Inference in second order identified models

Kleibergen, F.R.; Dovonon, Prosper; Hall, Alastair

Published in:
Journal of Econometrics

Inference in Second-Order Identified Models[1]

Prosper Dovonon
Concordia University[2]

Alastair R. Hall
University of Manchester[3]

and

Frank Kleibergen
University of Amsterdam[4]

September 25, 2018

[1]The paper has benefited from the comments of: the co-editors and two anonymous referees; participants at the 12-th Tinbergen Institute Conference in Amsterdam, the 2017 African Econometric Society meetings in Algiers, and the 2017 Canadian Econometric Society meetings in Toronto (especially the discussant Kyungchul Song); and also seminar participants at the University of Bristol.

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

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

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

Abstract

We explore the local power properties of different test statistics for conducting inference in moment condition models that only identify the parameters locally to second order. We consider the 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), KLM statistic (Kleibergen, 2002, 2005) and the GMM extension of Moreira's (2003) (GMM-M) conditional likelihood ratio statistic. The GAR, KLM and GMM-M statistics are so-called "identification robust" since their (conditional) limiting distribution is the same under first-order, weak and therefore also second order identification. For inference about the model specification, we consider the identification-robust J statistic (Kleibergen, 2005) and the GAR statistic. Interestingly, we find that the limiting distribution of the Wald statistic under local alternatives not only depends on the distance to the null hypothesis but also on the convergence rate of the Jacobian. We specifically analyze two empirically relevant models with second order identification. In the panel autoregressive model of order one, our analysis indicates that the Wald test of a unit root value of the autoregressive parameter has better power compared to the corresponding GAR test which, in turn, dominates the KLM, GMM-M 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

The 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. 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 generalized conditional likelihood ratio (GMM-M) 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[2]: 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, Lee and Chesher (1986), Jansen and *et al* (2006); ML estimation of production function models, Lee and Chesher (1986), Lee (1993); GMM estimation of conditionally heteroskedastic

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

[2] The notion of identification at a higher order can be extended to an order larger than two. We are not aware of any (non-linear) model where identification of some parameters occurs at say the third order so we refrained from such an extension. We note that our results and also tests for such instances can be developed along the lines of our current work and existing tests for identification.

factor models, Dovonon and Renault (2009, 2013); GMM estimation of panel data models using second moments, Madsen (2009), Bun and Kleibergen (2016); modified-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 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 (2018), 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.[3] 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, KLM and GMM-M 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. Interestingly, we find that the limiting distribution of the Wald statistic under local alternatives not only depends on the distance to the null hypothesis but also on the convergence rate of the Jacobian. 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 discriminatory power that 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. The power of the GMM-M statistic is slightly better than the power of the KLM statistic: this relative ranking can be explained by noting that the parameter of interest is (second-order) identified and so the conditioning statistic in GMM-M, which tests the significance of the Jacobian, slowly rises when the sample increases. For the resulting larger values of the conditioning statistic, the GMM-M statistic is known to be

[3] Kruiniger (2014) derives the limiting distribution of certain modified- ML estimators in a panel data model with second-order identification that can be viewed as special cases of GMM with just-identified parameters.

comparable to the KLM statistic. For all statistics, we observe that the finite sample power curves converge slowly to the local asymptotic power curve which results from the quartic root convergence rate. 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.

Our analysis contributes to the literature on inference based on the Wald statistic in situations where the standard first-order asymptotic framework does not apply: Gourieroux and Montfort (1995) analyze estimation and testing under inequality constraints; Andrews (2001, 2002) considers the case where the true parameter lies on the boundary of the parameter space; Gaffke, Steyer, and von Davier (1999) and Gaffke, Heiligers, and Offinger (1999) consider the case where the Jacobian of the restrictions being tested is singular at the true parameter value. In common with the scenario in our paper, the limiting distribution of the Wald statistic is non-standard under the null hypothesis in all these cases.

The paper is organized as follows. In the second section, we set up notation, introduce the concept of second order identification and present three examples of models with this identification pattern. 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 5 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:

$$E[f(X,\theta_0)] \;=\; 0, \tag{1}$$

where $f : \mathcal{X} \times \Theta \to \mathbb{R}^k$ 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: $G(\theta) = E\left[q(X,\theta)\right]$, $q(\bar{\theta}) = \partial f(X,\theta)/\partial\theta'\big|_{\theta=\bar{\theta}}$. Let $\{x_i, i = 1,\ldots,N\}$ 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:

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

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 for a second-step estimation. Defining

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

and

$$Q(\theta, \bar{\theta}) \;=\; N\bar{f}_N(\theta)'\hat{V}_{ff}(\bar{\theta})^{-1}\bar{f}_N(\theta), \tag{3}$$

the two-step GMM estimator is:

$$\hat{\theta}_N = \arg\min_{\theta\in\Theta} Q(\theta, \hat{\theta}_{1,s}). \tag{4}$$

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.[4] Hansen (1982) establishes the limiting properties of both these statistics under a set of regularity conditions.[5] Specifically, he shows that: $\hat{\theta}_N$ is consistent for θ_0; $N^{1/2}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0, V_\theta)$, where $V_\theta = \{G(\theta_0)'V_{ff}(\theta_0)^{-1}G(\theta_0)\}^{-1}$; and $Q(\hat{\theta}, \hat{\theta}_{1,s}) \xrightarrow{d} \chi^2_{k-p}$.

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 $m(\theta) = E\left[f(X, \theta)\right]$. 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) $rank\{G(\theta_0)\} = p$ so that θ_0 is first-order locally identified.[6] Of these three, the consistency of the GMM estimator only requires (i) and (ii) to hold; but the distributional results listed in the previous paragraph 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 locally identified at second order. To formally introduce this scenario, we let

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

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.

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

[5] If 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 $V_{ff} = \lim_{T\to\infty} Var\left[N^{-1/2}\sum_{t=1}^{N} f(X_t)\right]$ and $\hat{V}_{ff}(\theta)$ by a member of the class of Heteroskedasticty Autocorrelation Covariance (HAC) estimators, for example see Andrews (1991).

[6] Sometimes referred to as the rank condition for identification.

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* $G(\theta_0)$*,*

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

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). See Dovonon and Renault (2013) for further discussion.

2.1 Panel data example

Consider the first-order autoregressive linear dynamic panel data model

$$y_{i,t} = \quad c_i + \theta_0 y_{i,t-1} + u_{it} \qquad i = 1, \dots, N, \; t = 2, \dots, T, \tag{5}$$

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

$$\begin{aligned} E[u_{i,t}u_{i,s}] &= \quad 0, \qquad s \neq t; \; t = 2, \dots, T, \\ E[u_{i,t}c_i] &= \quad 0, \qquad t = 2, \dots, T, \\ E[u_{i,t}y_{i,1}] &= \quad 0, \qquad t = 2, \dots, T. \end{aligned} \tag{6}$$

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:

$$\bar{f}_N(\theta) \;=\; a\theta^2 \;+\; b\theta \;+\; d, \tag{7}$$

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:

$$
\begin{aligned}
E[a] = &\begin{pmatrix} E[(c_i - (1-\theta_0)y_{i,1})^2] + \sigma_2^2 \\ 0 \end{pmatrix}, \\
E[b] = &\begin{pmatrix} (1+\theta_0)^2 E[(c_i - (1-\theta_0)y_{i,1})^2) - \theta_0^2\sigma_2^2 - \sigma_3^2 \\ -\theta_0^2 E[(c_i - (1-\theta_0)y_{i,1})^2] \end{pmatrix}, \\
E[d] = &\begin{pmatrix} \theta_0(1+\theta_0+\theta_0^2)E[(c_i - (1-\theta_0)y_{i,1})^2] + \theta_0^2(\theta_0-1)\sigma_2^2 + \theta_0\sigma_3^2 \\ \theta_0^2 E[(c_i - (1-\theta_0)y_{i1})^2] \end{pmatrix}.
\end{aligned} \tag{8}
$$

with $\sigma_t^2 = E[u_{it}^2]$. If we assume mean-stationarity[7] - so that $E[(c_i - (1-\theta_0)y_{i1})^2] = 0$ - and the errors are homoskedastic - $\sigma_t^2 = \sigma^2$ - then these expected values simplify to

$$
E[a] = \sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad E[b] = -\sigma^2 \begin{pmatrix} \theta_0^2 + 1 \\ 0 \end{pmatrix}, \qquad E[d] = \sigma^2 \begin{pmatrix} \theta_0^2(\theta_0 - 1) + \theta_0 \\ 0 \end{pmatrix}. \tag{9}
$$

From (7) and (9), it follows that if $\theta_0 = 1$ then:

$$
m(\theta_0) = 0_{2\times 1}, \qquad G(\theta_0) = 0_{2\times 1}, \qquad H_1(\theta_0) = 2\sigma^2, \qquad H_2(\theta_0) = 0, \tag{10}
$$

where we have emphasized the dimensions of the null vectors for clarity. It can be seen from (10) 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.[8] 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}:

$$
E\left[Y_{t+1} \,|\, \mathfrak{F}_t\right] = 0, \tag{11}
$$

$$
Var\left[Y_{t+1} \,|\, \mathfrak{F}_t\right] = \Lambda D_t \Lambda' + \Omega, \tag{12}
$$

[7] See Blundell and Bond (1998).

[8] The 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 $\sigma^2_{\ell,t}$ for $\ell = 1, 2, \ldots, L$, 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 $\left\{\sigma^2_{\ell,t}\right\}_{1 \leq \ell \leq L, t \geq 0}$ are adapted with respect to the increasing filtration $\{\mathfrak{F}_t\}_{t \geq 0}$. It is assumed that $rank(\Lambda) = L$ and $Var[\sigma^2_{\ell,t}] > 0$ for all $\ell = 1, 2, \ldots, L$. 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

$$m(\theta_0) \;=\; 0 \tag{13}$$

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

$$f_t(\theta) = z_t\{(\theta' Y_{t+1})^2 \;-\; c(\theta)\}, \tag{14}$$

and $c(\theta) = E[(\theta' Y_{t+1})^2]$. Clearly (13) 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.[9] The population moment condition in (13) can be used as a basis for estimation of θ_0, and the existence of the common feature can be tested by testing whether (13) holds using the overidentifying restrictions statistic.

However, the population moment condition in (13) does not locally identify θ_0 at first order. Dovonon and Renault (2013) show that

$$G(\theta) \;=\; 2E\left[\,(z_t \,-\, E[z_t])\,\theta_0'\,(\Lambda D_t \Lambda' \,+\, \Omega)\,\right], \tag{15}$$

and that under the assumptions above,

$$E[(\theta_0' Y_{t+1})^2 \,|\, \mathfrak{F}_t] = \mu \qquad \Leftrightarrow \qquad \theta_0' \Lambda \;=\; 0. \tag{16}$$

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

$$H_s(\theta) \;=\; \Lambda' C_s \Lambda, \tag{17}$$

where C_s is the $L \times L$ diagonal matrix with ℓ^{th} main diagonal element equal to $Cov[z_{s,t}, \sigma^2_{\ell,t}]$. 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 $\left\{\sigma^2_{\ell,t}\right\}$.[10]

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 limiting distributions under the null

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 under this H_0 the value of θ_0 is completely specified. In the second type of hypothesis, the null

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

[10] Specifically, they assume $rank\{Cov[z_t, d_t]\} = L$ where $d_t = (\sigma^2_{\ell,t}, \sigma^2_{\ell,t}, \ldots, \sigma^2_{L,t})$.

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 the 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 $q_i(\bar{\theta}) = \partial f_i(\theta)/\partial\theta'\big|_{\theta=\bar{\theta}}$.

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:

$$\mathrm{Wald}_N(\theta_*) = N(\hat{\theta}_N - \theta_*)'\bar{q}_N(\hat{\theta}_N)'\hat{V}_{ff}(\hat{\theta}_N)^{-1}\bar{q}_N(\hat{\theta}_N)(\hat{\theta}_N - \theta_*), \tag{18}$$

and the LM statistic is,

$$LM(\theta_*) = N\bar{f}_N(\theta_*)'\hat{V}_{ff}(\theta_*)^{-1}\bar{q}_N(\theta_*)\left(\bar{q}_N(\theta_*)'\hat{V}_{ff}(\theta_*)^{-1}\bar{q}_N(\theta_*)\right)^{-1}\bar{q}_N(\theta_*)'\hat{V}_{ff}(\theta_*)^{-1}\bar{f}_N(\theta_*). \tag{19}$$

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^2_ρ where ρ is the number of restrictions which is p in our case here.

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

$$KLM(\theta_*) = N\bar{f}_N(\theta_*)'\hat{V}_{ff}(\theta_*)^{-1}\hat{D}_N(\theta_*)\left(\hat{D}_N(\theta_*)'\hat{V}_{ff}(\theta_*)^{-1}\hat{D}_N(\theta_*)\right)^{-1}\hat{D}_N(\theta_*)'\hat{V}_{ff}(\theta_*)^{-1}\bar{f}_N(\theta_*), \tag{20}$$

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

$$vec\left(\hat{D}_N(\theta)\right) = vec\left(\bar{q}_N(\theta)\right) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\bar{f}_N(\theta), \tag{21}$$

with $\hat{V}_{qf}(\theta) = N^{-1}\sum_{i=1}^N vec\left[q_i(\theta) - \bar{q}_N(\theta)\right]\left[f_i(\theta) - \bar{f}_N(\theta)\right]'$, where $vec(A)$ is the operator transforming the matrix A into a vector by stacking its columns. Kleibergen (2005) shows that $KLM(\theta_*)$ converges to a χ^2_p distribution under H_0 regardless of whether θ_0 is first order locally identified or weakly identified.

We also consider the conditional GMM statistic of Kleibergen (2005) (GMM-M(θ_*)) which is the GMM version of Moreira's (2003) conditional likelihood ratio (LR) statistic:

$$\begin{aligned} GMM - M(\theta_*) = \tfrac{1}{2}\Big(&KLM(\theta_*) + J(\theta_*) - rk(\theta_*) \\ &+\sqrt{[KLM(\theta_*) + J(\theta_*) + rk(\theta_*)]^2 - 4J(\theta_*)rk(\theta_*)}\Big), \end{aligned} \tag{22}$$

where $J(\theta)$ is defined in equation (24) below and $rk(\theta_*)$ is a test statistic that tests the hypothesis of a lower rank value of $G(\theta_*)$, $H_r : rank(G(\theta_*)) = p - 1$ and is function of $\hat{D}_N(\theta_*)$ and the inverse of $\hat{V}_{22.f}(\theta_*) \equiv \hat{V}_{22}(\theta_*) - \hat{V}_{2f}(\theta_*)\hat{V}_{ff}(\theta_*)^{-1}\hat{V}_{f2}(\theta_*)$; with $\hat{V}_{ab}(\theta_*)$ $(a, b = 2, f)$ a consistent estimator of $V_{ab}(\theta_*)$. Many examples of rank test statistics have this characterization. See Kleibergen (2005).

Stock and Wright (2000) propose using the GAR statistic:[11]

$$GAR(\theta_*) \; = \; Q(\theta_*, \theta_*), \tag{23}$$

and $Q(\cdot, \cdot)$ is defined in (3). Stock and Wright (2000) show that $GAR(\theta_*)$ converges to a χ^2_k distribution under H_0 regardless of whether θ_0 is first order locally identified or weakly identified. However, the implicit null of the 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

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

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^2_{k-p} 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 GAR statistic 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 $r = rank\{G(\theta_0)\}$. 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 $R = (R_1, R_2)$ such that the $p \times r$ matrix R_1 and $p \times (p - r)$ matrix R_2 satisfy:

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

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.

[11] Anderson 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.

Assumption 2. *θ_0 is an interior point of Θ.*

Let $\mathcal{N}_\epsilon$ denote an ϵ-neighbourhood of θ_0.

Assumption 3. *(i) $\|m(\theta)\| < \infty$, $\|G(\theta)\| < \infty$ and $\|H_s(\theta)\| < \infty$ for $s = 1, 2, \ldots, k$ 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 $\mathcal{N}_\epsilon$.*

Assumption 4.

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

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

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

Assumption 4 is a high-level condition that can apply whether the model involves a 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.[12]

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:

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

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

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

where

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

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.

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

The limiting distribution is evidently non-standard, reflecting the non-standard behaviour of the GMM estimator in this case (see Dovonon and Hall (2018)[Theorem 1] and equation (A.2) in the appendix). 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 (2018).[13] 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.*

Corollary 1 provides the limiting distribution of the Wald 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:

$$\text{Wald}_N(1) \;=\; N(\hat{\theta} - 1)\bar{q}_N(\hat{\theta})'\hat{V}_{ff}(\hat{\theta})^{-1}\bar{q}_N(\hat{\theta})(\hat{\theta} - 1). \tag{26}$$

Using a Mean Value expansion of $\bar{q}_N(\hat{\theta})$ around $\bar{q}_N(1)$, it can be shown that[14]

$$\text{Wald}_N(1) \;=\; N(\hat{\theta} - 1)^4 4\sigma^4 V_{1,1}^{-1} \tag{27}$$

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:

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

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

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

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

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

Note that even though the distribution of $\mathbb{W}_0(S)$ is nonstandard, its quantiles have a simple expression. Letting $\alpha \in (0, 1/2)$ and $c_{1-\alpha}$ be the $(1-\alpha)$-quantile of $\mathbb{W}_0(S)$, we can show that

[13] Dovonon and Hall (2018) also discuss at length how to estimate R_2.

[14] See the on-line appendix available from the authors upon request.

$c_{1-\alpha} = 4z^2_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution. The $(1-\alpha)$-quantile of the chi-squared distribution with one degree of freedom is $\chi^2_{1,1-\alpha} = z^2_{1-\alpha/2}$ and we can verify that $c_{1-\alpha} > \chi^2_{1,1-\alpha}$ for all $0 < \alpha < 0.30344$. This makes the Wald test based on $\mathbb{W}_0(S)$ typically conservative, hence valid, when used for models that are actually locally identified at first order.

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, GMM-M 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 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 (19).

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

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^2_p 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 (20) 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

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

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

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

and $\hat{V}_{2f}(\theta_0)$ be the sample counterpart of $V_{2f}(\theta_0)$ as defined in Assumption 4. Related to the asymptotic behaviour of GMM-M(θ_*), for any $\rho \in \mathbb{R}_+$, let

$$\Psi(\rho) = \frac{1}{2}\left(\psi_J + \psi_K - \rho + \sqrt{(\psi_J + \psi_K + \rho)^2 - 4\rho \times \psi_J}\right),$$

where ψ_J and ψ_K are independent χ^2_{k-p} and χ^2_p random variables. We have the following result:

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 $KLM(\theta_*) \xrightarrow{d} \chi^2_p$; also, for any $\rho \in \mathbb{R}_+$, GMM-M(θ_*) converges in distribution to $\Psi(\rho)$ conditionally on $rk(\theta_*) = \rho$, i.e. GMM-M$(\theta_*)\big|(rk(\theta_*) = \rho) \xrightarrow{d} \Psi(\rho)$; (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 $GAR(\theta_*) \xrightarrow{d} \chi^2_k$.*

From Theorem 3 it follows that the limiting distributions of the KLM, GMM-M and GAR statistics under second-order local identification are the same as under first-order local identification and weak identification. Therefore all three 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 $J(\theta_0) \xrightarrow{d} \chi^2_{k-p}$; (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 $GAR(\theta_0) \xrightarrow{d} \chi^2_k$.*

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^2_{k-q}, $q = 0, 1, \ldots, p$, 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

$$E_N\,[f(X_N, \theta_N)] \;=\; \mu_N, \tag{31}$$

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 $N \to \infty$ the following all hold: $P_N \to P$, $\theta_N \to \theta_0$ and $\mu_N \to 0_{k\times 1}$. 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:

$$m_N(\theta) = E_N\left[f(X_N,\theta)\right], \qquad G_N(\theta) \;=\; E_N\left[q(X_N,\theta)\right], \qquad H_N(\theta) \;=\; E_N\left[h(X_N,\theta)\right],$$

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

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 $\mathcal{N}_\epsilon$, 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.

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

where $R \equiv (R_1, R_2)$ is the nonsingular $p \times p$ matrix partitioned into r and $(p-r)$-column matrices R_1 and R_2 as defined by (25)*. $D = G(\theta_0)R_1$ 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 $\xi = 1/4$ is the appropriate choice in the panel data model in Section 2.1 while $\xi = 1/2$ is appropriate in the CHF model in Section 2.2. Under some mild smoothness condition, ξ is not expected to be smaller than $1/4$. For our analysis of tests of $H_0 : \theta_0 = \theta_*$, we restrict $\xi \geq 1/4$ to ensure that the drift in the Jacobian decreases at least as fast as 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.

$$\sqrt{N}\begin{pmatrix} \bar{f}_N(\theta_N) \;-\; \mu_N \\ vec\left[\bar{q}_N(\theta_N)R_2 - G_N(\theta_N)R_2\right] \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_f \\ vec(\psi_q) \end{pmatrix} \sim N(0,V)$$

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 alternatives 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 parameters 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; see equation (A.2) in the appendix. 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:

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

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 that[15]

$$\begin{array}{rcl} G_N(\theta_N) & = & \frac{c}{\sqrt[4]{N}}\sigma^2 \begin{pmatrix} 0 \\ 1/2 \end{pmatrix} + O(1/\sqrt{N}), \\ G_N(\theta_0) & = & \frac{c}{\sqrt[4]{N}}\sigma^2 \begin{pmatrix} 1 \\ 1/2 \end{pmatrix} + O(1/\sqrt{N}), \\ H_N(\theta_N) & = & H_N(\theta_0) \ = \ 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + O(1/\sqrt{N}). \end{array} \tag{34}$$

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

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

$$\begin{array}{rcl} \mathbb{W}_s(e) & = & (a(\zeta_s)+e_1)'\tilde{D}'\tilde{D}(a(\zeta_s)+e_1) + (\tilde{A}+\tilde{B}\zeta_s)'(\tilde{A}+\tilde{B}\zeta_s)(\zeta_s+e_2)^2 \\ & & +2(a(\zeta_s)+e_1)'\tilde{D}'(\tilde{A}+\tilde{B}\zeta_s)(\zeta_s+e_2), \end{array}$$

with

$$\begin{array}{rcl} a(\zeta) & = & -(\tilde{D}'\tilde{D})^{-1}\tilde{D}'\left(\tilde{X}+\tilde{A}\zeta+\frac{1}{2}\tilde{B}\zeta^2\right) \\ \zeta_s & \in & \arg\min_{u\in\mathbb{R}} \left(\tilde{X}+\tilde{A}u+\frac{1}{2}\tilde{B}u^2\right)' M_d \left(\tilde{X}+\tilde{A}u+\frac{1}{2}\tilde{B}u^2\right) \\ \tilde{X} & \sim & N(0, I_k) \\ \tilde{A} & = & V_{ff}(\theta_0)^{-1/2}A\mathbb{I}(\xi = 1/4) \end{array}$$

and M_d, $\tilde{B}$ and $\tilde{D}$ given just before Theorem 1.

[15] See on-line appendix available from the authors upon request.

In this theorem and as established by Theorem B.1, ζ_s represents the asymptotic distribution under P_N (along suitable subsequences) of $\hat{\theta} - \theta_N$ in the direction that is identified at second order, i.e. the last component of $R^{-1}(\hat{\theta} - \theta_N)$ scaled by $N^{1/4}$ while $a(\zeta_s)$ is the asymptotic distribution of $\hat{\theta} - \theta_N$ in the directions that are first-order identified, i.e. the first $p-1$ components of $R^{-1}(\hat{\theta} - \theta_N)$ scaled by $\sqrt{N}$. These two sets of components are the key ingredients of the asymptotic distribution of the Wald test statistic under local alternatives.

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

Corollary 2. *If the conditions of Theorem 5 hold and in addition, $r = 0$ and $p = 1$, then*

$$\mathbb{W}_s(e_2) = (\tilde{A} + \tilde{B}\zeta_s)'(\tilde{A} + \tilde{B}\zeta_s)(\zeta_s + e_2)^2,$$

with

$$\zeta_s \in \arg\min_{u \in \mathbb{R}} \left(\tilde{X} + \tilde{A}u + \frac{1}{2}\tilde{B}u^2 \right)' \left(\tilde{X} + \tilde{A}u + \frac{1}{2}\tilde{B}u^2 \right). \tag{35}$$

If, in addition $\xi > 1/4$, then

$$\mathbb{W}_s(e_2) = \tilde{B}'\tilde{B}\zeta_s^2(\zeta_s + e_2)^2,$$

with $\zeta_s^2 = -\frac{\tilde{B}'\tilde{X}}{\tilde{B}'\tilde{B}}\mathbb{I}(\tilde{B}'\tilde{X} \leq 0)$.

The panel data example in Section 2.1 fits into the setting of this corollary. From (34), we can claim that

$$\xi = 1/4, \;\; \tilde{A} = V_{ff}(\theta_0)^{-1/2}(\sigma^2 c/2)\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \;\; \tilde{B} = V_{ff}(\theta_0)^{-1/2}(2\sigma^2)\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \;\text{ and }\; \tilde{X} \sim N(0, I_2).$$

As we shall see in Proposition 1 below, the fact that $\tilde{A}$ and $\tilde{B}$ are not proportional ensures that the Wald test has discriminatory power.

The asymptotic distributions presented by Theorem 5 and Corollary 2 extend the results in Theorem 1 and Corollary 1 to local alternatives. Setting A and e to 0 in Theorem 5 and Corollary 2 yields the asymptotic distribution of the Wald statistic under the null as presented in the previous section. Corollary 2 sheds some light on a peculiar behaviour of the Wald test under local alternatives. It appears that the power of this test in large samples is not mainly determined by the distance between the parameter values under the null and the alternative but also by how fast the Jacobian matrix of the moment function evaluated under the alternative converges to 0. In the case of fast convergence of the Jacobian ($\xi > 1/4$), the asymptotic distribution of the test statistic has an atom mass at 0 regardless of the value of the localization parameter value e_2. This implies in particular that asymptotically, the probability of rejecting the null under local alternatives is always smaller than or equal to the atom mass $P(\tilde{B}'\tilde{X} > 0) = 0.5$ for all $e_2 \in \mathbb{R}$ whereas one should expect this rejection probability to tend to 1 as e_2 gets larger. When $\xi = 1/4$ so that the Jacobian drifts towards 0 at the rate $N^{-1/4}$, $\mathbb{W}_s(e_2)$ can still exhibit an atom mass at 0 but in more restrictive settings made precise by the following proposition.

Proposition 1. *Assume that the same conditions as in Theorem 5 hold with $r = 0$ and $p = 1$ and $\xi = 1/4$. Then,* *(a)* $[\forall e_2 \in \mathbb{R}, \quad P\left(\mathbb{W}_s(e_2) = 0\right) > 0]$ *if and only if* $[A = 0$ *or* $A = \delta B$ *for some* $\delta \in \mathbb{R} \setminus \{0\}]$.

(b) If $A = 0$ *then,* $\forall e_2 \in \mathbb{R}, \quad P\left(\mathbb{W}_s(e_2) = 0\right) \geq 1/2$.

(c) If $A = \delta B$ *for some* $\delta \in \mathbb{R} \setminus \{0\}$ *then,* $\forall e_2 \in \mathbb{R},\ P\left(\mathbb{W}_s(e_2) = 0\right) \geq 1 - \Phi\left(\frac{\delta^2}{2}\sqrt{\tilde{B}'\tilde{B}}\right)$, *where* Φ *is the standard normal cumulative distribution function.*

This proposition shows that when $\xi = 1/4$, the Wald test statistic has an atom at 0 asymptotically if and only if the localization vector A of the Jacobian matrix is 0 or is proportional to the vector of second-order derivatives B of the moment function at the true value. In any other case, $\mathbb{W}_s(e_2)$ shifts towards infinity as e_2 gets large; hence, showing evidence of power. The cases $A = 0$ and A proportional to B occur in a small set of nil Lebesgue measure in $\mathbb{R}^k$. Therefore, the power issue raised can be considered irrelevant when $\xi = 1/4$.

To present the limiting behaviour of the LM, KLM and GAR tests, we introduce some notation. Let

$$
\begin{aligned}
C(\theta) &= \left(\ vec\left(\frac{\partial^2 m_1(\theta)}{\partial\theta\partial\theta'}\right) \quad vec\left(\frac{\partial^2 m_2(\theta)}{\partial\theta\partial\theta'}\right) \quad \ldots \quad vec\left(\frac{\partial^2 m_k(\theta)}{\partial\theta\partial\theta'}\right) \ \right)', \\
\mu_\theta &= V_{ff}(\theta_*)^{-1/2}\left\{-De_1 + \tfrac{1}{2}\left[(R_2e_2)' \otimes I_k\right] H(\theta_*)(R_2e_2) - Ae_2\mathbb{I}(\xi = 1/4)\right\} \\
\mathbb{Q}(e_2) &= V_{ff}(\theta_*)^{-1/2}\left(\ D \vdots - C(\theta_*)\left[I_p \otimes (R_2e_2)\right] R_2 + A\mathbb{I}(\xi = 1/4) \ \right), \quad \text{and} \\
\mathbb{P}(e_2) &= \mathbb{Q}(e_2)\left(\mathbb{Q}(e_2)'\mathbb{Q}(e_2)\right)^{-1}\mathbb{Q}(e_2)'
\end{aligned}
$$

whenever this expression is well-defined.

$\mathbb{Q}(e_2)$ is the probability limit under P_N of the Jacobian matrix at θ_* of $\bar{f}_N$ suitably scaled and μ_θ represents the (normalized) deviation from 0 of the population mean of $\bar{f}_N(\theta_*)$ under P_N. In the expression of μ_θ, the first and second terms in curly brackets are the deviations induced by the first and second-order identified directions, respectively whereas the last term represents the shift induced by the rate of convergence to 0 of the population mean of the Jacobian matrix. (See Assumption 7.) Note that this term is present only if $\xi = 1/4$.

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

(a) If $\mathbb{Q}(e_2)$ *is full column rank,*

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

and, for any $\rho \in \mathbb{R}_+$,

$$GMM\text{-}M(\theta_*)\big|(rk(\theta_*) = \rho) \xrightarrow{d} \tilde{\Psi}(\rho) \geq \chi^2_p(\lambda_\theta)$$

under P_N, *with* $\lambda_\theta = \mu_\theta'\mathbb{P}(e_2)\mu_\theta$,

$$\tilde{\Psi}(\rho) = \frac{1}{2}\left(\tilde{\psi}_J + \tilde{\psi}_K - \rho + \sqrt{(\tilde{\psi}_J + \tilde{\psi}_K + \rho)^2 - 4\rho \times \tilde{\psi}_J}\right),$$

and $\tilde{\psi}_J \sim \chi^2_{k-p}(\mu'_\theta\mu_\theta - \lambda_\theta)$ independent of $\tilde{\psi}_K \sim \chi^2_p(\lambda_\theta)$.
If in addition, $\xi > 1/4$ or $A = 0$, then $\lambda_\theta = \mu'_\theta\mu_\theta > 0$. Otherwise, if in addition, $\xi = 1/4$ and $A \neq 0$, then $\lambda_\theta > 0$ for all $e \in \mathbb{R}^p$ such that $\mathbb{P}(e_2)V_{ff}(\theta_*)^{-1/2}Ae_2 \neq \mathbb{Q}(e_2)\begin{pmatrix} 2e_1 \\ e_2 \end{pmatrix}$.

(b)

$$GAR(\theta_*) \xrightarrow{d} \chi^2_k(\ell_\theta)$$

under P_N with $\ell_\theta = \mu'_\theta\mu_\theta$. If in addition, $\left(\xi = 1/4 \text{ and } A \notin \left\langle D \vdots -C(\theta_*)\left[I_p \otimes (R_2e_2)\right]R_2 \right\rangle\right)$ or $(\xi > 1/4)$ or $(A = 0)$ then $\ell_\theta > 0$ for all $e \neq 0$; where $\langle M \rangle$ is the space spanned by the columns of the matrix M.

This theorem shows that the LM and KLM statistics have the same limiting distribution under this sequence of local alternatives and this common distribution is a lower bound of the asymptotic distribution of GMM-M statistic. The full rank condition on $\mathbb{Q}(e_2)$ ensures that all these test statistics are well-defined with probability approaching one as the sample size grows. Since, except for some marginal configuration, both λ_θ and ℓ_θ are positive, it follows automatically from Theorems 3 and 6 and the properties of the chi-squared distribution that KLM, GMM-M and GAR statistics have non-trivial power against this alternative. 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^2_p 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 alternatives is given by (31) with $\theta_N = \theta_0$ and $\mu_N = c/\sqrt{N}$ for all N so that

$$m_N(\theta_0) = \frac{c}{\sqrt{N}}. \tag{36}$$

However, as noted above, the appropriate choice of ξ in Assumption 7 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 (12) can then be written as:

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

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

$N \to \infty$, $\lambda_{2,N} \to 0$. 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:[16]

$$m_N(\theta_0) = (\theta_0'\lambda_{2,N})^2 Cov[\sigma_{2,t}^2, z_t], \tag{37}$$

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 (37) may be of order $O\left(N^{-2\delta}\right)$ 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 $z_t = \left(vech(Y_{t-\tau}Y_{t-\tau}') : \tau = 0, \dots, h\right)'$, 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 can be shown that:[17]

$$m_N(\theta_0) \;=\; \begin{pmatrix} \lambda_{2,N,1}^2 \\ \lambda_{2,N,2}^2 \end{pmatrix} (\theta_0'\lambda_{2,N})^2 Cov\left[F_{2,t+1}^2, F_{2,t}^2\right], \tag{38}$$

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

$$G_N(\theta_0) \;=\; 2Cov[F_{2,t+1}^2, F_{2,t}^2](\theta_0'\lambda_{2,N}) \begin{pmatrix} \lambda_{2,N,1}^2 \\ \lambda_{2,N,2}^2 \end{pmatrix} \lambda_{2,N}'. \tag{39}$$

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

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

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 (36) 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

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

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:

[16]See the on-line appendix available from the authors upon request.

[17]Note that due to the necessary normalization only one element of θ has to be estimated; see discussion in Section 2.2. See the on-line appendix available from the authors upon request.

Theorem 7. *(i) Assume that $G_N(\theta_0) \to G(\theta_0)$ as $N \to \infty$ 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: $J(\theta_0) \xrightarrow{d} \chi^2_{k-p}(\lambda_m)$ under P_N, with $\lambda_m > 0$ almost surely; (ii) If $\sqrt{N}\left(\bar{f}_N(\theta_0) - c/\sqrt{N}\right) \xrightarrow{d} N(0, V_{ff}(\theta_0))$, under P_N, and $\hat{V}_{ff}(\theta_0)$ converges in probability (under P_N) to $V_{ff}(\theta_0)$ then $GAR(\theta_0) \xrightarrow{d} \chi^2_k(\nu)$ 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 $c \neq 0$.

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, GMM-M 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 $\theta_0 = 1$[18], the point of second order identification. We specify the local alternative as

$$\theta_N = 1 - \frac{c}{2\sqrt[4]{N}}, \tag{40}$$

with $c > 0$.

Corollary 2 characterizes the asymptotic distribution of the Wald statistic testing the null hypothesis that $\theta_0 = 1$. It is shown in the Supplementary appendix that under the local alternative

[18] In Bun and Kleibergen (2016), the maximal attainable power for testing a local alternative while the true value is one so the autoregressive parameter is second order identified is studied. This differs from the usual notion of power which we analyze here. The important difference between these two settings is that the Jacobian equals zero in the setting analyzed by Bun and Kleibergen (2016) while it is of the order $O_p(N^{-1/4})$ in our setting. The maximal attainable power curve from Bun and Kleibergen (2016) does therefore not apply to our setting.

in (40) the distribution of $\text{Wald}_N^*(1)$ is given by:

$$\text{Wald}_N(1) \xrightarrow{d} \quad \zeta^2 \frac{\left[\left(\sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta\right)' \left(\sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta\right)\right]^2}{\left(\sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta\right)' V_{ff}(1) \left(\sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta\right)} \tag{41}$$

where ζ is a root of the third order polynomial equation (D.13) stated in the Supplementary appendix. We compute the rejection frequency of H_0 under local alternatives for different sample sizes alongside the rejection frequency that results from the asymptotic distribution. 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 slowly converges to the asymptotic one which lies (primarily) on the right-hand side of the finite sample power curves. The 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 low power at small sample sizes which is as expected given the quartic root convergence rate. This can be further inferred from the values of θ when the drifting parameter c equals four. The power then exceeds 50%. A value of c equal to four corresponds with a value of θ of 0.25 ($N = 50$), 0.37 (100), 0.58 (500), 0.65 (1000), 0.77 (5000), 0.80 (10000) and 0.83 (20000).

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 to the $\chi_2^2(\ell_\theta)$ distribution. In the Supplementary appendix, it is shown that the non-centrality parameters are given by:

$$\begin{aligned} \lambda_\theta &= \tfrac{c^2\sigma^2}{16}\tbinom{1}{1}' V_{ff}(1)^{-1}\tbinom{1}{1/2}\left[\tbinom{1}{1/2}' V_{ff}(1)^{-1}\tbinom{1}{1/2}\right]^{-1}\tbinom{1}{1/2} V_{ff}(1)^{-1/2}\tbinom{1}{1} \\ \ell_\theta &= \tfrac{c^2\sigma^2}{16}\tbinom{1}{1}' V_{ff}(1)^{-1}\tbinom{1}{1}. \end{aligned} \tag{42}$$

Figures 2-5 show the finite sample and local asymptotic power curves of the GAR, KLM, GMM-M and LM tests for increasing number of observations. Figure 6 shows finite sample power curves of the GAR and Wald tests. The finite sample power curves of the GAR, KLM and LM tests all slowly move to the local asymptotic one when the number of observations increases. The slow convergence results from the quartic root convergence rate. 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. We did not show the local asymptotic power curve of the GMM-M statistic because of its conditional limiting distribution.

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. 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. This is also revealed by the finite sample power curves of the GMM-M statistic. It is known that the GMM-M statistic resembles the KLM statistic when identification is strong and the GAR statistic when identification is weak. Figure 4 shows that the GMM-M statistic has power more similar to the GAR statistic when sample size is small so identification is weak and power comparable to the KLM statistic when the sample size is large so the identification has improved. This combined effect leads to an oddly behaved power curve which suggests that the weighting on the conditioning statistic in the GMM-M statistic, which is optimal for the i.i.d. linear instrumental variables regression model, could be improved to obtain better power since now the GMM-M statistic has more power for small N, when it resembles the GAR statistic, compared to intermediate N, when it resembles the KLM statistic.

To our reading, the most striking feature of these results is the good performance of the Wald test as further reflected by Figure 6. 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 autoregressive 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 $\theta_0 = 1$.

It does have to be traded off though with the generality of the limiting distribution of the GAR statistic, which applies under a variety of settings of the nuisance parameters, while the limiting distribution of the Wald statistic is specific to the setting of the nuisance parameters at hand, i.e. constant variance over time and mean stationarity.

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,

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

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

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

The case of $c = 0$ 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 (13). 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 $\min_\theta \mathrm{GAR}(\theta)$ 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 7 shows the simulated rejection rates for the three tests under the null while Figure 8 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 7 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 $N = 1000$ and 5000, 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 100000, 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 8 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-J 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.

To conclude this section, we would like to reiterate that, unlike HS-J and min-GAR tests, K-J test is evaluated at the true parameter value. In empirical applications, this test can be used in this form only if there is a particular parameter value that arises from some theory supporting the moment condition model specification being tested.

6 Concluding remarks

We explore how to perform inference in moment condition models that only identify the parameters locally to second order. For inference on the parameters, we consider the conventional Wald and LM statistics, and also the identification-robust GAR, KLM and GMM-M 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 distributions under both null and local alternative hypotheses. The Wald statistic is shown to have a non-standard distribution under both null and local alternatives, which depends on the convergence rate of the Jacobian, 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 GMM-M statistics have a chi-squared distribution and non-central chi squared distribution under the null and alternatives respectively. The GMM-M statistic has a conditional distribution which mimicks the limit distribution of the KLM statistic under strong identification and that of the GAR statistic under weak identification. 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 model of order one estimated from a set of non-linear moment conditions, and the conditionally heteroskedastic factor model. In the panel autoregressive model with a unit root, the autoregressive parameter is only identified at second order, and we consider the use of Wald, LM, KLM, GMM-M and GAR statistics to test whether the autoregressive coefficient is one. Our results indicate that the Wald test has the best power properties, being matched by the GAR statistic in large samples and with both these tests exhibiting greater power than the KLM, GMM-M 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 under the local alternative the finite sample distributions of the Wald, GAR, KLM and LM statistics slowly converge to their limiting distributions. We conjecture this results from the quartic root convergence rate. Nevertheless, our results show that it is possible to conduct tests with meaningful power in second-order locally identified models.

A Proofs of results in Section 3: null hypothesis

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

$$E[f(X, R\eta)] \ = \ 0. \tag{A.1}$$

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

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

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

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

and

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

Using Assumption 1(b), it is not hard to verify that (A.1) identifies η_0 at the second order. If $r = p-1$, we can apply Theorem 1(b) of Dovonon and Hall (2018) and claim that:

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

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:

$$\begin{aligned}
\mathrm{Wald}_N(\theta_0) &= N(\hat{\eta} - \eta_0)'R'\bar{q}_N(\hat{\theta})'\hat{V}_{ff}(\hat{\theta})^{-1}\bar{q}_N(\hat{\theta})R(\hat{\eta} - \eta_0) \\
&= N\left((\hat{\eta}_1 - \eta_{0,1})'R_1'\bar{q}_N(\hat{\theta})' + (\hat{\eta}_2 - \eta_{0,2})'R_2'\bar{q}_N(\hat{\theta})'\right)\hat{V}_{ff}(\hat{\theta})^{-1} \times \\
&\quad \left(\bar{q}_N(\hat{\theta})R_1(\hat{\eta}_1 - \eta_{0,1}) + \bar{q}_N(\hat{\theta})R_2(\hat{\eta}_2 - \eta_{0,2})\right).
\end{aligned} \tag{A.3}$$

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

$$\bar{q}_N(\hat{\theta}) = \bar{q}_N(\theta_0) + \bar{C}_N(\dot{\theta})\left(I_p \otimes [R(\hat{\eta} - \eta_0)]\right), \tag{A.4}$$

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

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

Under Assumption 3, $\bar{C}_N(\dot{\theta})$ 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 (A.2), the expression of $\bar{q}_N(\hat{\theta})$ in (A.4) can be written as:

$$\bar{q}_N(\hat{\theta}) = \bar{q}_N(\theta_0) + C(\theta_0)\left(I_p \otimes R_2\right)(\hat{\eta}_2 - \eta_{02}) + o_P(N^{-1/4}).$$

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 (A.3) and after some simple calculations, we obtain:

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

where $V_{ff} \equiv V_{ff}(\theta_0)$. From (A.2), 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

$$\mathbb{W}_1 = \left(V_{ff}^{-1/2}\mathbb{Z}_0 - \tilde{B}\mathbb{V}/2\right)' P \left(V_{ff}^{-1/2}\mathbb{Z}_0 - \tilde{B}\mathbb{V}/2\right) \quad \textit{and} \quad \mathbb{W}_2 = \tilde{B}' M_d \tilde{B} \mathbb{V}^2. \tag{A.5}$$

It is easily verified that

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

and

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

Thus, we have

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

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

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

with $S_1 = V_{ff}^{-1/2}\mathbb{Z}_0 \sim N(0, I_k)$ 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 $\bar{q}_N(\theta_*)\left(R_1 \,\vdots\, \sqrt{N}R_2\right)$. Note also that, by Assumption 4, we have:

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

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 (20) stays unchanged if $\hat{D}_N(\theta_0)$ is replaced by

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

From Assumption 4, we have:

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

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^2_p distribution and the independence of $\bar{\psi}_q$ and ψ_f implies that this limit is unconditionally distributed as χ^2_p. The proof of the second part of (i) is given by Kleibergen (2005) and follows from the fact that $J(\theta_*)$, $KLM(\theta_*)$ and $\hat{D}_N(\theta_*)$ converge jointly in distribution and are pairwise asymptotically independent. (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^2_{k-p} distribution and the independence of $\bar{\psi}_q$ and ψ_f implies that this limit is unconditionally distributed as χ^2_{k-p}. (ii) See the proof of Theorem 3(ii). $\square$

B Proofs of results in Section 4: local alternatives

Theorem B.1. *Let $\hat{\eta} = R^{-1}\hat{\theta}$, $\eta_N = R^{-1}\theta_N$ and write $\eta = (\eta_1', \eta_2)' \in \mathbb{R}^{p-1}\times\mathbb{R}$. If Assumptions 1, 2, 6-8 (with $\mu_N = 0$ and $\xi \geq 1/4$) hold, Assumption 5 holds under P_N and $r = p-1$ then: any subsequence of $\hat{\eta}$ has a further subsequence with index say, $s(N)$ along which*

$$\begin{pmatrix} \sqrt{N}(\hat{\eta}_1 - \eta_{N,1}) \\ N^{1/4}(\hat{\eta}_2 - \eta_{N,2}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} a(\zeta_s) \\ \zeta_s \end{pmatrix}$$

under P_N, with $a(\zeta_s)$ and ζ_s given in Theorem 5.

Proof of Theorem B.1. We proceed in two steps. First, we show that

$$\sqrt{N}(\hat{\eta}_1 - \eta_{N,1}) = O_P(1) \quad \text{and} \quad N^{1/4}(\hat{\eta}_2 - \eta_{N,2}) = O_P(1)$$

under P_N. These orders of magnitude are then used in a second step to derive the claimed asymptotic distribution.

Step 1: The identification conditions in Assumption 1 along with the maintained regularity conditions are sufficient to show that the GMM estimator $\hat{\theta}$ is consistent under P_N. See Newey and McFadden (1994). As a result, $\hat{\eta} - \eta_N \xrightarrow{P} 0$ under P_N. By definition,

$$\hat{\eta} = \arg\min_{\eta \in \{R^{-1}\theta : \theta \in \Theta\}} \bar{g}_N(\eta)' \hat{V}_{ff}(\hat{\theta}_1)^{-1} \bar{g}_N(\eta),$$

with $\bar{g}_N(\eta) = \bar{f}_N(R\eta) = \bar{f}_N(\theta)$.

By a mean-value expansion of $\eta_1 \mapsto \bar{g}_N(\eta_1, \hat{\eta}_2)$ at $\hat{\eta}_1$ around $\eta_{N,1}$ and a second-order Taylor expansion of $\eta_2 \mapsto \bar{g}_N(\eta_{N,1}, \eta_2)$ at $\hat{\eta}_2$ around $\eta_{N,2}$, we have:

$$\bar{g}_N(\hat{\eta}) = \bar{g}_N(\eta_N) + \frac{\partial \bar{g}_N}{\partial \eta_1'}(\bar{\eta}_1, \hat{\eta}_2)(\hat{\eta}_1 - \eta_{N,1}) + \frac{\partial \bar{g}_N}{\partial \eta_2}(\eta_N)(\hat{\eta}_2 - \eta_{N,2}) + \frac{1}{2}\frac{\partial^2 \bar{g}_N}{\partial \eta_2^2}(\eta_{N,1}, \bar{\eta}_2)(\hat{\eta}_2 - \eta_{N,2})^2,$$

where $\bar{\eta}_1 \in (\eta_{N,1}, \hat{\eta}_1)$ and $\bar{\eta}_2 \in (\eta_{N,2}, \hat{\eta}_2)$ and both may differ by row. Let

$$\bar{D} = \bar{q}_N(\tilde{\theta}) R_1, \quad \bar{B} = \left(R_2' \frac{\partial^2 \bar{f}_{N,i}}{\partial \theta \partial \theta'}(\bar{\theta}) R_2 \right)_{i \le i \le k} \quad \text{and} \quad W_N = \hat{V}_{ff}(\hat{\theta}_1)^{-1},$$

with $\tilde{\theta} = R(\bar{\eta}_1', \hat{\eta}_2)'$ and $\bar{\theta} = R(\eta_{N,1}', \bar{\eta}_2)'$. Thanks to Assumption 8, we have:

$$\bar{g}_N(\hat{\eta}) = \bar{f}_N(\theta_N) + \bar{D}(\hat{\eta}_1 - \eta_{N,1}) + N^{-\xi} A(\hat{\eta}_2 - \eta_{N,2}) + \frac{1}{2}\bar{B}(\hat{\eta}_2 - \eta_{N,2})^2 + o_P(N^{-1/2}) + o_P(N^{-\xi})(\hat{\eta}_2 - \eta_{N,2}).$$

By pre-multiplying by $\bar{D}'W_N$ and solving in $\hat{\eta} - \eta_{N,1}$, we have:

$$\begin{aligned} \hat{\eta}_1 - \eta_{N,1} \quad &= \quad \left(\bar{D}'W_N\bar{D}\right)^{-1} \bar{D}'W_N \left(\bar{g}_N(\hat{\eta}) - \bar{f}_N(\theta_N) - N^{-\xi}A(\hat{\eta}_2 - \eta_{N,2}) - \tfrac{1}{2}\bar{B}(\hat{\eta}_2 - \eta_{N,2})^2\right) \\ &\quad + o_P(N^{-1/2}) + o_P(N^{-\xi})(\hat{\eta}_2 - \eta_{N,2}). \end{aligned} \tag{B.1}$$

Plugging this back into the expression of $\bar{g}_N(\hat{\eta})$ above yields:

$$\begin{aligned} \bar{g}_N(\hat{\eta}) \quad &= \quad W_N^{-1/2} M_N \left(N^{-\xi}\tilde{A}_N(\hat{\eta}_2 - \eta_{N,2}) + \tfrac{1}{2}\tilde{B}_N(\hat{\eta}_2 - \eta_{N,2})^2\right) \\ &\quad + \bar{f}_N(\theta_N) + \bar{D}\left(\bar{D}'W_N\bar{D}\right)^{-1}\bar{D}'W_N\left(\bar{g}_N(\hat{\eta}) - \bar{f}_N(\theta_N)\right) \\ &\quad + o_P(N^{-\xi})(\hat{\eta}_2 - \eta_{N,2}) + o_P(N^{-1/2}), \end{aligned} \tag{B.2}$$

with $\tilde{A}_N = W_N^{1/2}A$, $\tilde{B}_N = W_N^{1/2}\bar{B}$ and $M_N = I_k - W_N^{1/2}\bar{D}(\bar{D}'W_N\bar{D})^{-1}\bar{D}'W_N^{1/2}$. By definition,

$$\bar{g}_N(\hat{\eta})'W_N\bar{g}_N(\hat{\eta}) \le \bar{f}_N(\theta_N)'W_N\bar{f}_N(\theta_N) = O_P(N^{-1}),$$

where the order of magnitude is obtained from Assumption 8. Since W_N converges to a positive definite matrix, we can claim that $\bar{g}_N(\hat{\eta}) = O_P(N^{-1/2})$. Hence, by letting $z_N = N^{1/4}(\hat{\eta}_2 - \eta_{N,2})$, we have:

$$\begin{aligned} &\left(N^{-\xi-\frac{1}{4}}\tilde{A}_N z_N + \tfrac{1}{2}N^{-\frac{1}{2}}\tilde{B}_N z_N^2\right)' M_N \left(N^{-\xi-\frac{1}{4}}\tilde{A}_N z_N + \tfrac{1}{2}N^{-\frac{1}{2}}\tilde{B}_N z_N^2\right) \\ + \quad &\left(N^{-\xi-\frac{1}{4}}\tilde{A}_N z_N + \tfrac{1}{2}N^{-\frac{1}{2}}\tilde{B}_N z_N^2\right)' M_N \left(O_P(N^{-\frac{1}{2}}) + o_P(N^{-\xi-\frac{1}{4}})z_N\right) \\ + \quad &o_P(N^{-\xi-\frac{3}{4}})z_N + o_P(N^{-2\xi-\frac{1}{2}})z_N^2 = O_P(N^{-1}). \end{aligned}$$

By multiplying each side by N and after some simple expansions, we obtain:

$$\tfrac{1}{4}\tilde{B}_N' M_N \tilde{B}_N z_N^4 \le N^{\frac{1}{2}-2\xi} O_P(1) z_N^2 + N^{\frac{1}{4}-\xi} O_P(1)|z_N|^3 + N^{\frac{1}{4}-\xi} O_P(1)|z_N| + O_P(1) z_N^2 + O_P(1).$$

Thanks to the second-order local identification condition in Assumption 1, $[D\;\;B]$ is full-column rank therefore, $\tilde{B}' M_d \tilde{B} > 0$; where $\tilde{B}$ and M_d are probability limits of $\tilde{B}_N$ and M_N, respectively. This implies that, for $\gamma = \tilde{B}' M_d \tilde{B}/4$,

$$(\gamma + o_P(1)) z_N^4 \le N^{\frac{1}{2}-2\xi} O_P(1) z_N^2 + N^{\frac{1}{4}-\xi} O_P(1)|z_N|^3 + N^{\frac{1}{4}-\xi} O_P(1)|z_N| + O_P(1) z_N^2 + O_P(1).$$

Since the right hand side of this inequality is a polynomial function of $|z_N|$ of order less than 4 and with coefficients all of magnitude $O_P(1)$ by the fact that $\xi \ge 1/4$, we can claim that $z_N = O_P(1)$, i.e. $\hat{\eta}_2 - \eta_{N,2} = O_P(N^{-\frac{1}{4}})$. Then, using (B.1), we can claim that $\hat{\eta}_1 - \eta_{N,1} = O_P(N^{-\frac{1}{2}})$.

Step 2: Let us consider the parameterization $\eta_1 = \eta_{N,1} + \frac{u_1}{\sqrt{N}}$, $\eta_2 = \eta_{N,2} + \frac{u_2}{N^{1/4}}$, $(u_1, u_2) \in \mathbb{R}^{p-1} \times \mathbb{R}$ and write

$$\bar{h}_N(u) = \bar{g}_N\left(\eta_{N,1} + \frac{u_1}{\sqrt{N}}, \eta_{N,2} + \frac{u_2}{N^{1/4}}\right) = \bar{g}_N(\eta).$$

By definition,

$$\hat{u}_N \equiv \left(\sqrt{N}(\hat{\eta}_1 - \eta_{N,1}), N^{1/4}(\hat{\eta}_2 - \eta_{N,2})\right) \in \arg\min_u N\bar{h}_N(u)' W_N \bar{h}_N(u) \left(\equiv \mathbb{H}_N(u)\right).$$

Let $Z_N = \sqrt{N}\bar{g}_N(\eta_N)$. Similar expansions to those in Step 1 of $\bar{h}_N(u)$ around 0 yield:

$$\sqrt{N}\bar{h}_N(u) = Z_N + A\mathbb{I}(\xi = 1/4)u_2 + Du_1 + \frac{1}{2}Bu_2^2 + o_P(1) \equiv h(Z_N, u) + o_P(1), \tag{B.3}$$

where the $o_P(1)$ term is uniformly negligible under P_N over $u \in \mathbb{K}$, any compact subset of $\mathbb{R}^p$. Using (B.3), we can claim that $\mathbb{H}_N(u) = \mathbb{H}(Z_N, u) + o_P(1)$, with

$$\mathbb{H}(Z_N, u) = h(Z_N, u)' V_{ff}(\theta_0)^{-1} h(Z_N, u)$$

where the $o_P(1)$ term is uniformly negligible under P_N over any compact subset. Hence, if $\bar{u}_N$ and $\tilde{u}_N$ belong to the union of the sets of arguments of the minimum of $\mathbb{H}_N(u)$ and $\mathbb{H}(Z_N, u)$, and both are asymptotically stochastically bounded, we can claim that

$$\mathbb{H}_N(\bar{u}_N) = \mathbb{H}(Z_N, \tilde{u}_N) + o_P(1). \tag{B.4}$$

Note that, from Step 1, any sequence in the set of arguments of the minimum of $\mathbb{H}_N(u)$ is $O_P(1)$ and we can also proceed by the same way to show that any sequence in the set of arguments of the minimum of $\mathbb{H}(Z_N, u)$ is $O_P(1)$.

We can proof along similar lines to Dovonon and Renault (2013b, Lemma B.6) that:

$$\mathbb{H}(Z_N, \bullet) \xrightarrow{d} \mathbb{H}(Z, \bullet)$$

under P_N, uniformly over any compact subset, where $Z \sim N(0, V_{ff}(\theta_0))$. We can proceed along similar lines to Step 1 to show that the arguments of the minimum of $u \mapsto \mathbb{H}(Z, u)$ are all stochastically bounded. Therefore, by, Lemma B.5 of Dovonon and Renault (2013b), we can claim that

$$\min_u \mathbb{H}(Z_N, u) \xrightarrow{d} \mathbb{H}(Z, u).$$

Since $\hat{u}_N = O_P(1)$, any subsequence of $\hat{u}_N$ has a further subsequence, say $\hat{u}_{s(N)}$ along which, we have:

$$\begin{pmatrix} \min_u \mathbb{H}(Z_N, u) \\ Z_N \\ \hat{u}_N \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \min_u \mathbb{H}(Z, u) \\ Z \\ \hat{u}_s \end{pmatrix},$$

under P_N. From (B.4), we have $\mathbb{H}(Z_N, \hat{u}_N) = \min_u \mathbb{H}(Z_N, u) + o_P(1)$ and we can claim by the continuous mapping theorem that:

$$\min_u \mathbb{H}(Z, u) - \mathbb{H}(Z, \hat{u}_s) = 0,$$

that is $\hat{u}_s \in \arg\min_u \mathbb{H}(Z, u)$. Exploiting the first-order condition of this program in the direction of u_1, we can claim that any $(\hat{u}_1, \hat{u}_2) \in \arg\min_u \mathbb{H}(Z, u)$ can be written:

$$\hat{u}_1 = -(D'WD)^{-1}D'W\left(Z + A\mathbb{I}(\xi = 1/4)\hat{u}_2 + \frac{1}{2}B\hat{u}_2\right) = a(\hat{u}_2)$$

and

$$\hat{u}_2 \in \arg\min_{u \in \mathbb{R}} \left(Z + A\mathbb{I}(\xi = 1/4)u + \frac{1}{2}Bu^2\right)' W^{1/2} M_d W^{1/2} \left(Z + A\mathbb{I}(\xi = 1/4)u + \frac{1}{2}Bu^2\right),$$

with $W = V_{ff}(\theta_0)^{-1}$ $\square$

Proof of Theorem 5. 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

$$\bar{q}_N(\hat{\theta})R_1 = D + o_P(1), \quad \textit{and} \quad N^{1/4}\bar{q}_N(\hat{\theta})R_2 = A\mathbb{I}(\xi = 1/4) + BN^{1/4}(\hat{\eta}_2 - \eta_{N,2}) + o_P(1).$$

It follows that

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

with

$$W_{N,a} = \sqrt{N}(\hat{\eta}_1 - \eta_{0,1})' \tilde{D}' \tilde{D} \sqrt{N}(\hat{\eta}_1 - \eta_{0,1}),$$
$$W_{N,b} = 2\sqrt{N}(\hat{\eta}_1 - \eta_{0,1})' \tilde{D}' \left(\tilde{A} + \tilde{B}N^{1/4}(\hat{\eta}_2 - \eta_{N,2})\right) N^{1/4}(\hat{\eta}_2 - \eta_{0,2}),$$

and

$$W_{N,c} = \left(\tilde{A} + \tilde{B}N^{1/4}(\hat{\eta}_2 - \eta_{N,2})\right)' \left(\tilde{A} + \tilde{B}N^{1/4}(\hat{\eta}_2 - \eta_{N,2})\right) \sqrt{N}(\hat{\eta}_2 - \eta_{0,2})^2,$$

where (as before) $\tilde{D} = V_{ff}(\theta_0)^{-1/2}D$ and $\tilde{B} = V_{ff}(\theta_0)^{-1/2}B$. The expected result follows easily using the asymptotic distribution given by Theorem B.1 for the relevant subsequences of $\left(\sqrt{N}(\hat{\eta}_1 - \eta_{N,1}), N^{1/4}(\hat{\eta}_2 - \eta_{N,2})\right)$. $\square$

Proof of Proposition 1. (a) Let us assume that $P(\mathbb{W}_s(e_2) = 0) > 0$ for all $e_2 \in \mathbb{R}$ and $A \neq 0$ and show that $A = \delta B$ for some $\delta \neq 0$. We have: $\mathbb{W}_s(e_2) = 0$ if and only if $\tilde{A} + \tilde{B}\zeta_s = 0$ or $\zeta_s + e_2 = 0$. That is

$$P(\mathbb{W}_s(e_2) = 0) > 0 \Leftrightarrow P(\tilde{A} + \tilde{B}\zeta_s = 0) > 0 \text{ or } P(\zeta_s = -e_2) > 0.$$

Note that $P(\zeta_s = -e_2) > 0$ can hold only for at most a countably infinite number of values e_2. Thus for $P(\mathbb{W}_s(e_2) = 0) > 0$ to hold for all $e_2 \in \mathbb{R}$, we necessarily have $P(\tilde{A} + \tilde{B}\zeta_s = 0) > 0$. This means that there exists $\delta \neq 0$ such that $\tilde{A} = \delta\tilde{B}$ and $P(\zeta_s = -\delta) > 0$. This implies in particular that $A = \delta B$. The proofs of (b) and (c) below also establish the converse.

(b) If $A = 0$, simple derivations show that

$$\mathbb{W}_s(e_2) = \tilde{B}'\tilde{B}\zeta_s^2(\zeta_s + e_2)^2 \quad \text{with} \quad \zeta_s^2 = -\tilde{B}'\tilde{X}\mathbb{I}(\tilde{B}'\tilde{X} < 0)/\tilde{B}'\tilde{B} \quad \text{and} \quad \tilde{X} \sim N(0, I_k).$$

Thus, for all $e_2 \in \mathbb{R}$, $P(\mathbb{W}_s(e_2) = 0) \geq P(\zeta_s = 0) = 1/2$.

(c) If $A = \delta B$, $\delta \neq 0$, by some simple derivations, we have

$$\mathbb{W}_s(e_2) = \tilde{B}'\tilde{B}(\zeta_s + e_2)^2(\zeta_s + \delta)^2 \quad \text{and} \quad (\zeta_s + \delta)^2 = -\frac{B_0'X_0}{B_0'B_0}\mathbb{I}(B_0'X_0 < 0),$$

with $B_0 = \tilde{B}/2$ and $X_0 = \tilde{X} - (\delta^2/2)\tilde{B}$. Hence, for any $e_2 \in \mathbb{R}$,

$$P(\mathbb{W}_s(e_2) = 0) \geq P(B_0'X_0 \geq 0) = 1 - \Phi\left((\delta^2/2)\sqrt{\tilde{B}'\tilde{B}}\right). \quad \square$$

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 $\bar{q}_N(\theta_*)\Delta$ with Δ any nonsingular matrix. Here, we replace $\bar{q}_N(\theta_*)$ by $\bar{q}_N(\theta_*)\left(R_1 \;\; N^{1/4}R_2\right)$. A first-order mean value expansion of $\bar{q}_N(\theta_*)$ around θ_N similar to (A.4) gives:

$$\bar{q}_N(\theta_*) = \bar{q}_N(\theta_N) + \bar{C}_N(\dot{\theta})\left[I_p \otimes (\theta_* - \theta_N)\right] = \bar{q}_N(\theta_N) - \bar{C}_N(\dot{\theta})\left[I_p \otimes (R_1 e_{N,1} + R_2 e_{N,2})\right],$$

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 (A.4). 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,

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

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

$$N^{1/4}\bar{q}_N(\theta_*)R_2 = -C(\theta_*)[I_p \otimes (R_2 e_2)]R_2 + A\mathbb{I}(\xi = 1/4) + o_P(1).$$

Thus

$$\bar{q}_N(\theta_*)\left(R_1, \, N^{1/4}R_2\right) = \mathbb{Q}(e_2) + o_P(1).$$

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

$$\begin{aligned}
\bar{f}_N(\theta_*) &= \bar{f}_N(\theta_N) - \bar{q}_N(\theta_N)(\theta_N - \theta_*) + \tfrac{1}{2}\left[(\theta_N - \theta_*)' \otimes I_k\right] \bar{h}_N(\bar{\theta})(\theta_N - \theta_*), \\
&= \bar{f}_N(\theta_N) - \bar{q}_N(\theta_N) R_1 e_{N,1} - \bar{q}_N(\theta_N) R_2 e_{N,2} \\
&\quad + \tfrac{1}{2}\left[(R_1 e_{N,1} + R_2 e_{N,2})' \otimes I_k\right] \bar{h}_N(\bar{\theta})(R_1 e_{N,1} + R_2 e_{N,2}) \\
&= \bar{f}_N(\theta_N) - N^{-1/2} D e_1 - N^{-\xi - 1/4} A e_2 + \tfrac{1}{2} N^{-1/2}\left((R_2 e_2)' \otimes I_k\right) H(\theta_*)(R_2 e_2) \\
&\quad + o_P(N^{-1/2}) + o(N^{-\xi - 1/4}),
\end{aligned}$$

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 $\bar{q}_N(\theta_N) R_2 = N^{-\xi} A + o_P(N^{-1/4})$ under P_N. Thus $\sqrt{N}(V_{ff}(\theta_*)^{-1/2} \bar{f}_N(\theta_*))$ converges in distribution under P_N to $N(\mu_\theta, I_k)$ with

$$\mu_\theta = V_{ff}(\theta_*)^{-1/2}\left(-De_1 + \frac{1}{2}\left((R_2 e_2)' \otimes I_k\right) H(\theta_*)(R_2 e_2) - A e_2 \mathbb{I}(\xi = 1/4)\right).$$

Alternatively, we can write

$$\mu_\theta = -\mathbb{Q}(e_2)\begin{pmatrix} e_1 \\ \frac{1}{2} e_2 \end{pmatrix} - \frac{1}{2} V_{ff}(\theta_*)^{-1/2} A e_2 \mathbb{I}(\xi = 1/4).$$

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

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

Hence, we can claim that $LM(\theta_*)$ converges in distribution under P_N to $\chi^2_p(\lambda_\theta)$, with $\lambda_\theta = \mu_\theta' \mathbb{P}(e_2) \mu_\theta$.

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).

If $\xi > 1/4$ or $A = 0$, we can use the identity : $(e' \otimes I_k) H(\theta_*) e = C(\theta_*)(I_p \otimes e) e$, for all $e \in \mathbb{R}^p$ to show that $\mu_\theta = -\mathbb{Q}(e_2) \times \begin{pmatrix} e_1 \\ \frac{1}{2} e_2 \end{pmatrix}$. As a result, $\lambda_\theta = \mu_\theta' \mu_\theta$. Note also that the second-order identification condition in Assumption 1(b), ensures that $\mu_\theta \neq 0$ if $e \neq 0$. The latter is warranted by the full-rank condition on $\mathbb{Q}(e_2)$.

If $\xi = 1/4$ and $A \neq 0$, to conclude that $\lambda_\theta \neq 0$, it suffices to show that $\mathbb{P}(e_2)\mu_\theta \neq 0$. This latter holds if and only if

$$\mathbb{P}(e_2) V_{ff}(\theta_*)^{-1/2} A e_2 \neq \mathbb{Q}(e_2)\begin{pmatrix} 2e_1 \\ e_2 \end{pmatrix}.$$

The asymptotic distribution of GMM-M(θ_*) under P_N is obtained along similar lines as in the proof of Theorem 3. The asymptotic independence of $KLM(\theta_*)$, $J(\theta_*)$ and $\hat{D}_N(\theta_*)$ continues to hold under P_N with $KLM(\theta_*)$ and $J(\theta_*)$ asymptotically distributed as $\tilde{\psi}_K$ and $\tilde{\psi}_J$, respectively. To obtain the stochastic dominance claimed, note that

$$(\tilde{\psi}_J + \tilde{\psi}_K + \rho)^2 - 4\rho\tilde{\psi}_J = (\rho - \tilde{\psi}_J)^2 + \tilde{\psi}_K^2 + 2\tilde{\psi}_J\tilde{\psi}_K + 2\rho\tilde{\psi}_K.$$

Since $\tilde{\psi}_J$ and $\tilde{\psi}_K$ are nonnegative, this quantity is larger than or equal to

$$(\rho - \tilde{\psi}_J)^2 + \tilde{\psi}_K^2 + 2(\rho - \tilde{\psi}_J)\tilde{\psi}_K = (\rho - \tilde{\psi}_J + \tilde{\psi}_K)^2$$

and the result follows.

(b) From the asymptotic distribution of $\sqrt{N}\bar{f}_N(\theta_*)$ derived above, it is obvious that $\text{GAR}(\theta_*)$ converges in distribution under P_N to $\chi^2_k(\ell_\theta)$, where $\ell_\theta = \mu'_\theta \mu_\theta$.

If $\xi > 1/4$ or $A = 0$, $\mu_\theta \neq 0$ for any $e \neq 0$ as already shown. If $\xi = 1/4$ and $A \neq 0$, it is not hard to see that if $A \notin \left\langle D \,\vdots\, -C(\theta_*)\left[I_p \otimes (R_2 e_2)\right] R_2 \right\rangle$, $\;-De_1 + \frac{1}{2}\left((R_2e_2)' \otimes I_k\right) H(\theta_*)(R_2e_2) - Ae_2 \neq 0$ for all $e \neq 0$. Hence $\mu_\theta \neq 0$ for all $e \neq 0$. $\square$

Proof of Theorem 7. (i) Note that

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

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

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,

$$\sqrt{N}\bar{f}_N(\theta_0) \xrightarrow{d} \psi_f + c \sim N(c, V_{ff}(\theta_0)) \tag{B.5}$$

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,

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

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

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

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

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

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

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

under P_N with random non-centrality parameter $\lambda_m = c' V_{ff}(\theta_0)^{-1/2} M_{V_{ff}(\theta_0)^{-1/2}\bar{\psi}^a_q} V_{ff}(\theta_0)^{-1/2} c$. Clearly, if $c \notin \langle \bar{\psi}^a_q \rangle$ almost surely, then $\lambda_m > 0$ almost surely. (ii) Follows readily from (B.5). $\square$

C Figures

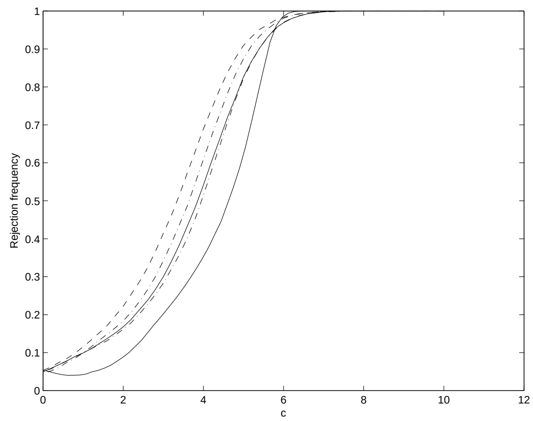

Figure 1: Local power curve of 5% tests of $H_0 : \theta = 1$ while the true value of $\theta = 1 - \frac{c}{2\sqrt[4]{N}}$ using the Wald statistic: Dashed: $N = 50$; Dash-dot: 500; Solid: 5000; Dash-dot: 20000 and Solid: asymptotic. For each type of line the one associated with the smaller sample size lies to the left.

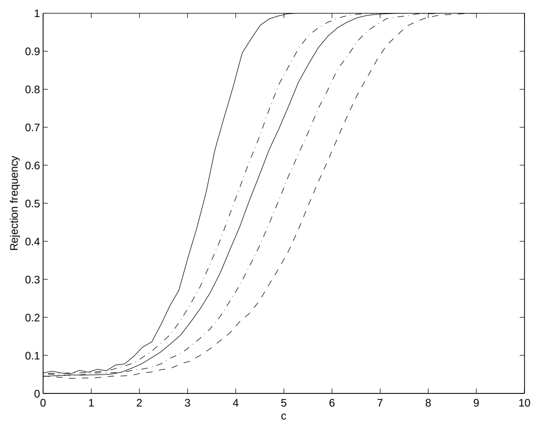

Figure 2: Local power curve of 5% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt[4]{N}}$ using the GAR statistic: Dashed: $N = 50$; Dash-dot: 500; Solid: 5000; Dash-dot: 20000 and Solid: asymptotic. For each type of line the one associated with the smaller sample size lies to the right.

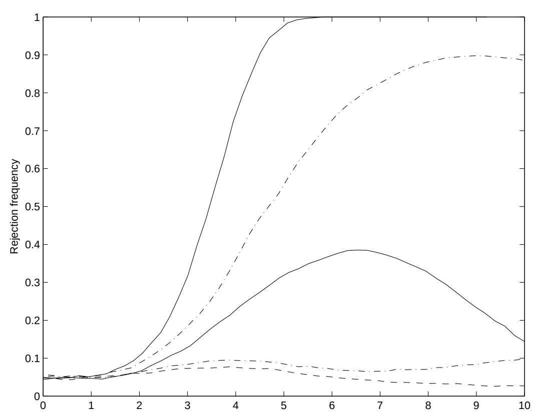

Figure 3: Local power curve of 5% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt[4]{N}}$ using the KLM statistic: Dashed: $N = 50$; Dash-dot: 500; Solid: 5000; Dash-dot: 20000 and Solid: asymptotic. For each type of line the one associated with the smaller sample size lies to the right.

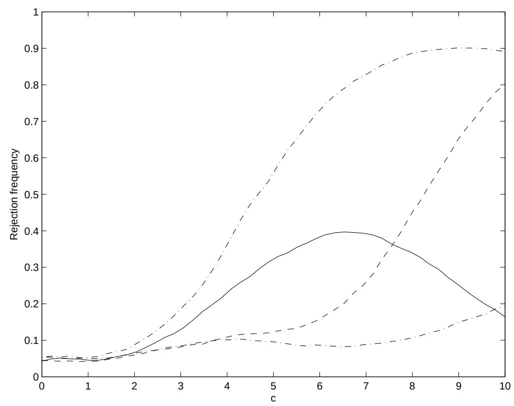

Figure 4: Local power curve of 5% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt[4]{N}}$ using the GMM-M statistic: Dashed: $N = 50$; Dash-dot: 500; Solid: 5000 and Dash-dot: 20000. For each type of line the one associated with the smaller sample size lies to the right.

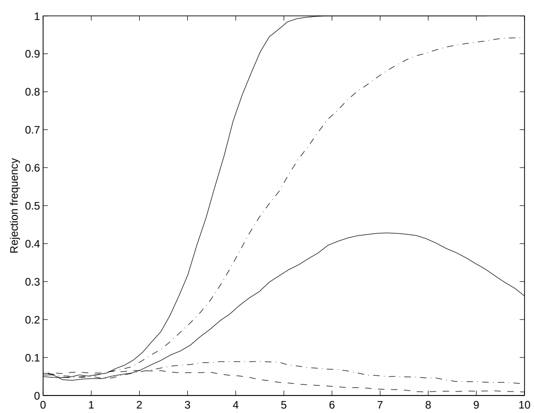

Figure 5: Local power curve of 5% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt[4]{N}}$ using the LM statistic: Dashed: $N = 50$; Dash-dot: 500; Solid: 5000; Dash-dot: 20000 and Solid: asymptotic. For each type of line the one associated with the smaller sample size lies to the right.

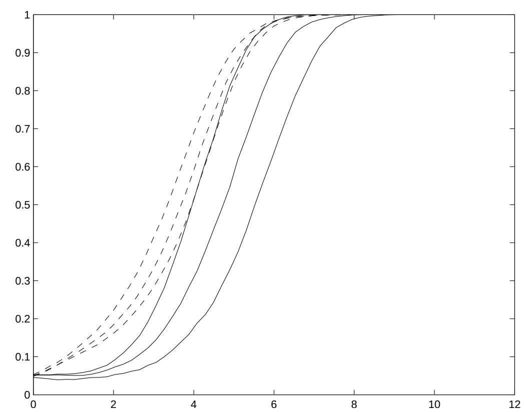

Figure 6: Local power curve of 5% tests of $H_0 : \theta = 1$ while true value of $\theta = 1 - \frac{c}{2\sqrt[4]{N}}$ using the Wald (dashed) and GAR statistics (solid) for $N = 50$, 1000, 20000. For the Wald (GAR) smaller sample sizes lie to the left (right).

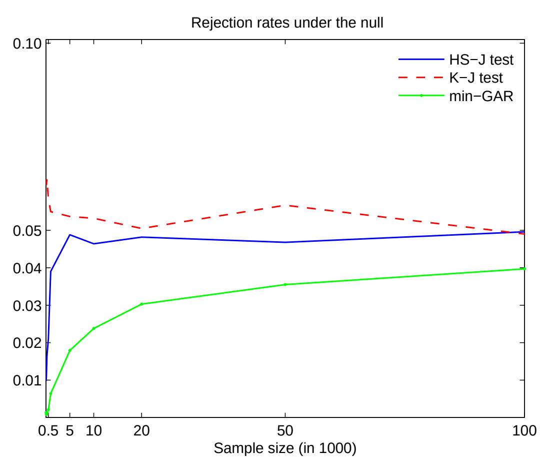

Figure 7: Rejection rates of the HS-J, K-J and min-GAR tests under the null; $10,000$ replications; $c = 0$. ($\alpha = 0.05$)

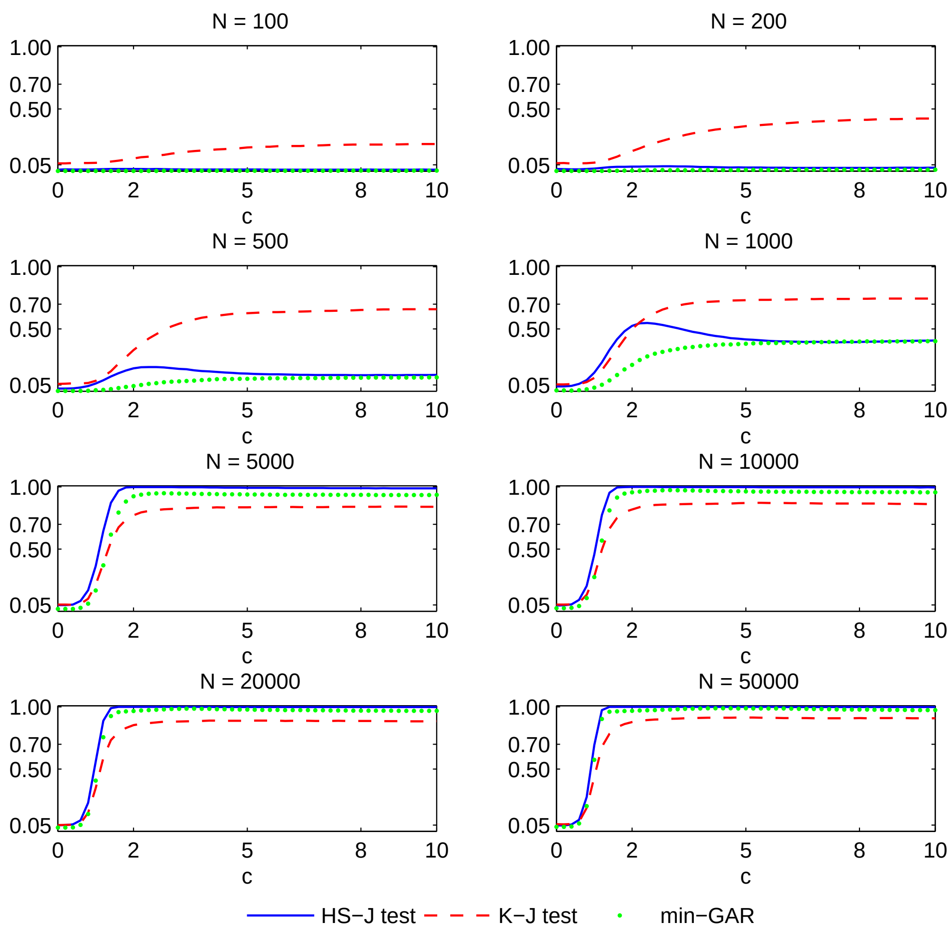

Figure 8: 10,000 replications; $c = 0 : 0.2 : 10$ ($\alpha = 0.05$)

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.

——— (2001). 'Testing when a parameter is on the boundary of the maintained hypothesis', *Econometrica*, 69: 683–734.

——— (2002). 'Generalized Method of Moments estimation when a parameter is on a boundary', *Journal of Business and Economic Statistics*, 20: 530–544.

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', *Scandanavian 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. (2018). 'The asymptotic properties of GMM and Indirect Inference under second-order identification', *Journal of Econometrics*, 205: 76–111.

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.

Gaffke, N., Heiligers, B., and Offinger, R. (1999). 'On the asymptotic null-distribution of the Wald statistic at singular parameter points', *Statistics & Decisions*, 20: 379–357.

Gaffke, N., Steyer, R., and von Davier, A. A. (1999). 'On the asymptotic null-distribution of the Wald statistic at singular parameter points', *Statistics & Decisions*, 17: 339–358.

Gourieroux, C., and Montfort, A. (1995). *Statistics and Econometric Models*. Cambridge University Press.

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.

Lee, L.-F. (1993). 'Asymptotic distribution of the Maximum Likelihood estimator for a stochastic frontier function model with a singular information matrix', *Econometric Theory*, 9: 413–430.

Lee, L.-F., and Chesher, A. (1986). 'Specification testing when score tests are identically zero', *Journal of Econometrics*, 31: 121–149.

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.

D Supplementary appendix

This appendix contains supplementary material that will be made available on line and is not to be included in the paper.

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

$$E(a) = \sigma^2 \begin{pmatrix} 1 + \frac{c^2}{4\sqrt{N}} \\ \\ 0 \end{pmatrix}, \quad E(b) = -\sigma^2 \begin{pmatrix} 2 - \frac{c}{\sqrt[4]{N}} + \frac{5c^2}{4\sqrt{N}} + \frac{c^4}{16N} - \frac{c^3}{2N^{3/4}} \\ \\ -\frac{c}{2\sqrt[4]{N}} + \frac{c^2}{4\sqrt{N}} - \frac{c^3}{8N^{3/4}} \end{pmatrix} \tag{D.1}$$

$$E(d) = \sigma^2 \begin{pmatrix} 1 - \frac{c}{\sqrt[4]{N}} + \frac{5c^2}{4\sqrt{N}} - \frac{7c^3}{8n^{3/4}} + \frac{c^4}{4N} - \frac{c^5}{32N^{5/4}} \\ \\ -\frac{c}{2\sqrt[4]{N}} + \frac{c^2}{2\sqrt{N}} - \frac{c^3}{4N^{3/4}} + \frac{c^4}{16N} \end{pmatrix}$$

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

$$m_N(\theta_N) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; G_N(\theta_N) = \sigma^2 \begin{pmatrix} -\frac{3c^2}{4\sqrt{N}} + \frac{c^3}{4n^{3/4}} - \frac{c^4}{16N} \\ \frac{c}{2N^{1/4}} - \frac{c^2}{4\sqrt{N}} + \frac{c^3}{8N^{3/4}} \end{pmatrix}, \; H_N(\theta_N) = 2\sigma^2 \begin{pmatrix} 1 + \frac{c^2}{4\sqrt{N}} \\ 0 \end{pmatrix} \tag{D.2}$$

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

$$m_N(\theta_0) = \frac{\sigma^2 c^2}{4\sqrt{N}} \begin{pmatrix} 1 - \frac{3c}{2N^{1/4}} + \frac{3c^2}{4\sqrt{N}} - \frac{c^3}{8N^{3/4}} \\ 1 - \frac{c}{2N^{1/4}} + \frac{c^2}{4\sqrt{N}} \end{pmatrix}, \; G_N(\theta_0) = \frac{\sigma^2 c}{N^{1/4}} \begin{pmatrix} 1 - \frac{3c}{4N^{1/4}} + \frac{c^2}{2\sqrt{N}} - \frac{c^3}{16N^{3/4}} \\ \frac{1}{2} - \frac{c}{4N^{1/4}} + \frac{c^2}{8\sqrt{N}} \end{pmatrix},$$

$$H_N(\theta_0) = H_N(\theta_N). \tag{D.3}$$

Therefore, 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 $\theta_N = 1 - \frac{c}{2\sqrt{N}}$, the expected values of a, b and d equal

$$E[a] = \sigma^2 \begin{pmatrix} 1 + \frac{c^2}{4N} \\ 0 \end{pmatrix}, \; E[b] = \sigma^2 \begin{pmatrix} -2 - \frac{c^4}{16N^2} + \frac{c^3}{2N^{3/2}} - \frac{5c^2}{4N} + \frac{c}{\sqrt{N}} \\ \frac{c^3}{8N^{3/2}} - \frac{c^2}{4N} + \frac{c}{2\sqrt{N}} \end{pmatrix} \tag{D.4}$$

$$E[d] = \sigma^2 \begin{pmatrix} 1 - \frac{c^5}{32N^{5/2}} + \frac{c^4}{4N^2} - \frac{7c^3}{8N^{3/2}} + \frac{5c^2}{4N} - \frac{c}{\sqrt{N}} \\ \frac{c^4}{16N^2} - \frac{c^3}{4N^{3/2}} + \frac{c^2}{2N} - \frac{c}{2\sqrt{N}} \end{pmatrix},$$

so

$$E_N[f(\theta_0)] = \sigma^2 \begin{pmatrix} -\frac{c^5}{32N^{5/2}} + \frac{3c^4}{16N^2} - \frac{3c^3}{8N^{3/2}} + \frac{c^2}{4N} \\ \frac{c^4}{16N^2} - \frac{c^3}{8N^{3/2}} + \frac{c^2}{4N} \end{pmatrix}, \tag{D.5}$$

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

Derivation of equation (39). We first derive equation (37). Using $\theta_0'\lambda_1 = 0$, we have under P_N that:

$$E_N\left[\theta_0' Y_{t+1} Y_{t+1}' \theta_0 | \mathfrak{F}_t\right] = \left(\theta_0'\lambda_{2N}\right)^2 \sigma_{2,t}^2 + \theta_0'\Omega\theta_0. \tag{D.6}$$

As in Section 2.2, let z_t be a relevant vector of instruments, then it follows from (D.6) that

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

which, together, imply

$$Cov\left(z_t, (\theta_0'Y_{t+1})^2\right) = (\theta_0'\lambda_{2N})^2 Cov(z_t, \sigma_{2,t}^2), \tag{D.7}$$

where $Cov[\,\cdot,\cdot\,]$ is relative to P_N. Using (14) and (D.7), we obtain (37). 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 (11)-(12):

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

where $Var(F_{t+1}|\mathfrak{F}_t) = D_t$, $Cov(F_{t+1}, U_{t+1}|\mathfrak{F}_t) = 0$, $E(F_{t+1}|\mathfrak{F}_t) = 0$, $E(U_{t+1}|\mathfrak{F}_t) = 0$. 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:

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

and so, using $z_t = (Y_{1t}^2, Y_{2t}^2)'$, obtain (38). Combining these results with (15), we obtain (39). $\square$

Derivation of equations (30), (41) and (42)

We consider the behaviour of the Wald test under P_N with $\theta_N = 1 - \frac{c}{2\sqrt[4]{N}}$. Assume that $N^{1/2}\left(a - E_N[a], b - E_N[b], c - E_N[c]\right)' \xrightarrow{d} (\psi_a, \psi_b, \psi_d)'$ 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:

$$\sqrt{N}\left[(b+2a) - \frac{c\sigma^2}{\sqrt[4]{N}}\begin{pmatrix}1\\ \frac{1}{2}\end{pmatrix}\right] \underset{d}{\to} \frac{1}{4}c^2\sigma^2\begin{pmatrix}-3\\ -1\end{pmatrix} + \psi_b + 2\psi_a \tag{D.8}$$

$$\sqrt{N}\bar{f}_N(1) = \sqrt{N}\left(a+b+d\right) \underset{d}{\to} \frac{1}{4}c^2\sigma^2\begin{pmatrix}1\\ 1\end{pmatrix} + \psi_a + \psi_b + \psi_d. \tag{D.9}$$

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 (D.3), it follows that under P_N, we have:

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

and

$$\sqrt[4]{N}\bar{q}_N(1) \xrightarrow{p} \sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}. \tag{D.10}$$

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

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

and so[19]

$$\sqrt[4]{N}\bar{q}_N(\hat{\theta}) \overset{a}{=} \sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2 c \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta, \tag{D.11}$$

where, as in the text, $\sqrt[4]{N}(\hat{\theta}_{1,s} - 1) = \zeta + o_p(1)$. Similarly, using (D.8), we have

$$\sqrt{N}\bar{f}_N(\hat{\theta}) \overset{a}{=} \psi \,+\, \frac{c^2}{4}\sigma^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \,+\, \sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta^2 + \sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} \zeta, \tag{D.12}$$

with $\psi = \psi_a + \psi_b + \psi_d$. Combining (D.11)-(D.12) with the first order condition for one step GMM, it can be seen that ζ is implicitly characterized by:

$$\left[\sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta\right]' \left[\psi \,+\, \frac{c^2}{4}\sigma^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \,+\, \sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta^2 + \sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} \zeta\right] \;=\; 0, \tag{D.13}$$

which can be re-written as

$$c_1\zeta^3 + c_2\zeta^2 + c_3\zeta + c_4 \;=\; 0, \tag{D.14}$$

with

$$\begin{aligned}
c_1 &= \; 2\sigma^4 \\
c_2 &= \; 2\sigma^4 c \begin{pmatrix} 1 \\ 0 \end{pmatrix}' \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + \sigma^4 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}' \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \; 3\sigma^4 c \\
c_3 &= \; 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}' \left[\psi \,+\, \tfrac{c^2}{4}\sigma^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] + \sigma^4 c^2 \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}' \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} \\
&= \; 2\sigma^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}' \psi + \tfrac{7}{4}c^2\sigma^4 \\
c_4 &= \; \sigma^2 c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}' \left[\psi \,+\, \tfrac{c^2}{4}\sigma^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] \\
&= \; \sigma^2 c \left[\begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}' \psi + \tfrac{3}{8}c^2\sigma^2\right].
\end{aligned}$$

[19]Using $\overset{a}{=}$ to denote equality up to $o_p(1)$.

The first order condition results from the fourth order polynomial, which is just the quadratic form of the moment equation, where the zero-th term equals the quadratic form of $\left[\psi + \frac{c^2}{4}\sigma^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right]$.

When $c = 0$, the above polynomial simplifies to

$$\begin{array}{l} c_1\zeta^3 + c_3\zeta = 0 \quad \Leftrightarrow \\ \zeta(\zeta^2 + \frac{c_3}{c_1}) = 0 \end{array} \tag{D.15}$$

with $\frac{c_3}{c_1} = \frac{\psi_1}{\sigma^2}$, so we can specify ζ as

$$\zeta = \mathbb{I}(\psi_1 < 0)\frac{1}{\sigma}\sqrt{|\psi_1|}.$$

For values of c different from zero, we have to solve for the root of the third order polynomial that minimizes the fourth order polynomial numerically or use an explicit for the roots of third order polynomials.

The expression for the one step Wald statistic is

$$\begin{aligned} \text{Wald}^*_N(\hat{\theta}) = \;& N(\hat{\theta} - 1)\left(\bar{q}_N(\hat{\theta})'\bar{q}_N(\hat{\theta})\right)\left[\bar{q}_N(\hat{\theta})'\hat{V}_{ff}(\hat{\theta})\bar{q}_N(\hat{\theta})\right]^{-1}\left(\bar{q}_N(\hat{\theta})'\bar{q}_N(\hat{\theta})\right)(\hat{\theta} - 1) \\ = \;& \left[\sqrt[4]{N}(\hat{\theta} - 1)\right]\left(\left(\sqrt[4]{N}\bar{q}_N(\hat{\theta})\right)'\left(\sqrt[4]{N}\bar{q}_N(\hat{\theta})\right)\right)\left[\left(\sqrt[4]{N}\bar{q}_N(\hat{\theta})\right)'\hat{V}_{ff}(\hat{\theta})\left(\sqrt[4]{N}\bar{q}_N(\hat{\theta})\right)\right]^{-1} \\ & \left(\left(\sqrt[4]{N}\bar{q}_N(\hat{\theta})\right)'\left(\sqrt[4]{N}\bar{q}_N(\hat{\theta})\right)\right)\left[\sqrt[4]{N}(\hat{\theta} - 1)\right] \\ \underset{d}{\rightarrow} \;& \zeta^2\frac{\left[\left(\sigma^2 c\begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2\begin{pmatrix} 1 \\ 0 \end{pmatrix}\zeta\right)'\left(\sigma^2 c\begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2\begin{pmatrix} 1 \\ 0 \end{pmatrix}\zeta\right)\right]^2}{\left(\sigma^2 c\begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2\begin{pmatrix} 1 \\ 0 \end{pmatrix}\zeta\right)' V_{ff}(1)\left(\sigma^2 c\begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 2\sigma^2\begin{pmatrix} 1 \\ 0 \end{pmatrix}\zeta\right)}. \end{aligned}$$

To derive, l_θ and λ_θ in Theorem 6 from (42), we first consider the GAR statistic:

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

Using (D.8), it follows that

$$GAR(1) \;\xrightarrow{d}\; (\psi + \frac{c^2}{4}\sigma^2\begin{pmatrix} 1 \\ 1 \end{pmatrix})'V_{ff}(1)^{-1}(\psi + \frac{c^2}{4}\sigma^2\begin{pmatrix} 1 \\ 1 \end{pmatrix}) \;=\; \chi^2_2(l_\theta)$$

with $l_\theta = \frac{c^2\sigma^2}{16}\binom{1}{1}'V_{ff}(1)^{-1}\binom{1}{1}$ and also stated in (42). Since

$$\hat{D}_N(1) = \bar{q}_N(1) - \hat{V}_{qf}(1)\hat{V}_{ff}(1)^{-1}\bar{f}_N(1),$$

and $\bar{q}_N(1) = O_p(N^{-1/4})$ and $\bar{f}_N(1) = O_p(N^{-1/2})$, the limit behavior of $\hat{D}_N(1)$ is identical to that of $\bar{q}_N(1)$ which is stated in (D.10). The non-centrality parameters of the non-central $\chi^2(1)$ asymptotic distributions of the KLM and LM statistics under local alternatives therefore coincide:

$$\lambda_\theta = \tfrac{c^2\sigma^2}{16}\tbinom{1}{1}'V_{ff}(1)^{-1}\tbinom{1}{1/2}\left[\tbinom{1}{1/2}'V_{ff}(1)^{-1}\tbinom{1}{1/2}\right]^{-1}\tbinom{1}{1/2}V_{ff}(1)^{-1/2}\tbinom{1}{1}.$$