

2012s-34

Testing for Common GARCH Factors

Prosper Dovonon, Éric Renault

Série Scientifique
Scientific Series

Montréal
Décembre 2012

© 2012 Prosper Dovonon, Éric Renault. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©.

Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.



Centre interuniversitaire de recherche en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère du Développement économique et régional et de la Recherche, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère du Développement économique et régional et de la Recherche, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO

Partenaire majeur

Ministère du Développement économique,
de l'Innovation et de l'Exportation

Partenaires corporatifs

Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque Laurentienne du Canada
Banque Nationale du Canada
Banque Royale du Canada
Banque Scotia
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Fédération des caisses Desjardins du Québec
Financière Sun Life, Québec
Gaz Métro
Hydro-Québec
Industrie Canada
Investissements PSP
Ministère des Finances du Québec
Power Corporation du Canada
Rio Tinto Alcan
State Street Global Advisors
Transat A.T.
Ville de Montréal

Partenaires universitaires

École Polytechnique de Montréal
HEC Montréal
McGill University
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.

ISSN 1198-8177

Partenaire financier
Développement
économique, Innovation
et Exportation
Québec 

Testing for Common GARCH Factors^{*}

Prosper Dovonon[†], Éric Renault[‡]

Résumé / Abstract

Cet article propose un test pour la détection de caractéristiques communes d'hétéroscédasticité conditionnelle (HC) dans des rendements d'actifs financiers. Conformément à Engle et Kozicki (1993), l'existence de caractéristiques communes HC est exprimée en termes de conditions de moment sur-identifiantes testables. Cependant nous montrons que ces conditions de moment ne sont pas localement linéairement indépendantes; la matrice Jacobienne est nulle à la vraie valeur des paramètres et, par conséquent, la théorie asymptotique de Hansen (1982) ne s'applique pas. Nous montrons dans ce contexte que la statistique de J-test de Hansen (1982) est distribuée asymptotiquement comme le minimum de la limite d'un processus empirique avec une distribution non standard. Quand on considère deux actifs, cette distribution asymptotique est un mélange à parts égales de χ_{H-1}^2 et χ_H^2 , où H est le nombre de conditions de moment, par opposition à χ_{H-1}^2 . Avec plus de deux actifs, cette distribution est comprise entre χ_{H-p}^2 et χ_H^2 (p , le nombre de paramètres). Ces résultats montrent que l'ignorance du défaut d'identification au premier ordre dans ce modèle de conditions de moments conduit à des tests qui rejettent trop souvent l'hypothèse nulle, le degré de sur-rejet étant croissant avec le nombre d'actifs. Une étude de Monte-Carlo illustre ces résultats..

This paper proposes a test for common conditionally heteroskedastic (CH) features in asset returns. Following Engle and Kozicki (1993), the common CH features 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_{H-1}^2 and χ_H^2 , where H is the number of moment conditions, as opposed to a χ_{H-1}^2 . With more than two assets, this distribution lies between the χ_{H-p}^2 and χ_H^2 (p , the number of parameters). 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.

Mots clés : Common features, GARCH factors, Nonstandard asymptotics, GMM, GMM overidentification test, identification, first order identification

* We would like to thank Manuel Arellano, Yves Atchadé, Valentina Corradi, Giovanni Forchini, Sílvia Gonçalves, Enrique Sentana and James Stock for very helpful comments and suggestions. The second author has benefited from the hospitality of CIRANO during the revision of this paper and is grateful for that.

[†] 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.

[‡] Brown University, CIRANO and CIREQ; Address: Department of Economics, 64 Waterman Street, Brown University, Providence, RI 02912; Phone: (401) 863-3519, Email: Eric.Renault@brown.edu.

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 exemplified 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 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 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 modeling 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, albeit seemingly related to the recent literature on weak identification, is not redundant with extant results. By contrast with the common weak identification setting (Staiger and Stock (1997), Stock and Wright (2000)), we share with the setting of Andrews and Cheng (2012) the fact that “in the present paper, the potential source of weak identification is an explicit part of the model”. Irrespective of the choice of instruments and regardless of any finite sample issue, the valid asymptotic distribution of the J -test statistic under the null involves a mixture of chi-squares with fewer degrees of freedom than the standard $\chi^2(H - n + 1)$. As a result, the rank deficiency leads to an oversized test when the J -test setting is compared to standard critical values. This is in sharp contrast with the common intuition (see e.g. Cragg and Donald (1993, 1996)) that rank deficiency should lead to conservative tests, since the restrictions under test would be less binding than they seem to be. In our case, all the parameters are actually identified and, due to the rank deficiency of the Jacobian matrix, the J -test statistic may not be as sensitive to parameter variation as it is in standard settings; chi-square distributions with fewer degrees of freedom show up as if some parameter were actually known.

The fact that the parameters of interest are always identified also implies that our setting, and the setting of Sargan (1983) as well, do not naturally fit into the general framework for identification put forward by Andrews and Cheng (2012). It would take a quite convoluted re-parameterization of our model to handle it with Andrews and Cheng’s (2012) toolbox of models for which some parameters are unidentified in some parts of the parameter space. In the context of maximum likelihood estimation (MLE), several authors have met a situation of local singularity similar to ours. Melino (1982), Lee and Chesher (1986) and Rotnitzky, Cox, Bottai and Robins (2000) have documented the non-standard rates of convergence of MLE implied by the singularity of the Fisher information matrix. Of course, the issue of singularity of the Jacobian matrix in GMM is germane to singularity of Fisher information matrix in MLE context, and, following Sargan (1983), we get non-standard rates of convergence of GMM estimators for quite similar reasons. We actually provide an interpretation in terms of random Fisher information matrix closely related to the analysis proposed by Andrews and Mikusheva (2012). However, our main focus of interest is not the asymptotic distribution of GMM estimators but 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. Moreover, our asymptotic result is new since it gives a well-defined asymptotic distribution for a test statistic while extant results could only acknowledge

that there is no such thing as a well-defined rate of convergence for estimators.

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 $\|A\| = (\text{tr}(AA'))^{1/2}$, 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 common CH features

A n -dimensional stochastic process $(Y_t)_{t \geq 0}$ is said to have $(n - K)$ time-invariant (Conditionnally Heteroskedastic) CH common features, $K < N$, if it has a conditional covariance matrix given by:

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

where:

- (i) D_t is a diagonal matrix of size K with diagonal coefficients σ_{kt}^2 , $k = 1, \dots, K$.
- (ii) Λ is a $n \times K$ matrix and Ω is an $n \times n$ symmetric positive semi-definite matrix.
- (iii) The stochastic processes $(Y_t)_{t \geq 0}$ and $(\sigma_{kt}^2)_{1 \leq k \leq K, t \geq 0}$ are adapted with respect to the increasing filtration $(\mathfrak{F}_t)_{t \geq 0}$.

In this context, CH common features are by definition any vector θ in \mathbb{R}^n such that $\text{Var}(\theta'Y_{t+1}|\mathfrak{F}_t)$ is constant. The decomposition (1) clearly warrants the existence of at least $(n - K)$ directions of

common features since the vectors θ such that $\theta'\Lambda = 0$ fulfill the required condition. Moreover, we will see that the dimension $(n - K)$ can be defined without ambiguity thanks to the following maintained assumption:

Assumption 1. (i) $\text{Rank}(\Lambda) = K$. (ii) $\text{Var}(\text{Diag}(D_t))$ is non-singular, where $\text{Diag}(D_t)$ is the K -dimensional vector with coefficients $D_{kk,t}(= \sigma_{kt}^2)$, $k = 1, \dots, K$.

Remark 2.1. A common intuition about the variance decomposition (1) is a (GAR)CH factor model with K factors and constant factor loadings:

$$Y_{t+1} = \mu_t + BF_{t+1} + u_{t+1},$$

$$\text{Var}(F_{t+1}|\mathfrak{F}_t) = V_t, E(F_{t+1}|\mathfrak{F}_t) = 0, \text{Var}(u_{t+1}|\mathfrak{F}_t) = \Omega, E(u_{t+1}|\mathfrak{F}_t) = 0, \text{and Cov}(F_{t+1}, u_{t+1}|\mathfrak{F}_t) = 0.$$

As recently developed by Hecq, Laurent and Palm (2012), this CH factor model can be tightly related to CH common features by a simple diagonalization of the conditional variance matrix of the factors:

$$V_t = P_t D_t P_t',$$

with D_t a diagonal matrix and P_t an orthogonal matrix. Then

$$\text{Var}(Y_{t+1}|\mathfrak{F}_t) = BP_t D_t P_t' B' + \Omega. \quad (2)$$

One may then see (2) as a convenient generalization of our model (1) by considering time varying factor loadings

$$\Lambda_t = BP_t.$$

However, this more general framework (where V_t and thus P_t are not diagonal) does not fit into our model of CH common features for the following reason. Our focus of interest is the set of portfolio returns $\theta'Y_{t+1}$, $\theta \neq 0$ with constant conditional variance. However, in the framework (2), since:

$$\text{Var}(\theta'Y_{t+1}|\mathfrak{F}_t) = \theta'BP_t D_t P_t' B' \theta + \theta' \Omega \theta$$

we see that it amounts to eliciting θ such that the vector $D_t^{1/2} P_t' B' \theta$ has a constant norm. We want to characterize the CH common feature as a simple algebraic property of the vector θ of portfolio weights. It would clearly take in general some highly convoluted assumptions about the joint dynamics of the coefficients of the matrix $D_t^{1/2} P_t'$ to deduce the required property of θ from the condition that the norm of $D_t^{1/2} P_t' B' \theta$ is constant. The only natural way to characterize easily the time-invariance of the above norm as an algebraic property of the vector θ of common features is to assume that $D_t^{1/2} P_t'$ is diagonal, that is P_t (and in turn V_t) is diagonal¹. In other words, we need to preclude conditional correlations between the latent GARCH factors. Then the change of basis P_t is immaterial and from now on,

¹All the results of this paper could be generalized to the case where no linear combination of $\text{Vech}(P_t D_t P_t')$ is constant. However, beyond the diagonal case with the maintained Assumption 1, such a restriction seems hard to interpret.

we can interpret the CH features as possibly (but not necessarily) underpinned by a (GAR)CH factor model with uncorrelated factors and constant factor loadings:

$$Y_{t+1} = \mu_t + \Lambda F_{t+1} + u_{t+1}, \quad (3)$$

$\text{Var}(F_{t+1}|\mathfrak{F}_t) = D_t$, $E(F_{t+1}|\mathfrak{F}_t) = 0$, $\text{Var}(u_{t+1}|\mathfrak{F}_t) = \Omega$, $E(u_{t+1}|\mathfrak{F}_t) = 0$, and $\text{Cov}(F_{t+1}, u_{t+1}|\mathfrak{F}_t) = 0$.

Then CH common features are vectors θ such that $\theta'Y_{t+1} = \theta'u_{t+1}$.

Remark 2.2. In the context of factor model (3), restricting Λ to be full column rank basically means that one cannot reduce the dimension K of the vector F_t of factors. However, since the testable implications of our model are encapsulated in the decomposition of conditional variance (1), irrespective of the latent factors F_t , we need to maintain instead that the conditional variances σ_{kt}^2 , $k = 1, 2, \dots, K$, cannot be linearly combined to erase conditional heteroskedasticity. Hence Assumption 1. Irrespective of its specific interpretation, our general framework (1) along with Assumption 1 allows us to characterize the CH-common features as the null space of the matrix Λ' :

Lemma 2.1. Under Assumption 1, the CH common features are the vectors θ , solution in \mathbb{R}^n , $\theta \neq 0$, of

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

Proof: See Appendix B.

The key idea of this paper is to test for the existence of CH-common features through the unpredictability of squared returns $(\theta'Y_{t+1})^2$, that is the null hypothesis:

H_0 : There exists $\theta \in \mathbb{R}^n$, $\theta \neq 0$ such that $E((\theta'Y_{t+1})^2|\mathfrak{F}_t)$ is constant.

Remark 2.3. It is worth noting that H_0 is equivalent to θ CH-common feature only if one assumes that $E(\theta'Y_{t+1}|\mathfrak{F}_t)$ is constant. This assumption does not preclude predictability of returns. It only maintains, in line with GARCH-in-mean modeling, that predictability of returns goes through conditional variance so that (see Doz and Renault (2006) for more discussion):

$$Y_{t+1} = \alpha + \Lambda\xi_t + \Lambda F_{t+1} + u_{t+1},$$

where ξ_t is a vector of risk premiums associated to the common factors. Then when θ is a CH-common feature, $E(\theta'Y_{t+1}|\mathfrak{F}_t) = \theta'\alpha$ is constant.

As usual, the null hypothesis H_0 will be tested through a test of its consequence $H_0(z)$ for a given choice of a H -dimensional vector z_t of instruments:

$H_0(z)$: There exists $\theta \in \mathbb{R}^n, \theta \neq 0$, such that : $E(z_t[(\theta'Y_{t+1})^2 - c(\theta)]) = 0$,
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 CH 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 CH common features model (1) gives:

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

and then, by a simple matrix manipulation,

$$\begin{aligned} E(z_t((\theta'Y_{t+1})^2 - c(\theta))) &= \text{Cov}(z_t, \text{tr}(\theta' \Lambda D_t \Lambda' \theta)) = \text{Cov}(z_t, \text{Diag}'(\Lambda' \theta \theta' \Lambda) \text{Diag}(D_t)) \\ &= \text{Cov}(z_t, \text{Diag}(D_t)) \text{Diag}(\Lambda' \theta \theta' \Lambda). \end{aligned} \tag{4}$$

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

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

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_{kt}^2 , $k = 1, \dots, K$, there exists at least one instrument z_{ht} , $h = 1, \dots, H$ correlated with this combination.

From (4), we see that under Assumptions 1 and 2, $H_0(z)$ amounts to:

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

and then implies that $\|\Lambda' \theta\|^2 = \text{tr}(\Lambda' \theta \theta' \Lambda) = 0$ 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 are the solutions θ in \mathbb{R}^n , $\theta \neq 0$, of the moment restrictions:

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

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

As in Engle and Kozicki (1993), CH 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.

Following Hansen (1982) as well as Engle and Kozicki (1993), we aim in this paper at testing the CH common features model through a J -test of overidentification applied to the moment conditions $H_0(z)$. In line with Hansen (1982), we will maintain the following assumption:

Assumption 3. (z_t, Y_t) is a stationary and ergodic process such that $E(\|z_t\|^2) < \infty$ and $E(\|Y_t\|^4) < \infty$. Moreover, both z_t and $\text{vec}(Y_t Y_t')$ fulfill a central limit theorem.

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 (such as $\sum_{i=1}^n \theta_i = 1$). 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 common feature. To see this, note that by virtue of the square integrability conditions in Assumption 3, we can change the order of expectation and differentiation in the following and write:

$$\begin{aligned} \Gamma(\theta) &= \frac{\partial}{\partial \theta'} E(z_t((\theta' Y_{t+1})^2 - c(\theta))) = E[z_t \{2(\theta' Y_{t+1}) Y_{t+1}' - 2E[(\theta' Y_{t+1}) Y_{t+1}']\}] \\ &= 2\text{Cov}(z_t, [Y_{t+1} Y_{t+1}'] \theta). \end{aligned}$$

Then by the law of iterated expectations,

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

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

Proposition 2.1. *Under Assumption 3, for any common feature θ ,*

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

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. Our main focus of interest will then be the impact of this rank failure on the behavior of the J -test statistic for $H_0(z)$, both asymptotically and in finite sample. However, it is worth stressing that, as discussed in Appendix A, a test for the CH common features model based on a more standard regression-based approach of testing for GARCH effects would not allow to circumvent the problem of the rank failure.

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 \mathcal{N} :

Assumption 4. $\theta \in \Theta_* \subset \mathbb{R}^n$ such that $\Theta^* = \Theta_* \cap \mathcal{N}$ is a compact set and

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

Remark 2.4. A typical normalization condition would be the unit cost condition to interpret $\theta'Y_{t+1}$ as a return per \$ invested:

$$\mathcal{N} = \left\{ \theta \in \mathbb{R}^n, \sum_{i=1}^n \theta_i = 1 \right\}.$$

This normalization can be maintained without loss of generality except if one wants to consider arbitrage portfolios for which $\sum_{i=1}^n \theta_i = 0$. Then, an alternative normalization condition would be:

$$\mathcal{N} = \{ \theta \in \mathbb{R}^n, \theta_1 = 1 \}.$$

Note that the latter choice implies that we know one particular asset, e.g. Asset 1, entering the common feature. To avoid this assumption, an alternative would be:

$$\mathcal{N} = \left\{ \theta \in \mathbb{R}^n, \sum_{i=1}^n \theta_i^2 = 1 \right\}.$$

This latter, albeit feasible, will not be explicitly considered to keep the simplicity of linear normalization. In this context, assumption 4 will be fulfilled with $\Theta_* = \Theta$, that is $\Theta^* = \Theta \cap \mathcal{N}$ in the setting of Engle and Kozicki (1993), that is $K = n - 1$. If more than one dimension of common features exist ($K < n - 1$), a practitioner may typically write some exclusion restrictions (like zero weight for some particular assets) to define a proper subset Θ_* of Θ such that Assumption 4 is fulfilled. Note that in this latter case, we are formally back to the particular case $K = n - 1$ by excluding the assets that do not enter into the definition of a given common feature. More precisely, if the time varying conditional heteroskedasticity of n assets can be captured with K factors, $K = n - 1 - q$, $q > 0$, a natural parsimonious approach leads to look for $q + 1$ linearly independent common features, each of them involving only $n - q$ assets. Then a normalization condition \mathcal{N} is sufficient for identification.

Under Assumptions 1, 2 and 4, global identification amounts to second-order identification:

Lemma 2.3. Under Assumptions 1, 2 and 4, with

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

we have, $\forall \theta \in \Theta^*$,

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

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:

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

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

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 key idea of Engle and Kozicki (1993) was to apply the theory of J -test for overidentification to the moment conditions defined by $H_0(z)$:

$$E(\psi_t(\theta)) = 0; \quad \psi_t(\theta) = z_t((\theta'Y_{t+1})^2 - c(\theta)); \quad \theta \in \Theta^*.$$

As already announced, the main point of this paper is that the standard asymptotic theory as derived by Hansen (1982) will not work due to failure of the rank condition (see Proposition 2.1):

$$E\left(\frac{\partial\psi_t(\theta)}{\partial\theta'}\Bigg|_{\theta=\theta^0}\right) = 0.$$

An additional issue worth addressing, albeit much simpler, is the need to replace, as Engle and Kozicki (1993) do, the above unknown function $c(\theta)$ by a feasible sample counterpart. We first sketch the relevant asymptotic theory for these two issues before focusing on the overidentification test of interest.

3.1 Feasible moment conditions

Throughout, we will rather work with the following feasible moment conditions:

$$E(\phi_{t,T}(\theta)) = 0; \quad \phi_{t,T}(\theta) = (z_t - \bar{z}_T)((\theta'Y_{t+1})^2 - \bar{c}_T(\theta)); \quad \theta \in \Theta^*, \quad (5)$$

where $\bar{z}_T = \frac{1}{T} \sum_{t=1}^T z_t$ and $\bar{c}_T(\theta) = \frac{1}{T} \sum_{t=1}^T (\theta'Y_{t+1})^2$.

As a result, the moment conditions have now a structure of double array and the GMM asymptotic distributional theory will then follow from a central-limit theorem applied to the sample mean of this double array:

$$\bar{\phi}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \phi_{t,T}(\theta).$$

Then, for any given θ , we can relate the sample mean $\bar{\phi}_T(\theta)$ of feasible moments to two sample means without any double array:

$$\sqrt{T}\bar{\phi}_T(\theta) = \sqrt{T}\bar{\psi}_T(\theta) + \sqrt{T}\bar{v}_T(\theta) + o_P(1), \quad (6)$$

where $\bar{v}_T(\theta) = \frac{1}{T} \sum_{t=1}^T v_t(\theta)$, $v_t(\theta) = \mu_z (c(\theta) - (\theta' Y_{t+1})^2)$ and $\mu_z = E(z_t)$. (See the explicit derivation of (6) as part of the proof of Corollary 3.1.)

Note in addition that it follows from Lemma 2.2 and Proposition 2.1 that, under the null H_0 , both $(\psi'_t(\theta^0), v'_t(\theta^0))'$ and $\left. \frac{\partial \psi_t}{\partial \theta'}(\theta) \right|_{\theta=\theta^0}$ are martingale difference sequences. Then, the central limit theorem of Billingsley (1961) for stationary ergodic martingale difference sequences implies that $\sqrt{T}(\bar{\psi}'_T(\theta^0), \bar{v}'_T(\theta^0))'$ and $\left. \sqrt{T} \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta) \right|_{\theta=\theta^0}$ are asymptotically normal. Note that $\left. \sqrt{T} \frac{\partial \bar{v}_T}{\partial \theta'}(\theta) \right|_{\theta=\theta^0} = \mu_z \cdot \sqrt{T} \left(\left. \frac{\partial c}{\partial \theta'}(\theta) \right|_{\theta=\theta^0} - 2\theta^{0'} \frac{1}{T} \sum_{t=1}^T Y_t Y_t' \right)$ is also asymptotically normal.

Overall, we will use the fact that, first $\sqrt{T}(\bar{\psi}'_T(\theta^0), \bar{v}'_T(\theta^0))'$ is asymptotically normal by the central limit theorem for martingale difference sequences and second $\sqrt{T} \left(\left. \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta) \right|_{\theta=\theta^0}, \left. \frac{\partial \bar{v}_T}{\partial \theta'}(\theta) \right|_{\theta=\theta^0} \right) = O_P(1)$ to conclude:

Corollary 3.1. *If Assumptions 1, 2, 3 and 4 hold, then $\sqrt{T}\bar{\phi}_T(\theta^0)$ is asymptotically normal with asymptotic variance $\Sigma(\theta^0)$ given by:*

$$E \left[(z_t - \mu_z)(z_t - \mu_z)' \left((\theta^{0'} Y_{t+1})^2 - c(\theta^0) \right)^2 \right]$$

and $\left. \sqrt{T} \frac{\partial \bar{\phi}_T}{\partial \theta'}(\theta) \right|_{\theta=\theta^0} = O_P(1)$.

Proof: See Appendix B.

Remark 3.1. *Note that if one interprets the null hypothesis H_0 as stemming from the GAR(CH) factor model (3) and one assumes in addition that the idiosyncratic terms u_t are independent from instruments z_t , then, by the law of iterative expectations, one can use an even simpler formula for the asymptotic variance matrix under H_0 :*

$$\Sigma(\theta^0) = \text{Var} \left((\theta^{0'} Y_{t+1})^2 \right) \text{Var}(z_t).$$

It is worth noting 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 $(\bar{\phi}_T(\theta))_{\theta \in \Theta}$. 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 4 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.

For the purpose of identification, we consider in the rest of the paper $K = n - 1$ along with the normalization $\mathcal{N} = \{\theta \in \mathbb{R}^n : \sum_{i=1}^n \theta_i = 1\}$. In the light of Remark 2.4, setting K to $n - 1$ is not

overly restrictive². We shall also mention that our results do not depend upon our particular choice of linear normalization as long as the identification of θ^0 along Assumption 4 is guaranteed.

Writing $\theta_n = 1 - \sum_{i=1}^{n-1} \theta_i$, the effective parameter set is given by the projection of Θ^* on its $n - 1$ first components. For economy of notation, we keep Θ to denote this parameter set and $\theta = (\theta_i)_{1 \leq i \leq n-1} \in \Theta \subset \mathbb{R}^{n-1}$ the parameter of interest. We shall consider the functions $\phi_t(\theta)$, $\bar{\phi}_T(\theta)$ and $\rho(\theta)$ as defined on $\Theta \subset \mathbb{R}^{n-1}$.

Thus, from now on, we define the GMM estimator $\hat{\theta}_T$ as

$$\hat{\theta}_T \equiv \arg \min_{\theta \in \Theta \subset \mathbb{R}^{n-1}} \bar{\phi}'_T(\theta) W_T \bar{\phi}_T(\theta). \quad (7)$$

For sake of notational simplicity, we will often denote by p ($= n - 1$) the dimension of the unknown vector θ of parameters of interest. We can now be more precise about the rate of convergence of this estimator.

3.2 Rates of convergence

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) $T^{1/2}$ consistent estimator for θ . The intuition of this result is quite simple. In the Taylor expansion of the sample average of $\phi_{t,T}(\theta)$, the first order term $(\partial \bar{\phi}_T(\theta^0) / \partial \theta')(\hat{\theta}_T - \theta^0)$, 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 $\|\hat{\theta}_T - \theta^0\|^2$. Hence the following result:

Proposition 3.1. *Under Assumptions 1, 2, 3, 4, if $\hat{\theta}_T$ is the GMM estimator as defined by Equation (7),*

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

Proof: See Appendix B.

Proposition 3.1 ensures a convergence rate for the GMM estimator $\hat{\theta}_T$ at least as fast as $T^{1/4}$ but possibly less than the standard $T^{1/2}$. The next Proposition 3.2 will complete this statement as follows. On the one hand, it will prove directly that, as expected from the Chamberlain's impossibility result, $T^{1/4}(\hat{\theta}_T - \theta^0)$ does not converge to zero in probability and thus the slow rate $T^{1/4}$ may prevail. On the other hand, it also proves that $T^{1/4}(\hat{\theta}_T - \theta^0)$ does converge to zero with a positive probability. In other words, depending on what part of the sample space the infinite "observed" sample lies in, the rate of convergence may be either $T^{1/4}$ or faster, possibly $T^{1/2}$.

Remark 3.2. *It is worth interpreting the heterogeneity of rates of convergence across the sample space in terms of randomness of a (population) matrix that may be seen as a Fisher information matrix.*

²It is worth mentioning that all of our results remain valid for $K < n - 1$ with a suitable definition of Θ_* . In this instance, the degrees of freedom derived in Theorem 3.2 need to be carefully adapted.

While randomness of the information matrix is known to occur in some non-ergodic settings, it has been recently considered as a possibility in the context of weak identification by Andrews and Mikusheva (2012), even though they eventually preclude this possibility by their maintained Assumption 1(b). By a slight abuse of language, we will use here their information theoretic terminology, even though we are in a GMM context more general than Maximum Likelihood for which Fisher information matrices are usually defined. The GMM analogs of the score vector and of the Hessian of the log-likelihood will be defined from the criterion function:

$$Q_T^W(\theta) = \frac{T}{2} \bar{\phi}_T'(\theta) W_T \bar{\phi}_T(\theta).$$

The GMM analog of the outer product of the score is then:

$$I_T^W(\theta) = \frac{1}{T} \frac{\partial Q_T^W(\theta)}{\partial \theta} \cdot \frac{\partial Q_T^W(\theta)}{\partial \theta'} = \frac{\partial \bar{\phi}_T'(\theta)}{\partial \theta} W_T \left(\sqrt{T} \bar{\phi}_T(\theta) \right) \left(\sqrt{T} \bar{\phi}_T'(\theta) \right) W_T \frac{\partial \bar{\phi}_T(\theta)}{\partial \theta'}$$

such that, under regularity conditions, we have at true value $\theta = \theta^0$:

$$I^W(\theta^0) = \lim_{T \rightarrow \infty} E(I_T^W(\theta^0)) = \Gamma'(\theta^0) W \Sigma(\theta^0) W \Gamma(\theta^0).$$

The GMM analog of the Hessian matrix of the log-likelihood is:

$$H_T^W(\theta) = \frac{1}{T} \frac{\partial^2 Q_T^W(\theta)}{\partial \theta \partial \theta'} = \frac{\partial \bar{\phi}_T'(\theta)}{\partial \theta} W_T \frac{\partial \bar{\phi}_T(\theta)}{\partial \theta'} + (h_{ijT}(\theta) W_T \bar{\phi}_T(\theta))_{1 \leq i, j \leq p}, \quad (8)$$

where:

$$h_{ijT}(\theta) = \frac{\partial^2 \bar{\phi}_T(\theta)}{\partial \theta_i \partial \theta_j}.$$

In particular, under regularity conditions, we have at true value $\theta = \theta^0$:

$$H^W(\theta^0) = \text{plim}_{T \rightarrow \infty} H_T^W(\theta^0) = \Gamma'(\theta^0) W \Sigma(\theta^0) W \Gamma(\theta^0).$$

In the (standard) strong identification case, $\Gamma(\theta^0)$ is full column rank. Then for large T , both the expected outer product matrix $E(I_T^W(\theta^0))$ and the Hessian matrix $H_T^W(\theta^0)$ are positive definite with probability one. Moreover, for the efficient choice $W = \Sigma^{-1}(\theta^0)$ of the weighting matrix, the difference $(E(I_T^W(\theta^0)) - H_T^W(\theta^0))$ converges to zero in probability. This generalization to non-maximum likelihood contexts of the so-called second informational equality has been put forward by Gourieroux and Monfort (1989) as the necessary and sufficient condition to keep the asymptotic equivalence between the standard asymptotic tests.

The situation is much different in the context of weak identification (drifting DGP such that $\Gamma(\theta^0) = O(1/\sqrt{T})$, see e.g. Kleibergen (2005)) or in our context ($\Gamma(\theta^0) = 0$). Then, as stressed by Andrews and Mikusheva (2012), “the difference between the two information matrices is asymptotically non-negligible compared with the information measure” itself. While they point this out in a maximum likelihood context, this statement remains true in a GMM context, regardless of the choice of the weighting matrix W . There is however an important difference between our setting and the common weak identification framework:

In the common “GMM with weak identification” asymptotics as developed by Stock and Wright (2000), the drifting DGP introduces a perverse factor ($1/\sqrt{T}$) at the level of the moment condition itself and this factor will go through all derivatives of moment conditions. Then, both $E(I_T^W(\theta^0))$, $H_T^W(\theta^0)$ and their difference as well are all of order $1/T$.

In our framework, while $E(I_T^W(\theta^0))$ is still of order $1/T$, the Hessian matrix $H_T^W(\theta^0)$ is now dominating since we can deduce from (8) that

$$\sqrt{T}H_T^W(\theta^0) \xrightarrow{d} Z(X),$$

where:

$$Z(X) = (h'_{ij}(\theta^0)WX)_{1 \leq i, j \leq p}, \quad h'_{ij}(\theta^0) = \text{plim}_{T \rightarrow \infty} h'_{ijT}(\theta^0),$$

and X is defined by the Gaussian limit in distribution of $\sqrt{T} \bar{\phi}_T(\theta^0)$:

$$\sqrt{T}\bar{\phi}_T(\theta^0) \xrightarrow{d} X \sim N(0, \Sigma(\theta^0)).$$

Note that since the variance matrix $\Sigma(\theta^0)$ is non-singular, $Z(X)$ is a non-degenerate random matrix, the coefficients of which are all Gaussian with zero mean. In particular, the matrix $Z(X)$ is positive semi-definite if and only if $\text{Vec}(Z(X))$ fulfills p multilinear inequalities corresponding to the non-negativity of the p leading principal minors of the matrix $Z(X)$ (see e.g. Horn and Johnson (1985), p 404). Therefore the random event “ $Z(X)$ positive semi-definite” denoted $(Z(X) \geq 0)$ throughout, will occur with a probability q , $0 < q < 1$. In particular when $Z(X)$ is positive definite, the rescaled Hessian matrix $\sqrt{T}H_T^W(\theta^0)$ is, for T sufficiently large (with probability 1) a positive definite matrix (with a random positive definite limit in distribution) while information measured by the outer product matrix $E(I_T^W(\theta^0))$ is negligible in front of the Hessian matrix. This property explains that we may get a root- T consistent estimator on this part of the sample space, by contrast with a common weak identification setting.

Proposition 3.2 below makes explicit the conclusion of Remark 3.2 above, with the same notations and the notation $\overline{(Z(X) \geq 0)}$ for the complement of the event $(Z(X) \geq 0)$:

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

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

Proof: See Appendix B.

Proposition 3.2 gives the asymptotic distribution of the GMM estimator as a mixture of two distributions, coming with two different rates of convergence. Up to the heterogeneity in rates of

convergence, the fact that the asymptotic distribution of an estimator may depend on the location of the (asymptotically) observed sample in the sample space is reminiscent of situations met in the context of pretest estimators. Typically, when the event of probability q defined by $(Z(X) \geq 0)$ occurs, all the component of $\hat{\theta}_T$ converge towards the true value θ^0 at a rate faster than $T^{1/4}$. For instance, this occurs with probability $q = 1/2$ in the case of only one parameter θ (case of a common feature in two asset returns as studied by Engle and Kozicki (1993)), since in this case $Z(X)$ is a real Gaussian variable with zero-mean. By contrast, when the event $(Z(X) \geq 0)$ does not occur, at least some linear combinations of $\hat{\theta}_T$ cannot converge faster than the slowest possible rate $T^{1/4}$. This classification of rates of convergence extends the one described by Sargan (1983) in the context of linear instrumental variables (with non-linearities with respect to parameters). Similar results have been derived by Rotnitzky et al. (2000) for maximum likelihood estimation with singularities in the information matrix. However, our main focus of interest here, specific to GMM, is the non-standard asymptotic distribution of the J -test statistic for overidentification induced by the non-standard asymptotic behavior of the GMM estimator.

Let us first consider a GMM estimator $\hat{\theta}_T$ associated to an arbitrary, albeit positive definite asymptotic weighting matrix W . The value function of the minimization problem (7) is then:

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

The key intuition is that, due to zero Jacobian and nonstandard rates of convergence derived in Propositions 3.1 and 3.2, a Taylor expansion of J_T^W around θ^0 will no longer depend primarily on first order terms (terms that are linear w.r.t. $(\hat{\theta}_T - \theta^0)$) but rather on second order terms. More precisely, if $\hat{v}_T = T^{1/4}(\hat{\theta}_T - \theta^0)$ 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 B), we have:

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

It is useful to introduce the \mathbb{R}^p -indexed empirical process

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

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

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

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

$$J^W(v) = X'WX + v'Z(X)v + \frac{1}{4}\text{Vec}'(vv')G'WG\text{Vec}(vv'), \quad v \in \mathbb{R}^p. \quad (9)$$

By construction, for each $v \in \mathbb{R}^p$, $\hat{J}^W(v)$ converges in distribution towards $J^W(v)$. Lemma B.5 in Appendix B shows that this convergence in distribution actually occurs uniformly on any compact

subset of \mathbb{R}^p . Upon the tightness of their respective minimizers, the minimum of $\hat{J}^W(v)$ converges in distribution towards the minimum of $J^W(v)$. This is formally stated in the following theorem:

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

Proof: See Appendix B.

Theorem 3.1 gives the asymptotic distribution of J_T^W as the minimum of the limiting process $J^W(v)$. This distribution is rather unusual since $J^W(v)$ is an even multivariate polynomial function of degree 4. In general, the minimum value of $J^W(v)$ 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 $J^W(v)$ makes the determination of critical values for asymptotic inferences involving J_T^W 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 $J^W(v)$. However, a brute force simulation approach is not trivial since it would involve plugging in first step estimators of nuisance parameters such as $\Sigma(\theta^0)$, W and G . Dovonon and Gonçalves (2012) have recently developed a bootstrap based alternative technique.

For the purpose of a non-simulation based asymptotic inference strategy, the next subsection shows that the standard choice of $W = \Sigma^{-1}(\theta^0)$ allows us to get an asymptotic distribution for J_T^W still tightly related to chi-square, albeit mixing several distributions, not only $\chi^2(H-p)$ but also $\chi^2(H)$ (and possibly some intermediate degrees). Note that our focus on the case $W = \Sigma^{-1}(\theta^0)$ is motivated by the search for close form formulas for the distribution of J^W and not by efficiency considerations. Since rates of convergence of any GMM estimator are heterogeneous depending on the occurrence of the event $(Z(X) \geq 0)$ (occurrence which itself depends upon the choice of W) there is no such thing as an efficient GMM estimator.

3.3 Overidentification test

The GMM overidentification test statistic based on the moment condition $E(\phi_{t,T}(\theta)) = 0$ is given by:

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

J_T is the minimum value of the GMM objective function using the so-called optimal weighting matrix defined as a consistent estimate of the inverse of the moment conditions' long run variance, i.e. $W^{-1} = \Sigma(\theta^0) \equiv \lim_{T \rightarrow \infty} \text{Var} \left(\sqrt{T} \bar{\phi}_T(\theta^0) \right)$. From Corollary 3.1, one will typically take:

$$W_T^{-1} = \frac{1}{T} \sum_{t=1}^{T-1} (z_t - \bar{z}_T)(z_t - \bar{z}_T)' \left(\left(\tilde{\theta}'_T Y_{t+1} \right)^2 - \bar{c}_T(\tilde{\theta}_T) \right)^2, \quad (10)$$

where $\tilde{\theta}_T$ denotes any consistent estimator of θ^0 , for instance a GMM estimator obtained with identity matrix as W_T . Note that, as far as controlling for the size of the test is concerned, we actually only need to estimate consistently the variance matrix $\Sigma(\theta^0)$ under the null. Then, a much simpler estimator is available in the context of Remark 3.1 by replacing population variances by sample counterparts.

Note that J_T stands for above J_T^W , while we erase for simplicity the upper index W for all quantities J when W is the limit of the above optimal choice of W_T . Recall that this specific choice of weighting matrix does not deserve anymore to be called optimal. However, it will allow us a direct comparison with two chi-square distributions, namely $\chi^2(H)$ and $\chi^2(H - p)$. First, we have by definition:

$$J_T \leq J_T(0) = T\bar{\phi}'_T(\theta^0)W_T\bar{\phi}_T(\theta^0) \xrightarrow{d} J(0) \sim \chi^2(H)$$

However, due to the nullity of the Jacobian matrix, the common asymptotic theory stating that J_T behave in large samples as a chi-square random variable with $H - p$ degrees of freedom (Hansen (1982)) is no longer valid. From the general result (3.1), we know that the limiting distribution of J_T is the distribution of:

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

Obviously:

$$J \geq L = \min_{u \in \mathbb{R}^{p^2}} \left\{ X'WX + X'WGu + \frac{1}{4}u'G'WGu \right\} \quad (11)$$

We will actually see below that $J > L$ with a positive probability while L will always follow a distribution $\chi^2(H - p)$. This is the key intuition of the reason why the standard J -test for identification, based on the statistic J_T converging to J but using critical values computed from $\chi^2(H - p)$ will be asymptotically oversized. To be more precise, it is first useful to get the following lemma:

Lemma 3.1. *With L defined in (11) (and $W = \Sigma^{-1}(\theta^0)$), we have the decomposition:*

$$J(0) = L + S,$$

L stochastically independent of $(S, Z(X))$ and: $L \sim \chi^2(H - p)$, $S \sim \chi^2(p)$, $J(0) \sim \chi^2(H)$, where $H = \dim(\bar{\phi}_T(\theta))$, and $p = \dim(\theta)$.

Proof: See Appendix B.

Lemma 3.1 is a key input to prove our main result as given by Theorem 3.2 below:

Theorem 3.2. *Under the same conditions as Theorem 3.1 with $W = \Sigma^{-1}(\theta^0)$, the overidentification test statistic J_T is asymptotically distributed as J which is such that:*

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

with:

(i) $Z(X) \geq 0 \Rightarrow J = J(0)$

(ii) $\forall c > 0,$

$$\text{Prob}(L > c) < \text{Prob}(J > c).$$

Proof: See Appendix B.

Remark 3.3. *The proof of Theorem 3.2 actually shows that:*

$$\text{Prob}(L > c, Z(X) \geq 0) < \text{Prob}(J > c, Z(X) \geq 0), \quad \forall c > 0.$$

This obviously implies the inequality given in Theorem 3.2(ii) above, since by virtue of the always valid inequality $L \leq J$, we have on any measurable part B of the sample space:

$$\text{Prob}((L > c) \cap B) \leq \text{Prob}((J > c) \cap B).$$

However, it is worth realizing that the part of the sample space where $Z(X) \geq 0$ is actually responsible for the strict inequality and, as a consequence (see Remark 3.4 below) for over-rejection of the J -test. When $Z(X) \geq 0$, all the components of θ are estimated at a rate faster than $T^{1/4}$. Then, since the Jacobian matrix is nil at the true value and only higher order terms matter in Taylor expansions, it is as if the parameters were actually known. This is the reason why the asymptotic distribution of the J -test statistics coincides in this case with $J(0)$, following $\chi^2(H)$. When the event $(Z(X) \geq 0)$ does not occur, depending on the part of the sample space, only a number $p_1 < p$ of components of θ are estimated at a rate faster than $T^{1/4}$. Then, we may expect in such cases that the J -test statistic asymptotically behaves as a $\chi^2(H - q)$, $q = p - p_1$. In other words, the eventual asymptotic distribution under the null of J_T should be a mixture of distributions $\chi^2(H - q)$, $q = 0, 1, \dots, p$. Then, it is not surprising that critical values computed from $\chi^2(H - p)$ lead to over-rejection as formally proved in Remark 3.4 below.

Remark 3.4. *By application of the Portmanteau lemma (see e.g. van der Vaart, Lemma 2.2.(v) page 6), the convergence in distribution of J_T towards J implies that:*

$$\liminf_{T \rightarrow \infty} \text{Prob}(J_T > \chi^2_{1-\alpha}(H - p)) \geq \text{Prob}(J > \chi^2_{1-\alpha}(H - p)) > \text{Prob}(L > \chi^2_{1-\alpha}(H - p)) = \alpha.$$

Hence, the standard J -test will be oversized, regardless of the nominal level.

The particular case $p = n - 1 = 1$ allows us to prove results that are even more explicit for two reasons:

First, there is no intermediate distribution $\chi^2(H - q)$, $q = 0, 1, \dots, p$ to consider between $\chi^2(H)$ and $\chi^2(H - 1)$.

Second, $Z(X)$ is now a univariate Gaussian variable, and thanks to the symmetry of the normal distribution with zero mean, conditioning by the event $(Z(X) \geq 0)$ has no impact on the distributions of interest and the conditioning event has always a probability one half.

In other words, we can show:

Corollary 3.2. *Under the same conditions as Theorem 3.2, if $p = 1$, the J -test statistic follows asymptotically under the null a half-half mixture of χ_H^2 and χ_{H-1}^2 . More precisely,*

$$J = I(z \geq 0)J(0) + I(z < 0)L,$$

where z is standard normal such that $I(z \geq 0)$ is independent of $J(0)$ and L , respectively; $I(A)$ denotes the usual indicator function.

Proof: See Appendix B.

Note that the case $p = 1$ is precisely the one studied by Engle and Kozicki (1993): out of two asset returns Y_{1t} and Y_{2t} , is it possible to find an homoskedastic portfolio return $(1 - \theta)Y_{1t} + \theta Y_{2t}$? The standard J -test based on quantiles of $\chi^2(H - 1)$ will over-reject this null hypothesis because an (asymptotically) exact test should use instead quantiles of the mixture $\frac{1}{2}\chi^2(H - 1) + \frac{1}{2}\chi^2(H)$. It is worth keeping in mind that, even though these two distributions may not be so different, their tails are different and thus the impact on corresponding critical values will be non negligible. Of course, this impact will be even more dramatic with large p (a large number $n = p + 1$ of assets at stake) since then, common critical values are based on while as shown above, the bound $\chi^2(H)$ is sharp. Note that this upper bound always allows us to define an asymptotically conservative test.

At this stage, it is worth reiterating 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 common feature 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. However, it is worth keeping in mind that the non-standard behaviours of the GMM usual statistics are not alleviated by large samples. It is then also important to simulate large samples to assess at what stage non-standard behaviours are prevalent.

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

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

according to five designs.

Design D1 generates a bivariate vector of two asset returns Y_{t+1} with a single conditionally heteroskedastic factor $f_{1,t+1}$ ($F_{t+1} = f_{1,t+1}$) following a Gaussian GARCH(1,1) dynamic, i.e.

$$f_{1,t+1} = \sigma_t \varepsilon_{t+1}, \quad \sigma_t^2 = \omega_1 + \alpha_1 f_{1,t}^2 + \beta_1 \sigma_{t-1}^2,$$

where $\varepsilon_{t+1} \sim \text{NID}(0, 1)$. We choose $\omega_1 = 0.2$, $\alpha_1 = 0.2$, and $\beta_1 = 0.6$. The factor loading vector is set to $\Lambda = (1, 0.5)'$ and the bivariate vector of idiosyncratic shocks $U_{t+1} \sim \text{NID}(0, 0.5Id_2)$. F_{t+1} and U_{t+1} are independent throughout.

Design D2 also simulates a bivariate vector of two asset returns Y_{t+1} but with two independent GARCH factors. The vector of conditionally heteroskedastic factors is $F_{t+1} = (f_{1,t+1}, f_{2,t+1})'$ where $f_{1,t+1}$ is independent of $f_{2,t+1}$, a GARCH process with $\omega_2 = 0.2$, $\alpha_2 = 0.4$, and $\beta_2 = 0.4$. The factor loading $\Lambda = Id_2$ and the idiosyncratic shocks $U_{t+1} \sim \text{NID}(0, 0.5Id_2)$.

Designs D3, D4 and D5 all generate trivariate vector of three asset returns Y_{t+1} with one, two and three GARCH factors, respectively. For all of them, the idiosyncratic shocks $U_{t+1} \sim \text{NID}(0, 0.5Id_3)$.

The single factor in Design D3 has the same GARCH dynamics as the factor in Design D1 with factor loading $\lambda_1 = (1, 1, 0.5)'$.

In Design D4, the dynamics of the two GARCH factors are the same as those in D2 with factor loading matrix $\Lambda = (\lambda_1 | \lambda_2)$; $\lambda_1 = (1, 1, 0.5)'$ and $\lambda_2 = (0, 1, 0.5)'$.

The return process in Design D5 is generated by the GARCH factors $F_{t+1} = (f_{1,t+1}, f_{2,t+1}, f_{3,t+1})'$; these factors are mutually independent and $\omega_3 = 0.1$, $\alpha_3 = 0.1$, and $\beta_3 = 0.8$ for $f_{3,t+1}$. The factor loading matrix is set to Id_3 .

Thanks to their respective parameter configurations, the GARCH factors $f_{i,t} : i = 1, 2, 3$ considered in these experiments are stationary ergodic with finite fourth moments so that the returns processes inherit the same properties. The sets of instruments z_{1t} and z_{2t} that we introduce below are also stationary and ergodic with finite second moment and the conditions in Assumption 3 are essentially fulfilled. (We refer to Bollerslev (1986) and Lindner (2009) for the conditions that guarantee these probabilistic properties for GARCH processes.)

Designs D1 and D4 correspond to the null hypothesis to be tested (with $p = n - 1$) including the parameter identification requirement. Designs D2 and D5 correspond to the alternative where the GARCH features in the simulated asset returns are not common. In Design D3, where the three simulated returns share one common GARCH factor, the parameters of one candidate common feature are globally unidentified since the space of CH common features is of dimension two.

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 10,000 times for each sample size T . The sample sizes that we consider are 50, 100, 1,000, 2,000, 5,000, 10,000, 20,000, 30,000, 40,000, 50,000, 100,000 and 200,000. 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 the asymptotic behaviours of interest to be confirmed for larger samples than those commonly relied upon.

For each simulated sample, we evaluate the GMM estimator according to (7). The weighting matrix W_T is set to the inverse of $\hat{\Sigma}$ given by (10) 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. While Engle and Susmel (1993) suggest to use as instruments, not only lagged squared returns but also lagged cross products, the latter ones do not appear to be needed to get the evidence we want to illustrate: over-rejection of the standard test under the null and excellent power of the conservative $\chi^2(H)$ test under the alternative.

The rate of convergence as derived in Propositions 3.1 and 3.2 will be illustrated by the GMM estimator from Design D1. From Theorem 3.2, the J -test statistic from this design is expected to follow asymptotically 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 D4 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 control the size of the test. Design D3 will give an idea about the behaviour of the J -test when the moment condition model is well-specified but non-identified, even globally. Designs D2 and D5 will illustrate the power of the test to detect the lack of factor structure.

The results are summarized in Tables I and II and Figures 1 and 2. The $T^{1/4}$ rate of convergence of the GMM estimator is confirmed by the fact that the simulated standard deviation of $T^{1/4}\hat{\theta}$ is stable around 2.5 as the sample size grows (see Table I and Figure 1). Evidence that the estimator cannot be \sqrt{T} -consistent lies in the fact that the standard deviation of $\sqrt{T}\hat{\theta}$ increases with the sample size. The rejection rate of the J -test for design D1 confirms that the half-half mixture of chi-squares derived as asymptotic distribution provides a quite accurate approximation (actual rejection rate between 4.5% and 5% for a nominal rate of 5%) for sample sizes larger or equal to 2,000. By contrast, for these sample sizes, the usual critical values computed from the standard $\chi^2(1)$ distribution lead to rejection rates that exceed the nominal one by 70% or more.

The rejection rate of the true model in Design D4 also confirms the results of Theorem 3.2. The rejection rate from the standard χ_1^2 almost triples (at about 13%) the nominal rate (5%) while the critical values from a χ_3^2 offer a conservative test.

Designs D2 and D5 highlight the power of our test for common conditionally heteroskedastic features. Under the alternative, as the sample size grows, the null is rejected even with the conservative bound in 100% of the cases. It is worth mentioning that this test converges slowly since sharp rejection rates are not obtained before $T = 2,000$.

In Design D3 where there is no-parameter-identification at all, the asymptotic distribution of the J -test seems to significantly squeeze to the left making both χ_1^2 and χ_3^2 useful to carry out tests with a correct size. An asymptotically correct critical value for the J -test (even one accounting for first order underidentification) would therefore under-reject. This suggests that the results of Cragg and Donald (1996) and Staiger and Stock (1997), namely that the GMM overidentification test is undersized in case of global identification failure, also hold in our context.

Table I: Simulated bias and standard deviation of the GMM estimator $\hat{\theta}$ in Design D1 and simulated rejection rate of the J -test for Designs D1 (2A-1F) and D2 (2A-2F) using critical values from χ_1^2 , χ_2^2 and the mixture $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$; 5%-nominal level.

	Design D1 (2A:1F)						Design D2 (2A:2F)		
	Bias	sd($T^{\frac{1}{4}}\hat{\theta}$)	sd($T^{\frac{1}{2}}\hat{\theta}$)	Rejection rate (in %) using					
				5%-critical value from:					
				χ_1^2	χ_2^2	Mixt.	χ_1^2	χ_2^2	Mixt.
50	1.21	2.00	5.32	3.05	0.41	0.85	3.6	0.4	1.1
100	1.16	2.33	7.36	3.38	0.67	1.44	4.2	1.1	1.8
500	0.88	3.23	15.29	4.90	1.29	2.30	31.9	13.7	19.8
1000	0.73	3.64	20.48	7.11	2.35	3.48	68.3	45.3	54.4
2000	0.57	2.67	17.82	8.90	3.15	4.87	94.9	86.3	90.3
5000	0.44	2.25	18.89	9.15	3.33	4.97	100.0	99.8	99.9
10000	0.36	2.34	23.41	9.04	3.13	4.65	100.0	100.0	100.0
20000	0.30	2.34	27.87	8.89	3.00	4.58	100.0	100.0	100.0
30000	0.26	2.41	31.78	9.16	3.34	4.95	100.0	100.0	100.0
40000	0.25	2.40	33.98	9.63	3.26	4.81	100.0	100.0	100.0
50000	0.23	2.42	36.18	9.18	3.18	4.80	100.0	100.0	100.0
100000	0.19	2.50	44.51	9.31	3.22	4.97	100.0	100.0	100.0
200000	0.15	2.54	53.81	10.05	3.39	5.10	100.0	100.0	100.0

‘Mixt.’ stands for $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$; the critical values of which are: 5.14(8.28) for a nominal level of 0.05(0.01).

‘sd’ for standard deviation, ‘A’ for Assets and ‘F’ for factor(s). The true value of θ in Design D1 is -1.0.

Table II: Simulated rejection rate of the J -test for Designs D3 (3A-1F), D4 (3A-2F) and D5 (3A-3F) using critical values from χ_1^2 and χ_3^2 ; 5%-nominal level.

T	Rejection rate (in %) using					
	5%-critical value from:					
	χ_1^2	χ_3^2	χ_1^2	χ_3^2	χ_1^2	χ_3^2
	<i>Design D4</i>		<i>Design D3</i>		<i>Design D5</i>	
	3A-2F		3A-1F		3A-3F	
50	2.06	0.06	2.16	0.05	2.5	0.1
100	2.04	0.07	1.87	0.03	2.6	0.1
500	3.69	0.29	2.17	0.09	9.6	1.0
1000	6.80	0.61	2.56	0.30	23.7	4.8
2000	9.25	1.16	3.08	0.20	52.1	20.4
5000	10.24	1.77	3.66	0.25	92.3	72.9
10000	11.06	2.08	3.79	0.35	99.8	98.2
20000	11.80	2.05	3.86	0.33	100.0	100.0
30000	12.61	2.25	4.17	0.26	100.0	100.0
40000	13.56	1.95	3.99	0.38	100.0	100.0
50000	13.04	2.50	4.53	0.41	100.0	100.0
100000	13.13	2.25	4.61	0.34	100.0	100.0
200000	13.57	2.16	4.52	0.38	100.0	100.0

‘A’ stands for Assets and ‘F’ for factor(s).

Figure 1: Simulated bias and standard deviation of the GMM estimator $\hat{\theta}$ in Design D1 (2A-1F) and simulated rejection rate of the J -test using critical values from χ_1^2 , χ_2^2 and the mixture $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$. Nominal levels 5% and 1% for Design D1 and 5% for D2 (2A-2F).

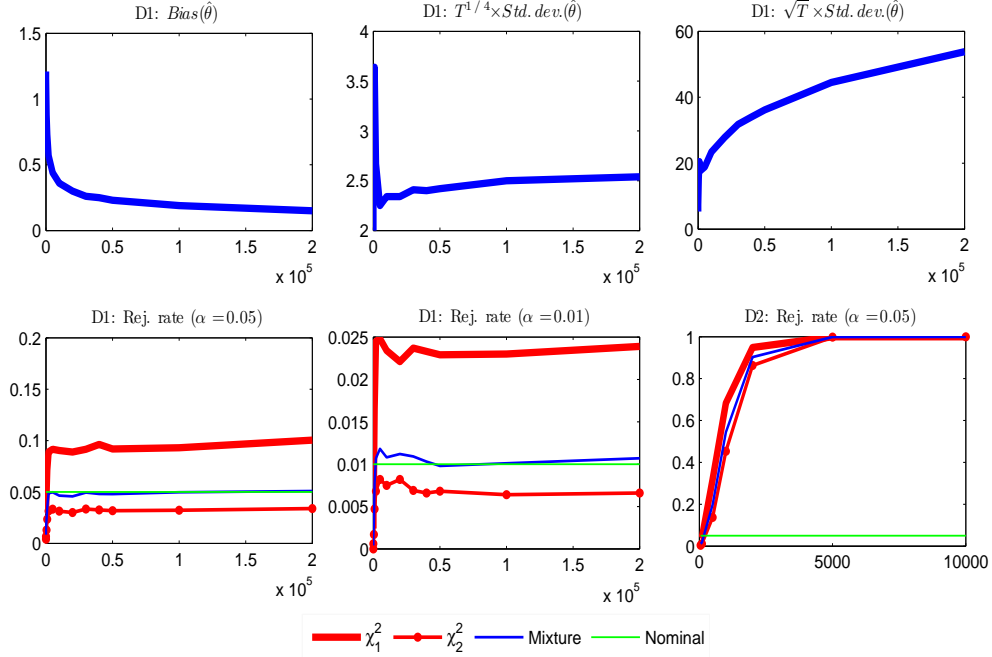
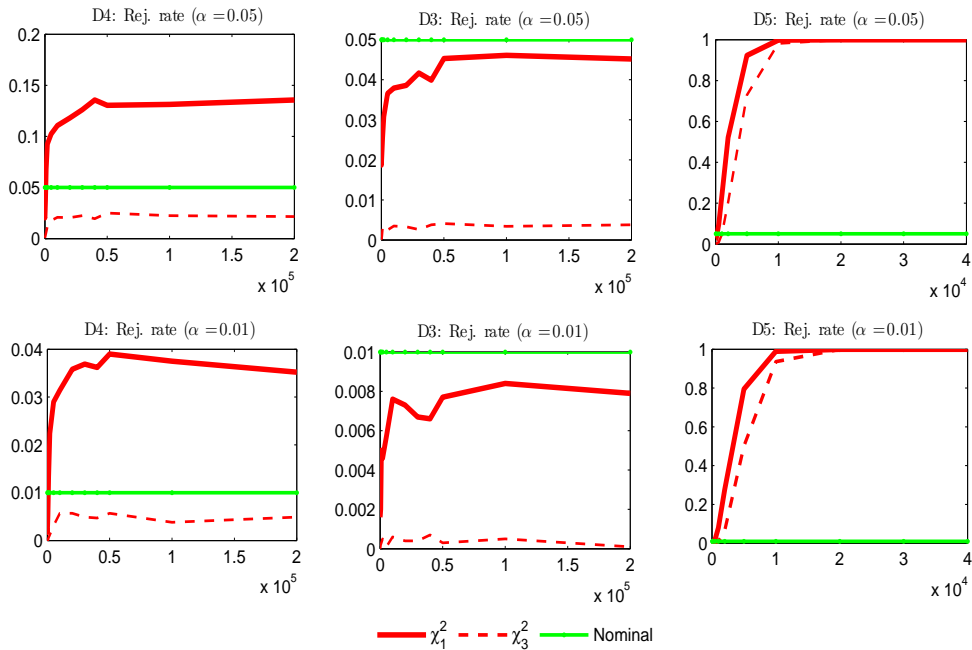


Figure 2: Simulated rejection rate of the J -test using critical values from χ_1^2 and χ_3^2 for Designs D3 (3A-1F), D4 (3A-2F) and D5 (3A-3F). Nominal levels 5% and 1%.



5 Conclusion

This paper sets the focus on an important albeit simple case of non-standard behaviour of the J -test statistic for overidentification. This is important because the detection of CH common features, or equivalently, of (GARCH) factor structures with fixed factor loadings, is crucial for modeling of multivariate conditional heteroskedasticity. However, the case considered in this paper is the simplest possible since second-order identification is addressed with quadratic functions. As documented at least in the MLE literature (see e.g. Rotnitzky et al. (2000) and Lee and Chesher (1986)), second-order identification, while first-order identification fails is at stake with more complicated non-quadratic settings.

Extensions of these results to a GMM framework is worthwhile for at least two reasons. First, as exemplified in the present paper, non-standard non-Gaussian asymptotic distribution of estimators may not prevent us from recovering asymptotic (mixtures) of chi-squares for the J -test of overidentification. This overidentification issue could not be addressed in the MLE context. Second, as mentioned in the paper, there is no such thing as an obvious efficiency argument to elicit a specific weighting matrix for GMM when first order identification fails. For these reasons, further work is warranted. Non-quadratic moment conditions with possible local singularities are actually pervasive in financial econometrics. An important example is inference about risk premium in equilibrium when investors have preferences for higher order moments.

Appendix A: Testing for GARCH effects on common features

A regression-based approach is akin to consider an instrumental heteroskedasticity model that can be written:

$$\xi_{t+1} = a + bz_t + \varepsilon_{t+1}, \quad (\text{A.1})$$

where ξ_{t+1} is a m -dimensional vector that gathers some coefficients of the matrix $Y_{t+1}Y'_{t+1}$, z_t is again a vector of H \mathfrak{F}_t -measurable instruments, $E(\varepsilon_{t+1}) = 0$, $\text{Cov}(\varepsilon_{t+1}, z_t) = 0$ and $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^{m \times H}$ are vectors of unknown parameters. It will be possible to check from this regression model that the portfolio $\theta'Y_{t+1}$ is a CH common feature insofar as:

$$(\theta'Y_{t+1})^2 = \theta'Y_{t+1}Y'_{t+1}\theta = \gamma'(\theta)\xi_{t+1}$$

for some known function γ from Θ to \mathbb{R}^{mH} . In other words, the regression model (1) must be rich enough to be such that the vector ξ_{t+1} gathers in particular the coefficients of $Y_{t+1}Y'_{t+1}$ that show up in $\theta'Y_{t+1}Y'_{t+1}\theta$. In these circumstances, the null hypothesis of interest for the test of CH common features will be:

$$H_0 : \exists \theta : \gamma'(\theta)b = 0.$$

Irrespective of the preferred test chosen in the trinity of asymptotic tests to test such a composite hypothesis (see Gourieroux and Monfort (1989)), the standard asymptotic chi-square distribution under the null will be warranted only if:

$$\left. \frac{\partial[\gamma'(\theta)b^0]}{\partial \theta} \right|_{\theta=\theta^0}$$

is full rank. However, from the regression model:

$$(\theta'Y_{t+1})^2 = \gamma'(\theta)a^0 + \gamma'(\theta)b^0z_t + \gamma'(\theta)\varepsilon_{t+1}$$

and thus:

$$\frac{\partial(\theta'Y_{t+1})^2}{\partial \theta} = \frac{\partial \gamma'(\theta)}{\partial \theta} a^0 + \frac{\partial \gamma'(\theta)}{\partial \theta} b^0 z_t + \frac{\partial \gamma'(\theta)}{\partial \theta} \varepsilon_{t+1}$$

and

$$\text{Cov} \left(\frac{\partial(\theta' Y_{t+1})^2}{\partial \theta}, z_t \right) = \frac{\partial \gamma'(\theta)}{\partial \theta} \text{Cov}(b^0 z_t, z_t).$$

therefore,

$$\frac{\partial \gamma'(\theta)}{\partial \theta} b^0 \Big|_{\theta=\theta^0} = \text{Cov} \left(\frac{\partial(\theta' Y_{t+1})^2}{\partial \theta}, z_t \right) \Big|_{\theta=\theta^0} (\text{Var}(z_t))^{-1} = 0$$

by Proposition 2.1 when θ^0 is a common feature.

Appendix B: Proofs

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

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

$p = n - 1$ 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,

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

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

Lemmas B.1-B.5 below connect as follows in proving the main results in the paper. Lemma B.1 is relevant thanks to lemma 2.3 and is useful to derive the rate of convergence as stated in Propositions 3.1 and 3.2. This lemma is also useful to establish part of Lemma B.5. Lemma B.2 is useful to establish Proposition 3.2. Lemma B.4 is useful to establish Lemma B.5; the part (iii) of which essentially proves Theorem 3.1, while Lemma B.3 is used in the proof of Theorem 3.2.

Lemma B.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 B.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 B.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,*

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

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

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

But

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

as $T \rightarrow \infty$. This establishes the result \square

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

We have

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

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: $\text{Vec}(e_i e'_i)$, $i = 1, \dots, p$, where $\{e_i : i = 1, \dots, p\}$ 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 \square

Lemma B.4. *Let $\hat{M}_T(v)$ 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) $\hat{M}_T(\cdot)$ converges in distribution towards $M(\cdot)$ in $\ell^\infty(\mathbb{K})$ for every compact $\mathbb{K} \subset \mathbb{R}^p$, where $\ell^\infty(\mathbb{K})$ is the set of all bounded real-valued functions on \mathbb{K} , endowed with the sup-norm,
- (ii) there exists $\hat{v}_T \in \arg \min_{v \in \mathbb{V}_T} \hat{M}_T(v)$ which is uniformly tight and
- (iii) there exists $\hat{v} \in \arg \min_{v \in \mathbb{R}^p} M(v)$ which is tight,

then,

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

Proof of Lemma B.4. We show that $\text{Prob}(\hat{M}_T(\hat{v}_T) \leq x) \rightarrow \text{Prob}(M(\hat{v}) \leq x)$ as $T \rightarrow \infty$ 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, $\hat{M}_T(\cdot)$ converges towards $M(\cdot)$ in distribution in $\ell^\infty(\{v : \|v\| \leq m_\epsilon\})$. Since the function \inf is continuous on $\ell^\infty(\mathbb{K})$, for any nonempty compact \mathbb{K} , we can apply the continuous mapping theorem and deduce that

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

Hence, from sample path continuity of $\hat{M}_T(\cdot)$ and $M(\cdot)$, we have:

$$\min_{\|v\| \leq m_\epsilon} \hat{M}_T(v) \xrightarrow{d} \min_{\|v\| \leq m_\epsilon} M(v).$$

Considering x as a continuity point for the cumulative distribution function of $\min_{\|v\| \leq m_\epsilon} M(v)$ (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

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

Clearly,

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

thus,

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

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

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

Therefore,

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

By the same way, we also have

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

Now, we observe that

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

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

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

- (i) *The stochastic process $\hat{J}^W(\cdot)$ converges in distribution towards $J^W(\cdot)$ in $\ell^\infty(\mathbb{K})$ for every compact $\mathbb{K} \subset \mathbb{R}^p$,*
- (ii) *$\hat{v}_T \equiv \arg \min_{v \in \mathbb{H}_T} \hat{J}^W(v)$ is uniformly tight and any $\hat{v} \in \arg \min_{v \in \mathbb{R}^p} J^W(v)$ is tight.*

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

Proof of Lemma B.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

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

Hence

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

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 $\hat{J}^W(\cdot)$ converges in distribution towards $J^W(\cdot)$ in $\ell^\infty(\mathbb{K})$, this theorem requires that:

- (a) The marginals $(\hat{J}^W(v_1), \dots, \hat{J}^W(v_k))$ converge in distribution towards $(J^W(v_1), \dots, J^W(v_k))$ for every finite subset $\{v_1, \dots, v_k\}$ of \mathbb{K} .
- (b) The empirical process $\hat{J}^W(\cdot)$ is asymptotically tight.

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

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

By the continuous mapping theorem, this latter converges in distribution towards $(J^W(v_1), \dots, J^W(v_k))$. 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 $\hat{J}^W(\cdot)$ to be asymptotically tight. From (a), $\hat{J}^W(v)$ converges in distribution towards $J^W(v)$, for any $v \in \mathbb{K}$. In addition, as a compact subset, \mathbb{K} equipped with the usual metric on \mathbb{R}^p is totally bounded. It remains to show that $\hat{J}^W(\cdot)$ is asymptotically uniformly equicontinuous in probability. That is for any $\epsilon, \eta > 0$, there exists $\delta > 0$ such that

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

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

$$|\hat{J}^W(v_1) - \hat{J}^W(v_2)| = X_T \|v_1 - v_2\| + o_P(1), \quad (\text{B.3})$$

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

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

We can safely ignore the $o_P(1)$ term in (B.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 $\hat{v} \in \arg \min_{v \in \mathbb{R}^p} J^W(v)$. Let $\epsilon > 0$. We have $0 \leq \min_{v \in \mathbb{R}^p} J^W(v) \leq J^W(0) = O_P(1)$, hence, there exists $m_1 > 0$ such that

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

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

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

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

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

Thus

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

This shows that \hat{v} is tight.

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

Proof of Lemma 2.1. Let $\theta \in \mathbb{R}^n$, $\theta \neq 0$. We know that

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

If $\Lambda' \theta = 0$, then $\text{Var}(\theta'Y_{t+1}|\mathfrak{F}_t) = cst$ and θ is a common feature. Conversely, if $\theta'\Lambda D_t \Lambda' \theta + \theta' \Omega \theta = cst$, writing $c = \Lambda' \theta$, we have

$$\sum_{k=1}^K c_k^2 D_{kk,t} = cst,$$

Hence, we have a linear combination of the terms in $\text{Diag}(D_t)$ that is constant. From Assumption 1-(ii), we necessarily have $c_k^2 = 0$; $k = 1, \dots, K$. Thus $\Lambda' \theta = 0$. \square

Proof of Corollary 3.1. Note that:

$$\psi_t(\theta) = z_t ((\theta'Y_{t+1})^2 - c(\theta)) = (z_t - \bar{z}_T) ((\theta'Y_{t+1})^2 - \bar{c}_T(\theta)) + (z_t - \bar{z}_T)(\bar{c}_T(\theta) - c(\theta)) + \bar{z}_T ((\theta'Y_{t+1})^2 - c(\theta))$$

and thus,

$$\begin{aligned} \sqrt{T}\bar{\psi}_T(\theta) &= \sqrt{T}\bar{\phi}_T(\theta) + \bar{z}_T \cdot \left(\frac{\sqrt{T}}{T} \sum_{t=1}^T [(\theta'Y_{t+1})^2 - c(\theta)]\right) \\ &= \sqrt{T}\bar{\phi}_T(\theta) + \mu_z \cdot \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T [(\theta'Y_{t+1})^2 - c(\theta)]\right) + (\bar{z}_T - \mu_z) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T [(\theta'Y_{t+1})^2 - c(\theta)]\right) \\ &= \sqrt{T}\bar{\phi}_T(\theta) - \sqrt{T}\bar{v}_T(\theta) + o_P(1), \end{aligned}$$

since $\sqrt{T}(\bar{z}_T - \mu_z) = O_P(1)$ by Assumption 3. Hence,

$$\begin{aligned} \sqrt{T}\bar{\phi}_T(\theta) &= \sqrt{T}\bar{\psi}_T(\theta) + \sqrt{T}\bar{v}_T(\theta) + o_P(1) \\ \sqrt{T}\frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta) &= \sqrt{T}\frac{\partial \bar{\phi}_T}{\partial \theta'}(\theta) + \sqrt{T}\frac{\partial \bar{v}_T}{\partial \theta'}(\theta) + o_P(1). \end{aligned} \tag{B.4}$$

The second equation in (B.4) is obtained similarly to the first one. For $\theta = \theta^0$, $\psi_t(\theta^0) = z_t \left((\theta^{0'} Y_{t+1})^2 - c(\theta^0) \right)$ and $v_t(\theta^0) = -\mu_z \left((\theta^{0'} Y_{t+1})^2 - c(\theta^0) \right)$ are two martingale difference sequences with respect to \mathfrak{F}_t . Hence, the asymptotic variance, Σ , of $\sqrt{T}\bar{\phi}_T(\theta^0)$ is equal to:

$$\begin{aligned}
\Sigma &= \text{Var}(\psi_t(\theta^0) + v_t(\theta^0)) \\
&= \text{Var}(\psi_t(\theta^0)) + \text{Var}(v_t(\theta^0)) + \text{Covar}(v_t(\theta^0), \psi_t(\theta^0)) + \text{Covar}(\psi_t(\theta^0), v_t(\theta^0)) \\
&= E(\psi_t(\theta^0)\psi_t'(\theta^0)) + E(v_t(\theta^0)v_t'(\theta^0)) + E(v_t(\theta^0)\psi_t'(\theta^0)) + E(\psi_t(\theta^0)v_t'(\theta^0)) \\
&= E\left(z_t z_t' \left((\theta^{0'} Y_{t+1})^2 - c(\theta^0) \right)^2\right) + \mu_z \mu_z' E\left((\theta^{0'} Y_{t+1})^2 - c(\theta^0) \right)^2 \\
&\quad - E\left(\left((\theta^{0'} Y_{t+1})^2 - c(\theta^0) \right)^2 z_t \right) \mu_z' - \mu_z E\left(\left((\theta^{0'} Y_{t+1})^2 - c(\theta^0) \right)^2 z_t' \right) \\
&= E\left((z_t - \mu_z)(z_t - \mu_z)' \left((\theta^{0'} Y_{t+1})^2 - c(\theta^0) \right)^2 \right).
\end{aligned}$$

Regarding the Jacobian, we mention that, under the conditions of the Corollary,

$$\frac{\partial \psi_t}{\partial \theta'}(\theta^0) = z_t \left((\theta^{0'} Y_{t+1}) Y_{t+1}' - E((\theta^{0'} Y_{t+1}) Y_{t+1}') \right)$$

is a martingale difference sequence and, thanks to Assumption 3, the central limit theorem of Billingsley (1961) for stationary and ergodic martingale difference sequences applies to $\sqrt{T} \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta^0)$ which is asymptotically normal, therefore $\sqrt{T} \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta^0) = O_P(1)$. \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,

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

From Corollary 3.1, $\sqrt{T}\bar{\phi}_T(\theta^0)$ and $\sqrt{T} \frac{\partial \bar{\phi}_T}{\partial \theta'}(\theta^0)$ are bounded in probability. 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

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

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:

$$\frac{T}{4}\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\Delta}(\hat{\theta}_T - \theta^0) \leq -T\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\phi}_T(\theta^0) + o_P\left(\|\sqrt{T}\bar{\Delta}(\hat{\theta}_T - \theta^0)\|\right) + o_P(1). \quad (\text{B.5})$$

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 (B.5), we can write

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

By the Cauchy-Schwarz inequality,

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

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

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

for some $\gamma_0, \gamma > 0$. Hence

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

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

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

and, for T large enough we can write

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

Hence, for large values of $\|\hat{\theta}_T\|^2$, the term $o_P(1)/\|\hat{\theta}_T\|^2$ stays asymptotically negligible in probability. Therefore, $\|\hat{\theta}_T\|^2$ is at most of the same asymptotic order of magnitude as $\|\sqrt{T} \bar{\phi}_T(\theta^0)\|$. This establishes that $\|\hat{\theta}_T\|^2 = O_P(1)$ or equivalently $\|\hat{\theta}_T\| = O_P(1)$ \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 3. Thus, these two random vectors, measurable (we implicitly assume $\hat{\theta}_T$ measurable—this is a common assumption in the literature on extremum estimators) on the same probability space, are jointly uniformly tight. Therefore, from the 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 $\text{Prob}(V = 0 | Z(X) \geq 0) = 1$. Since $\hat{\theta}_T - \theta^0 = O_P(T^{-1/4})$, we have

$$\sqrt{T} \bar{\phi}_T(\hat{\theta}_T) = \sqrt{T} \bar{\phi}_T(\theta^0) + \frac{1}{2} \sqrt{T} \left((\hat{\theta}_T - \theta^0)' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) (\hat{\theta}_T - \theta^0) \right)_{1 \leq h \leq H} + o_P(1) \quad (\text{B.6})$$

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 $\phi_{t,T}(\theta)$ by $W^{1/2}$. We keep $\phi_{t,T}(\theta)$ for $W^{1/2} \phi_{t,T}(\theta)$ in the rest of this proof for economy of notation. Thus

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

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

$$\Delta' \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right) \sqrt{T}\bar{\phi}_T(\theta^0) + \frac{1}{4}\Delta' \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right) \Delta \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right) \leq o_P(1) \quad (\text{B.7})$$

It is worth noting that

$$\Delta' \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right) \sqrt{T}\bar{\phi}_T(\theta^0) = \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right)' Z_T \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right). \quad (\text{B.8})$$

Actually, each hand side of (B.8) is equal to $\sum_{i,j=1}^p \sum_{h=1}^H \left(\frac{\partial^2 \rho_h(\theta^0)}{\partial \theta_i \partial \theta_j} \sqrt{T}\bar{\phi}_{T,h}(\theta^0) \right) (T^{1/4}(\hat{\theta}_{T,i} - \theta_i^0))(T^{1/4}(\hat{\theta}_{T,j} - \theta_j^0))$. Considering a subsequence of $\left(T^{1/4}(\hat{\theta}_T - \theta^0)', \text{Vec}'(Z_T) \right)'$ that converges in distribution towards a certain random vector $(V', \text{Vec}'(Z(X)))'$, we can write (for the sake of simplicity, we do not make explicit the notation for a subsequence):

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

From (B.7) and by Lemma B.2, we deduce that

$$\begin{aligned} \limsup_{T \rightarrow \infty} \text{Prob} \left(\Delta' \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right) \sqrt{T}\bar{\phi}_T(\theta^0) + \right. \\ \left. + \frac{1}{4}\Delta' \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right) \Delta \left(T^{1/4}(\hat{\theta}_T - \theta^0) \right) \leq \epsilon \right) = 1, \end{aligned}$$

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

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

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

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

In particular if $Z(X)$ is positive semi-definite

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

and thus

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

But, by Lemma B.1,

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

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

$$\text{Prob}(V = 0 | Z(X) \geq 0) = 1.$$

Now, let us establish that $\text{Prob}(V = 0 | \overline{(Z(X) \geq 0)}) = 0$.

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

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

This can be written

$$e' \left(\tilde{Z}_T + N_T \right) e \geq 0, \quad (\text{B.9})$$

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

$$\frac{\partial \bar{\phi}_T}{\partial \theta_i}(\hat{\theta}_T) = \frac{\partial^2 \bar{\phi}_T}{\partial \theta_i \partial \theta'}(\bar{\theta})(\hat{\theta}_T - \theta^0) + O_P(T^{-1/2}), \quad (\text{B.10})$$

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 (B.6), 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 $h_{ij} = \frac{\partial^2 \rho}{\partial \theta_i \partial \theta_j}(\theta^0)$,

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

Thus

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

and

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

From the inequality (B.9) 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:

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

By Lemma B.2,

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

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(vi) of van der Vaart (1998)), that

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

Thus, by right continuity of cumulative distribution functions,

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

and consequently,

$$\text{Prob} \left(\|V\|^2 \geq -\frac{e' Z(X) e}{A} \mid Z(X) = z \right) = 1, \quad P^Z \text{ a.s.} \quad (\text{B.11})$$

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

$$\text{Prob} (\|V\| > 0 \mid Z(X) = z) = 1.$$

Therefore $\text{Prob} \left(\|V\| > 0 \mid \overline{(Z(X) \geq 0)} \right) = 1 \quad \square$

Proof of Theorem 3.1. Follows from Lemma B.5-(iii) \square

Proof of Lemma 3.1. The first order condition associated to (11) is:

$$G' W X + \frac{1}{2} G' W G \dot{u} = 0.$$

Moreover, we know from Lemma B.3 that we can decompose $G = G_1 G_2$ with G_1 (resp. G_2) full column-rank (resp. full row-rank) p . Since G_2' is full column-rank, the above first order conditions are equivalent to:

$$G_1' W X + \frac{1}{2} G_1' W G_1 G_2 \hat{u} = 0$$

and, since G_1 is full column-rank, we deduce:

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

Defining:

$$\tilde{X} = W^{1/2} X, \quad \tilde{G}_1 = W^{1/2} G_1,$$

we see that:

$$\tilde{G}_1 G_2 \hat{u} = -2\tilde{G}_1 (\tilde{G}_1' \tilde{G}_1)^{-1} \tilde{G}_1' \tilde{X} = -2P_1 \tilde{X},$$

where P_1 stands for the matrix of orthogonal projection on the p -dimensional subspace of \mathbb{R}^H spanned by the columns of \tilde{G}_1 . Plugging in (11), we deduce:

$$\begin{aligned} L &= \tilde{X}' \tilde{X} + \tilde{X}' \tilde{G}_1 G_2 \hat{u} + \frac{1}{4} \hat{u}' G_2' \tilde{G}_1' \tilde{G}_1 G_2 \hat{u} \\ &= \tilde{X}' \tilde{X} - 2\tilde{X}' P_1 \tilde{X} + \tilde{X}' P_1 \tilde{X} = \tilde{X}' (Id - P_1) \tilde{X}. \end{aligned}$$

Thus:

$$L = \left\| (Id - P_1) \tilde{X} \right\|^2 \sim \chi^2(H - p)$$

since, for $W = \Sigma^{-1}(\theta^0)$, $\tilde{X} = W^{1/2} X$ is a standardized Gaussian vector. Since $J(0) = X' W X = \left\| \tilde{X} \right\|^2$:

$$S = J(0) - L = \left\| P_1 \tilde{X} \right\|^2 \sim \chi^2(p).$$

Moreover, since $P_1 \tilde{X}$ and $(Id - P_1) \tilde{X}$ are stochastically independent (orthogonal projections of standard Gaussian vectors on two orthogonal subspaces), L is independent of S and of course $J(0) = S + L \sim \chi^2(H)$.

In addition, elementary computations give:

$$\text{Vec}(Z(X)) = G_2' G_1' W X = G_2' (\tilde{G}_1' \tilde{G}_1) (\tilde{G}_1' \tilde{G}_1)^{-1} G_1' W X = G_2' \tilde{G}_1' P_1 \tilde{X}.$$

Therefore, L is actually jointly independent of $(S, Z(X))$. \square

Proof of Theorem 3.2. By definition of the minimization problems, we obviously have:

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

Moreover, from the alternative expression of $J(v)$ given by (9), one can easily see that, when $Z(X) \geq 0$, the minimum of $J(v)$ is reached at $v = 0$, leading to $J = J(0)$.

Part (ii) of Theorem 3.2 will be proved in two steps:

1st step: We show that there exists some $\varepsilon > 0$ such that:

$$\text{Prob}(J > L + \varepsilon, Z(X) \geq 0) > 0.$$

To see that, first note that since $S = J(0) - L \sim \chi^2(p)$, $\text{Prob}(J(0) > L) = 1$.

Thus:

$$\text{Prob}(J(0) > L, Z(X) \geq 0) = \text{Prob}(Z(X) \geq 0).$$

But we know that:

$$Z(X) \geq 0 \Rightarrow J = J(0).$$

Therefore,

$$\text{Prob}(J > L, Z(X) \geq 0) = \text{Prob}(Z(X) \geq 0).$$

However:

$$\text{Prob}(J > L, Z(X) \geq 0) = \text{Prob}\left(\bigcup_{n \geq 1} \left(J > L + \frac{1}{n}\right), Z(X) \geq 0\right) = \lim_{n \rightarrow \infty} \text{Prob}\left(J > L + \frac{1}{n}, Z(X) \geq 0\right).$$

Then:

$$\lim_{n \rightarrow \infty} \text{Prob}\left(J > L + \frac{1}{n}, Z(X) \geq 0\right) = \text{Prob}(Z(X) \geq 0) > 0$$

and we deduce that there exists $n \in \mathbb{N}$ such that:

$$\text{Prob}\left(J > L + \frac{1}{n}, Z(X) \geq 0\right) > 0.$$

2nd step: Following Remark 3.3, we actually show that:

$$\text{Prob}(L > c, Z(X) \geq 0) < \text{Prob}(J > c, Z(X) \geq 0), \quad \forall c > 0.$$

Since we always have $L \leq J$, we have for any measurable part B of the sample space:

$$\text{Prob}(L > c, (Z(X) \geq 0) \cap B) \leq \text{Prob}(J > c, (Z(X) \geq 0) \cap B), \quad \forall c > 0.$$

We will then obviously be able to deduce the announced strict inequality if we show that:

$$\text{Prob}(L > c, Z(X) \geq 0, J > L + \varepsilon) < \text{Prob}(J > c, Z(X) \geq 0, J > L + \varepsilon), \quad \forall c > 0.$$

But since again $L \leq J$:

$$\begin{aligned} & \text{Prob}(J > c, Z(X) \geq 0, J > L + \varepsilon) \\ &= \text{Prob}(L > c, Z(X) \geq 0, J > L + \varepsilon) + \text{Prob}(L \leq c, J > c, Z(X) \geq 0, J > L + \varepsilon). \end{aligned}$$

Hence we only need to show that:

$$\text{Prob}(L \leq c, J > c, Z(X) \geq 0, J > L + \varepsilon) > 0, \quad \forall c > 0.$$

Since when $Z(X) \geq 0$, $J = J(0)$ and thus $L = J - S$, we want to show that:

$$\text{Prob}(c - S < L \leq c, S > \varepsilon, Z(X) \geq 0) > 0.$$

Let $F_{L,S,Z(X)}(l, s, z)$ (resp. $F_L(l)$, and $F_{S,Z(X)}(s, z)$) be the joint distribution of $(L, S, Z(X))$ (resp. L , and $(S, Z(X))$), and $I(\cdot)$ the usual indicator function. We have:

$$\begin{aligned} & \text{Prob}(c - S < L \leq c, S > \varepsilon, Z(X) \geq 0) \\ &= \int I(c - s < l \leq c, s > \varepsilon, z \geq 0) dF_{L,S,Z(X)}(l, s, z) \\ &= \int I(c - s < l \leq c, s > \varepsilon, z \geq 0) dF_L(l) dF_{S,Z(X)}(s, z) \\ &= \int_{s > \varepsilon, z \geq 0} \left(\int_{c-s < l \leq c} dF_L(l) \right) dF_{S,Z(X)}(s, z) \\ &= \int_{s > \varepsilon, z \geq 0} (\text{Prob}(c - s < \chi^2(H - p) \leq c)) dF_{S,Z(X)}(s, z), \end{aligned}$$

where second equality follows from the independence of L and $(S, Z(X))$ and the last one follows from the fact that $L \sim \chi^2(H - p)$.

But,

$$\forall s > \varepsilon \geq 0, \text{ and } \forall c, \quad \text{Prob}(c - s < \chi^2(H - p) \leq c) \geq \text{Prob}(c - \varepsilon < \chi^2(H - p) \leq c).$$

Hence,

$$\begin{aligned}
& \int_{s>\varepsilon, z\geq 0} (\text{Prob}(c-s < \chi^2(H-p) \leq c)) dF_{S,Z(X)}(s, z) \\
& \geq \int_{s>\varepsilon, z\geq 0} (\text{Prob}(c-\varepsilon < \chi^2(H-p) \leq c)) dF_{S,Z(X)}(s, z) \\
& = \text{Prob}(c-\varepsilon < \chi^2(H-p) \leq c) \int_{s>\varepsilon, z\geq 0} dF_{S,Z(X)}(s, z) \\
& = \text{Prob}(c-\varepsilon < \chi^2(H-p) \leq c) \text{Prob}(S > \varepsilon, Z(X) \geq 0).
\end{aligned}$$

Since, by continuity and positivity on the positive half line of the $\chi^2(H-p)$ distribution, $\text{Prob}(c-\varepsilon < \chi^2(H-p) \leq c) > 0$, for all $c, \varepsilon > 0$ and $\text{Prob}(S > \varepsilon, Z(X) \geq 0) > 0$ from the *1st step*, we conclude that $\text{Prob}(c-S < L \leq c, S > \varepsilon, Z(X) \geq 0) > 0$ which concludes the proof. \square

Proof of Corollary 3.2. Since $p = 1$, G is the column vector $\frac{\partial^2 \rho}{\partial \theta^2}(\theta^0)$ and Lemma 2.3 guarantees that $G \neq 0$. Also, J now has the expression:

$$J = \min_{v \in \mathbb{R}} \left(X'WX + X'WGv^2 + \frac{1}{4}G'WGv^4 \right).$$

The first order necessary condition for optimality gives $v(2X'WG + G'WGv^2) = 0$ while the second order sufficient condition for $v = 0$ to be solution is $X'WG > 0$. If $X'WG < 0$, we can say, from the first order condition, that any solution v satisfies $2X'WG + G'WGv^2 = 0$. In the event that $X'WG = 0$, it appears that $v = 0$ is solution. In summary, we can write that: If $X'WG \geq 0$,

$$J = J(0) = X'WX$$

and if $X'WG < 0$, $J(v)$ is minimized at $v^2 = -2X'WG/G'WG$ so that

$$J = X'W^{1/2} \left(Id_H - W^{1/2}G(G'WG)^{-1}G'W^{1/2} \right) W^{1/2}X \equiv X'W^{1/2}\mathcal{P}W^{1/2}X \equiv L.$$

\mathcal{P} is the orthogonal projection matrix of the orthogonal of the subspace generated by the column vectors of $W^{1/2}G$.

Let $z = \frac{X'WG}{\sqrt{G'WG}}$. $z \sim N(0, 1)$ and clearly,

$$J = I(z \geq 0)J(0) + I(z < 0)L.$$

Now, we show that $I(z \geq 0)$ is independent of both $J(0)$ and L .

We have $Cov(z, \mathcal{P}W^{1/2}X) = Cov\left(\frac{X'WG}{\sqrt{G'WG}}, \mathcal{P}W^{1/2}X\right) = 0$. Thus, since X is a Gaussian vector, z is independent of $\mathcal{P}W^{1/2}X$ and so are $I(z \geq 0)$ and L .

To see that $I(z \geq 0)$ is independent of $J(0)$, we write $W^{1/2}X$ in the orthonormal basis

$$\left(W^{1/2}a_1, W^{1/2}a_2, \dots, W^{1/2}a_H \right)$$

of \mathbb{R}^H such that $a_1 = \frac{G}{\sqrt{G'WG}}$. (We choose a_1 such that the first component of $W^{1/2}X$ in this new basis is z). The coordinates of $W^{1/2}X$ in this basis are $(a'_1WX, a'_2WX, \dots, a'_HWX)$ and, by the invariance of the norm,

$$J(0) = X'WX = \sum_{h=1}^H (a'_hWX)^2.$$

Note that $Cov(a'_jWX, a'_iWX) = 0$ for $i \neq j$ so that $z = a'_1WX$ is independent of a'_jWX , $j = 2, \dots, H$. Hence, to claim that $I(z \geq 0)$ is independent of $J(0)$, it is sufficient to show that $(a'_1WX)^2 = z^2$ is independent of $I(z \geq 0)$. This becomes obvious once we see that $z \sim N(0, 1)$ has a symmetric distribution about the origin. This completes the proof. \square

References

- [1] Andrews, D. W. K. and X. Cheng, 2012. "Estimation and Inference with Weak, Semi-strong and Strong Identification," *Econometrica*, 80, 2153-2211.
- [2] Andrews, I. and A. Mikusheva, 2012. "Maximum Likelihood Inference in Weakly Identified DSGE Models," *Working Paper*, MIT, <http://economics.mit.edu/files/6593>.
- [3] Billingsley, P., 1961. "The Lindeberg-Lévy Theorem for Martingales," *Proc. Amer. Math. Soc.*, 12, 788-792.
- [4] Bollerslev, T., 1986. "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.
- [5] Chamberlain, G., 1986. "Asymptotic Efficiency in Semi-Parametric Models with Censoring," *Journal of Econometrics*, 32, 189-218.
- [6] Cragg, J. G. and S. G. Donald, 1993. "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222-240.
- [7] Cragg, J. C. and S. G. Donald, 1996. "Testing Overidentifying Restrictions in Unidentified Models," *Unpublished UBC discussion paper*, 96/20.
- [8] 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.
- [9] Dovonon, P. and S. Gonçalves, 2012. "Bootstrapping GMM Tests Under First Order Underidentification," *work in progress*, Concordia University.
- [10] 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.
- [11] Doz, C. and E. Renault, 2006. "Factor Volatility in Mean Models: a GMM Approach," *Econometric Reviews*, 25, 275-309.
- [12] Engle, R. F. and S. Kozicki, 1993. "Testing For Common Features," *Journal of Business and Economic Statistics*, 11(4), 369-395.
- [13] Engle, R. F. and J. Marcucci, 2006. "A long-run Pure Variance Common Features Model for the Common Volatilities of the Dow Jones," *Journal of Econometrics*, 132(1), 7-42.
- [14] 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.
- [15] Engle, R. F. and R. Susmel, 1993. "Common Volatility in International Equity Markets," *Journal of Business and Economic Statistics*, 11, 167-176.
- [16] Fiorentini, G., E. Sentana and N. Shephard, 2004. "Likelihood-based Estimation of Generalised ARCH Structures," *Econometrica*, 72, 1481-1517.
- [17] Gourieroux, C. and A. Monfort, 1989. "A General Framework for Testing a Null Hypothesis in a "Mixed" Form," *Econometric Theory*, 5(1), 63-82.
- [18] Hansen, L. P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.
- [19] Hecq, A., S. Laurent and F. C. Palm, 2012. "On the Univariate Representation of BEKK Models with Common Factors," *Working Paper No: RM/12/018*, Maastricht University.
- [20] Horn, A. R. and C. R. Johnson, 1985. "Matrix Analysis," *Cambridge University Press*.
- [21] Koul, H. L., 2002. "Weighted Empirical Processes in Dynamic Nonlinear Models," Springer-Verlag, New York.
- [22] Lee, L. F. and A. Chesher, 1986. "Specification Testing when Score Test Statistics are Identically Zero," *Journal of Econometrics*, 31, 121-149.

- [23] Lindner, A. M., 2009. "Stationarity, Mixing, Distributional Properties and Moments of GARCH(p,q)-Processes," *Handbook of Financial Time Series (eds T.G. Andersen, R.A. Davis, J.-P. Kreiß and T. Mikosch)*, 43-69. Springer-Verlag: New-York.
- [24] Melino, A., 1982. "Testing for Sample Selection Bias," *Review of Economic Studies*, 49, 151-153.
- [25] Rotnitzky, A., D. R. Cox, M. Bottai and J. Robins, 2000. "Likelihood-based Inference with Singular Information Matrix," *Bernoulli*, 6(2), 243-284.
- [26] Sargan, J. D., 1983. "Identification and lack of Identification," *Econometrica*, 51, 1605-1633.
- [27] Staiger, D. and J. H. Stock, 1997. "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.
- [28] Stock, J. H. and J. H. Wright, 2000. "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.
- [29] van der Vaart, A. W., 1998. "Asymptotic Statistics," *Cambridge University Press*.
- [30] van der Vaart, A. W. and J. A. Wellner, 1996. "Stochastic Convergence and Empirical Processes," Springer-Verlag, New York.