rank\{Cov[z_t, d_t]\} = L$ where $d_t = (\sigma_{\ell,t}^2, \sigma_{\ell,t}^2, \dots, \sigma_{L,t}^2)$.

under this H_0 the value of θ_0 is completely specified. In the second type of hypothesis, the null takes the form $H_0 : m(\theta_0) = 0$; tests of this hypothesis are often interpreted as tests of whether the model specification is correct. We first present all the test statistics and then provide their limiting distributions under their respective null hypotheses.

3.1 Test statistics and their null hypotheses

To present the statistics, we introduce the following notation: $\bar{q}_N(\theta) = N^{-1} \sum_{i=1}^N q_i(\theta)$ and $q_i(\bar{\theta}) = \partial f_i(\theta) / \partial \theta' |_{\theta=\bar{\theta}}$.

Test statistics for $H_0 : \theta_0 = \theta_$:*

Newey and West (1987) propose a number of statistics for testing whether θ_0 satisfies a set of nonlinear restrictions based on GMM estimators. Here we consider two: the Wald and Lagrange Multiplier (LM) statistics. Specializing to our null hypothesis, the Wald statistic is:

$$\text{Wald}_N(\theta_*) = N(\hat{\theta}_N - \theta_*)' \bar{q}_N(\hat{\theta}_N)' \hat{V}_{ff}(\hat{\theta}_N)^{-1} \bar{q}_N(\hat{\theta}_N)(\hat{\theta}_N - \theta_*), \quad (19)$$

and the LM statistic is,

$$LM(\theta_*) = N \bar{f}_N(\theta_*)' \hat{V}_{ff}(\theta_*)^{-1} \bar{q}_N(\theta_*) \left(\bar{q}_N(\theta_*)' \hat{V}_{ff}(\theta_*)^{-1} \bar{q}_N(\theta_*) \right)^{-1} \bar{q}_N(\theta_*)' \hat{V}_{ff}(\theta_*)^{-1} \bar{f}_N(\theta_*). \quad (20)$$

Under certain regularity conditions which include global identification and first-order local identification, Newey and West (1987) show that the Wald and LM statistics both converge to a χ_ρ^2 where ρ is the number of restrictions which is p in our case here.

Kleibergen (2005) introduces a modified version of the LM statistic:

$$KLM(\theta_*) = N \bar{f}_N(\theta_*)' \hat{V}_{ff}(\theta_*)^{-1} \hat{D}_N(\theta_*) \left(\hat{D}_N(\theta_*)' \hat{V}_{ff}(\theta_*)^{-1} \hat{D}_N(\theta_*) \right)^{-1} \hat{D}_N(\theta_*)' \hat{V}_{ff}(\theta_*)^{-1} \bar{f}_N(\theta_*), \quad (21)$$

where $\hat{D}_N(\theta)$ a $k \times p$ -dimensional matrix:

$$\text{vec} \left(\hat{D}_N(\theta) \right) = \text{vec} \left(\bar{q}_N(\theta) \right) - \hat{V}_{qf}(\theta) \hat{V}_{ff}(\theta)^{-1} \bar{f}_N(\theta), \quad (22)$$

with $\hat{V}_{qf}(\theta) = N^{-1} \sum_{i=1}^N \text{vec} [q_i(\theta) - \bar{q}_N(\theta)] [f_i(\theta) - \bar{f}_N(\theta)]'$. Kleibergen (2005) shows that $KLM(\theta_*)$ converges to a χ_p^2 distribution under H_0 regardless of whether θ_0 is first order locally identified or weakly identified.

Stock and Wright (2000) propose using the GAR statistic:⁹

$$GAR(\theta_*) = Q(\theta_*, \theta_*). \quad (23)$$

Stock and Wright (2000) show that $GAR(\theta_*)$ converges to χ_k^2 distribution under H_0 regardless of whether θ_0 is first order locally identified or weakly identified. However, the implicit null of the

⁹Anderson and Rubin (1949) introduce the statistic in the context of linear models, and Dufour (1997) and Staiger and Stock (1997) advocate using this original version of the statistic for inference in linear models with weak identification.

GAR statistic is larger than $H_0 : \theta_0 = \theta_*$ as we discuss below.

Test statistics for $H_0 : m(\theta_0) = 0$:

Kleibergen (2005) proposes testing this null using the statistic

$$J(\theta_0) = N \bar{f}_N(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1/2} M_{\hat{V}_{ff}(\theta_0)^{-1/2} \hat{D}_N(\theta_0)} \hat{V}_{ff}(\theta_0)^{-1/2} \bar{f}_N(\theta_0), \quad (24)$$

where $M_A = I_k - A(A'A)^{-1}A'$. Kleibergen (2005) shows that under H_0 the limiting distribution of $J(\theta_0)$ is χ_{k-p}^2 irrespective of whether θ_0 is first-order locally or weakly identified. The test is performed by searching to see if there are any values of θ_0 for which $J(\theta_0)$ is less than the appropriate critical value.

As noted by Kleibergen (2005),

$$GAR(\theta) = KLM(\theta) + J(\theta)$$

and so the GMM-AR can be viewed as a joint test of $\theta_0 = \theta_*$ and $m(\theta_0) = 0$.

3.2 Limiting distributions under the null

For our analysis of both types of statistics, the structure of the Jacobian is important. We define $r = \text{rank}\{G(\theta_0)\}$. Since our focus is on cases where θ_0 is globally identified and only locally identified at second order, we assume $r < p$ and that the model satisfies Assumption 1. Note that if $0 < r < p$ then there exists a nonsingular $p \times p$ matrix $R = (R_1 \dot{R}_2)$ such that the $p \times r$ matrix R_1 and $p \times (p - r)$ matrix R_2 satisfy:

$$\text{rank}\{G(\theta_0)R_1\} = r \quad \text{and} \quad G(\theta_0)R_2 = 0. \quad (25)$$

The matrices R_1 and R_2 are key to our analysis below because they give respectively the directions of possible fast convergence estimation and the directions of slower convergence estimation. If $r = 0$ (as in the CHF example) then we set $R = R_2 = I_p$ and $R_1 = 0$. In the subsequent analysis, we set $D = G(\theta_0)R_1$.

We also impose the following conditions.

Assumption 2. θ_0 is an interior point of Θ .

Let \mathcal{N}_ϵ denote an ϵ -neighbourhood of θ_0 .

Assumption 3. (i) $\|m(\theta)\| < \infty$, $\|G(\theta)\| < \infty$ and $\|H_s(\theta)\| < \infty$ for $s = 1, 2, \dots, k$ for all $\theta \in \mathcal{N}_\epsilon$;
(ii) $\bar{f}_N(\theta)$ converges uniformly in probability to $m(\theta)$ and the partial derivatives up to order 2 of $\bar{f}_N(\theta)$ converge in probability uniformly to those of $m(\theta)$ over \mathcal{N}_ϵ .

Assumption 4.

$$\sqrt{N} \begin{pmatrix} \bar{f}_N(\theta_0) \\ \text{vec}(\bar{q}_N(\theta_0)R_2) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_f \\ \text{vec}(\psi_q) \end{pmatrix} \sim N(0, V),$$

where $V = \begin{pmatrix} V_{ff}(\theta_0) & V_{f2}(\theta_0) \\ V_{2f}(\theta_0) & V_{22}(\theta_0) \end{pmatrix}$, with

$$\begin{aligned} V_{2f}(\theta) &= E \{ [(q_i(\theta) - \mu_q(\theta))R_2][f_i(\theta) - \mu_f(\theta)]' \}, & V_{f2}(\theta) &= V_{2f}(\theta)', \\ V_{22}(\theta) &= E \{ [(q_i(\theta) - \mu_q(\theta))R_2R_2'[q_i(\theta) - \mu_q(\theta)]] \}, & \mu_f(\theta) &= E(f_i(\theta)) \quad \text{and} \quad \mu_q(\theta) = E(q_i(\theta)). \end{aligned}$$

Assumption 4 is a high-level condition that can apply whether the model involves random vector X or a time series process X_t , in the latter case V is the long run variance of the relevant random vector.

Under Assumptions 1(a) and certain other regularity conditions, $\hat{\theta}_{1,s}$ and $\hat{\theta}_N$ are consistent. Since this is not the focus of our analysis, we do not document the required conditions here, and instead adopt the following high-level assumption.¹⁰

Assumption 5. $\hat{\theta}_{1,s} \xrightarrow{P} \theta_0$ and $\hat{\theta}_N \xrightarrow{P} \theta_0$.

Given the consistency of $\hat{\theta}_N$, it follows from Assumption 3 that $\bar{q}_N(\hat{\theta}_N)R_1 \xrightarrow{P} D$.

We now present the limiting distributions of the test statistics presented in Section 3.1.

Test statistics for $H_0 : \theta_0 = \theta_$:*

For the Wald statistic, we consider only the case where $r = p - 1$ because to our knowledge this is the only case for which the limiting distribution of the GMM estimator is tractable. For what follows, it is useful to introduce the following additional notation:

$$\begin{aligned} P &= \tilde{D} \left(\tilde{D}' \tilde{D} \right)^{-1} \tilde{D}', & \tilde{D} &= V_{ff}(\theta_0)^{-1/2} D, & M_d &= I_k - P, \\ B &= (R_2' H_s(\theta_0) R_2)_{1 \leq s \leq k}, & \tilde{B} &= V_{ff}(\theta_0)^{-1/2} B, & \text{and} \quad \alpha &= \frac{\tilde{B}}{\sqrt{\tilde{B}' M_d \tilde{B}}}. \end{aligned}$$

Theorem 1. *If Assumptions 1-5 hold, $r = p - 1$ and $\theta_0 = \theta_*$ then*

$$\text{Wald}_N(\theta_*) \xrightarrow{d} \mathbb{W}$$

where

$$\mathbb{W} = \mathbb{W}_0(S, S_1) \equiv (S_1 + \alpha S \mathbb{I}(S \leq 0))' P (S_1 + \alpha S \mathbb{I}(S \leq 0)) + 4S^2 \mathbb{I}(S \leq 0),$$

and: $S_1 \sim N(0, I_k)$, $S \sim N(0, 1)$, S_1 and S are independent and $\mathbb{I}(\cdot)$ is the usual indicator function.

The limiting distribution is evidently non-standard, reflecting the non-standard behaviour of the GMM estimator in this case (see Dovonon and Hall (2016)[Theorem 1]). Although non-standard this distribution can easily be simulated, along similar lines to the method proposed for simulating the distribution of the GMM estimator in Dovonon and Hall (2016).¹¹ In the special case when $r = 0$ and $p = 1$ then the distribution simplifies. In this case, we set $D = 0$, $P = 0$ and $B = (H_s(\theta_0))_{1 \leq s \leq k}$, and the distribution of the Wald test is as follows.

Corollary 1. *If the conditions of Theorem 1 hold and in addition $r = 0$ and $p = 1$ then $\mathbb{W} = \mathbb{W}_0(S) \equiv 4S^2 \mathbb{I}(S \leq 0)$ where S is defined in Theorem 1.*

¹⁰For example, see Hansen (1982), Newey and McFadden (1994) or Hall (2005)[Chapter 3].

¹¹Dovonon and Hall (2016) also discuss at length how to estimate R_2 .

Corollary 1 provides the limiting distribution of the Wald test for the test of $H_0 : \theta_0 = 1$ in our panel data example in Section 2.1. Notice that this limiting distribution involves a point mass of 0.5 for the event $Wald_N(\theta_*) = 0$. We can use our panel data example to provide some intuition for why the distribution takes the form it does. In this setting, the Wald statistic is:

$$Wald_N(1) = N(\hat{\theta} - 1)\bar{q}_N(\hat{\theta})'\hat{V}_{ff}(\hat{\theta})^{-1}\bar{q}_N(\hat{\theta})(\hat{\theta} - 1). \quad (26)$$

Using a Mean Value expansion of $\bar{q}_N(\hat{\theta})$ around $\bar{q}_N(1)$, it can be shown that¹²

$$Wald_N(1) = N(\hat{\theta} - 1)^4 4\sigma^4 V_{1,1}^{-1} \quad (27)$$

where $V_{1,1}^{-1}$ is the (1, 1) element of $\{V_{ff}(1)\}^{-1}$. If we define ζ via $N^{1/4}(\hat{\theta} - 1) = \zeta + o_p(1)$ and set $e = V_{1,1}^{-1}$ then it is shown in the mathematical appendix that, under H_0 , the first order conditions of the GMM estimation imply that ζ satisfies the following condition:

$$\zeta \left(\zeta^2 + \frac{1}{e^{1/2}\sigma^2} S \right) = 0. \quad (28)$$

If $S > 0$ then there is no real value of ζ that can set the term in parentheses to zero, and so the solution must be $\zeta = 0$. However, if $S < 0$ then

$$\zeta^2 = \frac{1}{e^{1/2}\sigma^2} |S|,$$

sets the term in parentheses to zero. Thus, we have

$$\zeta^2 = \mathbb{I}(S \leq 0) \frac{1}{e^{1/2}\sigma^2} |S|, \quad (29)$$

Using (29) in (27), it follows that

$$\begin{aligned} Wald_N(1) &= \zeta^4 4\sigma^4 e + o_p(1), \\ Wald_N(1) &\xrightarrow{d} \left\{ \mathbb{I}(S \leq 0) \frac{1}{e^{1/2}\sigma^2} |S| \right\}^2 4\sigma^4 e = 4S^2 \mathbb{I}(S \leq 0). \end{aligned} \quad (30)$$

The Wald test principle is based on testing whether the unrestricted estimator satisfies the restrictions in question. In contrast, the test principles behind the LM, KLM and GAR statistics are based on the restricted model. In our case here, the null hypothesis completely specifies the value of θ_0 and so calculation of these statistics does not involve a GMM estimation *per se*. Therefore, while our analysis assumes identification fails locally at first order in an arbitrary number of directions, it does not require the parameters to be locally identified at second order - although the results still hold if that is the case.

The following theorem gives the limiting distribution of the LM statistic in (20).

Theorem 2. *If Assumption 4 holds, $\tilde{\psi}'_q \tilde{\psi}_q$ is nonsingular with probability one, $\hat{V}_{ff}(\theta_0)$ and $\bar{q}_N(\theta_0)R_1$ converge in probability to $V_{ff}(\theta_0)$ and D , respectively, and $\theta_0 = \theta_*$ then:*

$$LM(\theta_*) \xrightarrow{d} \mathbb{L} \equiv \psi'_f V_{ff}(\theta_*)^{-1} \tilde{\psi}_q \left(\tilde{\psi}'_q V_{ff}(\theta_*)^{-1} \tilde{\psi}_q \right)^{-1} \tilde{\psi}_q V_{ff}(\theta_*)^{-1} \psi_f,$$

where $\tilde{\psi}_q = (D : \psi_q)$. If in addition ψ_f and ψ_q are uncorrelated, then $\mathbb{L} = \chi_p^2$.

¹²See equation (58) in the mathematical appendix.

Theorem 2 gives the asymptotic distribution of the LM statistic under H_0 when the first order local identification condition is violated. Only in the special case where $\sqrt{N}\bar{q}_N(\theta_0)R_2$ and $\sqrt{N}\bar{f}_N(\theta_0)$ are asymptotically uncorrelated (and hence independent) is this distribution χ_p^2 and so the same as would be the case if θ_0 is identified locally at first order. A comparison of Theorems 1 and 2 indicates that the limiting distributions of the Wald and LM statistics are different if identification fails locally at first order but holds at second order. In contrast, Newey and West (1987) show the two statistics are asymptotically equivalent under the null when θ_0 is first order locally identified.

The following theorem gives the limiting distributions of the KLM and GAR statistics in (21) and (23) respectively. We first introduce some notation. Let $\hat{\psi}_q$ be the $k \times p$ matrix with its (l, m) -entry given by

$$\hat{\psi}_{q,lm} = Cov[q_{i,lm}(\theta_0), f_i(\theta_0)]\{V_{ff}(\theta_0)\}^{-1}\psi_f,$$

$l = 1, \dots, k$ and $m = 1, \dots, p$. Let $\varepsilon_q = \psi_q - \hat{\psi}_q R_2$,

$$\bar{\psi}_q = \begin{cases} \varepsilon_q & \text{if } r = 0 \\ (D \dot{\varepsilon}_q) & \text{if } r > 0 \end{cases}$$

and $\hat{V}_{2f}(\theta_0)$ be the sample counterpart of $V_{2f}(\theta_0)$ as defined in Assumption 4. We have:

Theorem 3. (i) If Assumption 4 holds, $\bar{\psi}'_q \bar{\psi}_q$ is nonsingular with probability one, $\hat{V}_{2f}(\theta_0)$, $\hat{V}_{ff}(\theta_0)$ and $\bar{q}_N(\theta_0)R_1$ converge in probability to $V_{2f}(\theta_0)$, $V_{ff}(\theta_0)$ and D , respectively, and $\theta_0 = \theta_*$ then $KLM(\theta_*) \xrightarrow{d} \chi_p^2$; (ii) If $\sqrt{N}\bar{f}_N(\theta_0) \xrightarrow{d} N(0, V_{ff}(\theta_0))$, $\hat{V}_{ff}(\theta_0)$ converges in probability to $V_{ff}(\theta_0)$ and $\theta_0 = \theta_*$ then $GAR(\theta_*) \xrightarrow{d} \chi_k^2$.

From Theorem 3 it follows that the limiting distributions of the KLM and GAR statistics under second-order local identification are the same as under first-order local identification and weak identification. Therefore both statistics are robust to all three forms of identification.

Test statistics for $H_0 : m(\theta_0) = 0$:

The following theorem presents the limiting distributions of $J(\theta_0)$ and $GAR(\theta_0)$ under this null hypothesis.

Theorem 4. (i) If Assumptions 4 holds, $\bar{\psi}'_q \bar{\psi}_q$ is nonsingular with probability one, $\hat{V}_{2f}(\theta_0)$, $\hat{V}_{ff}(\theta_0)$ and $\bar{q}_N(\theta_0)R_1$ converge in probability to $V_{2f}(\theta_0)$, $V_{ff}(\theta_0)$ and D , respectively, then $J(\theta_0) \xrightarrow{d} \chi_{k-p}^2$; (ii) If $\sqrt{N}\bar{f}_N(\theta_0) \xrightarrow{d} N(0, V_{ff}(\theta_0))$ and $\hat{V}_{ff}(\theta_0)$ converges in probability to $V_{ff}(\theta_0)$ then $GAR(\theta_0) \xrightarrow{d} \chi_k^2$.

From Theorem 4 it follows that the limiting distribution of the K-J statistic under second-order local identification is the same as under first-order local identification and weak identification, and so it is robust to all three forms of identification. This contrasts with Hansen's (1982) overidentifying restrictions test statistic which Dovonon and Renault (2013) show converges in distribution to a mixture of χ_{k-q}^2 , $q = 0, 1, \dots, p$, distributions if θ_0 is only locally identified at second order. The limiting distribution of the $GAR(\theta_0)$ follows trivially from the asymptotic normality of $\sqrt{N}\bar{f}_N(\theta_0)$.

4 The large sample behaviour of the test statistics under local alternatives

In this section, we explore the local power properties of the tests. To this end, we index the data generation process by N and so now replace X by X_N . The distribution of X_N is denoted by P_N and this distribution implies the population moment condition

$$E_N [f(X_N, \theta_N)] = \mu_N, \quad (31)$$

where $E_N[\cdot]$ denotes expectation under P_N , $\{\theta_N\}$ is a sequence of parameter values and $\{\mu_N\}$ is a sequence of $k \times 1$ vectors. It is assumed that as $N \rightarrow \infty$ the following all hold: $P_N \rightarrow P$, $\theta_N \rightarrow \theta_0$ and $\mu_N \rightarrow 0_{k \times 1}$. Recall that P is the probability distribution of X in Section 2, and so the limit process satisfies the population moment condition (1). As in Section 2, it is assumed further that under P , θ_0 is identified locally at second order.

To analyze the behaviour of the tests under local alternatives, we must also modify certain of the assumptions. To this end, we introduce the following definitions:

$$m_N(\theta) = E_N [f(X_N, \theta)], \quad G_N(\theta) = E_N [q(X_N, \theta)], \quad H_N(\theta) = E_N [h(X_N, \theta)],$$

$$H(\theta) = E [h(X, \theta)], \quad h(X_N, \dot{\theta}) = \left. \frac{\partial \text{vec}\{q(X_N, \theta)\}}{\partial \theta'} \right|_{\theta=\dot{\theta}}, \quad \bar{h}_N(\theta) = N^{-1} \sum_{i=1}^N h(x_i, \theta).$$

We replace Assumption 3 by the following condition.

Assumption 6. (i) $\|m_N(\theta)\| < \infty$, $\|G_N(\theta)\| < \infty$, $\|H_N(\theta)\| < \infty$ for $\theta \in \mathcal{N}_\epsilon$; (ii) over a neighborhood \mathcal{N}_ϵ , the following hold: $\bar{f}_N(\theta)$, $m_N(\theta)$ converge uniformly (in probability P_N for the former) to $m(\theta)$; $\bar{q}_N(\theta)$, $G_N(\theta)$ converge uniformly (in probability P_N for the former) to $G(\theta)$, $\bar{h}_N(\theta)$, $H_N(\theta)$ converge uniformly (in probability P_N for the former) to $H(\theta)$.

We must also modify our assumptions about the behaviour of the Jacobian. It is worth mentioning that, even if the rank property of the Jacobian at θ_0 under P (the data distribution under the null) is known, this does not necessarily imply the rank property under θ_N because of the lack of continuity of the rank function.

Assumption 7.

$$G_N(\theta_N)R_1 = D + o(1), \quad \text{and} \quad G_N(\theta_N)R_2 = N^{-\xi}A,$$

where $R \equiv (R_1 \dot{=} R_2)$ is the nonsingular $p \times p$ matrix partitioned into r and $(p-r)$ -column matrices R_1 and R_2 as defined by (25). D is a $p \times r$ matrix of rank r , A is a $k \times p-r$ matrix and $\xi > 0$.

Under this assumption, the Jacobian is local to zero in the directions of the parameter that are identified locally only at the second order. The specific choice of ξ likely depends on the model in question. We show below that $\xi = 1/2$ is the appropriate choice in both our examples in Section 2. For our analysis of tests of $H_0 : \theta_0 = \theta_*$, we restrict $\xi > 1/4$ to ensure that the drift in the Jacobian decreases faster than the rate of convergence of the second order identified parameters. Such a restriction is particularly useful to derive the asymptotic distribution of the Wald test statistic. Finally, we replace Assumption 4 by the following condition.

Assumption 8.

$$\sqrt{N} \begin{pmatrix} \bar{f}_N(\theta_N) - \mu_N \\ \text{vec} [\bar{q}_N(\theta_N)R_2 - N^{-\xi}A] \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_f \\ \text{vec}(\psi_q) \end{pmatrix} \sim N(0, V)$$

under P_N , with V given in Assumption 4.

Section 4.1 covers tests of $H_0 : \theta_0 = \theta_*$; Section 4.2 considers the tests of $H_0 : m(\theta_0) = 0$.

4.1 Local power of tests of $H_0 : \theta_0 = \theta_*$

For this null hypothesis, the natural sequence of local is given by (31) with $\mu_N = 0$ for all N . In this case, the population moment condition is satisfied at a different parameter value for each N that is,

$$m_N(\theta_N) = 0. \tag{32}$$

To explicitly define the sequence of parameter θ_N under the local alternative, we take into account the rate of convergence of estimators under the null. Under the second-order identification condition, we know that the directions of the parameters that are identified at the first order are estimated at the standard \sqrt{N} -rate whereas the directions that are identified only at the second order are estimated at a slower $N^{1/4}$ -rate. In particular, considering R as defined by Equation (25), we know that the first r components of $R^{-1}\theta$ are estimated at \sqrt{N} -rate whereas the remaining components are estimated at the $N^{1/4}$ -rate. In the light of this, we define θ_N such that:

$$\theta_N - \theta_* = Re_N, \tag{33}$$

where the first r and the last $(p - r)$ components of $e_N \in \mathbb{R}^p$, denoted respectively $e_{N,1}$ and $e_{N,2}$ are such that:

$$e_{N,1} = \frac{e_1}{\sqrt{N}} \quad \text{and} \quad e_{N,2} = \frac{e_2}{\sqrt[4]{N}},$$

with e_1 and e_2 are nonzero vectors of size r and $p - r$, respectively.

Before presenting the limiting distributions of our test statistics, it is instructive to use our panel data example to motivate the behaviour of the Jacobian specified in Assumption 7. Recall from Section 2.1 that θ is a scalar and is only locally identified at second order. Therefore, in view of the remarks in the preceding paragraph, we set $\theta_N = 1 - \frac{c}{2\sqrt[4]{N}}$. In this case, it can be shown that¹³

$$G_N(\theta_N) = -\frac{c^2}{4\sqrt{N}}\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad H_1(\theta_0) = 2\sigma^2, \quad H_2(\theta_0) = 0, \tag{34}$$

This setting is covered by Assumption 7 with $\xi = 1/2$ and $A = -(\sigma^2 c^2/4)[1, 0]'$.

Theorem 5. *If Assumptions 1, 2, 6-8 (with $\mu_N = 0$ and $\xi > 1/4$) hold, Assumption 5 holds under P_N , $\theta_0 = \theta_*$ and $r = p - 1$ then: any subsequence of $\text{Wald}_N(\theta_*)$ has a further subsequence with index say, $s(N)$, that converges in distribution under P_N to \mathbb{W}_s , defined by:*

$$\mathbb{W}_s = \mathbb{W}_0(S, S_1) + \mathbb{W}_1(S, S_1, s) + \lambda_1 + \lambda_2(S, s),$$

¹³See mathematical appendix.

for

$$\begin{aligned}\mathbb{W}_1(S, S_1, s) &= -2(S_1 - \alpha S\mathbb{I}(S \leq 0))' P \left(\tilde{D}e_1 + ae_2\alpha\mathbb{X}(s) \right) \\ &\quad + 2e_1'\tilde{D}'\alpha(-2S\mathbb{I}(S \leq 0) + ae_2\mathbb{X}(s)), \\ \lambda_2(S, s) &= -2ae_2\alpha'(2\mathbb{X}(s) + e_2)S\mathbb{I}(S \leq 0) \\ \lambda_1 &= e_1'\tilde{D}'\tilde{D}e_1.\end{aligned}$$

where $S_1 \sim N(0, I_k)$, $S \sim N(0, 1)$ and S is independent of S_1 . $\mathbb{X}(s)$ is a random variable depending on the subsequence $s(N)$ and satisfies for any subsequence s : $\mathbb{X}(s)^2 = -(2/a)S\mathbb{I}(S \leq 0)$. $a = \sqrt{\tilde{G}'M_d\tilde{G}}$ and $\mathbb{W}_0(S, S_1)$, α , M_d , \tilde{G} and \tilde{D} are given in Theorem 1.

For the case in which $r = 0$ and $p = 1$, this results specializes as follows.

Corollary 2. *If $r = 0$ and $p = 1$, then*

$$\mathbb{W}_s = \mathbb{W}_0(S) - 2ae_2(2\mathbb{X}(s) + e_2)S\mathbb{I}(S \leq 0),$$

where $S \sim N(0, 1)$, $\mathbb{X}(s)^2 = -(2/a)S\mathbb{I}(S \leq 0)$, $a = \sqrt{\tilde{G}'\tilde{G}}$, and $\mathbb{W}_0(S)$ is given in Corollary 1

Notice that in this case the power against local alternatives is capped at 0.5 asymptotically; we return to this issue in Section 5.1.

To present the limiting behaviour of the LM, KLM and GAR tests, we define $C(\theta)$ to be the $k \times p^2$ matrix:

$$C(\theta) = \left(\text{vec} \left(\frac{\partial^2 m_1(\theta)}{\partial \theta \partial \theta'} \right) \quad \text{vec} \left(\frac{\partial^2 m_2(\theta)}{\partial \theta \partial \theta'} \right) \quad \dots \quad \text{vec} \left(\frac{\partial^2 m_k(\theta)}{\partial \theta \partial \theta'} \right) \right)'$$

Theorem 6. *If Assumptions 1(b), 2, 6-8 (with $\mu_N = 0$ and $\xi > 1/4$) hold, and $\theta_0 = \theta_*$ then:*

(a) *If the $k \times p$ matrix $Q(e_2)$ defined by:*

$$Q(e_2) = \left(D \quad ; \quad -C(\theta_*) [I_p \otimes (R_2 e_2)] R_2 \right)$$

is full column rank, then:

$$LM(\theta_*), KLM(\theta_*) \xrightarrow{d} \chi_p^2(\lambda_\theta)$$

under P_N , with $\lambda_\theta = \mu'_\theta V_{ff}(\theta_)^{-1} \mu_\theta > 0$ iff $(e_1, e_2) \neq 0$; $\mu_\theta = -De_1 + \frac{1}{2} [(R_2 e_2)' \otimes I_k] H(\theta_*)(R_2 e_2)$.*

(b)

$$GAR(\theta_*) \xrightarrow{d} \chi_k^2(\lambda_\theta)$$

under P_N with λ_θ as given above.

This theorem shows that the LM and KLM statistics have the same limiting distribution under this sequence of local alternatives. Since $\lambda_\theta > 0$, it follows automatically from Theorems 3 and 6 and the properties of the chi-squared distribution that both the KLM and GAR statistics have

non-trivial power against this alternative and also that the KLM statistic is the more powerful. The relative performance of the LM statistic is less clear. Theorem 2 indicates that in general the LM statistic has a non-standard limiting distribution under the null, but does have the (standard) limiting χ_p^2 distribution in the special case where $\sqrt{N}\bar{q}_N(\theta_0)R_2$ and $\sqrt{N}\bar{f}_N(\theta_0)$ are asymptotically independent. In the former case, it is not possible to make a power comparison with the KLM and GAR statistics analytically. It is worth noting that the differences in the distributions of the LM statistic under null and local alternative can be rationalized as follows. Under the null, the large sample behaviour of $LM(\theta_*)$ depends on $\sqrt{N}\bar{q}_N(\theta_0)R_2$ which is random in the limit, and may or (most likely) may not be asymptotically independent of $\sqrt{N}\bar{f}_N(\theta_0)$. Under the local alternative, the large sample behaviour of $LM(\theta_*)$ depends on $N^{1/4}\bar{q}_N(\theta_0)R_2$ which converges in probability to a constant, and so is trivially independent of $\sqrt{N}\bar{f}_N(\theta_0)$.

4.2 Local power of tests of $H_0 : m(\theta_0) = 0$

For this null hypothesis, the natural sequence of local is given by (31) with $\theta_N = \theta_0$ and $\mu_N = c/\sqrt{N}$ for all N so that

$$m_N(\theta_0) = \frac{c}{\sqrt{N}}. \quad (35)$$

However, as noted above, the appropriate choice of ξ in (35) depends on the model in question. To illustrate, we consider the CHF model in Section 2.2 with two assets.

Under the alternative of no-common conditionally heteroskedastic factors structure, each asset brings a specific dimension for conditional heteroskedasticity so that two factors are present. The volatility factor model in (13) can then be written as:

$$E[Y_{t+1}Y'_{t+1}|\mathfrak{F}_t] = \lambda_1\lambda'_1\sigma_{1,t}^2 + \lambda_{2,N}\lambda'_{2,N}\sigma_{2,t}^2 + \Omega.$$

A natural way to create a local alternative to a single common factor is to assume that the return process is generated for a given sample size N from a probability distribution P_N such that, as $N \rightarrow \infty$, $\lambda_{2,N} \rightarrow 0$. Therefore, the common conditionally heteroskedastic factor structure holds in the limit but not in finite samples. Let θ_0 be the co-feature vector associated to the limit model. Then $\theta'_0\lambda_1 = 0$ and under P_N , we have:¹⁴

$$m_N(\theta_0) = (\theta'_0\lambda_{2,N})^2 Cov[\sigma_{2,t}^2, z_t], \quad (36)$$

where $Cov[\cdot, \cdot]$ here denotes the covariance operator relative to P_N . Suppose now that $\lambda_{2,N} = \lambda/N^\delta$, with $\lambda \in \mathbb{R}^2$. The right hand side of (36) may be of order $O(N^{-2\delta})$ so long as $\theta'_0\lambda_{2,N} \neq 0$ and $Cov[\sigma_{2,t}^2, z_t] \neq 0$. However, the order of magnitude of this latter term depends on that of $\lambda_{2,N}$ through the choice of the vector of instruments z_t . The most common choice of instruments is $z_t = (vech(Y_{t-\tau}Y'_{t-\tau}) : \tau = 0, \dots, h)'$, for some $h \in \mathbb{N}$. To simplify, let us consider $z_t = (Y_{1t}^2, Y_{2t}^2)'$. Under certain commonly invoked assumptions about the asset return process, it is shown in the mathematical appendix that:¹⁵

$$m_N(\theta_0) = \begin{pmatrix} \lambda_{2,N,1}^2 \\ \lambda_{2,N,2}^2 \end{pmatrix} (\theta'_0\lambda_{2,N})^2 Cov[F_{2,t+1}^2, F_{2,t}^2], \quad (37)$$

¹⁴See mathematical appendix

¹⁵Note that due to the necessary normalization only one element of θ has to be estimated; see discussion in Section 2.2.

where $\lambda_{2,N} = (\lambda_{2,N,1}, \lambda_{2,N,1})'$, and

$$G_N(\theta_0) = 2Cov[F_{2,t+1}^2, F_{2,t}^2](\theta_0' \lambda_{2,N}) \begin{pmatrix} \lambda_{2,N,1}^2 \\ \lambda_{2,N,2}^2 \end{pmatrix} \lambda_{2,N}'. \quad (38)$$

Assuming $Cov[F_{2,t+1}^2, F_{2,t}^2] \neq 0$ - a reasonable assumption as the factors are assumed conditionally heteroskedastic - it follows that:

$$m_N(\theta_0) = \frac{c}{N^{4\delta}}, \quad \text{and} \quad G_N(\theta_0) = \frac{A}{N^{4\delta}}$$

where c is a 2×1 non-zero vector of constants, and A is a non-null 2×2 matrix of constants. Thus setting $\delta = 1/8$ to ensure $\mu_N = c/N^{4\delta} = c/\sqrt{N}$, we also obtain $\xi = 1/2$.

While the \sqrt{N} -rate for the drifting sequence in (35) is convenient to obtain a non-trivial behaviour of the test statistics of interest under local alternatives as we shall see, the following result allows for the Jacobian of the moment function at θ_0 under P_N to converge to 0 in some directions at any rate N^ξ , $\xi > 0$. To derive the asymptotic distribution of the specification test statistics $J(\theta_0)$ and $GAR(\theta_0)$ under local alternatives, we introduce some notation.

Let $\hat{\psi}_q^a$ be the $k \times p$ matrix with its (l, m) -entry given by

$$\hat{\psi}_{q,lm}^a = Cov[q_{i,lm}(\theta_0), f_i(\theta_0)] \{V_{ff}(\theta_0)\}^{-1} (\psi_f + c),$$

$l = 1, \dots, k$ and $m = 1, \dots, p$. Let

$$\varepsilon_q^a = \begin{cases} \psi_q + A - \hat{\psi}_q^a R_2 & \text{if } \xi = \frac{1}{2} \\ A & \text{if } 0 < \xi < \frac{1}{2} \\ \psi_q - \hat{\psi}_q^a R_2 & \text{if } \xi > \frac{1}{2} \end{cases}, \quad \bar{\psi}_q^a = \begin{cases} \varepsilon_q^a & \text{if } r = 0 \\ (D : \varepsilon_q^a) & \text{if } r > 0 \end{cases}$$

and $\lambda_m = c' V_{ff}(\theta_0)^{-1/2} M_{V_{ff}(\theta_0)^{-1/2} \bar{\psi}_q^a} V_{ff}(\theta_0)^{-1/2} c$. Letting $\langle M \rangle$ denote the column span of M , We have:

Theorem 7. (i) Assume that $G_N(\theta_0) \rightarrow G(\theta_0)$ as $N \rightarrow \infty$ and $rank(G(\theta_0)) = r < p$. If Assumptions 7 and 8 (with $\theta_N = \theta_0$, $\mu_N = c/N^{1/2}$, $c \in \mathbb{R}^k$) hold, $\hat{V}_{2f}(\theta_0)$, $\hat{V}_{ff}(\theta_0)$ and $\bar{q}_N(\theta_0) R_1$ converge in probability (under P_N) to $V_{2f}(\theta_0)$, $V_{ff}(\theta_0)$ and D , respectively, $\bar{\psi}_q^a$ is full column rank with probability one and $P(c \in \langle \bar{\psi}_q^a \rangle) = 0$, then: $J(\theta_0) \xrightarrow{d} \chi_{k-p}^2(\lambda_m)$ under P_N , with $\lambda_m > 0$ almost surely; (ii) If $\sqrt{N}(\bar{f}_N(\theta_0) - c/\sqrt{N}) \xrightarrow{d} N(0, V_{ff}(\theta_0))$, under P_N , and $\hat{V}_{ff}(\theta_0)$ converges in probability (under P_N) to $V_{ff}(\theta_0)$ then $GAR(\theta_0) \xrightarrow{d} \chi_k^2(\nu)$ under P_N , with $\nu = c' V_{ff}(\theta_0)^{-1} c$.

The first part of this theorem shows that the K-J statistic is asymptotically distributed as a noncentral chi-squared with $k - p$ degrees of freedom and non-centrality λ_m which is random if $\xi \geq 0.5$. The randomness of λ_m stems from the fact that the estimated Jacobian matrix of the estimating function in the parameter directions that are not (locally) identified at first order is asymptotically random. This non-centrality parameter is almost surely positive and therefore warrants non trivial power for the test under local alternatives if the drift parameter c does not fall into the column-span of the limiting distribution of the Jacobian with positive probability. The second part of the theorem establishes that the GAR test also has non trivial power against local

alternatives since $\nu > 0$ so long as $c \neq 0$. A power ranking of the two tests is possible if $V_{ff}(\theta_0)^{-1/2}c$ is an element of the orthogonal complement of $V_{ff}(\theta_0)^{-1/2}\bar{\psi}_q^a$ almost surely then $\lambda_m = \nu$ and so the K-J statistic is unambiguously more powerful than the GAR statistic against the local alternative considered here.

5 Simulation evidence

In this section we explore the finite sample power properties of the tests analyzed in Section 3 and 4. Section 5.1 explores the power properties of the Wald, LM, KLM and GAR statistics for testing $H_0 : \theta_0 = 1$ in the panel data example in Section 2.1. Section 5.2 explores the power properties of the K-J and GAR statistics for testing $H_0 : m(\theta_0)$ in the CHF model in Section 2.2, and also compares their properties to those of Hansen's (1982) overidentifying restrictions statistic.

5.1 Testing for a unit root in the panel data model

We study inference on the autoregressive parameter of a panel autoregressive model of order one identified by the moment conditions from Section 2.1 under local alternatives to $\theta_0 = 1$, the point of second order identification. We specify the local alternative as

$$\theta_N = 1 - \frac{c}{2\sqrt[4]{N}}, \quad (39)$$

with $c > 0$.

Recall from the discussion following Corollary 1 that under the null hypothesis that $\theta_0 = 1$, the first order conditions imply a solution $N^{1/4}(\hat{\theta}_N - 1)$ if $S \geq 0$ and a solution for $N^{1/2}(\hat{\theta}_N - 1)^2$ if $S < 0$. Under P_N , the situation becomes more complicated. In this case, if $S \geq 0$ then the first order conditions imply a solution for $N^{1/4}(\hat{\theta}_N - 1)$, but if $S < 0$ then $N^{1/4}(\hat{\theta}_N - 1)$ satisfies a quadratic equation the roots of which do not imply a unique value of $N^{1/2}(\hat{\theta}_N - 1)^2$. Here we consider the local power curve implied by choosing the smallest root of the aforementioned quadratic equation as this maximizes $N^{1/2}(\hat{\theta}_N - 1)^2$ and hence the limiting value of the Wald statistic, making it the root with the largest asymptotic power. Let $\text{Wald}_N^*(1)$ denote the Wald statistic evaluated at the solution for θ just described. It is shown in the mathematical appendix that under the local alternative in (39) the distribution of $\text{Wald}_N^*(1)$ is given by:

$$\begin{aligned} \text{Wald}_N^*(1) &\xrightarrow{d} |S|(c_{2s,\lambda} + 2\sqrt{|S|})^2 I(S < 0) \\ &= |S|(4|S| + c_{2s,\lambda}^2 + 4c_{2s,\lambda}|S|^{1/2}) \mathbb{I}(S < 0) \end{aligned} \quad (40)$$

with S a standard normal random variable and

$$c_{2s,\lambda} = \frac{c\sigma}{\sqrt[4]{V_{11.2}}}$$

where $V = V_{ff}(1)$, $V_{11.2} = V_{11} - V_{12}V_{22}^{-1}V_{12}$, and V_{ij} the $i - j^{\text{th}}$ element of V . As noted in Section 4.1, the maximal local asymptotic power is 50%. We therefore compute the rejection frequency of H_0 under local alternatives for different sample sizes to determine if they are also at most 50%. Figure 1 shows the distribution of the Wald statistic for different sample sizes as a function of the localizing parameter c . It uses 10^4 simulations and a value of σ^2 equal to one with normal errors.

The local power curves of the Wald statistic in Figure 1 show that the finite sample discriminatory power can be much larger than 50% and even equal to one. Figure 1 also shows that the power curves slowly move to the right when the sample size increases so they might eventually coincide with the asymptotic local power curve from Corollary 2. This moderate convergence of the finite sample distributions of the Wald statistic results from the quartic root convergence rate. Interestingly, the convergence towards the limiting distribution when the null hypothesis holds is much faster since we do not observe any size distortions. The power curves are all very similar and show that the Wald statistic has adequate power at small sample sizes. This can be further inferred from the values of θ when the drifting parameter c equals two. The power then exceeds 50%. A value of c equal to two corresponds with a value of θ of 0.6239 ($N = 50$), 0.6838 (100), 0.7885 (500), 0.8222 (1000), 0.8811 (5000), 0.9000 (10000) and 0.9159 (20000). This suggests that - as emphasized by the name - the local power results are only a guide to behaviour in a small neighbourhood around the null hypothesis value.

Specializing Theorem 6 to the model here, it follows that the KLM and LM statistics both converge to the $\chi_1^2(\lambda_\theta)$ distribution, and the GAR statistic converges $\chi_2^2(\lambda_\theta)$ distribution. In the mathematical appendix, it is shown that the non-centrality parameter is given by:

$$\lambda_\theta = \frac{1}{16}\sigma^4 c^4 \begin{pmatrix} 1 \\ 0 \end{pmatrix}' V_{ff}(\theta_0)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (41)$$

Figures 2, 3 and 4 show the local power curves of the GAR, KLM and LM tests for increasing number of observations. Figure 5 shows local power curves of the GAR and Wald tests. The power curves of the GAR, KLM and LM tests all move to the left when the number of observations increases. It shows again the slow convergence rates of the statistics towards their limiting distributions under the local alternative. All statistics are size correct under the null hypothesis where their limiting distributions are standard χ_1^2 or χ_2^2 , in case of the GAR statistic, distributions.

The power curve of the GAR statistic shows that it has decent power while the power of the KLM and LM statistics only becomes reasonable when there are many observations. This is unlike the power of the Wald statistic which already has adequate power for small numbers of observations. It is interesting to relate the behaviour of the KLM and GAR statistics to previous analyses of these tests in other identification scenarios. If identification is weak then it has been found that the KLM statistic is size correct but has low power, and the GAR statistic is both size correct and also has good power compared to other weak identification robust procedures, see e.g. Andrews, Moreira, and Stock (2006) and Kleibergen (2005). However, if identification is strong then the KLM test dominates. Therefore, the relative performance of the KLM and GAR tests under second-order identification is more in line with what has been observed under weak identification.

To our reading, the most striking feature of these results is the superior performance of the Wald test as further reflected by Figure 5. It not only dominates the others but exhibits reasonable power as a test for a unit root in this model. These results also show an advantage to basing inference about a unit root value of the AR parameter on the moment conditions in Bun and Kleibergen (2016) as opposed to more popular choices of moments such as those proposed by Arellano and Bond (1991) or Blundell and Bond (1998) with which identification either fails or is problematic at $\theta_0 = 1$.

5.2 Testing for common conditionally heteroskedastic factors

In this section, we explore the finite sample performance of the K-J statistic under the null of correct model specification and under local alternatives. We also consider the Hansen-Sargan's overidentification test (HS-J test, hereafter) and the GAR test. Example 2 on conditionally heteroskedastic factor models offers a suitable framework for this investigation. We consider a bi-variate vector Y_t of two asset return processes with the representation

$$Y_{t+1} = \Lambda_N F_{t+1} + U_{t+1},$$

where Λ_N is the 2×2 matrix of factor loadings, F_{t+1} is the bivariate vector of conditionally heteroskedastic and mutually independent factors and U_{t+1} , the bivariate vector of idiosyncratic shocks. We let $U_{t+1} \sim i.i.d.N(0, 0.5I_2)$, where I_2 denotes the identity matrix of size 2. The generic component f_{t+1} of F_{t+1} follows a Gaussian-GARCH model,

$$f_{t+1} = \sigma_t \varepsilon_{t+1}, \quad \sigma_t^2 = \omega + \alpha f_t^2 + \beta \sigma_{t-1}^2; \quad \omega, \alpha, \beta > 0 \text{ and } \varepsilon_t \sim i.i.d. N(0, 1).$$

The processes ε_t and U_t are mutually independent and independent of $\{F_\tau, Y_\tau : \tau \leq t\}$. We set $(\omega, \alpha, \beta) = (0.2, 0.2, 0.6)$ and $(0.2, 0.4, 0.4)$, respectively for the first and second component of F_{t+1} . With N being the sample size, we set

$$\Lambda_N = \begin{pmatrix} 1 & 0 \\ 0.5 & \frac{c}{N^{1/8}} \end{pmatrix}; \quad c = 0, 0.2, 0.4, \dots, 10.$$

$c = 0$ corresponds to the null hypothesis of the existence of a common conditionally heteroskedastic factor structure for the components of Y_t that can be tested by either of the three tests under consideration when applied to the moment restriction (14). We use $z_t = (Y_{1,t}^2, Y_{2,t}^2)'$ as vector of instruments in the simulations. The local approximation to the null value is given by $\lambda_N = c/N^{1/8}$; $c \neq 0$. The rate $N^{1/8}$ is chosen such that the resulting moment function under local alternatives is proportional to $N^{-1/2}$, the local approximation of the moment function under which the local alternative distribution of K-J test statistic is derived in Theorem 7.

For global identification of the moment condition model, we follow Dovonon and Renault (2013) and re-parameterize the co-feature vector as $(\theta_0, 1 - \theta_0)$, $\theta_0 \in \mathbb{R}$. Under H_0 in our simulations, $\theta_0 = -1$. The test statistics considered are specifically: $J(\theta_0)$ for the K-J test, the two-step GMM overidentification test statistic for HS-J test and $\min_\theta \text{GAR}(\theta)$ for the GAR test that we denote min-GAR. From Dovonon and Renault (2013), the last two test statistics are asymptotically distributed as a 50-50 mixture of χ_1^2 and χ_2^2 under the null whereas Theorem 4 states that the first one is asymptotically distributed as a χ_1^2 .

Figure 6 shows the simulated rejection rates for the three tests under the null while Figure 7 plots the power curves of these tests for sample sizes $N = 100; 200; 500; 1000; 5000; 10000; 20000$ and 50000. Rejection rates are obtained for 10000 Monte Carlo replications.

It appears from the display in Figure 6 that if the null hypothesis is true then all the three tests have rejection rates closer to nominal ($\alpha = 0.05$) as the sample size increases. The HS-J and min-GAR tests are significantly below the nominal rejection level for small sample sizes but the HS-J test seems to converge to nominal rejection rate faster than the min-GAR. For instance, for $N = 1,000$ and $5,000$, the rejection rate of the HS-J test is 3.9% and 4.88%, respectively whereas that of the min-GAR test is 0.064% and 1.79%, respectively. For N as large as 100,000, the rejection rate of the min-GAR is about 4.0%. The reality is different for the K-J test which has rejection

rates closer to 5% across the sample sizes considered. For $N = 50$ and 100 , this rate is at 6.31 and 6.22%, respectively and falls below 6% from $N = 500$ onwards.

The power curves of these tests displayed by Figure 7 show contrasting performance of the three tests depending on sample sizes. For sample sizes equal or below 200, the power curves of the HS-J and min-GAR tests are flat and even below nominal level (recall that these two tests barely reject the null under H_0 for such sample sizes) whereas the K-J test shows some moderate power. For $N = 500$ and 1000 , the K-J test seems to outperform the other tests which now show some power for large values of c even though the rejection rates do not exceed 50%. From $N = 5000$ the performance ranking is reversed with the HS-test performing slightly better than the min-GAR test, and both having higher rejection rates than the K-J test. For $c = 10$, with $N = 5000$ and 50000 , this latter test has 84.0% and 90.84% rejection rates, respectively while the HS-J test has 98.93% and 99.95%, respectively and the min-GAR 93.6% and 97.43%, respectively.

These results suggest that in small samples, these tests are not reliable and even more so for the HS-J and min-GAR tests compared to the K-J test evaluated at the true value. This may be connected to the local identification pattern of the model under the null. As the sample size increases, all the three tests show evidence of power against local alternatives as expected from our asymptotic theory in Section 4.2 for the K-J test. It is worth mentioning that the powers of the HS-J and min-GAR tests seem to converge to one faster than that of the K-J test.

6 Concluding remarks

In this paper, we explore how to perform inference in moment condition models that only identify the parameters locally to second order. For hypotheses about the parameters, we consider inference based on conventional Wald and LM statistics, and also the identification robust GAR and KLM statistics. For inference about the model specification, we consider the identification-robust K-J statistic and the GAR statistic. In each case, we derive the limiting distribution of statistics under both null and local alternative hypotheses. The Wald statistic is shown to have a non-standard distribution under both null and local alternatives, but the distribution under the null is easily simulated making inference practicable. The LM statistic also has a non-standard distribution under the null in the general case, but has a non-central chi squared distribution under local alternatives. Unlike in the case of strong (first-order) local identification, the Wald and LM statistics have different distributions in the limit. The GAR, KLM and K-J statistics have a chi-squared distribution and non-central chi squared distribution under the null and alternatives respectively. These distributions are exactly the same as those obtained under weak or strong identification, and thus the identification robustness of these tests extends to second-order identified models.

We also explore the finite sample behaviour of the tests in detail in two empirically relevant models with second-order identification: the panel autoregressive (AR) model of order one estimated from a set of non-linear moment conditions, and the conditionally heteroskedastic factor model. In the panel AR model with a unit root, the AR parameter is only identified at second order, and we consider the use of Wald, LM, KLM and GAR statistics to test whether the AR coefficient is one. Our results indicate that the Wald test has the best power properties, being matched by the GAR statistic only in large samples and with both these tests exhibiting greater power than the KLM and LM. In the conditionally heteroskedastic factor model, the moment condition in question only identifies the parameters at second order over the entire parameter space. In this context, the key issue is testing whether the moment condition is valid. In this context, we examine the power properties of the K-J and GAR statistics, and compare them to those of Hansen's (1982)

overidentifying restrictions test (previously analyzed in this setting by Dovonon and Renault, 2013). Here the ranking of the tests is sensitive to the sample size: the K-J test dominates in moderate sized samples, but the overidentifying restrictions test dominates in large samples.

Comparing our theoretical results with the simulations, we find that the analytical local power curves are not always very indicative of the power in finite sample settings. For example, we find that, in our panel data model, the Wald statistic has much better finite sample power than is suggested by its limiting distributions under the local alternative. Similarly, we find that under the local alternative the finite sample distributions of the GAR, KLM and LM statistics only converge very slowly to their limiting distributions. We conjecture this results from the quartic root convergence rate that occurs in second-order locally identified models. Nevertheless, our results show that it is possible to conduct tests with meaningful power in second-order locally identified models.

A Mathematical appendix

Proof of Theorem 1. Consider model (1) with the re-parameterization $\theta = R\eta$, with parameter η :

$$E[f(X, R\eta)] = 0. \quad (42)$$

The true parameter value is clearly $\eta_0 = R^{-1}\theta_0$. Also, so long as the same weighting matrix is used at the first step, the two-step GMM estimators satisfy the relation : $\hat{\eta} = R^{-1}\hat{\theta}$, where for notational brevity we have set $\hat{\theta} = \hat{\theta}_N$. Note that

$$\text{rank} \left(E \left(\frac{\partial}{\partial \eta'} f(x_i, R\eta) \Big|_{\eta=\eta_0} \right) \right) = \text{rank}\{G(\theta_0)R\} = \text{rank}\{G(\theta_0)\} = r.$$

Partitioning η into η_1 and η_2 , its first r and last $p - r$ components, we have:

$$\text{Rank} \left(E \left(\frac{\partial}{\partial \eta_1'} f(x_i, R\eta) \Big|_{\eta=\eta_0} \right) \right) = \text{rank}\{G(\theta_0)R_1\} = r$$

and

$$E \left(\frac{\partial}{\partial \eta_2'} f(x_i, R\eta) \Big|_{\eta=\eta_0} \right) = G(\theta_0)R_2 = 0.$$

Using Assumption 1(b), it is not hard to verify that (42) identifies η_0 at the second order. If $r = p - 1$, we can apply Theorem 1(b) of Dovonon and Hall (2016) and claim that:

$$\sqrt{N} \begin{pmatrix} \hat{\eta}_1 - \eta_{0,1} \\ (\hat{\eta}_2 - \eta_{0,2})^2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathcal{H}\mathbb{Z}_0 + \mathcal{H}B\mathbb{V}/2 \\ \mathbb{V} \end{pmatrix}, \quad (43)$$

with $\mathcal{H} = -(D'V_{ff}(\theta_0)^{-1}D)^{-1}D'V_{ff}(\theta_0)^{-1}$, $\mathbb{V} = -2\frac{\mathbb{Z}\mathbb{I}(\mathbb{Z}<0)}{\tilde{B}'M_d\tilde{B}}$, $\mathbb{Z} = \tilde{B}'M_dV_{ff}(\theta_0)^{-1/2}\mathbb{Z}_0$, and $\mathbb{Z}_0 \sim N(0, V_{ff}(\theta_0))$.

We can write:

$$\begin{aligned} \text{Wald}_N(\theta_0) &= N(\hat{\eta} - \eta_0)'R'\bar{q}_N(\hat{\theta})'\hat{V}_{ff}(\hat{\theta})^{-1}\bar{q}_N(\hat{\theta})R(\hat{\eta} - \eta_0) \\ &= N \left((\hat{\eta}_1 - \eta_{0,1})'R_1'\bar{q}_N(\hat{\theta}) + (\hat{\eta}_2 - \eta_{0,2})'R_2'\bar{q}_N(\hat{\theta}) \right) \hat{V}_{ff}(\hat{\theta})^{-1} \times \\ &\quad \left(\bar{q}_N(\hat{\theta})R_1(\hat{\eta}_1 - \eta_{0,1}) + \bar{q}_N(\hat{\theta})R_2(\hat{\eta}_2 - \eta_{0,2}) \right). \end{aligned} \quad (44)$$

By first-order mean-value expansions, we have:

$$\bar{q}_N(\hat{\theta}) = \bar{q}_N(\theta_0) + \bar{C}_N(\hat{\theta}) (I_p \otimes [R(\hat{\eta} - \eta_0)]), \quad (45)$$

where $\hat{\theta} \in (\hat{\theta}, \theta_0)$ and may differ from row to row and $\bar{C}_N(\theta)$ is the $k \times p^2$ matrix defined by:

$$\bar{C}_N(\theta) = \left(\text{vec} \left(\frac{\partial^2 \bar{f}_{N,1}(\theta)}{\partial \theta \partial \theta'} \right) \quad \text{vec} \left(\frac{\partial^2 \bar{f}_{N,2}(\theta)}{\partial \theta \partial \theta'} \right) \quad \dots \quad \text{vec} \left(\frac{\partial^2 \bar{f}_{N,k}(\theta)}{\partial \theta \partial \theta'} \right) \right)'.$$

Under Assumption 3, $\bar{C}_N(\hat{\theta})$ converges in probability to $C(\theta_0)$ where $C(\theta)$ is defined like $\bar{C}_N(\theta)$ but with sample means replaced by population means. Using (43), the expression of $\bar{q}_N(\hat{\theta})$ in (45) can be written as:

$$\bar{q}_N(\hat{\theta}) = \bar{q}_N(\theta_0) + C(\theta_0) (I_p \otimes R_2) (\hat{\eta} - \eta_0) + o_P(N^{-1/4}).$$

By the law of large number and also noting that $[C(\theta_0) (I_p \otimes R_2)] R_2 = B$, we have:

$$\bar{q}_N(\hat{\theta})R_1 = D + o_P(1), \quad \text{and} \quad \bar{q}_N(\hat{\theta})R_2 = B(\hat{\eta}_2 - \eta_{0,2}) + o_P(N^{-1/4}).$$

Substituting the latter results into (44) and after some simple calculations, we obtain:

$$\begin{aligned} \text{Wald}_N(\theta_0) &= \sqrt{N}(\hat{\eta}_1 - \eta_{0,1})' D' V_{ff}^{-1} D \sqrt{N}(\hat{\eta}_1 - \eta_{0,1}) + B' V_{ff}^{-1} D \sqrt{N}(\hat{\eta}_1 - \eta_{0,1}) \sqrt{N}(\hat{\eta}_2 - \eta_{0,2})^2 \\ &\quad + B' V_{ff}^{-1} D \sqrt{N}(\hat{\eta}_1 - \eta_{0,1}) \sqrt{N}(\hat{\eta}_2 - \eta_{0,2})^2 + B' V_{ff}^{-1} B N(\hat{\eta}_2 - \eta_{0,2})^4 + o_P(1), \end{aligned}$$

where $V_{ff} \equiv V_{ff}(\theta_0)$. From (43), this converges in distribution to

$$\mathbb{W} = (\mathbb{Z}_0 + B\mathbb{V}/2)' \mathcal{H}' D' V_{ff}^{-1} D \mathcal{H} (\mathbb{Z}_0 + B\mathbb{V}/2) + 2B' V_{ff}^{-1} D \mathcal{H} (\mathbb{Z}_0 + B\mathbb{V}/2) \mathbb{V} + B' V_{ff}^{-1} B \mathbb{V}^2.$$

After some simple algebra, we have

$$\mathbb{W} = \mathbb{W}_1 + \mathbb{W}_2,$$

with

$$\mathbb{W}_1 = \left(V_{ff}^{-1/2} \mathbb{Z}_0 - \tilde{B}\mathbb{V}/2 \right)' P \left(V_{ff}^{-1/2} \mathbb{Z}_0 - \tilde{B}\mathbb{V}/2 \right) \quad \text{and} \quad \mathbb{W}_2 = \tilde{B}' M_d \tilde{B} \mathbb{V}^2. \quad (46)$$

It is easily verified that

$$\mathbb{W}_2 = 4S^2 \mathbb{I}(S \leq 0), \quad \text{with} \quad S = \frac{\tilde{B}' M_d V_{ff}^{-1/2} \mathbb{Z}_0}{\sqrt{\tilde{B}' M_d \tilde{B}}} \sim N(0, 1),$$

and

$$V_{ff}^{-1/2} \mathbb{Z}_0 - \tilde{B}\mathbb{V}/2 = V_{ff}^{-1/2} \mathbb{Z}_0 + \alpha S \mathbb{I}(S \leq 0).$$

Thus, we have

$$\mathbb{W}_1 = \left(V_{ff}^{-1/2} \mathbb{Z}_0 + \alpha S \mathbb{I}(S \leq 0) \right)' P \left(V_{ff}^{-1/2} \mathbb{Z}_0 + \alpha S \mathbb{I}(S \leq 0) \right).$$

Since $PV_{ff}^{-1/2} \mathbb{Z}_0$ is independent of $M_d V_{ff}^{-1/2} \mathbb{Z}_0$, it is also independent of S and we can claim that:

$$\mathbb{W}_1 = (S_1 + \alpha S \mathbb{I}(S \leq 0))' P (S_1 + \alpha S \mathbb{I}(S \leq 0)),$$

with $S_1 \sim N(0, I_k)$ independent of S . \square

Proof of Theorem 2. Notice that the value of $LM_N(\theta_*)$ is unchanged by replacing $\bar{q}_N(\theta_*)$ by $\bar{q}_N(\theta_*)A$ with A any nonsingular matrix. In particular, this statistic stays the same when this quantity is replaced by $\bar{q}_N(\theta_*) \left(R_1 \dot{\vdash} \sqrt{N} R_2 \right)$. Note also that, by Assumption 4, we have:

$$\bar{q}_N(\theta_*) \left(R_1 \dot{\vdash} \sqrt{N} R_2 \right) = \left(\bar{q}_N(\theta_*) R_1 \dot{\vdash} \sqrt{N} \bar{q}_N(\theta_*) R_2 \right) \xrightarrow{d} \tilde{\psi}_q \equiv \left(D \dot{\vdash} \psi_q \right),$$

where D is constant and ψ_q is a Gaussian matrix defined in Assumption 4. The result then follows directly. \square

Proof of Theorem 3. (i) Similarly to the LM test statistic, $KLM(\theta_*)$ in (21) stays unchanged if $\hat{D}_N(\theta_0)$ is replaced by

$$\hat{D}_N(\theta_*) \left(R_1 \dot{\vdash} \sqrt{N} R_2 \right) = \left(\hat{D}_N(\theta_*) R_1 \dot{\vdash} \sqrt{N} \hat{D}_N(\theta_*) R_2 \right).$$

From Assumption 4, we have:

$$\hat{D}_N(\theta_*) R_1 \xrightarrow{P} D, \quad \text{and} \quad \sqrt{N} \hat{D}_N(\theta_*) R_2 \xrightarrow{d} \varepsilon_q.$$

Since (ψ_q, ψ_f) is Gaussian, ε_q is independent of ψ_f . Under the non-singularity assumption for $\bar{\psi}_q' \bar{\psi}_q$, $\hat{V}_{ff}(\theta_*)^{-1/2} \hat{D}_N(\theta_*) \left(\hat{D}_N(\theta_*)' \hat{V}_{ff}(\theta_*)^{-1} \hat{D}_N(\theta_*) \right)^{-1} \hat{D}_N(\theta_*)' \hat{V}_{ff}(\theta_*)^{-1/2}$ is well-defined in large samples and the continuous mapping theorem ensures that $KLM(\theta_*)$ converges in distribution to

$$\psi_f' V_{ff}(\theta_*)^{-1/2} \left(I_k - M_{V_{ff}(\theta_*)^{-1/2} \bar{\psi}_q} \right) V_{ff}(\theta_*)^{-1/2} \psi_f.$$

Conditionally on $\bar{\psi}_q$, this limit follows χ_p^2 distribution and the independence of $\bar{\psi}_q$ and ψ_f implies that this limit is unconditionally distributed as χ_p^2 . (ii) The result for the GAR statistic is immediate under the stated conditions. \square

Proof of Theorem 4. (i) Similarly to the proof of Theorem 3, we can claim that $J(\theta_0)$ converges in distribution to

$$\psi_f' V_{ff}(\theta_0)^{-1/2} M_{V_{ff}(\theta_0)^{-1/2} \bar{\psi}_q} V_{ff}(\theta_0)^{-1/2} \psi_f.$$

Conditionally on $\bar{\psi}_q$, this limit follows χ_{k-p}^2 distribution and the independence of $\bar{\psi}_q$ and ψ_f implies that this limit is unconditionally distributed as χ_{k-p}^2 . (ii) See the proof of Theorem 3(ii). \square

Derivation of equation (34). If $\theta_N = 1 - \frac{c}{2\sqrt[4]{N}}$ then it can be shown that

$$E[a] = \sigma^2 \binom{1}{0} \quad E[b] = -\sigma^2 \left(2 - \frac{c}{\sqrt[4]{N}} + \frac{c^2}{4\sqrt[4]{N}} \right) \quad E[d] = \sigma^2 \left(1 - \frac{c}{\sqrt[4]{N}} + \frac{c^2}{2\sqrt[4]{N}} - \frac{c^3}{8N^{3/4}} \right), \quad (47)$$

where a , b and d are defined in Section 2.1, and so

$$m_N(\theta_N) = \binom{0}{0}, \quad G_N(\theta_N) = -\frac{c^2}{4\sqrt[4]{N}} \sigma^2 \binom{1}{0}, \quad H_N(\theta_N) = 2\sigma^2 \binom{1}{0} \quad (48)$$

It is also instructive to explore the population moment, Jacobian and Hessian evaluated at θ_0 under P_N . Using similar arguments, it can be shown that

$$m_N(\theta_0) = \left(\frac{c^2}{4\sqrt[4]{N}} - \frac{c^3}{8N^{3/4}} \right) \sigma^2 \binom{1}{0}, \quad G_N(\theta_0) = \left(\frac{c}{\sqrt[4]{N}} - \frac{c^2}{4\sqrt[4]{N}} \right) \sigma^2 \binom{1}{0}, \quad H_N(\theta_0) = 2\sigma^2 \binom{1}{0}. \quad (49)$$

Therefore, under under this sequence of local alternatives, the rate of decrease of $E_N[f(\theta_0)]$ is proportional to the random component in the sample moment. If we set the rate differently say at $\theta_N = 1 - \frac{c}{2\sqrt[4]{N}}$, the expected values of a , b and d equal

$$E[a] = \sigma^2 \binom{1}{0} \quad E[b] = -\sigma^2 \left(2 - \frac{c}{\sqrt[4]{N}} + \frac{c^2}{4\sqrt[4]{N}} \right) \quad E[d] = \sigma^2 \left(1 - \frac{c}{\sqrt[4]{N}} + \frac{c^2}{2\sqrt[4]{N}} - \frac{c^3}{8N^{3/4}} \right), \quad (50)$$

so

$$E_N[f(\theta_0)] = \left(\frac{c^2}{4N} - \frac{c^3}{8N^{3/2}} \right) \sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (51)$$

which shows that the rate is too fast as it sits below the rate of the random component of the sample moment.

Proof of Theorem 5. Similarly to the proof of Theorem 1, we consider the re-parameterization $R\eta = \theta$. Let $\eta_0 = R^{-1}\theta_0$ and $\eta_N = R^{-1}\theta_N$. We have: $E_N[f(X_N, R\eta_N)] = 0$ and, from Assumption 7, and assuming that we can interchange E_N and derivatives freely, we have

$$\frac{\partial}{\partial \eta_1'} E_N[f(X_N, R\eta_N)] = G_N(\theta_N)R_1 = D + o(1) \quad \text{and} \quad \frac{\partial}{\partial \eta_2} E_N[f(X_N, R\eta_N)] = G_N(\theta_N)R_2 = O(N^{-\xi}).$$

The fact that the Jacobian in the direction of η_2 is $O(N^{-\xi})$ and not exactly 0 make the current configuration slightly different than the assumptions of Theorem 1 of Dovonon and Hall (2016). However, the fact that $\xi > 1/4$ allows the conclusions of the parts (a) and (b) of that theorem to stand with ϕ_0 replaced by η_N as we now show. Let $\hat{\eta}$ be the GMM estimator of η_N . First, we observe using Assumption 8 (with $\mu_N = 0$) that, under P_N , $\bar{f}_N(R\eta_N) = O_P(N^{-1/2})$ and

$$\left. \frac{\partial \bar{f}_N(R\eta)}{\partial \eta_2} \right|_{\eta=\eta_N} = \bar{q}_N(\theta_N)R_2 = (\bar{q}_N(\theta_N) - N^{-\xi}A)R_2 + N^{-\xi}AR_2 = O_P(N^{-1/2}) + O(N^{-\xi}).$$

Via similar expansions as those in the proof of Theorem 1(a) of Dovonon and Hall (2016) and leading to their equations (34) and (35), we have:

$$\begin{aligned} \bar{f}_N(\hat{\eta}) &= \bar{f}_N(\eta_N) + D(D'V_{ff}^{-1}D)^{-1}V_{ff}^{-1}(\bar{f}_N(\hat{\eta}) - \bar{f}_N(\eta_N)) + \frac{1}{2}V_{ff}^{1/2}M_d\tilde{B}(\eta_2 - \eta_{N,2})^2 \\ &\quad + (\eta_2 - \eta_{N,2})^2 o_P(1) + |\eta_2 - \eta_{N,2}| O_P(N^{-\xi}) + O_P(N^{-1/2}), \end{aligned}$$

where by an abuse of notation, we set $\bar{f}_N(\eta) \doteq \bar{f}_N(R\eta)$ and $V_{ff} \doteq V_{ff}(\theta_0)$ and use the fact that $\bar{f}_N(\hat{\eta}) = O_P(N^{-1/2})$ under P_N . This latter follows from the fact that the GMM norm of $\bar{f}_N(\hat{\eta})$ is smaller or equal to that of $\bar{f}_N(\eta_N)$ by definition. Hence,

$$\begin{aligned} &\bar{f}_N(\hat{\eta})' \hat{V}_{ff}(\hat{\theta}_{1,s})^{-1} \bar{f}_N(\hat{\eta}) \\ &= \bar{f}_N(\eta_N)' \hat{V}_{ff}(\hat{\theta}_{1,s})^{-1} \bar{f}_N(\eta_N) + \frac{1}{4} \tilde{B}' M_d \tilde{B} (\hat{\eta}_2 - \eta_{N,2})^4 + (\hat{\eta}_2 - \eta_{N,2})^4 o_P(1) + O_P(N^{-1}) \\ &\quad + (\hat{\eta}_2 - \eta_{N,2})^2 \left(O_P(N^{-\frac{1}{2}}) + O_P(N^{-2\xi}) \right) + |\hat{\eta}_2 - \eta_{N,2}| O_P(N^{-\frac{1}{2}-\xi}) + |\hat{\eta}_2 - \eta_{N,2}|^3 O_P(N^{-\xi}). \end{aligned}$$

By the definition of the GMM estimator, this quantity is less or equal to $\bar{f}_N(\eta_N)' \hat{V}_{ff}(\hat{\theta}_{1,s})^{-1} \bar{f}_N(\eta_N)$. Let $z_N = N^{1/4}|\hat{\eta}_2 - \eta_{N,2}|$ and $\gamma = \frac{1}{4} \tilde{B}' M_d \tilde{B}$. After multiplying each side of the previous equation by N and since $\xi > 1/4$, we can claim that:

$$\gamma z_N^4 + z_N^4 o_P(1) \leq O_P(1) + z_N^2 O_P(1) + z_N o_P(1) + z_N^3 o_P(1).$$

Since $\gamma > 0$, this shows that $N^{1/4}(\hat{\eta}_2 - \eta_{N,2}) = O_P(1)$ under P_N and using the analogue of Equation (35) of Dovonon and Hall (2016), we obtain $\sqrt{N}(\hat{\eta}_1 - \eta_{N,1}) = O_P(1)$. Using these rates

of convergence, the steps of the proof of Theorem 1(b) of Dovonon and Hall (2016) follow readily (only taking Taylor expansions around η_N) and we obtain:

$$\sqrt{N} \begin{pmatrix} \hat{\eta}_1 - \eta_{N,1} \\ (\hat{\eta}_2 - \eta_{N,2})^2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbb{X}_1 \\ \mathbb{V} \end{pmatrix} \equiv \begin{pmatrix} \mathcal{H}\mathbb{Z}_0 + \mathcal{H}B\mathbb{V}/2 \\ \mathbb{V} \end{pmatrix}, \quad (52)$$

under P_N with asymptotic distribution as given by (43).

Note that, since $(\sqrt{N}(\hat{\eta}_1 - \eta_{N,1}), N^{1/4}(\hat{\eta}_2 - \eta_{N,2})) = O_P(1)$ under P_N , by the Prokhorov theorem, any of its subsequence has a further subsequence, indexed say by $s(N)$, that converges in distribution under P_N to say, $(\mathbb{X}_1(s), \mathbb{X}(s))$ which, by (52), are such that $\mathbb{X}_1(s) = \mathbb{X}_1$ and $\mathbb{X}(s)^2 = \mathbb{V}$ for any converging subsequence $s(N)$.

Similar derivations as those in Theorem 1 yield:

$$\text{Wald}_N(\theta_0) = N(\hat{\eta} - \eta_0)' R' \bar{q}_N(\hat{\theta})' \hat{V}_{ff}(\hat{\theta})^{-1} \bar{q}_N(\hat{\theta}) R(\hat{\eta} - \eta_0)$$

and

$$\bar{q}_N(\hat{\theta}) R_1 = D + o_P(1), \quad \text{and} \quad \bar{q}_N(\hat{\theta}) R_2 = B(\hat{\eta}_2 - \eta_{N,2}) + o_P(N^{-1/4}).$$

It follows that

$$\text{Wald}_N(\theta_0) = W_{N,a} + W_{N,b} + W_{N,c} + o_P(1),$$

with

$$\begin{aligned} W_{N,a} &= \sqrt{N}(\hat{\eta}_1 - \eta_{0,1})' D' V_{ff}^{-1} D \sqrt{N}(\hat{\eta}_1 - \eta_{0,1}) \\ W_{N,b} &= 2\sqrt{N}(\hat{\eta}_1 - \eta_{0,1})' D' V_{ff}^{-1} B N^{1/4}(\hat{\eta}_2 - \eta_{0,2}) N^{1/4}(\hat{\eta}_2 - \eta_{N,2}) \end{aligned}$$

and

$$W_{N,c} = B' V_{ff}^{-1} B \sqrt{N}(\hat{\eta}_2 - \eta_{0,2})^2 \sqrt{N}(\hat{\eta}_2 - \eta_{N,1})^2.$$

We have:

$$\begin{aligned} W_{N,a} &= N \left((\hat{\eta}_1 - \eta_{N,1})' \tilde{D}' \tilde{D}(\hat{\eta}_1 - \eta_{N,1}) + 2(\hat{\eta}_1 - \eta_{N,1})' \tilde{D}' \tilde{D}(\eta_{N,1} - \eta_{0,1}) \right. \\ &\quad \left. + (\eta_{N,1} - \eta_{0,1})' \tilde{D}' \tilde{D}(\eta_{N,1} - \eta_{0,1}) \right) \\ &\xrightarrow{d} (a) \equiv \mathbb{X}'_1 \tilde{D}' \tilde{D} \mathbb{X}_1 + 2\mathbb{X}'_1 \tilde{D}' \tilde{D} e_1 + e'_1 \tilde{D}' \tilde{D} e_1. \end{aligned}$$

Similar calculation show that:

$$W_{N,b} \xrightarrow{d} (b) \equiv 2\mathbb{X}'_1 \tilde{D}' \tilde{B} \mathbb{X}(s)^2 + 2\mathbb{X}'_1 \tilde{D}' \tilde{B} \mathbb{X}(s) e_2 + 2e'_1 \tilde{D}' \tilde{G} \mathbb{X}(s)^2 + 2e'_1 \tilde{D}' \tilde{B} \mathbb{X}(s) e_2$$

and

$$W_{N,c} \xrightarrow{d} (c) \equiv \tilde{B}' \tilde{B} \mathbb{X}(s)^2 (\mathbb{X}(s)^2 + 2e_2 \mathbb{X}(s) + e_2^2).$$

The convergence of $W_{N,a}$, $W_{N,b}$ and $W_{N,c}$ holds jointly and as a result, $W_N(\theta_0)$ converges in distribution to $(a) + (b) + (c)$ under P_N . Note that simple expansions yield:

$$\begin{aligned} (a) &= \left(V_{ff}^{-1/2} \mathbb{Z}_0 + \tilde{B} \mathbb{V} / 2 \right)' P \left(V_{ff}^{-1/2} \mathbb{Z}_0 + \tilde{B} \mathbb{V} / 2 \right) - 2 \left(V_{ff}^{-1/2} \mathbb{Z}_0 + \tilde{B} \mathbb{V} / 2 \right)' P \tilde{D} e_1 + e'_1 \tilde{D}' \tilde{D} e_1 \\ (b) &= -2 \left(V_{ff}^{-1/2} \mathbb{Z}_0 + \tilde{B} \mathbb{V} / 2 \right)' P \tilde{B} \mathbb{X}(s)^2 - 2 \left(V_{ff}^{-1/2} \mathbb{Z}_0 + \tilde{B} \mathbb{V} / 2 \right)' P \tilde{B} \mathbb{X}(s) e_2 \\ &\quad + 2e'_1 \tilde{D}' \tilde{B} \mathbb{X}(s)^2 + 2e'_1 \tilde{D}' \tilde{B} \mathbb{X}(s) e_2 \\ (c) &= \tilde{B}' \tilde{B} \mathbb{X}(s)^2 (\mathbb{X}(s)^2 + 2e_2 \mathbb{X}(s) + e_2^2). \end{aligned}$$

To obtain the form of the asymptotic distribution given in the theorem, write this limit as $\pi_A + \pi_B + \pi_C + \pi_F$ with:

$$\begin{aligned}\pi_A &= \left(V_{ff}^{-1/2}\mathbb{Z}_0 + \tilde{B}\mathbb{V}/2\right)' P \left(V_{ff}^{-1/2}\mathbb{Z}_0 + \tilde{B}\mathbb{V}/2\right) - 2 \left(V_{ff}^{-1/2}\mathbb{Z}_0 + \tilde{B}\mathbb{V}/2\right)' P \tilde{B}\mathbb{X}(s)^2 + \tilde{B}' \tilde{B}\mathbb{X}(s)^4 \\ \pi_B &= -2 \left(V_{ff}^{-1/2}\mathbb{Z}_0 + \tilde{B}\mathbb{V}/2\right)' P \left(\tilde{D}e_1 + \tilde{B}\mathbb{X}(s)e_2\right) + 2e_1' \tilde{D}' \tilde{B} (\mathbb{X}(s)^2 + \mathbb{X}(s)e_2) \\ \pi_C &= \tilde{B}' \tilde{B}\mathbb{X}(s)^2 (2e_2\mathbb{X}(s) + e_2^2) \\ \pi_F &= e_1' \tilde{D}' \tilde{D}e_1.\end{aligned}$$

Note that $\mathbb{V} = -\frac{2}{a}(S\mathbb{I}(S \leq 0))$, with $S = \tilde{G}'M_dV_{ff}^{-1/2}\mathbb{Z}_0/a$, with $a = \sqrt{\tilde{B}'M_d\tilde{B}}$ and $\mathbb{Z}_0 \sim N(0, V_{ff})$.

Some simple calculations yields:

$$\begin{aligned}\pi_A &= \left(V_{ff}^{-1/2}\mathbb{Z}_0 + \alpha S\mathbb{I}(S \leq 0)\right)' P \left(V_{ff}^{-1/2}\mathbb{Z}_0 + \alpha S\mathbb{I}(S \leq 0)\right) + 4S^2\mathbb{I}(S \leq 0). \\ \pi_B &= -2 \left(V_{ff}^{-1/2}\mathbb{Z}_0 - \alpha S\mathbb{I}(S \leq 0)\right)' P \left(\tilde{D}e_1 + \alpha a\mathbb{X}(s)e_2\right) \\ &\quad + 2e_1' \tilde{D}' \alpha (-2S\mathbb{I}(S \leq 0) + a\mathbb{X}(s)e_2) \\ \pi_C &= -2ae_2\alpha'(2\mathbb{X}(s) + e_2)S\mathbb{I}(S \leq 0).\end{aligned}$$

Since $PV_{ff}^{-1/2}\mathbb{Z}_0$ is independent of S , we can replace $V_{ff}^{-1/2}\mathbb{Z}_0$ in π_A and π_B by $S_1 \sim N(0, I_k)$ independent of S which gives the stated result. \square

Proof of Theorem 6. As noted in the proof of Theorem 2, the value of $LM(\theta_*)$ is unchanged by replacing $\bar{q}_N(\theta_*)$ by $\bar{q}_N(\theta_*)A$ with A any nonsingular matrix. Here, we replace $\bar{q}_N(\theta_*)$ by $\bar{q}_N(\theta_*) (R_1 \ N^{1/4}R_2)$.

A first-order mean value expansion of $\bar{q}_N(\theta_*)$ around θ_N similar to (45) gives:

$$\bar{q}_N(\theta_*) = \bar{q}_N(\theta_N) + \bar{C}_N(\dot{\theta}) [I_p \otimes (\theta_* - \theta_N)] = \bar{q}_N(\theta_N) - \bar{C}_N(\dot{\theta}) [I_p \otimes (R_1e_{N,1} + R_2e_{N,2})],$$

where $\dot{\theta} \in (\theta_*, \theta_N)$ and may differ by entry of $\bar{q}_N(\theta_*)$ and with $\bar{C}_N(\theta)$ defined as in (45). Under Assumption 6, $\bar{C}_N(\dot{\theta})$ converges in probability P_N to $C(\theta_*)$ and thanks to Assumptions 6 and 7, we have:

$$\bar{q}_N(\theta_*)R_1 = D + o_P(1).$$

Also,

$$\bar{q}_N(\theta_N)R_2 = (\bar{q}_N(\theta_N)R_2 - N^{-\xi}A) + N^{-\xi}A = O_P(N^{-1/2}) + O(N^{-\xi}) = o_P(N^{-1/4}),$$

where the stochastic orders are with respect to P_N . As a result, we also have, with respect to P_N ,

$$N^{1/4}\bar{q}_N(\theta_*)R_2 = -C(\theta_*)[I_p \otimes (R_2e_2)]R_2 + o_P(1).$$

Thus

$$\bar{q}_N(\theta_*) \begin{pmatrix} R_1 \\ N^{1/4}R_2 \end{pmatrix} = \begin{pmatrix} D \\ -C(\theta_*)[I_p \otimes (R_2e_2)]R_2 \end{pmatrix} + o_P(1) = Q(e_2) + o_P(1).$$

By a second-order mean-value expansion of $\bar{f}_N(\theta_*)$ around θ_N , we have:

$$\begin{aligned}
\bar{f}_N(\theta_*) &= \bar{f}_N(\theta_N) - \bar{q}_N(\theta_N)(\theta_N - \theta_*) + \frac{1}{2} [(\theta_N - \theta_*)' \otimes I_k] \bar{h}_N(\bar{\theta})(\theta_N - \theta_*), \\
&= \bar{f}_N(\theta_N) - \bar{q}_N(\theta_N)R_1e_{N,1} - \bar{q}_N(\theta_N)R_2e_{N,2} \\
&\quad + \frac{1}{2} [(R_1e_{N,1} + R_2e_{N,2})' \otimes I_k] \bar{h}_N(\bar{\theta})(R_1e_{N,1} + R_2e_{N,2}) \\
&= \bar{f}_N(\theta_N) - N^{-1/2}De_1 + \frac{1}{2}N^{-1/2} ((R_2e_2)' \otimes I_k) H(\theta_*)(R_2e_2) + o_P(N^{-1/2}),
\end{aligned}$$

where $\bar{\theta} \in (\theta_*, \theta_N)$ and may differ by equation. We use in this expansion the fact that $\bar{H}_N(\bar{\theta})$ converges in probability P_N to $H(\theta_*)$ and the fact that $\bar{q}_N(\theta_N)R_2 = o_P(N^{-1/4})$ under P_N . Thus $\sqrt{N}\bar{f}_N(\theta_*)$ converges in distribution under P_N to $N(\mu_\theta, V_{ff}(\theta_*))$ with

$$\mu_\theta = -De_1 + \frac{1}{2} ((R_2e_2)' \otimes I_k) H(\theta_*)(R_2e_2).$$

Thanks to the identity: $(e' \otimes I_k)H(\theta_*)e = C(\theta_*)(I_p \otimes e)e$, for all $e \in \mathbb{R}^p$, we have $\mu_\theta = -Q(e_2) \times \begin{pmatrix} e_1 \\ \frac{1}{2}e_2 \end{pmatrix}$ which then belongs to the range of $Q(e_2)$. Note also that thanks to the second-order identification condition in Assumption 1(b), $(e_1, e_2) \neq 0$ implies that $\mu_\theta \neq 0$.

To prove (a), letting $V_{ff} = V_{ff}(\theta_*)$, note that

$$LM(\theta_*) = \sqrt{N}(V_{ff}^{-1/2}\bar{f}_N(\theta_*))'V_{ff}^{-1/2}Q(e_2) \left(Q(e_2)'V_{ff}^{-1}Q(e_2) \right)^{-1} Q(e_2)'V_{ff}^{-1/2}\sqrt{N}(V_{ff}^{-1/2}\bar{f}_N(\theta_*)) + o_P(1)$$

converges in distribution under P_N to $\chi_p^2(\lambda_\theta)$, with $\lambda_\theta = \mu_\theta'V_{ff}^{-1}\mu_\theta$. Since $Q(e_2)$ is nonsingular, $e_2 \neq 0$ and as a result, $\lambda_\theta > 0$.

Regarding $KLM(\theta_*)$, since $\bar{f}_N(\theta_*) = O_P(N^{-1/2})$ under P_N , $\bar{q}_N(\theta_*)$ is the leading term of $\hat{D}_N(\theta_*)$. Thus, $KLM(\theta_*) = LM(\theta_*) + o_P(1)$ under P_N and this concludes (a).

(b) From the asymptotic distribution of $\sqrt{N}\bar{f}_N(\theta_*)$ derived above, it is obvious that $\text{GAR}(\theta_0)$ converges in distribution under P_N to $\chi_k^2(\lambda_\theta)$. As mentioned above, $\lambda_\theta > 0$ if $(e_1, e_2) \neq 0$. \square

Derivation of equation (38). We first derive equation (36). Using $\theta'_0\lambda_1 = 0$, we have under P_N that:

$$E_N [\theta'_0 Y_{t+1} Y'_{t+1} \theta_0 | \mathfrak{F}_t] = (\theta'_0 \lambda_{2N})^2 \sigma_{2,t}^2 + \theta'_0 \Omega \theta_0. \quad (53)$$

As in Section 2.2, let z_t be a relevant vector of instruments, then it follows from (53) that

$$E_N [z_t (\theta'_0 Y_{t+1})^2] = (\theta'_0 \lambda_{2N})^2 E_N [z_t \sigma_{2,t}^2] + \theta'_0 \Omega \theta_0 E_N [z_t]; \quad \theta'_0 \Omega \theta_0 = E_N [(\theta'_0 Y_{t+1})^2] - (\theta'_0 \lambda_{2N})^2 E_N [\sigma_{2,t}^2]$$

which, together, imply

$$Cov(z_t, (\theta'_0 Y_{t+1})^2) = (\theta'_0 \lambda_{2N})^2 Cov(z_t, \sigma_{2,t}^2), \quad (54)$$

where $Cov[\cdot, \cdot]$ is relative to P_N . Using (15) and (54), we obtain (36). To evaluate $Cov(z_t, \sigma_{2,t}^2)$, it is useful to consider a factor representation of the returns that is in line with (12)-(13):

$$Y_{t+1} = \Lambda_N F_{t+1} + U_{t+1},$$

where $\text{Var}(F_{t+1}|\mathfrak{F}_t) = D_t$, $\text{Cov}(F_{t+1}, U_{t+1}|\mathfrak{F}_t) = 0$, $E(F_{t+1}|\mathfrak{F}_t) = 0$, $E(U_{t+1}|\mathfrak{F}_t) = 0$. We set $\Lambda_N = (\lambda_1, \lambda_{2,N})$. Following Doz and Renault (2006), we further assume that $(F_{1,t}^2, U_{1,t}^2, U_{2,t}^2, F_{1,t}U_{1,t}, F_{1,t}U_{2,t})$ is uncorrelated with $\sigma_{2,t}^2$ and $(U_t', F_{1,t})$ is uncorrelated with $\sigma_{2,t}^2 F_{2,t}$. After some simple expansions, we have:

$$\text{Cov}[\sigma_{2,t}^2, Y_{j,t}^2] = \lambda_{2,N,j}^2 \text{Cov}[\sigma_{2,t}^2, F_{2,t}^2] = \lambda_{2,N,j}^2 \text{Cov}[F_{2,t+1}^2, F_{2,t}^2],$$

and so, using $z_t = (Y_{1,t}^2, Y_{2,t}^2)'$, obtain (37). Combining these results with (16), we obtain (38). \square

Proof of Theorem 7. (i) Note that

$$\hat{D}_N(\theta_0) = \bar{q}_N(\theta_0) - \left[\widehat{\text{Cov}}(q_{i,lm}(\theta_0), f_i(\theta_0)) \hat{V}_{ff}(\theta_0)^{-1} \bar{f}_N(\theta_0) \right]_{1 \leq l \leq k, 1 \leq m \leq p},$$

with $\widehat{\text{Cov}}(q_{i,lm}(\theta_0), f_i(\theta_0)) = \frac{1}{N} \sum_{i=1}^N q_{i,lm}(\theta_0) f_i(\theta_0)' - \bar{q}_{N,lm} \bar{f}_N(\theta_0)'$.

But, $\bar{f}_N(\theta_0) = \left(\bar{f}_N(\theta_0) - \frac{c}{\sqrt{N}} \right) + \frac{c}{\sqrt{N}} = O_P(N^{-1/2})$ under P_N . In fact,

$$\sqrt{N} \bar{f}_N(\theta_0) \xrightarrow{d} \psi_f + c \sim N(c, V_{ff}(\theta_0)) \quad (55)$$

under P_N . Thus, $\hat{D}_N(\theta_0)R_1 = \bar{q}_N(\theta_0)R_1 + o_P(1) = D + o_P(1)$, under P_N . Also,

$$\hat{D}_N(\theta_0)R_2 = \left(\bar{q}_N(\theta_0)R_2 - \frac{A}{N\xi} \right) + \frac{A}{N\xi} - \left[\widehat{\text{Cov}}(q_{i,lm}(\theta_0), f_i(\theta_0)) \hat{V}_{ff}(\theta_0)^{-1} \bar{f}_N(\theta_0) \right]_{1 \leq l \leq k, 1 \leq m \leq p} R_2.$$

Letting $\delta = \frac{1}{2}\mathbb{I}(\xi \geq \frac{1}{2}) + \xi\mathbb{I}(0 < \xi < \frac{1}{2})$, it is not hard to see that

$$N^\delta \hat{D}_N(\theta_0)R_2 \xrightarrow{d} \varepsilon_q^a$$

under P_N . The statistic $J(\theta_0)$ is unchanged if $\hat{D}_N(\theta_0)$ is replaced by $\hat{D}_N(\theta_0) \left(R_1 \vdash N^\delta R_2 \right)$ which converges in distribution to $\bar{\psi}_q^a$ under P_N . Under the full column rank assumption, $M_{V_{ff}(\theta_0)^{-1/2} \bar{\psi}_q^a}$ is well-defined and the continuous mapping theorem ensures that

$$J(\theta_0) \xrightarrow{d} (\psi_f + c)' V_{ff}(\theta_0)^{-1/2} M_{V_{ff}(\theta_0)^{-1/2} \bar{\psi}_q^a} V_{ff}(\theta_0)^{-1/2} (\psi_f + c)$$

under P_N . From the independence of ψ_f and $\bar{\psi}_q^a$, we can claim that

$$J(\theta_0) \xrightarrow{d} \chi_{k-p}^2(\lambda_m)$$

under P_N with random non-centrality parameter $\lambda_m = c' V_{ff}(\theta_0)^{-1/2} M_{V_{ff}(\theta_0)^{-1/2} \bar{\psi}_q^a} V_{ff}(\theta_0)^{-1/2} c$. Clearly, if $c \notin \langle \bar{\psi}_q^a \rangle$ almost surely, then $\lambda_m > 0$ almost surely. (ii) Follows readily from (55). \square

Derivation of equations (30), (40) and (41) We consider the behaviour of the Wald test under P_N with $\theta_N = 1 - \frac{c}{2\sqrt{N}}$. Assume that $N^{1/2}(a - E_N[a], b - E_N[b], c - E_N[c])' \xrightarrow{d} (\psi_a, \psi_b, \psi_d)'$ under P_N where $(\psi_a, \psi_b, \psi_d)'$ have a normal distribution with mean zero. Define $\psi = \psi_a + \psi_b + \psi_d$ and let ψ_i denote the i^{th} element of ψ . For brevity but with an abuse of notation, let $V_{ff}(\theta_0) = V$.

For ease of notation, let $\hat{V} = \hat{V}_{ff}(\hat{\theta}_{1,s})$ and $V = V_{ff}(\theta_0)$. The two-step GMM estimator is defined as:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} N \times \bar{f}_N(\theta)' \hat{V}^{-1} \bar{f}_N(\theta),$$

and the associated first order conditions are

$$N \bar{q}_N(\hat{\theta})' \hat{V}^{-1} \bar{f}_N(\hat{\theta}) = 0,$$

where

$$\begin{aligned} \bar{f}_N(\theta) &= a\theta^2 + b\theta + d = a + b + d + a(\theta - 1)^2 + (b + 2a)(\theta - 1), \\ \bar{q}_N(\theta) &= 2a\theta + b = 2a(\theta - 1) + b + 2a. \end{aligned}$$

Note that under P_N , we have:

$$\sqrt{N} \left((b + 2a) - \frac{1}{\sqrt[4]{N}} \sigma^2 \begin{pmatrix} c - \frac{c^2}{4\sqrt[4]{N}} \\ 0 \end{pmatrix} \right) \xrightarrow{d} 2\psi_a + \psi_b \quad (56)$$

$$\sqrt{N} \bar{f}_N(1) = \sqrt{N}(a + b + d) \xrightarrow{d} \sigma^2 \begin{pmatrix} \frac{c^2}{4} \\ 0 \end{pmatrix} + \psi_a + \psi_b + \psi_d. \quad (57)$$

Taking a mean value expansion of $\bar{q}_N(\hat{\theta})$ around $\bar{q}_N(1)$ (and recalling that $\theta_0 = 1$)

$$\bar{q}_N(\hat{\theta}) = \bar{q}_N(1) + H_N(\bar{\theta})(\hat{\theta} - 1),$$

with $\bar{\theta}$ an intermediate value between 1 and $\hat{\theta}$. From (49), it follows that under P_N , we have:

$$H_N(\bar{\theta}) \xrightarrow{p} 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

and so

$$\sqrt[4]{N} \bar{q}_N(1) \xrightarrow{p} \sigma^2 \begin{pmatrix} c \\ 0 \end{pmatrix}.$$

Therefore, since $\hat{\theta} = O_p(N^{-1/4})$, we have

$$\bar{q}_N(\hat{\theta}) = \begin{pmatrix} O_p(N^{-1/4}) \\ O_p(N^{-1/2}) \end{pmatrix},$$

and so¹⁶

$$\sqrt[4]{N} \bar{q}_N(\hat{\theta}) \stackrel{a}{=} \sigma^2 \begin{pmatrix} c \\ 0 \end{pmatrix} + 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta, \quad (58)$$

where, as in the text, $\sqrt[4]{N}(\hat{\theta}_{1,s} - 1) = \zeta + o_p(1)$. Similarly, using (57), we have

$$\sqrt{N} \bar{f}_N(\hat{\theta}) \stackrel{a}{=} \psi + \frac{c^2}{4} \sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta^2 + \sigma^2 \begin{pmatrix} c \\ 0 \end{pmatrix} \zeta. \quad (59)$$

¹⁶Using $\stackrel{a}{=}$ to denote equality up to $o_p(1)$.

Combining (58)-(59) with the first order conditions and $\hat{V} \xrightarrow{p} V$, it can be seen that ζ is implicitly characterized by:

$$\left[\sigma^2 \begin{pmatrix} c \\ 0 \end{pmatrix} + 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta \right]' V^{-1} \left[\psi + \begin{pmatrix} \sigma^2 c^2/4 \\ 0 \end{pmatrix} + \begin{pmatrix} \sigma^2 \\ 0 \end{pmatrix} \zeta^2 + \sigma^2 \begin{pmatrix} c \\ 0 \end{pmatrix} \zeta \right] = 0,$$

which can be re-written as

$$2\sigma^4 e \zeta^3 + 3\sigma^4 c e \zeta^2 + \sigma^2 \left[2V_1^{-1} \psi + c^2 e \sigma^2 \frac{3}{2} \right] \zeta + c \sigma^2 [V_1^{-1} \psi + \frac{1}{4} e \sigma^2 c^2] = 0, \quad (60)$$

where e is the (1, 1) element of V^{-1} and V_1^{-1} is the first row of V^{-1} . Equation (60) implies:

$$2\sigma^4 e \left(\zeta + \frac{c}{2} \right) \left(\left\{ \zeta + \frac{c}{2} \right\}^2 + \frac{1}{e \sigma^2} V_1^{-1} \psi \right) = 0, \quad (61)$$

Using (61) and noticing that ζ is a real-valued root of the above third-order polynomial, we obtain a twofold solution for $\zeta + c/2$. The first solution occurs when the quadratic polynomial contained in the second set of parentheses in (61) only has complex roots: in this case the solution is obtained from the term in the first set of parentheses in (61). The second solution occurs when the quadratic polynomial in the second set of parentheses in (61) has real roots. In the latter case, the two roots imply different values for ζ^2 - unless $c = 0$ - and so for $c \neq 0$ we choose the root that maximizes ζ^2 and therefore leads to the largest asymptotic power. The solution to (61) just described is:

$$\zeta_* = -\frac{c}{2} - \mathbb{I}(V_1^{-1} \psi < 0) \frac{1}{\sigma} \sqrt{\frac{1}{e} |V_1^{-1} \psi|}.$$

Define h_* via $\zeta_* = N^{1/4}(h_* - 1)$, and consider

$$\text{Wald}_N^*(1) = N(h_* - 1) \bar{q}_N(h_*)' \hat{V}_{ff}(h_*)^{-1} \bar{q}_N(h_*) (h_* - 1).$$

Using $\hat{V} \xrightarrow{p} V$ (as $h_* \xrightarrow{p} 1$), we obtain from (58) that

$$\begin{aligned} \sqrt{N} q_N(h_*)' V(h_*)^{-1} q_N(h_*) &\xrightarrow{d} 0 && V_1^{-1} \psi \geq 0 \\ &\xrightarrow{d} 4\sigma^2 |V_1^{-1} \psi| && V_1^{-1} \psi < 0. \end{aligned}$$

It therefore follows that

$$\begin{aligned} \text{Wald}_N^*(1) &\xrightarrow{d} 0 && V_{ff}(1)_1^{-1} \psi \geq 0 \\ &\xrightarrow{d} 4\sigma^2 |V_{ff}(1)_1^{-1} \psi| \times && \\ &\quad \left(\frac{c}{2} + \frac{1}{\sigma} \sqrt{\frac{1}{e} |V_1^{-1} \psi|} \right)^2 && V_1^{-1} \psi < 0. \end{aligned}$$

Since

$$4\sigma^2 |V_1^{-1} \psi| \left(\frac{c}{2} + \frac{1}{\sigma} \sqrt{\frac{1}{e} |V_1^{-1} \psi|} \right)^2 = \left| \frac{1}{\sqrt{e}} V_1^{-1} \psi \right| \left(\sqrt[4]{e} c \sigma + 2 \sqrt[4]{e} \sqrt{\frac{1}{\sqrt{e}} \frac{1}{\sqrt{e}} |V_1^{-1} \psi|} \right)^2,$$

and $V_1^{-1} V V_1^{-1'} = e$, it follows that $e^{-1/2} V_1^{-1} \psi \sim N(0, 1)$, and so the limiting distribution of $\text{Wald}_N^*(1)$ can be written as in (40). Equation (30) follows by setting $c = 0$ in the above analysis.

To derive, λ_θ in (41), it suffices to consider the GAR statistic (as from Theorem 6 the non-centrality parameter is the same for all three tests),

$$GAR(1) = N\bar{f}_N(1)' \hat{V}_{ff}(1)^{-1} \bar{f}_N(1)$$

Using (57), it follows that

$$GAR(1) \xrightarrow{d} \left(\psi + \frac{c^2}{4}\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)' V_{ff}(1)^{-1} \left(\psi + \frac{c^2}{4}\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \chi_2^2(\lambda_\theta)$$

with λ_θ given in (41).

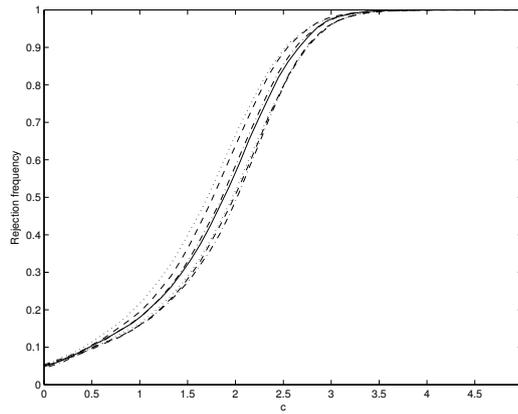


Figure 1: Local power curve of 95% tests of $H_0 : \theta = 1$ while the true value of $\theta = 1 - \frac{c}{2\sqrt[3]{N}}$ using the Wald statistic: Dotted: $N = 50, 5000$; Dashed: 100, 10000; Dash-dot: 500, 20000; Solid (1000)

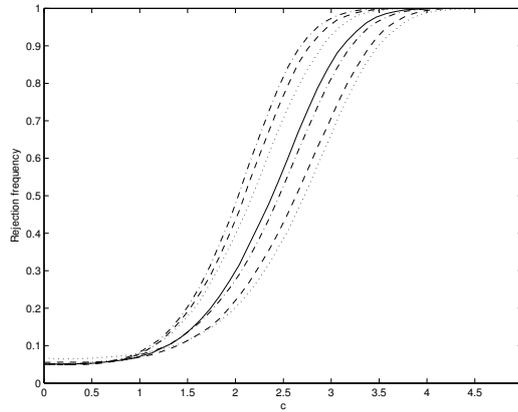


Figure 2: Local power curve of 95% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt[3]{N}}$ using the GAR statistic: Dotted: $N = 50, 5000$; Dashed: 100, 10000; Dash-dot: 500, 20000; Solid 1000.

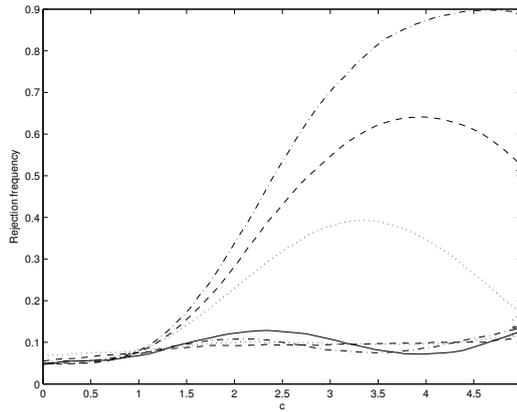


Figure 3: Local power curve of 95% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt{N}}$ using the KLM statistic: Dotted: $N = 50, 5000$; Dashed: 100, 10000; Dash-dot: 500, 20000; Solid 1000.

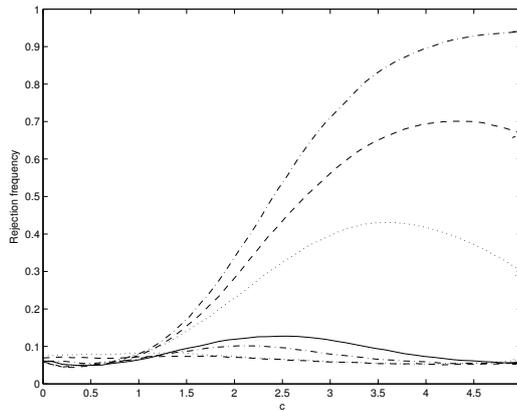


Figure 4: Local power curve of 95% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt{N}}$ using the LM statistic: Dotted: $N = 50, 5000$; Dashed: 100, 10000; Dash-dot: 500, 20000; Solid 1000.

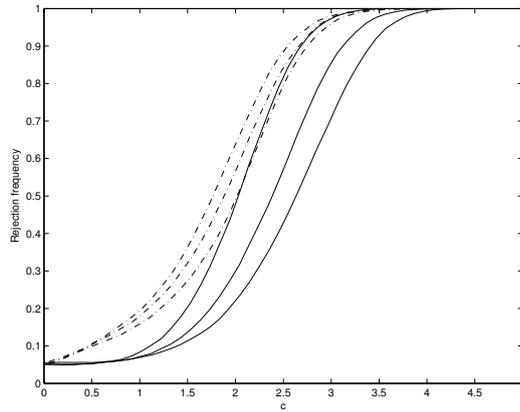


Figure 5: Local power curve of 95% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt{N}}$ using the Wald (dash-dot) and GAR statistics (solid) for $N = 50, 1000, 20000$.

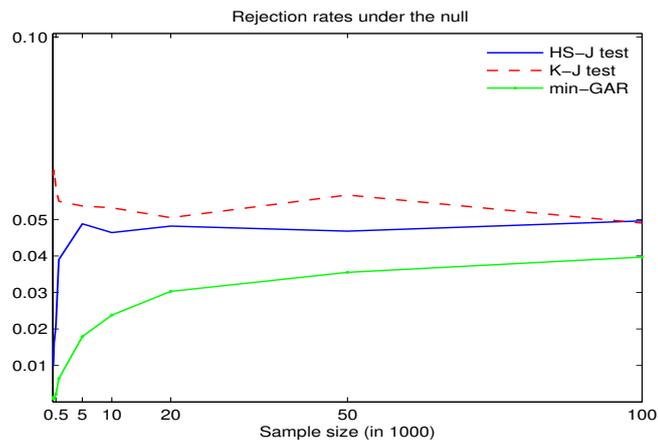


Figure 6: Rejection rates of the HS-J, K-J and min-GAR tests under the null; 10,000 replications; $c = 0$. ($\alpha = 0.05$)

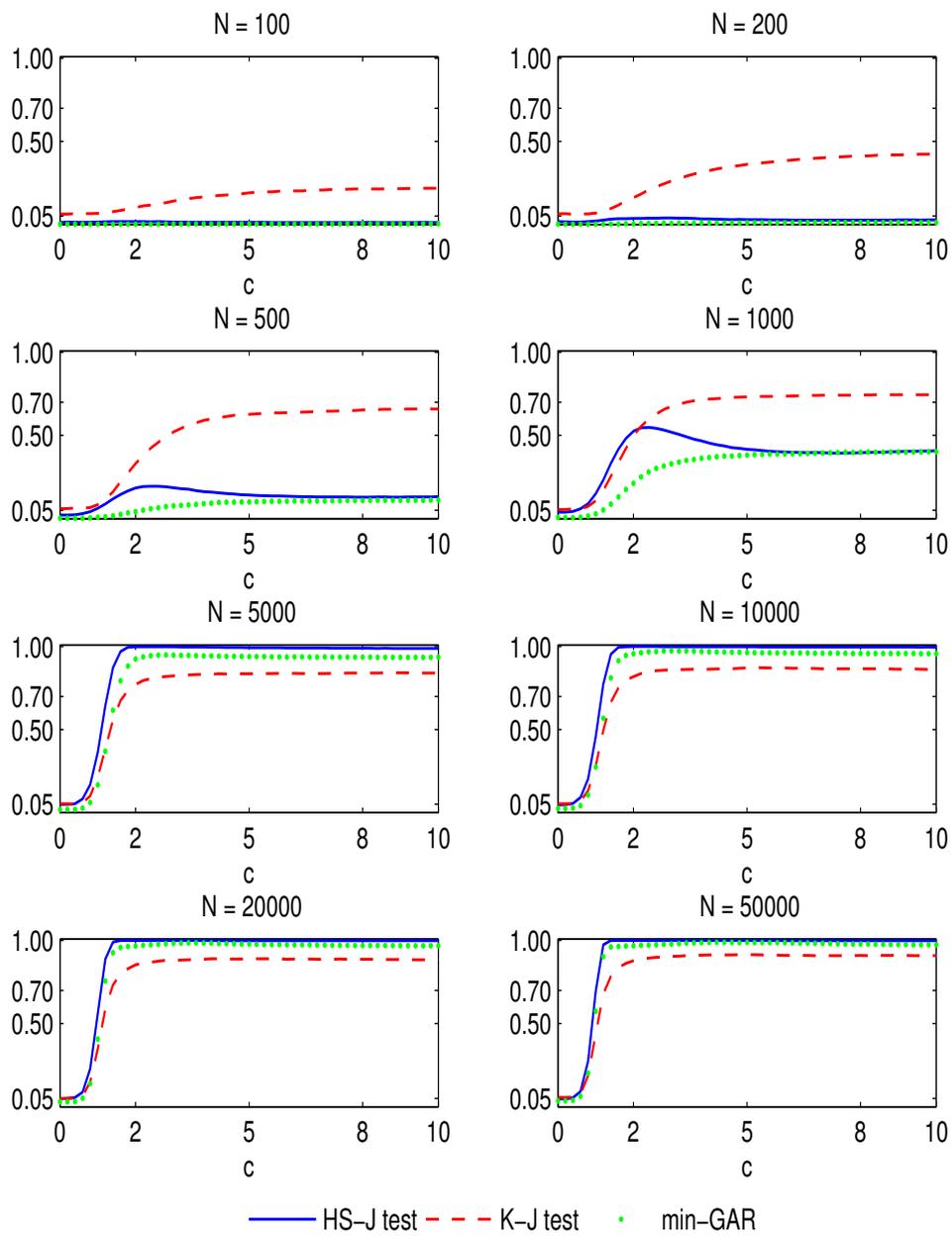


Figure 7: 10,000 replications; $c = 0 : 0.2 : 10$ ($\alpha = 0.05$)

References

- Ahn, S. C., and Schmidt, P. (1995). ‘Efficient estimation of models for dynamic panel data’, *Journal of Econometrics*, 68: 29–52.
- Anderson, T. W., and Hsiao, C. (1981). ‘Estimation of dynamic models with error components’, *Journal of the American Statistical Association*, 76: 46–63.
- Anderson, T. W., and Rubin, H. (1949). ‘Estimation of the parameters of a single equation in a complete system of stochastic equations’, *Annals of Mathematical Statistics*, 20: 46–63.
- Andrews, D. W. K. (1991). ‘Heteroscedasticity and autocorrelation consistent covariance matrix estimation’, *Econometrica*, 59: 817–858.
- Andrews, D. W. K., Moreira, M., and Stock, J. (2006). ‘Optimal two-sided invariant similar tests for Instrumental Variables regression’, *Econometrica*, 74: 715–752.
- Arellano, M., and Bond, S. R. (1991). ‘Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations’, *Review of Economic Studies*, 58: 277–297.
- Azzalini, A. (2005). ‘The skew-normal distribution and related multivariate families’, *Scandinavian Journal of Statistics*, 32: 159–188.
- Blundell, R., and Bond, S. R. (1998). ‘Initial conditions and moment restrictions in dynamic panel data models’, *Journal of Econometrics*, 87: 115–143.
- Bun, M., and Kleibergen, F. (2016). ‘Identification and inference in moments based analysis of linear dynamic panel data models’, Discussion paper, Department of Economics, University of Amsterdam, Amsterdam, NL.
- Diebold, F. X., and Nerlove, M. (1989). ‘The dynamics of exchange rate volatility: a multivariate latent factor ARCH model’, *Journal of Applied Econometrics*, 4: 1–22.
- Dovonon, P., and Hall, A. R. (2016). ‘The asymptotic properties of GMM and Indirect Inference under second-order identification’, Discussion paper, Department of Economics, University of Manchester, Manchester, UK.
- Dovonon, P., and Renault, E. (2009). ‘GMM overidentification test with first order underidentification’, Discussion paper, Department of Economics, Concordia University, Montreal, Canada.
- (2013). ‘Testing for common conditionally heteroscedastic factors’, *Econometrica*, 81: 2561–2586.
- Doz, C., and Renault, E. (2006). ‘Factor volatility in mean models: a GMM approach’, *Econometric Reviews*, 25: 275–309.
- Dufour, J.-M. (1997). ‘Some impossibility theorems in econometrics with applications to structural and dynamic models’, *Econometrica*, 65: 1365–1387.
- Engle, R., Ng, V. K., and Rothschild, M. (1990). ‘Asset pricing with a factor-ARCH covariance structure: empirical estimates for Treasury Bills’, *Journal of Econometrics*, 45: 213–237.

- Engle, R. F., and Kozicki, S. (1993). ‘Testing for common features’, *Journal of Business and Economic Statistics*, 11: 369–395.
- Fiorentini, G., Sentana, E., and Shephard, N. (2004). ‘Likelihood-based estimation of generalised ARCH structures’, *Econometrica*, 72: 1481–1517.
- Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press, Oxford, U.K.
- (2015). ‘Econometricians have their moments: GMM at 32’, *Economic Record*, 91, S1: 1–24.
- Hansen, L. P. (1982). ‘Large sample properties of Generalized Method of Moments estimators’, *Econometrica*, 50: 1029–1054.
- Jansen, I., and *et al* (2006). ‘The nature of sensitivity in monotone missing not at random models’, *Computational Statistics and Data Analysis*, 50: 830–858.
- Kleibergen, F. (2002). ‘Pivotal statistics for testing structural parameters in instrumental variables regression’, *Econometrica*, 70: 1781–1803.
- (2005). ‘Testing parameters in GMM without assuming that they are identified’, *Econometrica*, 73: 1103–1124.
- Kruiniger, H. (2014). ‘A further look at Modified ML estimation of the panel AR(1) model with fixed effects and arbitrary initial conditions’, Discussion paper, University of Durham, unpublished mimeo.
- Madsen, E. (2009). ‘GMM-based inference in the AR(1) panel data model for parameter values where local identification fails’, Discussion paper, Centre for Applied Microeconometrics, Department of Economics, University of Copenhagen, Copenhagen, Denmark.
- Moreira, M. J. (2003). ‘A conditional likelihood ratio test for structural models’, *Econometrica*, 71: 1027–1048.
- Newey, W. K. (1985). ‘Generalized Method of Moments specification testing’, *Journal of Econometrics*, 29: 229–256.
- Newey, W. K., and McFadden, D. L. (1994). ‘Large sample estimation and hypothesis testing’, in R. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2113–2247. Elsevier Science Publishers, Amsterdam, The Netherlands.
- Newey, W. K., and West, K. D. (1987). ‘Hypothesis testing with efficient method of moments testing’, *International Economic Review*, 28: 777–787.
- Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000). ‘Likelihood-based inference with singular information matrix’, *Bernoulli*, 6: 243–284.
- Sargan, J. D. (1983). ‘Identification and lack of identification’, *Econometrica*, 51: 1605–1633.
- Staiger, D., and Stock, J. (1997). ‘Instrumental variables regression with weak instruments’, *Econometrica*, 65: 557–586.

Stingo, F. C., Stanghellini, E., and Capobianco, R. (2011). ‘On the estimation of a binary response model in a selected population’, *Journal of Statistical Planning and Inference*, 141: 3293–3303.

Stock, J., and Wright, J. (2000). ‘GMM with weak identification’, *Econometrica*, 68: 1055–1096.