Testing for Common GARCH Factors

Dovonon, Prosper and Renault, Eric

Concordia University, CIREQ

June 2011

Testing for Common GARCH Factors*

Prosper DOVONON[†] and Éric RENAULT[‡]

June 6, 2011

Abstract

This paper proposes a test for common GARCH factors in asset returns. Following Engle and Kozicki (1993), the common GARCH factors property is expressed in terms of testable overidentifying moment restrictions. However, as we show, these moment conditions have a degenerate Jacobian matrix at the true parameter value and therefore the standard asymptotic results of Hansen (1982) do not apply. We show in this context that the Hansen's (1982) J-test statistic is asymptotically distributed as the minimum of the limit of a certain empirical process with a markedly nonstandard distribution. If two assets are considered, this asymptotic distribution is a half-half mixture of χ^2_{H-1} and χ^2_H, where H is the number of moment conditions, as opposed to a χ^2_{H-1}. With more than two assets, this distribution lies between the χ^2_{H-p} and χ^2_H (p, the number of parameters) and both bounds are conditionally sharp. These results show that ignoring the lack of first order identification of the moment condition model leads to oversized tests with possibly increasing over-rejection rate with the number of assets. A Monte Carlo study illustrates these findings.

Keywords: GARCH factors, Nonstandard asymptotics, GMM, GMM overidentification test, identification, first order identification.

1 Introduction

Engle and Kozicki (1993) have given many examples of the following interesting question : are some features that are detected in several single economic time series actually *common* to all of them? Following their definition, "a feature will be said to be *common* if a linear combination of the series fails to have the feature even though each of the series individually has the feature". They propose testing procedures to determine whether features are common. The null hypothesis under test is the existence of common features. As nicely examplified by Engle and Kozicki (1993), an unified testing framework is provided by the Hansen (1982) J-test for overidentification in the context of Generalized Method of Moments (GMM). Under the null, the J-test statistic is supposed to have a limiting chi-square distribution with degrees of freedom equal to the number of overidentifying restrictions. After normalization, a common feature to n individual time series is defined by a vector of $(n-1)$ unknown

*We would like to thank the co-editor (James Stock), Manuel Arellano, Yves Atchadé, Valentina Corradi, Giovanni Forchini, Sílvia Gonçalves and Enrique Sentana for very helpful comments and suggestions.

†Concordia University, CIRANO and CIREQ; Address: Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada; tel: (514) 848-2424 (ext.3479), fax: (514) 848-4536, email: prosper.dovonon@concordia.ca.

‡University of North Carolina at Chapel Hill (USA), CIRANO and CIREQ (Canada); Email: renault@email.unc.edu.

parameters and the limiting distribution under the null will be $\chi^2(H-n+1)$ where H stands for the number of moment restrictions deduced from the common features property. Engle and Kozicki (1993) successfully apply this testing strategy to several common features issues of interest (regression common feature, cofeature rank, Granger causality and cointegration). When they come to the common GARCH features, they acknowledge that it is their first non-linear example. Unfortunately, they do not realize that, as already pointed out by Sargan (1983) in the context of Instrumental Variables (IV) estimation, non-linearities may give rise to non-standard asymptotic behavior of GMM estimators when an estimating equation, seen as function of the unknown parameters, may have a zero derivative at the true value, although this function is never flat. It turns out that, as shown in the next section, this is precisely the case in the "Test for Common GARCH Factors" which motivates the test for common GARCH features.

While Sargan (1983) focuses on non-standard asymptotic distributions of GMM estimators in the context of linear instrumental variables estimation with some non-linearities (and associated singularities) with respect to the parameters, we rather set the focus in this paper on the testing procedure for common GARCH features. The reason why it is important is twofold.

First, detecting a factor structure is a key issue for multivariate modelling of volatility of financial asset returns. Without such a structure (or alternatively ad hoc assumptions about the correlations dynamics) there is an inflation of the number of parameters to estimate and nobody can provide reliable estimators of joint conditional heteroskedasticity of a vector of more than a few (10 or even 5) asset returns. Many factor models of conditional heteroskedasticity have been studied in the literature since the seminal paper of Diebold and Nerlove (1989). Let us mention among others Engle, Ng and Rothschild (1990), Fiorentini, Sentana and Shephard (2004) and Doz and Renault (2006). In all these models, it is assumed that the factors have conditional heteroskedasticity but the idiosyncracies do not. The test for common GARCH features is then a universal tool for detecting any of these factor structures.

Second, the singularity issue a la Sargan (1983) that we point out for the estimation of common features parameters has perverse consequences for testing for the factor structure. We show that the test computed with the standard critical value provided by a $\chi^2(H-n+1)$ will be significantly oversized. In other words, the mechanical application of Hansen (1982) J-testing procedure will lead the empirical researcher to throw away too often hypothetical factor structures that are actually valid. The main purpose of this paper is to characterize the degree of over-rejection and give ways to compute correct critical values, or at least valid bounds for a conservative testing approach.

The issue addressed in this paper appears to be new and quite different from seemingly related issues previously considered in the literature.

Cragg and Donald (1996) set the focus on testing for overidentifying restrictions in a linear IV context, when the instruments are weak. Weakness is meant either in the sense of Phillips (1989) when the structural parameters are not identified because the rank condition fails or in the sense of Staiger and Stock (1997) because the reduced form matrix, albeit fulfilling the rank condition, converges with increases in the sample size to a matrix of smaller rank. In both cases, Cragg and Donald (1996) are able to use general results of Cragg and Donald (1993) and also Schott (1984) to show that the actual size of the overidentification test is strictly smaller than the nominal one given by the standard chi-square critical value. The overidentification test is conservative.

The case considered in the paper may look at first sight quite similar since we consider cases where the Jacobian matrix of the moment conditions does not fulfill the rank condition. However, this rank deficiency in our case is due to local singularities produced by non-linearities while the global identification is ensured. This difference has dramatic consequences regarding the actually misleading intuition of similarity with weak identification settings. We show that, in sharp contrast with the cases considered by Cragg and Donald (1996), the rank deficiency in our case will lead to an oversized test, instead of a conservative one. Therefore, the discrepancy with the standard chi-square distribution under the null is much more harmful. The intuition for this difference of results is the following. It is of course quite intuitive that, when they are identification failures, the actual degree of overidentification is not as high as one may believe and thus the naive overidentification test is conservative. On the contrary, when global identification is ensured but the Jacobian displays some rank deficiencies, the degree of overidentification becomes to some extent sample dependent. The structural parameters may indeed be more or less accurately estimated, depending on the location of the data sequence in the sample space. More precisely, there is a positive probability that the estimators of some parameters behave as root-T consistent estimators. Moreover, due to the rank deficiency of the Jacobian matrix, the J-test statistic may not be as sensitive to parameter variation as it should be. Then, when estimators are converging as fast as square-root-T, it is as if the true values were actually known. Then, the right chi-square distribution to consider should not be $\chi^2(H-n+1)$ but rather $\chi^2(H-q)$ for some $q < n-1$. Consequently, the actual distribution of the J-test statistic under the null is somewhere between a $\chi^2(H-n+1)$ and a $\chi^2(H)$, because it involves with positive probabilities some $\chi^2(H-q)$ components for $0 \leq q < n-1$ and the use of the critical value based on $\chi^2(H-n+1)$ leads to over-rejection. Finally, it is worth realizing that by contrast with the most common weak identification phenomenon (see e.g. Staiger and Stock (1997) and also Stock and Wright (2000) for non-linear GMM), the issue we point out is fundamentally an issue of the model. Irrespective of the choice of instruments and independently of any finite sample issue, the valid asymptotic distribution of the J-test statistic under the null involves a mixture of chi-square distributions.

While the focus of this paper is on the overidentification test which is key to detect a factor GARCH structure, the underlying estimation issue must be related to some extant literature. To the best of our knowledge, Sargan (1983) is the only one to have addressed this estimation issue in a GMM context, at least for the particular case of linear (in variables) IV with non-linearities with respect to the parameters. However, in the context of maximum likelihood estimation (MLE), several authors have met a similar situation of local singularity. More precisely, when MLE is seen as a Method of Moments based on the score, the GMM Jacobian matrix corresponds to the Fisher information matrix. The fact that singularity of the Fisher information matrix (when global identification is warranted) may lead to MLE with non-standard rates of convergence has been documented in particular by Melino (1982), Lee and Chesher (1986) and Rotnitzky, Cox, Bottai and Robins (2000). The estimation of sample selectivity bias is a leading example of these three papers. We face in the present paper non standard rates of convergence for GMM estimators of GARCH common features for quite similar reasons. However, our focus is not on the asymptotic distribution of these estimators but rather on the impact of it for the distribution of the J-test statistic for overidentification. This issue could not be addressed in the MLE context since the first order conditions of likelihood maximization are by definition just identified estimating equations.

The paper is organized as follows. The issue of testing for factor GARCH and the intrinsic singularity which comes with it is analyzed in section 2. Section 3 provides the relevant asymptotic theory for the J-test statistic of the null of common GARCH features. Since we will show that the standard J-test is oversized, our focus of interest is more on size than power. We show why the right asymptotic distribution for the J-test statistic under the null involves some $\chi^2(H-q)$ for $q < n-1$ and thus why the use of the critical value based on $\chi^2(H-n+1)$ leads to over-rejection. By contrast, the distribution $\chi^2(H)$ always provides a conservative upper bound. Since the correct asymptotic distribution involves some $\chi^2(H-q)$ for $q < n-1$, very large samples (as often available in finance) are not a solution to the problem pointed out in this paper, quite the contrary indeed. This prediction is confirmed by the small Monte Carlo study provided in section 4. This Monte Carlo study also indicates that the asymptotic results are helpful in evaluating likely finite-sample performance and in providing more correct critical values. It is in particular worth realizing that the size of the test is related to the tail behavior of the distribution of the test statistic under the null. In this respect, even a relatively small mistake on the number of degrees of freedom of the chi-square at play may make a big difference in terms of probability of rejection. Section 5 concludes and sketch other possible contexts of application of the general testing methodology put forward in this paper. Technical proofs are included in an appendix.

Throughout the paper $\|\cdot\|$ denotes not only the usual Euclidean norm but also a matrix norm $\|A\| = (tr(AA'))^{1/2}$, where tr is the usual trace function of square matrices. By the Cauchy-Schwarz inequality, it has the useful property that, for any vector x and any conformable matrix A, $\|Ax\| \leq \|A\|\|x\|$.

2 Testing for Factor GARCH

An n-dimensional stochastic process $(Y_t)_{t\geq 0}$ is said to have a factor GARCH structure with K factors ($K < n$) if it has a conditional covariance matrix given by:

$$\mathrm{Var}\,(Y_{t+1}|\mathfrak{F}_t) = \Lambda D_t \Lambda' + \Omega, \tag{1}$$

where

- D_t is a diagonal matrix of size K with coefficients $\sigma^2_{kt}, k = 1, \ldots, K$, and
- The stochastic processes $(Y_t)_{t\geq 0}$, $(\sigma^2_{kt})_{1\leq k\leq K, t\geq 0}$ are adapted with respect to the increasing filtration $(\mathfrak{F}_t)_{t\in\mathbb{N}}$.

The following assumption is standard and can be maintained without loss of generality:

Assumption 1. *(i)* Rank$(\Lambda) = K$*, (ii)* Var(Diag(D_t)) *is non-singular where* Diag(D_t) *is the K-dimensional vector with coefficients $\sigma^2_{kt}, k = 1, \ldots, K$.*

Assumption 1-*(i)* means that we cannot build a factor structure with $(K-1)$-factors by expressing a column of the matrix Λ of factor loadings as linear combination of the other columns. Assumption 1-*(ii)* means that we cannot build a factor structure with $(K-1)$-factors by expressing one variance component σ^2_{kt} as an affine function of the other components.

For the sake of expositional simplicity, we will assume throughout that:

$$E\left(Y_{t+1}|\mathfrak{F}_t\right) = 0.$$

One may typically see Y_{t+1} as the vector of innovations in a vector r_{t+1} of n asset returns

$$Y_{t+1} = r_{t+1} - E\left(r_{t+1}|\mathfrak{F}_t\right).$$

The way to go in practice from data on r_{t+1} to a consistent estimation of Y_{t+1} through a forecasting model of returns is beyond the scope of this paper.

Following Engle and Kozicki (1993) a GARCH common feature is a portfolio whose return $\theta' Y_{t+1}$, $\sum_{i=1}^n \theta_i = 1$, has no conditional heteroskedasticity :

$$\mathrm{Var}(\theta' Y_{t+1}|\mathfrak{F}_t) \text{ is constant.}$$

Since, by virtue of the factor structure (1),

$$\mathrm{Var}(\theta' Y_{t+1}|\mathfrak{F}_t) = \theta'\Lambda D_t \Lambda'\theta + \theta'\Omega\theta$$

we can see, from Assumption 1-*(ii)*, that $\mathrm{Var}(\theta' Y_{t+1}|\mathfrak{F}_t)$ will be constant if and only if $\theta'\Lambda = 0$:

Lemma 2.1. *The GARCH common features are the vectors $\theta \in \mathbb{R}^n$ solution of*

$$\Lambda'\theta = 0.$$

Lemma 2.1 shows that, irrespective of the detailed specification of a multivariate model of heteroskedasticity, we can test for the existence of a factor structure by simply devising a test of the null hypothesis:

H_0: There exists $\theta \in \mathbb{R}^n$ such that $\mathrm{Var}(\theta' Y_{t+1}|\mathfrak{F}_t)$ is constant.

It is then natural to devise a test of the null H_0 through a test of its consequence $H_0(z)$ for a given choice of a H-dimensional vector z_t of instruments:

$$H_0(z): \;\; E\left(z_t((\theta' Y_{t+1})^2 - c(\theta))\right) = 0, \text{ where } c(\theta) = E((\theta' Y_{t+1})^2).$$

$H_0(z)$ is implied by H_0 insofar as the variables z_t are valid instruments, i.e. are $\mathfrak{F}_t$-measurable. Besides validity, the instruments z_t must identify the GARCH common features θ in order to devise a test $H_0(z)$ from Hansen (1982) theory of the J-test for overidentification.

By the law of iterated expectations, the factor structure (1) gives:

$$E\left(z_t((\theta' Y_{t+1})^2 - c(\theta))\right) = E\left((z_t - Ez_t)\theta'(\Lambda D_t \Lambda' + \Omega)\theta\right)$$

and then, by a simple matrix manipulation,

$$\begin{aligned} E\left(z_t((\theta' Y_{t+1})^2 - c(\theta))\right) &= \mathrm{Cov}\left(z_t, tr(\theta'\Lambda D_t \Lambda'\theta)\right) = \mathrm{Cov}\left(z_t, \mathrm{Diag}'(\Lambda'\theta\theta'\Lambda)\mathrm{Diag}(D_t)\right) \\ &= \mathrm{Cov}(z_t, \mathrm{Diag}(D_t))\mathrm{Diag}(\Lambda'\theta\theta'\Lambda). \end{aligned} \tag{2}$$

The convenient identification assumption about the vector z_t of instruments is then:

Assumption 2. *(i) z_t is $\mathfrak{F}_t$-measurable and* $\mathrm{Var}(z_t)$ *is non-singular, (ii)* $\mathrm{Rank}\left[\mathrm{Cov}(z_t, \mathrm{Diag}(D_t))\right] = K$*.*

Assumption 2-*(i)* is standard. Assumption 2-*(ii)* is non-restrictive, by virtue of Assumption 1-*(ii)*, insofar as we choose a sufficiently rich set of H instruments, $H \geq K$. Sufficiently rich means here that, for any linear combination of K volatility factors σ^2_{kt}, $k = 1, \ldots, K$, there exists at least one instrument z_{ht}, $h = 1, \ldots, H$ correlated with this combination.

From (2), we see that under Assumptions 1 and 2, $H_0(z)$ amounts to:

$$\mathrm{Diag}(\Lambda'\theta\theta'\Lambda) = 0$$

and then, implies:

$$\|\Lambda'\theta\|^2 = tr(\Lambda'\theta\theta'\Lambda) = 0$$

that is θ is a common feature. Conversely, any common feature clearly fulfills the condition of $H_0(z)$. We have thus proved:

Lemma 2.2. *Under Assumptions 1 and 2, the common features $\theta \in \mathbb{R}^n$ are the solutions of the moment restrictions:*

$$\rho(\theta) \equiv E\left(z_t((\theta' Y_{t+1})^2 - c(\theta))\right) = 0,$$

where $c(\theta) = E((\theta' Y_{t+1})^2)$.

As in Engle and Kozicki (1993), GARCH common features are thus identified by moment restrictions $H_0(z)$. $H_0(z)$ will then be considered as the null hypothesis under test in order to test for common features. Engle and Kozicki (1993) focus on the particular case $K = n-1$ in order to be sure that the moment restrictions of $H_0(z)$ (under the null hypothesis that they are valid) define a unique true unknown value θ^0 of the common feature θ, up to a normalization condition (like $\sum_{i=1}^n \theta_i = 1$). Irrespective of a choice of such exclusion/normalization condition to identify a true unknown value θ^0, we show that the standard GMM inference theory will not work for moment restrictions $H_0(z)$. This issue comes from the nullity of the moment Jacobian at the true value, that is at any GARCH common feature. To see this, note that:

$$\begin{aligned}\Gamma(\theta) &= \tfrac{\partial}{\partial\theta'} E\left(z_t((\theta' Y_{t+1})^2 - c(\theta))\right) = E\left[z_t\left\{2(\theta' Y_{t+1})Y'_{t+1} - 2E[(\theta' Y_{t+1})Y'_{t+1}]\right\}\right] \\ &= 2\mathrm{Cov}\left(z_t, [Y_{t+1}Y'_{t+1}]\theta\right).\end{aligned}$$

Then by the law of iterated expectations,

$$\Gamma(\theta) = 2E\left((z_t - E(z_t))\theta'(\Lambda D_t \Lambda' + \Omega)\right) = 0$$

when $\theta'\Lambda = 0$, that is when θ is a common cofeature:

Proposition 2.1. *For any common feature θ,*

$$\Gamma(\theta) \equiv \frac{\partial}{\partial\theta'} E\left(z_t((\theta' Y_{t+1})^2 - c(\theta))\right) = 0.$$

For the application of the GMM asymptotic theory, we then face a singularity issue that is, as announced in the introduction, an intrinsic property of the common GARCH factor model. Irrespective of the quality of the instruments, the sample size and/or the identification restrictions about the common features θ, any choice of a true unknown value θ^0 will lead to a zero Jacobian matrix at θ^0. The rank condition fails by definition.

For the purpose of any asymptotic theory of estimators and testing procedures local identification must then be provided by higher order derivatives. Since our moment conditions of interest $H_0(z)$ are second order polynomials in the parameter θ, the only non-zero higher order derivatives are of order two. Let us assume that exclusion restrictions characterize a set $\Theta_* \subset \mathbb{R}^n$ of parameters which contains at most only one unknown common feature θ^0, up to a normalization condition:

Assumption 3. *$\theta \in \Theta_* \subset \mathbb{R}^n$ such that $\Theta^* = \{\theta \in \Theta_* : \sum_{i=1}^n \theta_i = 1\}$ is a compact set and*

$$\left(\theta \in \Theta^* \text{ and } \theta'\Lambda = 0\right) \Leftrightarrow (\theta = \theta^0).$$

Recall that Assumption 3 is actually implied by Assumptions 1 and 2 in the setting of Engle and Kozicki (1993), that is $K = n-1$. This setting may naturally arise along ascending model choice procedure where it is observed that adding one financial asset always implies adding one common factor.

Under Assumptions 1, 2 and 3, global identification amounts to second-order identification:

Lemma 2.3. *Under Assumptions 1, 2 and 3, with*

$$\rho_h(\theta) \equiv E\left(z_{ht}((\theta' Y_{t+1})^2 - c(\theta))\right), \;\; h = 1, \dots, H,$$

we have

$$\left((\theta - \theta^0)' \frac{\partial^2 \rho_h}{\partial\theta\partial\theta'}(\theta^0)(\theta - \theta^0)\right)_{1 \le h \le H} = 0 \Leftrightarrow (\theta = \theta^0).$$

Note that Lemma 2.3 is a direct consequence of Lemmas 2.1, 2.2 and Proposition 2.1 thanks to the following polynomial identity:

$$\rho(\theta) = \rho(\theta^0) + \frac{\partial\rho}{\partial\theta'}(\theta^0)(\theta - \theta^0) + \frac{1}{2}\left((\theta - \theta^0)' \frac{\partial^2 \rho_h}{\partial\theta\partial\theta'}(\theta^0)(\theta - \theta^0)\right)_{1 \le h \le H},$$

where $\rho(\theta) = (\rho_h(\theta))_{1 \le h \le H}$.

Of course, since $\rho(\theta)$ is a polynomial of degree 2 in θ, the Hessian matrix does not depend on θ^0. However, we maintain the general notation since we refer to a concept of second order identification which may be useful in more general settings (see Dovonon and Renault (2009)). Moreover, the interest of revisiting global identification in terms of second order identification is to point out the rate of convergence we can expect for GMM estimators. The nullity of the Jacobian matrix implies that the square-root-T rate of convergence is not warranted. However, since second order identification is ensured by Lemma 2.3, we expect the GMM estimators not to converge at a slower rate than $T^{1/4}$. We will actually show in Section 3 that $T^{1/4}$ is only a lower bound while faster rates may sometimes occur.

3 Asymptotic theory

The first step is to ensure the announced minimum rate of convergence $T^{1/4}$ for any GMM estimator of interest. This result comes from the standard regularity conditions maintained in the vectorial process of moment functions:

$$\phi_t(\theta) = z_t\left((\theta' Y_{t+1})^2 - c(\theta)\right)$$

and its sample mean:

$$\bar{\phi}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \phi_t(\theta) = \left(\bar{\phi}_{h,T}(\theta)\right)_{1 \le h \le H}.$$

Assumption 4. *In the context of Assumptions 1 to 3, (z_t, Y_t) is a stationary and ergodic process such that $\phi_t(\theta^0)$ is square integrable with a non-singular variance matrix $\Sigma(\theta^0)$.*

Note in addition that it follows from Lemma 2.2 and Proposition 2.1 that both $\phi_t(\theta^0)$ and $\partial\phi_t(\theta^0)/\partial\theta'$ are martingale difference sequences. Then the central limit theorem of Billingsley (1961) for stationary ergodic martingales implies that $\sqrt{T}\bar{\phi}_T(\theta^0)$ and $\sqrt{T}\partial\bar{\phi}_T(\theta^0)/\partial\theta'$ are asymptotically normal. Note that, by contrast with the weak identification literature (Stock and Wright (2000)), we do not need a functional central limit theorem for the empirical process $\left(\bar{\phi}_T(\theta)\right)_{\theta\in\Theta}$. Moreover, we assume throughout that the stationary and ergodic process (z_t, Y_t) fulfills the integrability conditions needed for all the laws of large numbers of interest. Thanks to the polynomial form of the moment restrictions, they will ensure the relevant uniform laws of large numbers for $\bar{\phi}_T(\theta)$ and its derivatives. In particular, any GMM estimator will be consistent under Assumptions 1, 2 and 3 if we define a GMM estimator as

$$\hat{\theta}_T \equiv \arg\min_{\theta\in\Theta^*} \bar{\phi}'_T(\theta) W_T \bar{\phi}_T(\theta),$$

where W_T is a sequence of positive definite random matrices such that $\text{plim}(W_T) = W$ is positive definite.

It is worth noting that the minimization over Θ^* amounts to optimizing with respect to a vector $\theta = h(\theta_{)n(})$ with

$$\theta_{)n(} = (\theta_i)_{1\le i\le n-1},\ \ h(\theta_{)n(}) = \left(\theta'_{)n(}, 1 - \sum_{i=1}^{n-1}\theta_i\right)'.$$

Note that $\theta_{)n(}$ lies in the compact subset of $\mathbb{R}^{n-1}$ obtained by projecting Θ^* on its $n-1$ first components. For the sake of notational simplicity, we let Θ denote this parameter set and $\theta \in \Theta \subset \mathbb{R}^{n-1}$ denote the parameter of interest. We consider the functions $\phi_t(\theta)$, $\bar{\phi}_T(\theta)$ and $\rho(\theta)$ as defined on $\Theta \subset \mathbb{R}^{n-1}$. We also define the GMM estimator $\hat{\theta}_T$ as

$$\hat{\theta}_T \equiv \arg\min_{\theta\in\Theta\subset\mathbb{R}^{n-1}} \bar{\phi}'_T(\theta) W_T \bar{\phi}_T(\theta). \tag{3}$$

We implicitly assume in the rest of the paper that any $\hat{\theta}_T$ defined by Equation (3) is a measurable random vector. This assumption is quite common in the literature on extremum estimators. (See e.g. van der Vaart (1998).) We can prove as already announced that:

Proposition 3.1. *Under Assumptions 1, 2, 3, 4, if $\hat{\theta}_T$ is the GMM estimator as defined by Equation* (3),

$$\|\hat{\theta}_T - \theta^0\| = O_P(T^{-1/4}).$$

Proof: See Appendix.

Proposition 3.1 ensures a convergence at the rate $T^{1/4}$ for the GMM estimator $\hat{\theta}_T$ as opposed to the usual faster rate $T^{1/2}$. Following Chamberlain (1986), it could be deduced from Proposition 2.1 that the partial information matrix for θ is zero. Therefore (see Chamberlain's Theorem 2) there is no (regular) square-root-Tconsistent estimator for θ. The intuition of this result is quite simple. The slope (linear) term appearing in the Taylor expansion of the sample average of $\phi_t(\theta)$, $(\partial\bar{\phi}_T(\theta^0)/\partial\theta')(\hat{\theta}_T - \theta^0)$, has a smaller order of magnitude than $\bar{\phi}_T(\theta^0)$ (the intercept term) and disappears in front of the curvature (quadratic) terms which then determine the asymptotic order of magnitude of $\hat{\theta}_T - \theta^0$. Because these quadratic terms are of order $T^{1/2}$, we can only extract an order $T^{1/2}$ for $\|\hat{\theta}_T - \theta^0\|^2$. Without using Chamberlain (1986), we confirm this result in Proposition 3.2 below by showing that $T^{1/4}(\hat{\theta}_T - \theta^0)$

does not converge to zero with probability 1. However, we also show that there is a positive probability to get $T^{1/4}(\hat{\theta}_T - \theta^0)$ asymptotically equal to zero, that is to have a rate of convergence faster than $T^{1/4}$, typically $T^{1/2}$. As already pointed out by Sargan (1983) in a context of linear instrumental variables, this heterogeneity of convergence rates over the sample space is characterized by sign restrictions on some multilinear functions of components of a Gaussian vector with zero mean.

This vector will be defined from the limit behavior of a sequence of symmetric random matrices Z_T of size $p = n - 1$ with coefficients (i, j), $i, j = 1, \ldots, p$ equal to:

$$\frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}(\theta^0) W \sqrt{T} \bar{\phi}_T(\theta^0)$$

By Assumption 4, the sequence Z_T converges in distribution towards a random matrix Z with Gaussian coefficients:

$$\frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}(\theta^0) W X$$

where $X \sim N(0, \Sigma(\theta^0))$. For this random symmetric matrix Z, we denote $(Z \geq 0)$ the event "Z is positive semidefinite" and $\overline{(Z \geq 0)}$ its complement. We can then state:

Proposition 3.2. *If Assumptions 1, 2, 3, 4 hold and θ^0 is an interior point of Θ, then, the sequence $\left(T^{1/4}(\hat{\theta}_T - \theta^0)', \mathrm{Vec}'(Z_T)\right)'$ has at least one subsequence that converges in distribution and for any such subsequence with limit distribution $(V', \mathrm{Vec}'(Z))'$, we have:*

$$\mathrm{Prob}\,(V = 0 | Z \geq 0) = 1 \text{ and } \mathrm{Prob}\left(V = 0 \middle| \overline{(Z \geq 0)}\right) = 0.$$

Proof: See Appendix.

Note that $\mathrm{Vec}(Z)$ is by definition a zero-mean Gaussian distribution linear function of the limit distribution $N(0, \Sigma(\theta^0))$ of $\sqrt{T}\bar{\phi}_T(\theta^0)$. It is in particular important to realize that Z is positive definite if and only if $\mathrm{Vec}(Z)$ fulfills p multilinear inequalities corresponding to the positivity of the p leading principal minors of the matrix Z (see e.g. Horn and Johnson (1985, p. 404)). Therefore, the probability q_1 of the event $(Z \geq 0)$ is strictly positive but strictly smaller than one. In particular, $q_1 = 0.5$ if $p = 1$. This case corresponds to testing for common GARCH factors in two asset returns and

$$Z_T = \frac{\partial^2 \rho'}{\partial \theta^2}(\theta^0) W \sqrt{T} \bar{\phi}_T(\theta^0).$$

Here, Z corresponds to the (non degenerate) zero-mean univariate normal asymptotic distribution of Z_T. Proposition 3.2 states that the rate of convergence of $\hat{\theta}_T$ is $T^{1/4}$ or more depending on the sign of Z. More generally, the message of Proposition 3.2 is twofold. First, in the part of the sample space where Z is positive semi-definite, all the components of $\hat{\theta}_T$ converge at a rate faster than $T^{1/4}$.

Besides, $T^{1/4}(\hat{\theta}_T - \theta^0)$ must have a non-zero limit in the part of the sample space where Z is not positive semi-definite. As already mentioned, this classification of rates of convergence for GMM estimators in the case of lack of first order identification has clearly been pointed out by Sargan (1983) in the particular context of instrumental variables estimation. It is also related to the result of Rotnitzky et al. (2000) for the maximum likelihood estimation. This mixture of rate of convergence is the cause of the nonstandard asymptotic distribution of the J-test statistic as we see next.

The GMM overidentification test statistic based on the moment condition $E(\phi_t(\theta)) = 0$ is given by:

$$J_T = T\bar{\phi}_T'(\hat{\theta}_T)W_T\bar{\phi}_T(\hat{\theta}_T).$$

We recall that the above moment condition fails to identify the true parameter value at the first order but locally identifies the true parameter value at the second order. (See Proposition 2.1 and Lemma 2.3.) J_T is the minimum value of the GMM objective function using the optimal weighting matrix defined as a consistent estimate of the inverse of the moment conditions' long run variance, i.e. $W^{-1} = \Sigma(\theta^0) \equiv \lim_{T\to\infty} \mathrm{Var}\left(\sqrt{T}\bar{\phi}_T(\theta^0)\right)$. This specific choice of weighting matrix ensures the required normalization of the moment functions that makes J_T behave in large samples as a chi-square random variable with $H-p$ degrees of freedom (Hansen (1982)) when the moment conditions are valid and the first order local identification condition holds.

The next result gives the asymptotic distribution of J_T in our lack of first order identification framework. From the rate of convergence derived in Propositions 3.1 and 3.2, we can see, after some straightforward calculation that

$$J_T = T\bar{\phi}_T'(\theta^0)W\bar{\phi}_T(\theta^0) + T^{1/2}\bar{\phi}_T'(\theta^0)WG\mathrm{Vec}(\hat{v}_T\hat{v}_T') + \frac{1}{4}\mathrm{Vec}'(\hat{v}_T\hat{v}_T')G'WG\mathrm{Vec}(\hat{v}_T\hat{v}_T') + o_P(1),$$

where $\hat{v}_T = T^{1/4}(\hat{\theta}_T - \theta^0)$ and G is a (H, p^2) matrix gathering the second derivatives of the moment conditions with respect to the p components of θ (see Appendix). For our approach to deriving the asymptotic distribution of J_T, it is useful to introduce the $\mathbb{R}^p$-indexed empirical process

$$\hat{J}(v) = T\bar{\phi}_T'\left(\theta^0 + T^{-1/4}v\right)W_T\bar{\phi}_T\left(\theta^0 + T^{-1/4}v\right),$$

where $v \in \mathbb{R}^p$ is implicitly defined as $v = T^{1/4}(\theta - \theta^0)$. By definition, $J_T = \hat{J}(\hat{v}_T) = \min_{v\in\mathbb{H}_T}\hat{J}(v)$, where $\mathbb{H}_T = \left\{v \in \mathbb{R}^p : v = T^{1/4}(\theta - \theta^0), \theta \in \Theta\right\}$. Let $J(v)$ be the $\mathbb{R}^p$-indexed random process defined by:

$$J(v) = X'WX + X'WG\mathrm{Vec}(vv') + \frac{1}{4}\mathrm{Vec}'(vv')G'WG\mathrm{Vec}(vv'),\ v \in \mathbb{R}^p,$$

where $X \sim N(0, \Sigma(\theta^0))$. Note that $X'WG\mathrm{Vec}(vv') = v'Zv$ so that $J(v)$ can also be written:

$$J(v) = X'WX + v'Zv + \frac{1}{4}\mathrm{Vec}'(vv')G'WG\mathrm{Vec}(vv'),\ v \in \mathbb{R}^p.$$

By construction, for each $v \in \mathbb{R}^p$, $\hat{J}(v)$ converges in distribution towards $J(v)$. Lemma A.5 in Appendix shows that this convergence in distribution actually occurs in $\ell^\infty(K)$ for any compact subset K of $\mathbb{R}^p$. Upon the tightness of their respective minimizers, the minimum of $\hat{J}(v)$ converges in distribution towards the minimum of $J(v)$. This is formally stated in the following theorem:

Theorem 3.1. *If Assumptions 1, 2, 3, 4 hold, θ^0 is an interior point of Θ, and $W^{-1} = \Sigma(\theta^0)$, then $J_T = \min_{v\in\mathbb{H}_T} \hat{J}(v)$ converges in distribution towards $\min_{v\in\mathbb{R}^p} J(v)$.*

Proof: See Appendix.

Theorem 3.1 gives the asymptotic distribution of J_T as the minimum of the limiting process $J(v)$. This distribution is rather unusual since $J(v)$ is an even multivariate polynomial function of degree 4. In general, the minimum value of $J(v)$ does not have a close form expression. In usual cases polynomial of degree 2 are often derived as limiting process yielding the usual chi-square distribution. (See e.g. Koul (2002) for the treatment of minimum distance estimators derived from Locally Asymptotically Normal Quadratic dispersions that include the Locally Asymptotically Normal models as particular case as well as the usual GMM framework when the local identification condition holds.) This peculiarity of $J(v)$ makes the determination of critical values for asymptotic inferences involving J_T rather difficult. One possible way may consist on simulating a large number of realizations of X and get an empirical distribution of the minimum value of $J(v)$. But this simulation approach would require an estimation of some nuisance parameters such as $\Sigma(\theta^0)$, W and G. This estimation's effect on the simulated tests' outcome would need a thorough investigation to make this approach useful. Another possible and more promising approach is through some bootstrap techniques (see Dovonon and Gonçalves (2011)).

The next result gives some further and more practical characterization of the asymptotic distribution of J_T.

Theorem 3.2. *Under the same conditions as Proposition 3.2 and Theorem 3.1, the overidentification test statistic J_T is asymptotically distributed as a mixture*

$$J = 1_{[Z\geq 0]}J^{(1)} + (1 - 1_{[Z\geq 0]})J^{(2)}$$

with $J^{(1)} \sim \chi^2_H$, and $\chi^2_{H-p} \leq J^{(2)} < \chi^2_H$ and $J^{(2)} \sim \chi^2_{H-p}$ with positive probability (where $H = \dim(\rho(\theta))$, $p = \dim(\theta)$, and 1_A denotes the usual indicator function.)

In particular, if $p = 1$, J_T is asymptotically distributed as the mixture

$$\frac{1}{2}\chi^2_{H-1} + \frac{1}{2}\chi^2_H.$$

Proof: See Appendix.

Theorem 3.2 confirms the non-standard nature of the asymptotic distribution of J_T. The χ^2_{H-p} which is expected in the standard case to be the asymptotic distribution of J_T is now a lower bound of this asymptotic distribution which also behaves as a χ^2_H with positive probability $q_1 = \text{Prob}(Z \geq 0)$. The interpretation of this result is the following. Considering the parts of the sample space where Z is positive semidefinite, the only minimizer of $J(v)$ is actually 0 and the scaled GMM estimator $\hat{v}_T$ converges in probability to 0. This means that the GMM estimator $\hat{\theta}_T$ converges at a faster rate than its unconditional rate and therefore behaves for J_T as though it was not estimated, thus the χ^2_H. But, when Z is not positive semidefinite, which means for $p = 1$ that Z is negative, $\hat{v}_T$ is no longer necessarily asymptotically degenerate and the estimation cost appears to discount the degrees of freedom of J_T which then has the standard asymptotic distribution, χ^2_{H-1} in this particular case of $p = 1$.

This result also shows that J_T has asymptotically larger quantiles than usual. In the univariate case where $p = 1$, its asymptotic distribution is fully derived but for $p > 1$, Theorem 3.2 provides an upper bound for the asymptotic distribution of J_T (χ^2_H) conservative enough to allow for tests with the correct size asymptotically. Both the lower and upper bounds are shown to be conditionally sharp in the sense that J_T actually behaves asymptotically as a χ^2_{H-p} and χ^2_H with positive probabilities conditionally on some regions of the sample space. In any case, ignoring the first order lack of local identification may lead to possibly severely oversized tests.

At this stage, it is worth reminding that the asymptotic results obtained by Propositions 3.1 and 3.2 and Theorems 3.1 and 3.2 stand regardless of the choice of linear exclusion/normalization condition imposed to identify the true cofeature vector. Our derivations are based upon a portfolio weights constraint that sets the sum of weights to one. But these results are also valid for the types of normalization that set a certain component of the cofeature vector to one as in Engle and Kozicki (1993).

4 Monte Carlo evidence

The Monte Carlo experiments in this section investigate the finite sample performance of the GMM overidentification test proposed in this paper for testing for common GARCH factors. We mainly confirm the non-standard asymptotic distribution of the test statistic as expected from our main result in the previous section.

We simulate an asset return vector process Y_{t+1} as:

$$Y_{t+1} = \Lambda F_{t+1} + U_{t+1}$$

according to two designs. The first one (*Design* D_1) includes two assets so that Y_{t+1} is a bivariate return vector. Y_{t+1} is generated by a single conditionally heteroskedastic factor f_{t+1} ($F_{t+1} = f_{t+1}$) following a Gaussian GARCH(1,1) dynamic, i.e.

$$f_{t+1} = \sigma_t \varepsilon_{t+1}, \quad \sigma_t^2 = \omega + \alpha f_t^2 + \beta \sigma_{t-1}^2,$$

where $\varepsilon_{t+1} \sim \text{NID}(0,1)$. We choose $\omega = 0.2$, $\alpha = 0.2$, and $\beta = 0.6$. The factor loading vector is set to $\Lambda = (1, 0.5)'$ and the bivariate vector of idiosyncratic shocks $U_{t+1} \sim \text{NID}(0, 0.5Id_2)$.

The second design (*Design* D_2) includes three assets and Y_{t+1} is a trivariate return process generated by two independent Gaussian GARCH(1,1) factors $F_{t+1} = (f_{1t+1}, f_{2t+1})'$ where f_{1t+1} is generated with the parameters values $(\omega, \alpha, \beta) = (0.2, 0.2, 0.6)$ and f_{2t+1} is generated with the parameters values

$(\omega, \alpha, \beta) = (0.2, 0.4, 0.4)$. We consider the factor loading matrix $\Lambda = (\lambda_1|\lambda_2)$, with $\lambda_1 = (1, 1, 0.5)'$ and $\lambda_2 = (0, 1, 0.5)'$. The idiosyncratic shocks $U_{t+1} \sim \text{NID}(0, 0.5Id_3)$.

The parameters values considered in these designs match those found in empirical applications for monthly returns and are also used by Fiorentini, Sentana and Shephard (2004) in their Monte Carlo experiments. Each design is replicated 5,000 times for each sample size T. The sample sizes that we consider are 1,000, 2,000, 5,000, 10,000, 20,000, 30,000, and 40,000. We include such large sample sizes in our experiments because of the slower rate of convergence of the GMM estimator. Since the unconditional rate of convergence of this estimator is $T^{1/4}$ and not $\sqrt{T}$ as usual, we expect that the asymptotic behaviours of interest become perceptible for larger samples than those commonly used for such studies.

For each simulated sample, we evaluate the GMM estimator according to (3). The efficient weighting matrix W_T is the inverse of the sample second moment of the moment conditions computed at the first stage GMM estimator of θ associated to the identity weighting matrix. We use a set of two instruments $z_{1t} = (y_{1t}^2, y_{2t}^2)'$ to test for common GARCH factors for the bivariate simulated returns and $z_{2t} = (y_{1t}^2, y_{2t}^2, y_{3t}^2)'$ to test for common GARCH factors for the trivariate simulated returns.

Since these data generating processes satisfy the null hypothesis of common GARCH factors for the respective return vector processes, we expect from Theorem 3.2 that the J-test statistic yielded by *Design* D_1 is asymptotically distributed as a half-half mixture of χ_1^2 and χ_2^2 instead of a χ_1^2 as one would get under standard settings where there is first order local identification. The J-test statistic from *Design* D_2 is expected to lead to substantial over-rejection if the critical values of χ_1^2 (the usual asymptotic distribution of J_T) are used while the critical values of χ_3^2 would permit a test that controls the size of the test.

Table I: Simulated rejection rates of the test for common GARCH factors for *Designs* D_1 and D_2. This test is carried out at 5% level.

	Rejection rate (in %) using 5%-critical value from:					
	χ_1^2	χ_2^2	mixt1	χ_1^2	χ_3^2	mixt2
T	*Design* D_1			*Design* D_2		
1,000	6.84	2.20	3.36	6.88	0.78	1.74
2,000	8.48	3.08	4.62	10.08	1.18	3.06
5,000	8.86	3.32	4.86	12.20	2.00	4.04
10,000	9.28	3.24	4.82	15.18	2.90	5.50
20,000	9.02	2.90	4.72	14.40	2.54	5.04
30,000	8.84	3.06	4.54	12.66	2.42	4.24
40,000	9.48	3.26	4.84	12.54	2.02	4.24

mixt1 stands for $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$ and mixt2 for $\frac{1}{4}\chi_1^2 + \frac{1}{2}\chi_2^2 + \frac{1}{4}\chi_3^2$.

Table I displays the simulated rejection rates of the test for common GARCH factors at the nominal level $\alpha = 0.05$. For *Design* D_1, this table shows the rejection rates when the critical values of a χ_1^2, χ_2^2 and $0.5\chi_1^2 + 0.5\chi_2^2$ are used. These critical values are 3.84, 5.99 and 5.13, respectively. For *Design* D_2, the simulated rejection rates related to the critical values from a χ_1^2 (3.84), χ_3^2 (7.82) and $0.25\chi_1^2 + 0.5\chi_2^2 + 0.25\chi_3^2$ (6.25) are displayed.

As expected for *Design* D_1, the critical value of χ_1^2 leads to an over-rejection of the null of common GARCH factor. For large samples, the rejection rate typically doubles the nominal level of the test. Also, we can see that the critical value from a χ_2^2 is conservative and confirms the result of Theorem 3.2. Furthermore, since only one parameter is involved in the model, the asymptotic distribution of the test statistic is a half-half mixture of χ_1^2 and χ_2^2. This is also confirmed by Table I. We can see that the simulated rejection rates in the column corresponding to the mixture closely match the nominal level of the test as the sample size grows.

The testing results for *Design* D_2 also confirm our main result. The χ_1^2 critical value lead to over-rejection while the critical value of χ_3^2 yields a test with a correct level. In addition, it is worth mentioning that the rejection rate from the standard asymptotic distribution (χ_1^2) is much larger than the over-rejection from standard asymptotic distribution from *Design* D_1. This means that, as we increase the number of assets, the standard asymptotic results are more and more likely to fail to detect common GARCH factors. This is also suggested by our theory. Actually, as the size of the return vector gets larger, the whole asymptotic distribution of J_T shifts farther to the right of the standard asymptotic distribution (χ^2_{H-p}) while still being bounded by a χ^2_H which is attained with positive probability, conditionally on certain regions of the sample space. For the sake of illustration, we also give in Table I for *Design D2* the rejection rate when the critical value is computed from a mixture $0.25\chi_1^2 + 0.5\chi_2^2 + 0.25\chi_3^2$. Although we have no theoretical result to prove the asymptotic validity of this precise mixture, it seems to be a fairly accurate approximation in the context of our Monte Carlo experiments.

5 Conclusion

This paper proposes a test for common GARCH factors in asset returns. Following Engle and Kozicki (1993) the test statistic is conformable to a GMM overidentification test (J-test) of the moment conditions resulting from the factor GARCH structure. However, we claim that the critical value of this J-test must not be computed as usual because the set of moment conditions is first order underidentified in the sense that the Jacobian matrix of the moment conditions evaluated at the true parameter value is not of full rank; it is actually identically zero regardless of the true parameter value in the parameter space and how strong the instruments are. A Jacobian matrix of full rank at the true parameter value is referred to in the literature as a local identification condition. This is required for moment condition models for the usual asymptotic results of Hansen (1982) for the J-test to apply.

We study the J-test for common GARCH factors under this local identification condition failure while maintaining the global identification condition. The asymptotic distribution of the J-test statistic is markedly nonstandard. We show that it corresponds to the minimum of a certain limiting stochastic process that does not yield the usual chi-square distribution. A further characterization of this distribution shows, for the case of two assets, that it is a half-half mixture of chi-squares while, the complexity of the distribution in the case of more than two assets means that we can only provide some bounds. We show that the upper bound distribution, which is a chi-square is useful for testing the null hypothesis of common GARCH factors even if such tests are meant to be conservative.

The exploration of these asymptotic results also reveals that ignoring the first order underidentification, and hence using the standard asymptotic results, leads to over-rejecting tests. Our Monte Carlo results suggest that this over-rejection should become even more severe as we increase the number of assets. An interesting extension of this work may consist on studying the validity of some resampling techniques such as the bootstrap to approximate the asymptotic distribution of the test statistic instead of relying on conservative bounds for testing. This is the main focus of Dovonon and Gonçalves (2011).

It is worth recalling that the asymptotic results obtained in this paper are related to the case where the local identification failure is due to a *null* Jacobian of the moment condition at the true parameter value. Also, the moment condition functions involved are quadratic so that they match their own higher order expansions. An interesting generalization that is the focus of interest of Dovonon and Renault (2009) is to study the GMM asymptotic properties when the Jacobian is rank deficient and the moment functions are not necessarily quadratic.

Appendix

Throughout this appendix, we denote Δ and $\bar{\Delta}$ the $\mathbb{R}^H$-valued functions defined by

$$\Delta(v) = \left(v'\frac{\partial^2 \rho_h}{\partial\theta\partial\theta'}(\theta^0)v\right)_{1\leq h\leq H} \quad \text{and} \quad \bar{\Delta}(v) = \left(v'\frac{\partial^2 \bar{\phi}_{h,T}}{\partial\theta\partial\theta'}(\theta^0)v\right)_{1\leq h\leq H}, \forall v \in \mathbb{R}^p,$$

$p = n-1$ and $n = \dim(Y_t)$. We let G and $\bar{G}$ be two (H, p^2) matrices defined such that $\Delta(v) = G\text{Vec}(vv')$ and $\bar{\Delta}(v) = \bar{G}\text{Vec}(vv')$, for all $v \in \mathbb{R}^p$. By definition,

$$G = \left(\text{Vec}\left(\frac{\partial^2 \rho_1}{\partial\theta\partial\theta'}(\theta^0)\right), \text{Vec}\left(\frac{\partial^2 \rho_2}{\partial\theta\partial\theta'}(\theta^0)\right), \cdots, \text{Vec}\left(\frac{\partial^2 \rho_H}{\partial\theta\partial\theta'}(\theta^0)\right)\right)'$$

and $\bar{G}$ has the same expression but with $\bar{\phi}_{h,T}$ instead of ρ_h, $h = 1, \ldots, H$.

Lemma A.1. *If* $(\Delta(v) = 0) \Rightarrow (v = 0))$, *then there exists* $\gamma > 0$ *such that for any* $v \in \mathbb{R}^p$,

$$\Delta(v) \geq \gamma\|v\|^2.$$

Proof of Lemma A.1. $\Delta(v)$ is an homogeneous function of degree 2 with respect to v. Therefore, for all $v \in \mathbb{R}^p$,

$$\|\Delta(v)\| = \|v\|^2 \left\|\Delta\left(\frac{v}{\|v\|}\right)\right\|.$$

Define $\gamma = \inf_{\|v\|=1}\|\Delta(v)\|$. From the compactness of $\{v \in \mathbb{R}^p : \|v\| = 1\}$ and the continuity of $\Delta(v)$, there exists v^* such that $\|v^*\| = 1$ and $\gamma = \|\Delta(v^*)\|$. $\Delta(v^*) \neq 0$ since $v^* \neq 0$ and this shows the expected result.$\square$

Lemma A.2. *Let $\{X_T : T \in \mathbb{N}\}$ and $\{\varepsilon_T : T \in \mathbb{N}\}$ be two sequences of real valued random variables such that ε_T converges in probability towards 0 and for all T, $X_T \leq \varepsilon_T$, a.s. Then,*

$$\limsup_{T\to\infty} \text{Prob}\left(X_T \leq \epsilon\right) = 1, \quad \forall \epsilon > 0.$$

Proof of Lemma A.2. Let $\epsilon > 0$. We have

$$\limsup_{T\to\infty} \text{Prob}\left(X_T \leq \epsilon\right) = 1 - \liminf_{T\to\infty} \text{Prob}\left(X_T > \epsilon\right).$$

But

$$\inf_{n\geq T} \text{Prob}\left(X_n > \epsilon\right) \leq \text{Prob}\left(X_T > \epsilon\right) \leq \text{Prob}\left(\varepsilon_T > \epsilon\right) \to 0$$

as $T \to \infty$. This establishes the result□

Lemma A.3. *Under the same conditions as Theorem 3.2, there exists an (H, p) matrix G_1 ($p = n - 1$) and a (p, p^2) matrix G_2 such that*

$$G = G_1 G_2 \quad \text{and} \quad \text{Rank}(G) = \text{Rank}(G_1) = \text{Rank}(G_2) = p.$$

Proof of Lemma A.3. Let $\theta_* = \left(\theta', 1 - \sum_{i=1}^{n-1} \theta_i\right)'$, $\theta \in \mathbb{R}^{n-1}$. We recall that $\rho(\theta) = E\left[z_t\left((\theta_*' Y_{t+1})^2 - c(\theta_*)\right)\right]$. We have

$$\begin{aligned}
\rho(\theta) &= E[(z_t - E(z_t))(\theta_*' Y_{t+1})^2] = E[(z_t - E(z_t))(\theta_*' Y_{t+1} Y_{t+1}' \theta_*)] \\
&= E[(z_t - E(z_t))E(\theta_*' Y_{t+1} Y_{t+1}' \theta_* | \mathfrak{F}_t)] = E[(z_t - E(z_t))\theta_*' \Lambda D_t \Lambda' \theta_*] \\
&= E[(z_t - E(z_t))tr(D_t \Lambda' \theta_* \theta_*' \Lambda) = E[(z_t - E(z_t))\text{Diag}'(D_t)\text{Diag}(\Lambda' \theta_* \theta_*' \Lambda) \qquad \text{(A.1)} \\
&= \text{Cov}(z_t, \text{Diag}(D_t))\text{Diag}(\Lambda' \theta_* \theta_*' \Lambda) \\
&\equiv G_1 \text{Diag}(\Lambda' \theta_* \theta_*' \Lambda)
\end{aligned}$$

where $G_1 = \text{Cov}(z_t, \text{Diag}(D_t))$ is a (H, p) matrix of rank p by Assumption 2.

Then, By computing the second order derivatives at θ^0, we deduce that

$$G = G_1 G_2$$

for some (p, p^2) matrix G_2. We now show that G_2 has full row rank p. We proceed by contradiction. If G_2 does not have full row rank, G itself would be of rank smaller than p and the null space of G would be of dimension larger than $p^2 - p$. This cannot be true since, by Lemma 2.3.,

$$G\text{Vec}(vv') = 0 \Rightarrow v = 0$$

and clearly, none of the p linearly independent vectors: $\text{Vec}(e_i e_i')$, $i = 1, \ldots, p$, where $\{e_i : i = 1, \ldots, p\}$ is the canonical basis of $\mathbb{R}^p$ (all the components of e_i are zero except the i-th one equal to 1), belongs to the null space of G□

Lemma A.4. *Let $\hat{M}(v)$ and $M(v)$ be two real-valued stochastic processes with continuous sample paths indexed by $\mathbb{R}^p$ and $\{\mathbb{V}_T : T \in \mathbb{N}\}$ a non-decreasing sequence of subsets of $\mathbb{R}^p$ such that $\bigcup_{T\geq 0} \mathbb{V}_T = \mathbb{R}^p$. If*

(i) $\hat{M}(\cdot)$ converges in distribution towards $M(\cdot)$ in $\ell^\infty(K)$ for every compact $K \subset \mathbb{R}^p$, where $\ell^\infty(K)$ is the set of all uniformly bounded real-valued functions on K,

(ii) there exists $\hat{v}_T \in \arg\min_{v\in\mathbb{V}_T} \hat{M}(v)$ which is uniformly tight and

(iii) there exists $\hat{v} \in \arg\min_{v\in\mathbb{R}^p} M(v)$ which is tight,

then,

$$\hat{M}(\hat{v}_T) \xrightarrow{d} M(\hat{v}).$$

Proof of Lemma A.4. We show that $\text{Prob}(\hat{M}(\hat{v}_T) \le x) \to \text{Prob}(M(\hat{v}) \le x)$ as $T \to \infty$ for any continuity point x of the cumulative distribution of $M(\hat{v})$. Let $x \in \mathbb{R}$ be such a point and $\epsilon > 0$. Since $\hat{v}_T$ is uniformly tight and $\hat{v}$ is tight, there exists $m_\epsilon > 0$ such that

$$\sup_T \text{Prob}(\|\hat{v}_T\| > m_\epsilon) < \frac{\epsilon}{3} \quad \text{and} \quad \text{Prob}(\|\hat{v}\| > m_\epsilon) < \frac{\epsilon}{3}$$

and from Condition *(i)* of the Lemma, $\hat{M}(\cdot)$ converges towards $M(\cdot)$ in distribution in $\ell^\infty(\{v : \|v\| \le m_\epsilon\})$. Since the function inf is continuous on $\ell^\infty(K)$, for any nonempty compact K, we can apply the continuous mapping theorem and deduce that

$$\inf_{\|v\|\le m_\epsilon} \hat{M}(v) \xrightarrow{d} \inf_{\|v\|\le m_\epsilon} M(v).$$

Considering x as a continuity point for the cumulative distribution function of $\inf_{\|v\|\le m_\epsilon} M(v)$ (if not, considering that $\hat{v}$ is tight, we can make m_ϵ large enough so that this is true), we can write that there exists T_ϵ such that for all $T > T_\epsilon$, $\{v : \|v\| < m_\epsilon\} \subset \mathbb{V}_T$ and

$$\left|\text{Prob}\left(\inf_{\|v\|\le m_\epsilon} \hat{M}(v) \le x\right) - \text{Prob}\left(\inf_{\|v\|\le m_\epsilon} M(v) \le x\right)\right| < \frac{\epsilon}{3}.$$

Clearly,

$$\begin{aligned}
(\hat{M}(\hat{v}_T) \le x) &= \left(\hat{M}(\hat{v}_T) \le x; \|\hat{v}_T\| \le m_\epsilon\right) \bigcup \left(\hat{M}(\hat{v}_T) \le x; \|\hat{v}_T\| > m_\epsilon\right) \\
&= \left(\inf_{\|v\|\le m_\epsilon} \hat{M}(v) \le x; \|\hat{v}_T\| \le m_\epsilon\right) \bigcup \left(\hat{M}(\hat{v}_T) \le x; \|\hat{v}_T\| > m_\epsilon\right) \\
&= \left[\left(\inf_{\|v\|\le m_\epsilon} \hat{M}(v) \le x\right) \setminus \left(\inf_{\|v\|\le m_\epsilon} \hat{M}(v) \le x; \|\hat{v}_T\| > m_\epsilon\right)\right] \bigcup \left(\hat{M}(\hat{v}_T) \le x; \|\hat{v}_T\| > m_\epsilon\right)
\end{aligned}$$

thus,

$$\text{Prob}\left(\hat{M}(\hat{v}_T) \le x\right) - \text{Prob}\left(\inf_{\|v\|\le m_\epsilon} \hat{M}(v) \le x\right) \le \text{Prob}(\|\hat{v}_T\| > m_\epsilon).$$

We can actually replace $\hat{M}(\hat{v}_T)$ by $\inf_{\|v\|\le m_\epsilon} \hat{M}(v)$ in the previous set operations and deduce that

$$\text{Prob}\left(\inf_{\|v\|\le m_\epsilon} \hat{M}(v) \le x\right) - \text{Prob}\left(\hat{M}(\hat{v}_T) \le x\right) \le \text{Prob}(\|\hat{v}_T\| > m_\epsilon).$$

Therefore,

$$\left|\text{Prob}(\hat{M}(\hat{v}_T) \le x) - \text{Prob}\left(\inf_{\|v\|\le m_\epsilon} \hat{M}(v) \le x\right)\right| \le \text{Prob}(\|\hat{v}_T\| > m_\epsilon) < \frac{\epsilon}{3}.$$

By the same way, we also have

$$\left|\text{Prob}(M(\hat{v}) \le x) - \text{Prob}\left(\inf_{\|v\|\le m_\epsilon} M(v) \le x\right)\right| \le \text{Prob}(\|\hat{v}\| > m_\epsilon) < \frac{\epsilon}{3}.$$

Now, we observe that

$$\begin{aligned}
\left|\text{Prob}(\hat{M}(\hat{v}_T) \le x) - \text{Prob}(M(\hat{v}) \le x)\right| &\le \left|\text{Prob}(\hat{M}(\hat{v}_T) \le x) - \text{Prob}(\inf_{\|v\|<m_\epsilon} \hat{M}(v) \le x)\right| \\
&+ \left|\text{Prob}(\inf_{\|v\|<m_\epsilon} \hat{M}(v) \le x) - \text{Prob}(\inf_{\|v\|<m_\epsilon} M(v) \le x)\right| \\
&+ \left|\text{Prob}(\inf_{\|v\|<m_\epsilon} M(v) \le x) - \text{Prob}(M(\hat{v}) \le x)\right|.
\end{aligned}$$

Hence, for any $T > T_\epsilon$, $\left|\text{Prob}(\hat{M}(\hat{v}_T) \le x) - \text{Prob}(M(\hat{v}) \le x)\right| < 3\epsilon/3$. This completes the proof$\square$

Lemma A.5. *Under the same conditions as Theorem 3.1, we have*

(i) The stochastic process $\hat{J}(\cdot)$ *converges in distribution towards* $J(\cdot)$ *in* $\ell^\infty(K)$ *for every compact* $K \subset \mathbb{R}^p$,

(ii) $\hat{v}_T \equiv \arg\min_{v\in\mathbb{H}_T} \hat{J}(v)$ *is uniformly tight and any* $\hat{v} \in \arg\min_{v\in\mathbb{R}^p} J(v)$ *is tight.*

(iii) In particular, $\hat{J}(\hat{v}_T) \xrightarrow{d} J(\hat{v})$.

Proof of Lemma A.5. We have

$$\bar{\phi}_T\left(\theta + T^{-1/4}v\right) = \bar{\phi}_T(\theta^0) + T^{-1/4}\frac{\partial\bar{\phi}_T}{\partial\theta'}(\theta^0)v + \frac{1}{2}T^{-1/2}\bar{\Delta}(v)$$

and

$$\begin{aligned}
\hat{J}(v) &= T\bar{\phi}_T'\left(\theta + T^{-1/4}v\right)W_T\bar{\phi}_T\left(\theta + T^{-1/4}v\right)\\
&= T\bar{\phi}_T'(\theta^0)W_T\bar{\phi}_T(\theta^0) + 2T^{1/2}\bar{\phi}_T'(\theta^0)W_T T^{1/4}\tfrac{\partial\bar{\phi}_T}{\partial\theta'}(\theta^0)v\\
&\quad + T^{1/2}\bar{\phi}_T'(\theta^0)W_T\bar{G}\mathrm{Vec}(vv') + T^{1/2}v'\tfrac{\partial\bar{\phi}_T'}{\partial\theta}(\theta^0)W_T\tfrac{\partial\bar{\phi}_T}{\partial\theta'}(\theta^0)v\\
&\quad + T^{1/4}v'\tfrac{\partial\bar{\phi}_T'}{\partial\theta}(\theta^0)W_T\bar{G}\mathrm{Vec}(vv') + \tfrac{1}{4}\mathrm{Vec}'(vv')\bar{G}'W_T\bar{G}\mathrm{Vec}(vv').
\end{aligned}$$

Hence

$$\hat{J}(v) = T\bar{\phi}_T'(\theta^0)W\bar{\phi}_T(\theta^0) + T^{1/2}\bar{\phi}_T'(\theta^0)WG\mathrm{Vec}(vv') + \frac{1}{4}\mathrm{Vec}'(vv')G'WG\mathrm{Vec}(vv') + o_P(1), \tag{A.2}$$

where the $o_P(1)$ term is in fact uniformly negligible over any compact subset of $\mathbb{R}^p$.

(i) We apply Theorem 1.5.4 of van der Vaart and Wellner (1996). To deduce that the stochastic process $\hat{J}(\cdot)$ converges in distribution towards $J(\cdot)$ in $\ell^\infty(K)$, this theorem requires that:

(a) The marginals $(\hat{J}(v_1), \ldots, \hat{J}(v_k))$ converge in distribution towards $(J(v_1), \ldots, J(v_k))$ for every finite subset $\{v_1, \ldots, v_k\}$ of K.

(b) The empirical process $\hat{J}(\cdot)$ is asymptotically tight.

To show *(a)*, we observe that, since the $o_P(1)$ terms in (A.2) is uniformly negligible over any compact, $(\hat{J}(v_1), \ldots, \hat{J}(v_k))$ is asymptotically equivalent to a continuous function of $\sqrt{T}\bar{\phi}_T(\theta^0)$ whose components are

$$T\bar{\phi}_T'(\theta^0)W\bar{\phi}_T(\theta^0) + T^{1/2}\bar{\phi}_T'(\theta^0)WG\mathrm{Vec}(v_iv_i') + \frac{1}{4}\mathrm{Vec}'(v_iv_i')G'WG\mathrm{Vec}(v_iv_i'), \quad i = 1, \ldots, k.$$

By the continuous mapping theorem, this latter converges in distribution towards $(J(v_1), \ldots, J(v_k))$. This establishes *(a)*.

To establish *(b)*, we rely on Theorem 1.5.7 of van der Vaart and Wellner (1996). This theorem gives some sufficient conditions for the empirical process $\hat{J}(\cdot)$ to be asymptotically tight. From *(a)*, $\hat{J}(v)$ converges in distribution towards $J(v)$, for any $v \in K$. In addition, as a compact subset, K equipped with the usual metric on $\mathbb{R}^p$ is totally bounded. It remains to show that $\hat{J}(\cdot)$ is asymptotically uniformly equicontinuous in probability. That is for any $\epsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\limsup_T \mathrm{Prob}\left(\sup_{v_1,v_2\in K:\|v_1-v_2\|<\delta}\left|\hat{J}(v_1) - \hat{J}(v_2)\right| > \epsilon\right) < \eta.$$

From (A.2), $\hat{J}(v)$ is essentially a polynomial function of v and since K is bounded, we can write

$$|\hat{J}(v_1) - \hat{J}(v_2)| = X_T\|v_1 - v_2\| + o_P(1), \tag{A.3}$$

where $X_T = O_P(1)$. Let $\epsilon, \eta > 0$. Since $X_T = O_P(1)$, there exists $m_\eta > 0$ such that $\sup_T \mathrm{Prob}(|X_T| > m_\eta) < \eta$. Let $\delta = \epsilon/(2m_\eta)$ and $A_T = \left(\sup_{v_1,v_2\in K:\|v_1-v_2\|<\delta}\left|\hat{J}(v_1) - \hat{J}(v_2)\right| > \epsilon\right)$. We have

$$A_T = (A_T, |X_T| > m_\eta)\bigcup(A_T, |X_T| \leq m_\eta).$$

We can safely ignore the $o_P(1)$ term in (A.3) and write

$$(A_T, |X_T| \leq m_\eta) \subset \left(\sup_{\|v_1 - v_2\| < \delta} |X_T| \|v_1 - v_2\| > \epsilon, |X_T| \leq m_\eta \right) \subset (|X_T| > 2m_\eta, |X_T| \leq m_\eta) = \emptyset.$$

Thus

$$\text{Prob}(A_T) \leq \text{Prob}(|X_T| > m_\eta) < \eta.$$

As a result, $\limsup_T \text{Prob}(A_T) < \eta$ and this completes the proof of *(b)*; thus *(i)*.

(ii) By definition, $\hat{v}_T = T^{1/4}(\hat{\theta}_T - \theta^0)$ and the uniform tightness of $\hat{v}_T$ follows from Proposition 3.1. Next, consider $\hat{v} \in \arg\min_{v \in \mathbb{R}^p} J(v)$. Let $\epsilon > 0$. We have $0 \leq \min_{v \in \mathbb{R}^p} J(v) \leq J(0) = O_P(1)$, hence, there exists $m_1 > 0$ such that

$$\text{Prob}\left(\min_{v \in \mathbb{R}^p} J(v) > m_1 \right) < \frac{\epsilon}{2}.$$

Note that the leading term in $J(v)$ is $\text{Vec}'(vv')G'WG\text{Vec}(vv')$ and we know from Lemma A.1 that $\gamma\|v\|^4 \leq \text{Vec}'(vv')G'WG\text{Vec}(vv')$, $\gamma > 0$. Therefore, for $\|v\|$ large enough, we can make $J(v)$ as large as we want. That is:

$$\forall \alpha, \beta > 0, \ \exists m_2 > 0: \ \text{Prob}\left(\inf_{\|v\| > m_2} J(v) > \alpha \right) > 1 - \beta.$$

We apply this with $\alpha = m_1$ and $\beta = \frac{\epsilon}{2}$ and observe that

$$(\|\hat{v}\| > m_2) = (\|\hat{v}\| > m_2, J(\hat{v}) > m_1) \bigcup (\|\hat{v}\| > m_2, J(\hat{v}) \leq m_1).$$

Thus

$$\text{Prob}(\|\hat{v}\| > m_2) \leq \text{Prob}(J(\hat{v}) > m_1) + \text{Prob}\left(\inf_{\|v\| > m_2} J(v) \leq m_1 \right) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This shows that $\hat{v}$ is tight.

(iii) This last point follows from Lemma A.4 since θ^0 is an interior point for Θ, the sequence $\mathbb{H}_T$ verifies the condition of this lemma$\square$

Proof of Proposition 3.1. We want to show that $\hat{v}_T = T^{1/4}(\hat{\theta}_T - \theta^0)$ is bounded in probability. We observe that as a second order polynomial,

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\phi}_T(\theta^0) + \sqrt{T}\frac{\partial \bar{\phi}_T}{\partial \theta'}(\theta^0)(\hat{\theta}_T - \theta^0) + \frac{1}{2}\sqrt{T}\bar{\Delta}(\hat{\theta}_T - \theta^0).$$

We remind that from Assumption 4, $\sqrt{T}\bar{\phi}_T(\theta^0)$ and $\sqrt{T}\partial\bar{\phi}_T(\theta^0)/\partial\theta'$ are bounded in probability since asymptotically normal. Hence,

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\phi}_T(\theta^0) + \frac{1}{2}\sqrt{T}\bar{\Delta}(\hat{\theta}_T - \theta^0) + o_P(1)$$

and

$$T\bar{\phi}_T'(\hat{\theta}_T)W_T\bar{\phi}_T(\hat{\theta}_T) = T\bar{\phi}_T'(\theta^0)W_T\bar{\phi}_T(\theta^0) + \frac{T}{4}\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\Delta}(\hat{\theta}_T - \theta^0) + T\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\phi}_T(\theta^0) + o_P(1).$$

By definition,

$$T\bar{\phi}_T'(\theta^0)W_T\bar{\phi}_T(\theta^0) - T\bar{\phi}_T'(\hat{\theta}_T)W_T\bar{\phi}_T(\hat{\theta}_T) \geq 0$$

and we can write:

$$\frac{T}{4}\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\Delta}(\hat{\theta}_T - \theta^0) \leq -T\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\phi}_T(\theta^0) + o_P(1). \tag{A.4}$$

Let $\hat{\delta} \equiv \text{Vec}((\hat{\theta}_T - \theta^0)(\hat{\theta}_T - \theta^0)')$. By definition, $\bar{\Delta}(\hat{\theta}_T - \theta^0) = \bar{G}\hat{\delta}$ and we have

$$\begin{aligned} \bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\Delta}(\hat{\theta}_T - \theta^0) &= \hat{\delta}'\bar{G}'W_T\bar{G}\hat{\delta} \\ &= \hat{\delta}'G'WG\hat{\delta} + \hat{\delta}'(\bar{G} - G)'W_T\bar{G}\hat{\delta} + \hat{\delta}'G'(W_T - W)\bar{G}\hat{\delta} + \hat{\delta}'G'W(\bar{G} - G)\hat{\delta} \end{aligned}$$

and from (A.4), we can write

$$\begin{aligned}\tfrac{T}{4}\hat{\delta}'G'WG\hat{\delta} &\leq -T\hat{\delta}'(\bar{G}-G)'W_T\bar{\phi}_T(\theta^0) - T\hat{\delta}'G'(W_T-W)\bar{\phi}_T(\theta^0) - T\hat{\delta}'G'W\bar{\phi}_T(\theta^0)\\ &\quad -\tfrac{T}{4}\hat{\delta}'(\bar{G}-G)'W_T\bar{G}\hat{\delta} - \tfrac{T}{4}\hat{\delta}'G'(W_T-W)\bar{G}\hat{\delta} - \tfrac{T}{4}\hat{\delta}'G'W(\bar{G}-G)\hat{\delta} + o_P(1).\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}\tfrac{T}{4}\hat{\delta}'G'WG\hat{\delta} &\leq \sqrt{T}\|\hat{\delta}\|\|\bar{G}-G\|\|W_T\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\| + \sqrt{T}\|\hat{\delta}\|\|G\|\|W_T-W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\|\\ &\quad +\sqrt{T}\|\hat{\delta}\|\|G\|\|W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\| + \tfrac{T}{4}\|\hat{\delta}\|^2\|\bar{G}'-G'\|\|W_T\|\|\bar{G}\|\\ &\quad +\tfrac{T}{4}\|\hat{\delta}\|^2\|G\|\|W_T-W\|\|\bar{G}\| + o_P(1).\end{aligned}$$

Noting that $\|\hat{\delta}\| = \|\hat{\theta}_T - \theta^0\|^2$, and W is symmetric positive definite and also using Lemma A.1, we can write

$$\hat{\delta}'G'WG\hat{\delta} \geq \gamma_0\|\hat{\delta}'G'G\hat{\delta}\| = \gamma_0\|\Delta(\hat{\theta}_T-\theta^0)\|^2 \geq \gamma\|\hat{\theta}_T-\theta^0\|^4,$$

for some γ_0, $\gamma > 0$. Hence

$$\gamma\|\hat{v}_T\|^4 \leq 4\|\hat{v}_T\|^2\|G\|\|W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\| + \|\hat{v}_T\|^2 o_P(1) + \|\hat{v}_T\|^4 o_P(1) + o_P(1).$$

Dividing each side by $\|\hat{v}_T\|^2$ and after some re-arrangements, we have

$$\|\hat{v}_T\|^2(\gamma + o_P(1)) \leq 4\|G\|\|W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\| + \frac{o_P(1)}{\|\hat{v}_T\|^2} + o_P(1)$$

and, for T large enough we can write

$$\|\hat{v}_T\|^2 \leq \frac{4}{\gamma}\|G\|\|W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\| + \frac{o_P(1)}{\|\hat{v}_T\|^2} + o_P(1).$$

Hence, for large values of $\|\hat{v}_T\|^2$, the term $o_P(1)/\|\hat{v}_T\|^2$ stays asymptotically negligible in probability. Therefore, $\|\hat{v}_T\|^2$ is at most of the same asymptotic order of magnitude as $\|\sqrt{T}\bar{\phi}_T(\theta^0)\|$. This establishes that $\|\hat{v}_T\|^2 = O_P(1)$ or equivalently $\|\hat{v}_T\| = O_P(1)$ $\square$

Proof of Proposition 3.2. Since Z_T is a continuous function of $\sqrt{T}\bar{\phi}_T(\theta^0)$ it suffices to show that the sequence $\left(T^{1/4}(\hat{\theta}_T-\theta^0)', \sqrt{T}\bar{\phi}_T(\theta^0)'\right)'$ has a subsequence that converges in distribution. From Proposition 3.1, $T^{1/4}(\hat{\theta}_T-\theta^0)$ is uniformly tight and $\sqrt{T}\bar{\phi}_T(\theta^0)$ is also uniformly tight following Assumption 4. Thus, these two quantities defined measurable on the same probability space are jointly uniformly tight. Therefore, Prohorov's theorem (see Theorem 2.4 of van der Vaart (1998)), the joint sequence has a subsequence that converges in distribution. This establishes the first part of the Proposition.

Next, we show that Prob $(V = 0|Z \geq 0) = 1$. Since $\hat{\theta}_T - \theta^0 = O_P(T^{-1/4})$, we have

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\phi}_T(\theta^0) + \frac{1}{2}\sqrt{T}\left((\hat{\theta}_T-\theta^0)'\frac{\partial^2\rho_h}{\partial\theta\partial\theta'}(\theta^0)(\hat{\theta}_T-\theta^0)\right)_{1\leq h\leq H} + o_P(1) \tag{A.5}$$

In particular $\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = O_P(1)$ and thus:

$$J_T = T\bar{\phi}_T'(\hat{\theta}_T)W\bar{\phi}_T(\hat{\theta}_T) + o_P(1).$$

For the sake of expositional simplicity, we will consider $W = Id_H$. This is not restrictive as it amounts to rescaling $\phi_t(\theta)$ by $W^{1/2}$. We keep $\phi_t(\theta)$ for $W^{1/2}\phi_t(\theta)$ in the rest of this proof for economy of notation. Thus

$$\begin{aligned}J_T &= T\bar{\phi}_T'(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T) + o_P(1)\\ &= T\bar{\phi}_T'(\theta^0)\bar{\phi}_T(\theta^0) + \Delta'\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\sqrt{T}\bar{\phi}_T(\theta^0) + \tfrac{1}{4}\Delta'\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\Delta\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right) + o_P(1).\end{aligned}$$

By definition, $J_T \leq T\bar{\phi}_T'(\theta^0)\bar{\phi}_T(\theta^0)$. Hence

$$\Delta'\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\sqrt{T}\bar{\phi}_T(\theta^0)+\frac{1}{4}\Delta'\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\Delta\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\leq o_P(1) \tag{A.6}$$

It is worth noting that

$$\Delta'\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\sqrt{T}\bar{\phi}_T(\theta^0)=\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)'Z_T\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right), \tag{A.7}$$

Considering a subsequence of $\left(T^{1/4}(\hat{\theta}_T-\theta^0)', \mathrm{Vec}'(Z_T)\right)'$ that converges in distribution towards a certain random vector $(V', \mathrm{Vec}'(Z))'$, we can write (for the sake of simplicity, we do not make explicit the notation for a subsequence):

$$\Delta'\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\sqrt{T}\bar{\phi}_T(\theta^0)\xrightarrow{d}V'ZV.$$

From (A.6) and by Lemma A.2, we deduce that

$$\limsup_{T\to\infty}\mathrm{Prob}\Big(\Delta'\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\sqrt{T}\bar{\phi}_T(\theta^0)+$$
$$+\tfrac{1}{4}\Delta'\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\Delta\left(T^{1/4}(\hat{\theta}_T-\theta^0)\right)\leq\epsilon\Big)=1,$$

for any $\epsilon>0$. And, by the Portmanteau Lemma (Lemma 2.2 of van der Vaart (1998)), we have

$$\mathrm{Prob}\left(V'ZV+\frac{1}{4}\Delta'(V)\Delta(V)\leq\epsilon\right)=1,\quad\forall\epsilon>0.$$

We deduce, by right continuity of cumulative distribution functions, that

$$\mathrm{Prob}\left(V'ZV+\frac{1}{4}\Delta'(V)\Delta(V)\leq 0\right)=1.$$

In particular if Z is positive semi-definite

$$\Delta'(V)\Delta(V)=0,\ \text{almost surely}.$$

and thus

$$\|\Delta(V)\|=0,\ \text{almost surely}.$$

But, by Lemma A.1,

$$\|\Delta(V)\|\geq\gamma\|V\|^2.$$

Thus $V=0$, almost surely. In other words, we have shown that

$$\mathrm{Prob}\left(V=0|Z\geq 0\right)=1.$$

Now, we establish that $\mathrm{Prob}\left(V=0\,\middle|\,\overline{(Z\geq 0)}\right)=0$.

The necessary second order condition for an interior solution for a minimization problem implies that for any vector $e\in\mathbb{R}^p$:

$$e'\left(\frac{\partial^2}{\partial\theta\partial\theta'}\left(\bar{\phi}_T'(\theta)\bar{\phi}_T(\theta)\right)\Big|_{\theta=\hat{\theta}_T}\right)e\geq 0.$$

This can be written

$$e'\left(\tilde{Z}_T+N_T\right)e\geq 0, \tag{A.8}$$

where

$$\tilde{Z}_T=\left(\frac{\partial^2\bar{\phi}_T'}{\partial\theta_i\partial\theta_j}(\hat{\theta}_T)\sqrt{T}\bar{\phi}_T(\hat{\theta}_T)\right)_{1\leq i,j\leq p}$$

and

$$N_T=\sqrt{T}\frac{\partial\bar{\phi}_T'}{\partial\theta}(\hat{\theta}_T)\frac{\partial\bar{\phi}_T}{\partial\theta'}(\hat{\theta}_T).$$

By a mean value expansion, we have

$$\frac{\partial \bar{\phi}_T}{\partial \theta_i}(\hat{\theta}_T) = \frac{\partial^2 \bar{\phi}_T}{\partial \theta_i \partial \theta'}(\bar{\theta})(\hat{\theta}_T - \theta^0) + O_P(T^{-1/2}), \tag{A.9}$$

with $\bar{\theta} \in (\theta^0, \hat{\theta}_T)$ and may differ from row to row and $i = 1, \dots, p$. On the other hand, thanks to Equation (A.5), we have

$$\frac{\partial^2 \bar{\phi}'_T}{\partial \theta_i \partial \theta_j}(\hat{\theta}_T)\, \bar{\phi}_T(\hat{\theta}_T) = \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}(\theta^0) \left(\bar{\phi}_T(\theta^0) + \frac{1}{2}\Delta(\hat{\theta}_T - \theta^0)) \right) + o_P(T^{-1/2}).$$

Hence, with $a_{ij} = \frac{\partial^2 \rho}{\partial \theta_i \partial \theta_j}(\theta^0)$,

$$\frac{\partial^2 \bar{\phi}'_T}{\partial \theta_i \partial \theta_j}(\hat{\theta}_T)\, \sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = a'_{ij}\sqrt{T}\bar{\phi}_T(\theta^0) + \frac{1}{2}a'_{ij}\Delta \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right) + o_P(1).$$

Thus

$$\tilde{Z}_T = Z_T + \frac{1}{2}\left(a'_{ij}\Delta(T^{1/4}(\hat{\theta}_T - \theta^0)) \right)_{1 \le i,j \le p} + o_P(1)$$

and

$$N_T = \left(\ T^{1/4}(\hat{\theta}_T - \theta^0)' \tfrac{\partial^2 \rho'}{\partial \theta_i \partial \theta}(\theta^0) \tfrac{\partial^2 \rho}{\partial \theta_j \partial \theta'}(\theta^0) T^{1/4}(\hat{\theta}_T - \theta^0)\ \right)_{1 \le i,j \le p} + o_P(1).$$

From the inequality (A.8) and some successive applications of the Cauchy-Schwarz inequality, we can find a deterministic constant real number $A > 0$ such that for any vector $e \in \mathbb{R}^p$ with unit norm:

$$-e'Z_T e \le A\sqrt{T}\|\hat{\theta}_T - \theta^0\|^2 + o_P(1),$$

By Lemma A.2,

$$\limsup_{T \to \infty} \text{Prob}\left(-e'Z_T e - A\sqrt{T}\|\hat{\theta}_T - \theta^0\|^2 \le \epsilon \right) = 1, \quad \forall \epsilon > 0.$$

Considering again a subsequence along which $(T^{1/4}(\hat{\theta}_T - \theta^0)', \sqrt{T}\bar{\phi}_T(\theta^0)')'$ converges in distribution, we can write, using the Portmanteau Lemma (Lemma 2.2 of van der Vaart (1998)), that

$$\text{Prob}\left(-e'Ze - A\|V\|^2 \le \epsilon \right) = 1, \quad \forall \epsilon > 0.$$

Thus, by right continuity of cumulative distribution functions,

$$\text{Prob}\left(-e'Ze - A\|V\|^2 \le 0 \right) = 1$$

and consequently,

$$\text{Prob}\left(\|V\|^2 \ge -\frac{e'Ze}{A} \,\middle|\, Z = z \right) = 1,\ P^Z a.s. \tag{A.10}$$

In particular, when $Z = z$ non positive semidefinite, we can find a vector $e \in \mathbb{R}^p$ with unit norm and such that $e'Ze < 0$ and thus:

$$\text{Prob}\left(\|V\| > 0 \,|\, Z = z \right) = 1$$

Therefore $\text{Prob}\left(\|V\| > 0 \,\middle|\, \overline{(Z \ge 0)} \right) = 1$ □

Proof of Theorem 3.1. Follows from Lemma A.5-*(iii)*□

Proof of Theorem 3.2. From Lemma A.5, the limiting distribution of J_T is

$$J = \min_{v \in \mathbb{R}^p} \left(X'WX + \text{Vec}'(vv')G'_2G'_1WX + \frac{1}{4}\text{Vec}'(vv')G'_2G'_1WG_1G_2\text{Vec}(vv') \right),$$

where X is the limiting distribution of $\sqrt{T}\bar{\phi}_T(\theta^0)$. Let

$$L = \min_{u\in\mathbb{R}^{p^2}} \left(X'WX + u'G_2'G_1'WX + \frac{1}{4}u'G_2'G_1'WG_1G_2u \right).$$

By definition, along any sample path, we have

$$L \leq J \leq J(0).$$

It is clear that

$$J(0) = X'WX \sim \chi^2_H,$$

since $X \sim N(0, W^{-1})$. Also, along any sample path, the first order condition associated to L is

$$G_2'G_1'WX + \frac{1}{2}G_2'G_1'WG_1G_2\hat{u} = 0.$$

Since G_1 and G_2' have full column rank, we can write

$$G_2\hat{u} = -2(G_1'WG_1)^{-1}G_1'WX.$$

Plugging this back in the definition of L yields

$$L = X'WX - X'WG_1(G_1'WG_1)^{-1}G_1'WX = X'W^{1/2}\left(Id_H - W^{1/2}G_1(G_1'WG_1)^{-1}G_1'W^{1/2}\right)W^{1/2}X \sim \chi^2_{H-p}.$$

Now, we show that the two distributional bounds are conditionally sharp. By some straightforward manipulations, one can verify that

$$\text{Vec}'(vv')G_2'G_1'WX = \text{Vec}'(vv')G'WX = v'Zv$$

so that

$$J = X'WX + \min_{v\in\mathbb{R}^p}\left(v'Zv + \frac{1}{4}\text{Vec}'(vv')G_2'G_1'WG_1G_2\text{Vec}(vv')\right).$$

Hence, conditional on $(Z \geq 0)$, it appears that $J(v)$ is minimized at $v = 0$ and $J = X'WX \sim \chi^2_H$ and we can claim that $J \sim \chi^2_H$ with probability at least equal to $\text{Prob}(Z \geq 0)$. The probability of $J \sim \chi^2_H$ is actually exactly equal to $\text{Prob}(Z \geq 0)$ because we can show that, when Z is not positive semidefinite, $J < \chi^2_H$.

To see that, it is sufficient to show that when Z is not positive semidefinite, we can find $v \in \mathbb{R}^p$ such that:

$$v'Zv + \frac{1}{4}\text{Vec}'(vv')G_2'G_1'WG_1G_2\text{Vec}(vv') < 0.$$

This is true because, not only we can of course find $v \in \mathbb{R}^p$ such that $v'Zv < 0$, but we can also impose:

$$-v'Zv > g(v) = \frac{1}{4}\text{Vec}'(vv')G_2'G_1'WG_1G_2\text{Vec}(vv')$$

since, if not true for v, it will be true for $v(\lambda) = \lambda v$ insofar as:

$$-v'Zv > \lambda^2 g(v)$$

which will always be true for sufficiently small $\lambda \in \mathbb{R}$.

Finally, we show that $J \sim \chi^2_{H-p}$ with positive probability.

Note that conditional on $(G_2\hat{u} \in S \equiv \{G_2\text{Vec}(vv') : v \in \mathbb{R}^p\})$, $J = L$ and $J \sim \chi^2_{H-p}$. Let us evaluate $\text{Prob}(G_2\hat{u} \in S)$. From (A.1), the columns of G_2 are the column vectors

$$\frac{\partial^2 \text{Diag}(\Lambda'\theta_*\theta_*'\Lambda)}{\partial\theta_i\partial\theta_j};\ 1 \leq i, j \leq p.$$

By some tedious but straightforward calculations, we have

$$\frac{\partial^2 \text{Diag}(\Lambda'\theta_*\theta_*'\Lambda)}{\partial\theta_i\partial\theta_j} = 2\left\{[\Lambda_{i\cdot}' - \Lambda_{n\cdot}'] \odot [\Lambda_{j\cdot}' - \Lambda_{n\cdot}']\right\}$$

where $\Lambda_{i\cdot}, i = 1, \ldots, p$ is the i-th row of Λ and $\odot$ denotes the Hadamard element-by-element product of vectors. Hence $\forall v \in \mathbb{R}^p$,

$$G_2 \text{Vec}(vv') = 2 \left(\left[\sum_{i=1}^{p} (\Lambda_{ik} - \Lambda_{nk}) v_i \right]^2 \right)_{1 \le k \le p}.$$

From Lemma 2.3, since G_1 has full column rank, we have $(G_2 \text{Vec}(vv') = 0 \Leftrightarrow v = 0)$ so that the linear function

$$v \mapsto \left(\sum_{i=1}^{p} (\Lambda_{ik} - \Lambda_{nk}) v_i \right)_{1 \le k \le p}$$

is a one-to-one mapping of $\mathbb{R}^p$ on itself. Therefore, $S = \{G_2 \text{Vec}(vv') : v \in \mathbb{R}^p\} = \mathbb{R}^p_+$.

Also, note that $G_2 \hat{u} \sim N(0, 4(G_1' W G_1)^{-1})$ and, since $G_1' W G_1$ is positive definite, $\text{Prob}(G_2 \hat{u} \in \mathbb{R}^p_+) > 0$. We can conclude that J is distributed as a χ^2_{H-p} with probability $q_2 \ge \text{Prob}(G_2 \hat{u} \in \mathbb{R}^p_+) > 0$. This completes the proof of the first part of the theorem.

Next, we complete the derivation of the asymptotic distribution of J_T when $p = 1$. In this case, G is an H-vector defined by $G = \partial^2 \rho(\theta^0)/\partial \theta^2$, $Z = G' W X$ and $\hat{u} = -2Z/(G' W G)$. So, $(\hat{u} > 0) = (Z < 0)$ and $\text{Prob}(Z \ge 0) = \text{Prob}(Z < 0) = 1/2$. From the previous lines, conditional on $(Z \ge 0)$, $J \sim \chi^2_H$ and conditional on $(\hat{u} > 0)$, $J \sim \chi^2_{H-1}$. Hence, J is distributed as a χ^2_H with probability 1/2 and as a χ^2_{H-1} also with probability 1/2. We can therefore claim that

$$J \sim \frac{1}{2}\chi^2_{H-1} + \frac{1}{2}\chi^2_H$$

□

References

[1] Billingsley, P., 1961. "The Lindeberg-Lévy Theorem for Martingales," *Proc. Amer. Math. Soc.*, 12, 788-792.

[2] Chamberlain, G., 1986. "Asymptotic Efficiency in Semi-Parametric Models with Censoring," *Journal of Econometrics*, 32, 189-218.

[3] Cragg, J.G. and S. G. Donald, 1993. "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

[4] Cragg, J.C. and S.G. Donald, 1996. "Testing Overidentifying Restrictions in Unidentified Models," *Unpublished UBC discussion paper*, 96/20.

[5] Diebold, F. and M. Nerlove, 1989. "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor Arch Model," *Journal of Applied Econometrics*, 4, 1-21.

[6] Dovonon, P. and S. Gonçalves, 2011. "Bootstrapping Gmm Tests Under First Order Underidentification," *work in progress*, Concordia University.

[7] Dovonon, P. and E. Renault, 2009. "Gmm Overidentification Test with First Order Underidentification," *working paper*, UNC, http://www.unc.edu/depts/econ/profs/renault/ J_testDR20090824.pdf.

[8] Doz, C. and E. Renault, 2006. "Factor Volatility in Mean Models: a Gmm Approach," *Econometric Reviews*, 25, 275-309.

[9] Engle, R. F., V. K. Ng and M. Rothschild, 1990. "Asset Pricing with a Factor-Arch Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, 45, 213-237.

[10] Engle, R. F. and S. Kozicki, 1993. "Testing For Common Features," *Journal of Business and Economic Statistics*, 11(4), 369-395.

[11] Fiorentini, G., E. Sentana and N, Shephard, 2004. "Likelihood-based Estimation of Generalised arch Structures," *Econometrica*, 72, 1481-1517.

[12] Hansen, L. P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[13] Horn, A. R. and C. R. Johnson 1985. "Matrix Analysis," *Cambridge University Press.*

[14] Koul, H. L., 2002. "Weighted Empirical Processes in Dynamic Nonlinear Models," Springer-Verlag, New York.

[15] Lee, L. F. and A. Chesher, 1986. "Specification Testing when Score Test Statistics are Identically Zero," *Journal of Econometrics*, 31, 121-149.

[16] Melino, A., 1982. "Testing for Sample Selection Bias," *Review of Economic Studies*, 49, 151-153.

[17] Phillips, P. C. B., 1989. "Partially Identified Econometric Models," *Econometric Theory*, 5, 151-240.

[18] Rotnitzky, A., D. R. Cox, M. Bottai and J. Robins, 2000. "Likelihood-based Inference with Singular Information Matrix," *Bernoulli*, 6(2), 243-284.

[19] Sargan, J. D., 1983. "Identification and lack of Identification," *Econometrica*, 51, 1605-1633.

[20] Schott, J. R., 1984. "Optimal Bounds for the Distribution of some Test Criteria for Dimensionality," *Biometrica*, 71, 561-567.

[21] Staiger, D. and J. H. Stock, 1997. "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

[22] Stock, J. H. and J. H. Wright, 2000. "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.

[23] van der Vaart, A. W., 1998. "Asymptotic Statistics," *Cambridge University Press.*

[24] van der Vaart, A. W. and J. A. Wellner, 1996. "Stochastic Convergence and Empirical Processes," Springer-Verlag, New York.