TESTING FOR COMMON CONDITIONALLY
HETEROSKEDASTIC FACTORS

PROSPER DOVONON
*Concordia University, Montreal, Quebec, H3G 1M8 Canada, and CIRANO, and
CIREQ*

ERIC RENAULT
Brown University, Providence, RI 02912, U.S.A., and CIRANO, and CIREQ

TESTING FOR COMMON CONDITIONALLY
HETEROSKEDASTIC FACTORS

By Prosper Dovonon and Eric Renault[1]

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 Hansen's (1982) J-test statistic is asymptotically distributed as the minimum of the limit of a certain random process with a markedly nonstandard distribution. If two assets are considered, this asymptotic distribution is a fifty–fifty mixture of χ^2_{H-1} and χ^2_H, where H is the number of moment conditions, as opposed to a χ^2_{H-1}. With more than two assets, this distribution lies between the χ^2_{H-p} and χ^2_H (p denotes 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 a possibly increasing overrejection rate with the number of assets. A Monte Carlo study illustrates these findings.

Keywords: Common features, GARCH factors, nonstandard asymptotics, GMM, GMM overidentification test, identification, first-order identification.

1. INTRODUCTION

Engle and Kozicki (1993) have given many examples of the 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 proposed 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), a 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-squared 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 applied this testing strategy to several common features of interest (regression common feature,

cofeature rank, Granger causality, and cointegration). When they came to the common conditionally heteroskedastic (CH) features, they acknowledged that it was their first nonlinear example. Unfortunately, they did not realize that, as already pointed out by Sargan (1983) in the context of instrumental variables (IV) estimation, nonlinearities may give rise to nonstandard 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 CH factors" that motivates the test for common CH features.

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

First, detecting a factor structure is a key issue for multivariate modeling of the 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 CH features is then a universal tool for detecting any of these factor structures.

Second, the singularity issue à 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 noticeably oversized. In other words, the mechanical application of Hansen's (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 overrejection and give ways to compute correct critical values or at least valid bounds for a conservative testing approach. More precisely, we show that the right asymptotic distribution for the J-test statistic under the null involves some $\chi^2(H - q)$ for $q < n - 1$ and this is the reason why the use of the critical value based on $\chi^2(H - n + 1)$ leads to overrejection. By contrast, the distribution $\chi^2(H)$ always provides a conservative upper bound.

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 more 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 to 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-squared distributions with larger 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 nonstandard rates of convergence of the MLE implied by the singularity of the Fisher information matrix. Of course, the issue of the singularity of the Jacobian matrix in the GMM is germane to the singularity of the Fisher information matrix in the MLE context and, following Sargan (1983), we get nonstandard rates of convergence of GMM estimators for quite similar reasons. We actually provide an interpretation in terms of the random Fisher information matrix closely related to the analysis proposed by Andrews and Mikusheva (2012) as well as Qu (2011). However, our main focus 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 CH factors and the intrinsic singularity that 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 CH features. Our main focus of interest is the degree of overrejection implied by the use of standard critical values. Since the correct asymptotic null distribution involves some $\chi^2(H - q)$ for $q < n - 1$, very large samples (as

are 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-squared at play may make a big difference in terms of the probability of rejection. Section 5 concludes and sketches other possible contexts of application of the general testing methodology put forward in this paper.

Appendix A in the Supplemental Material (Dovonon and Renault (2013)) discusses alternative regression-based approaches for testing for common CH factors. It shows that such alternative approaches do not protect against singularities discussed in this paper and regardless of the approach, tests based on standard chi-squared asymptotic distributions are not valid anymore. In addition, we make explicit in Appendix A an interpretation in terms of the Fisher information matrix for the kind of singularity as well as associated nonstandard asymptotic inference that are the focus of interest in this paper. This analysis, and in particular a comparison with Andrews and Mikusheva (2012), allows us to make even more explicit the difference between our setting and the kind of singularity implied by weak identification in the context of maximum likelihood, as discussed by Andrews and Mikusheva (2012) and also Qu (2011). All proofs appear in Appendix B of the Supplemental Material (Dovonon and Renault (2013)).

2. TESTING FOR COMMON CONDITIONALLY HETEROSKEDASTIC FEATURES

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

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

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 semidefinite 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 $\theta \neq 0$ in $\mathbb{R}^n$ such that $\mathrm{Var}(\theta' Y_{t+1} | \mathfrak{F}_t)$ 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) *We have* $\text{Rank}(\Lambda) = K$. (ii) *The term* $\text{Var}(\text{Diag}(D_t))$ *is nonsingular, where* $\text{Diag}(D_t)$ *is the K-dimensional vector with coefficients* $D_{kk,t}$ $(= \sigma_{kt}^2)$, $k = 1, \ldots, K$.

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

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

$\text{Var}(F_{t+1}|\mathfrak{F}_t) = V_t$, $E((F'_{t+1}, u'_{t+1})'|\mathfrak{F}_t) = 0$, $\text{Var}(u_{t+1}|\mathfrak{F}_t) = \Omega$, and $\text{Cov}(F_{t+1}, u_{t+1}|\mathfrak{F}_t) = 0$.

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

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

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 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 diagonal.[2] In other words, we need to preclude conditional correlations between the latent CH factors. Then the change of basis

---

[2]All 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.

P_t is immaterial and from now on, we can interpret the CH features as possibly (but not necessarily) underpinned by a CH factor model with uncorrelated factors and constant factor loadings:

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

$\mathrm{Var}(F_{t+1}|\mathfrak{F}_t) = D_t$, $E((F'_{t+1}, u'_{t+1})'|\mathfrak{F}_t) = 0$, $\mathrm{Var}(u_{t+1}|\mathfrak{F}_t) = \Omega$, and $\mathrm{Cov}(F_{t+1}, u_{t+1}|\mathfrak{F}_t) = 0$.

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, $k = 1, 2, \ldots, K$, 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 Λ'.

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

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

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

$$(4) \qquad H_0: \text{there exists } \theta \in \mathbb{R}^n, \theta \neq 0 \text{ such that } E\big((\theta' Y_{t+1})^2 | \mathfrak{F}_t\big) \text{ is constant.}$$

For the sake of expositional simplicity, we make sure that for all $\theta \in \mathbb{R}^n$,

$$E\big((\theta' Y_{t+1})^2 | \mathfrak{F}_t\big) = \text{constant} \quad \Leftrightarrow \quad \Lambda' \theta = 0$$

by maintaining the following assumption.

ASSUMPTION 2: *We have* $E(Y_{t+1}|\mathfrak{F}_t) = 0$.

Under Assumption 2, the existence of CH common features is obviously equivalent to the null hypothesis (4). For all practical purpose, this approach will work insofar as we are ready to assume that the vector Y_{t+1} of returns is unpredictable, that is, $E(Y_{t+1}|\mathfrak{F}_t)$ is constant. A preliminary demeaning would then allow us to be back to the framework of Assumption 2.

REMARK 2.3: It is worth noting that the point made by this paper is even more general. According to a common view of risk compensation (see Doz and Renault (2006) for more discussion), if there is some return predictability, it should go through risk premium and variance predictability. Then, with a standard linear risk compensation in the spirit of GARCH-in-mean modeling, the factor model (1) implies

$$(5) \qquad Y_{t+1} = \alpha + \Lambda \xi_t + \Lambda F_{t+1} + u_{t+1}.$$

Then we obviously have

$$(6) \qquad \Lambda' \theta = 0 \quad \Rightarrow \quad E\big((\theta' Y_{t+1})^2 | \mathfrak{F}_t \big) = \text{constant}.$$

The main goal of this paper is to show that a test of (4) performed as an overidentification test with standard critical values would be oversized. Implication (6) shows that the same test used to test for the existence of common features would be a fortiori oversized in the context of return predictability conformable to (5).

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

$$H_0(z) : \text{there exists } \theta \in \mathbb{R}^n, \ \theta \neq 0,$$
$$\text{such that } E\big(z_t [(\theta' Y_{t+1})^2 - c(\theta)] \big) = 0,$$

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

The consequence $H_0(z)$ is implied by H_0 insofar as the variables z_t are valid instruments, that is, are $\mathfrak{F}_t$-measurable. Besides validity, the instruments z_t must identify the CH common features θ if one aims to devise a test of $H_0(z)$ based on Hansen's (1982) theory of the J-test for overidentification.

By the law of iterated expectations, the CH common features model (1) gives

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

and then, by a simple matrix manipulation,

$$(7) \qquad E\big(z_t ((\theta' Y_{t+1})^2 - c(\theta)) \big) = \text{Cov}\big(z_t, \text{tr}(\theta' \Lambda D_t \Lambda' \theta) \big)$$
$$= \text{Cov}\big(z_t, \text{Diag}'(\Lambda' \theta \theta' \Lambda) \text{Diag}(D_t) \big)$$
$$= \text{Cov}\big(z_t, \text{Diag}(D_t) \big) \text{Diag}(\Lambda' \theta \theta' \Lambda).$$

The convenient identification assumption about the vector z_t of instruments is then given by the following assumption.

ASSUMPTION 3: (i) *The vector z_t is $\mathfrak{F}_t$-measurable and* $\mathrm{Var}(z_t)$ *nonsingular*; (ii) $\mathrm{Rank}[\mathrm{Cov}(z_t, \mathrm{Diag}(D_t))] = K$.

Assumption 3(i) is standard. Assumption 3(ii) is nonrestrictive 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, $k = 1, \ldots, K$, there exists at least one instrument z_{ht}, $h = 1, \ldots, H$, that is correlated with this combination.

From (7), we see that under Assumptions 1, 2, and 3, $H_0(z)$ amounts to $\mathrm{Diag}(\Lambda' \theta \theta' \Lambda) = 0$ and then implies that $\|\Lambda' \theta\|^2 = \mathrm{tr}(\Lambda' \theta \theta' \Lambda) = 0$, that is, θ is a common feature. Conversely, any common feature clearly fulfills the condition of $H_0(z)$. We have thus proved the following lemma.

LEMMA 2.2: *Under Assumptions 1, 2, and 3, the common features are the solutions θ in $\mathbb{R}^n$, $\theta \neq 0$, of the moment restrictions*

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

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

As in Engle and Kozicki (1993), CH common features are thus identified by moment restrictions $H_0(z)$. Then $H_0(z)$ will be considered as the null hypothesis under test so as to test for common features.

Following Hansen (1982) as well as Engle and Kozicki (1993), we aim in this paper to test 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 maintain the following assumption.

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

Engle and Kozicki (1993) focused on the particular case $K = n - 1$ so as 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 an 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 4, we can change the order of expectation and differentiation, and

write

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

$$= E\big[z_t\{2(\theta' Y_{t+1}) Y'_{t+1} - 2E\big[(\theta' Y_{t+1}) Y'_{t+1}\big]\}\big]$$

$$= 2\,\mathrm{Cov}\big(z_t, \big[Y_{t+1} Y'_{t+1}\big]\theta\big).$$

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 Assumptions* 2 *and* 4, *for any common feature* θ,

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

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 CH 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 CH effects would not allow us to circumvent the problem of 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 nonzero higher order derivatives are of order 2. Let Θ be the parameter space and let us assume that exclusion restrictions characterize a set $\Theta_* \subset \Theta \subset \mathbb{R}^n$ of parameters that contains at most only one unknown common feature θ^0, up to a normalization condition $\mathcal{N}$.

ASSUMPTION 5: *We have* $\theta \in \Theta_* \subset \mathbb{R}^n$ *such that* $\Theta^* = \Theta_* \cap \mathcal{N}$ *is a compact set and*

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

REMARK 2.4: A typical normalization condition would be the unit cost condition to interpret $\theta' Y_{t+1}$ as a return per dollar 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, for example, Asset 1, entering the common feature. To avoid this assumption, an alternative would be $\mathcal{N} = \{\theta \in \mathbb{R}^n, \sum_{i=1}^{n} \theta_i^2 = 1\}$. This latter alternative, albeit feasible, will not be explicitly considered to keep the simplicity of linear normalization. In this context, Assumption 5 will be fulfilled with $\Theta_* = \Theta$, that is, $\Theta^* = \Theta \cap \mathcal{N}$ in the setting of Engle and Kozicki (1993), that is, $K = n - 1$. In this case, Θ may be an arbitrarily large compact subset of $\mathbb{R}^n$. If more than one dimension of common features exists ($K < n - 1$), 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 5 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 us 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 1–5, global identification amounts to second-order identification.

LEMMA 2.3: *Let $\rho_h(\theta) \equiv E(z_{ht}((\theta' Y_{t+1})^2 - c(\theta)))$ and let $h = 1, \ldots, H$ be the hth component of $\rho(\theta)$. Under Assumptions 1–5, we have*

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

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

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

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 that 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 $H_0(z)$:

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

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 $c(\theta)$ with 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 feasible moment conditions

$$(8) \qquad E\big(\phi_{t,T}(\theta)\big) = 0, \quad \phi_{t,T}(\theta) = (z_t - \bar{z}_T)\big((\theta' Y_{t+1})^2 - \bar{c}_T(\theta)\big), \quad \theta \in \Theta^*,$$

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 now have a double array structure 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,

$$(9) \qquad \sqrt{T}\bar{\phi}_T(\theta) = \sqrt{T}\bar{\psi}_T(\theta) + \sqrt{T}\bar{v}_T(\theta) + o_P(1),$$

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 (9) 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 $\frac{\partial\psi_t}{\partial\theta'}(\theta)|_{\theta=\theta^0}$ 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 $\sqrt{T}\frac{\bar{\psi}_T}{\partial\theta'}(\theta)|_{\theta=\theta_0}$ are asymptotically normal. Note that $\sqrt{T}\frac{\partial\bar{v}_T}{\partial\theta'}(\theta)|_{\theta=\theta^0} = \mu_z \cdot \sqrt{T}(\frac{\partial c}{\partial\theta'}(\theta)|_{\theta=\theta^0} - 2\theta^{0'}\frac{1}{T}\sum_{t=1}^{T} Y_t Y'_t)$ is also asymptotically normal.

Overall, we will use the facts 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, $\sqrt{T}(\frac{\partial\bar{\psi}_T}{\partial\theta'}(\theta)|_{\theta=\theta^0}, \frac{\partial\bar{v}_T}{\partial\theta'}(\theta)|_{\theta=\theta^0}) = O_P(1)$ to draw the following conclusion.

COROLLARY 3.1: *If Assumptions 1–5 hold, then $\sqrt{T}\bar{\phi}_T(\theta^0)$ is asymptotically normal with asymptotic variance $\Sigma(\theta^0)$ given by*

$$E\big((z_t - \mu_z)(z_t - \mu_z)'\big((\theta^{0'}Y_{t+1})^2 - c(\theta^0))^2\big) \quad and$$

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

REMARK 3.1: Note that if one interprets the null hypothesis H_0 as stemming from the 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: $\Sigma(\theta^0) = \mathrm{Var}((\theta^{0'}Y_{t+1})^2)\,\mathrm{Var}(z_t)$.

We assume throughout that the asymptotic variance $\Sigma(\theta^0)$ is nonsingular.

It is worth noting that, in contrast to the weak identification literature (Stock and Wright (2000)), we do not need a functional central limit theorem for the empirical process $(\sqrt{T}(\bar{\phi}_T(\theta) - E\bar{\phi}_T(\theta)))_{\theta\in\Theta}$. 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, 3, and 5 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 $\mathrm{plim}(W_T) = W$ 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 light

of Remark 2.4, setting K to $n-1$ is not overly restrictive.[3] We also mention that our results do not depend on our particular choice of linear normalization as long as the identification of θ^0 via Assumption 5 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 $\theta = (\theta_i)_{1 \leq i \leq n-1} \in \Theta \subset \mathbb{R}^{n-1}$ denotes the parameter of interest. We shall consider the functions $\phi_t(\theta)$, $\bar{\phi}_T(\theta)$, and $\rho(\theta)$ as defined on $\Theta \subset \mathbb{R}^{n-1}$. For the sake of notational simplicity, we will often denote by p $(= n-1)$ the dimension of the unknown vector θ of parameters of interest. We will maintain throughout the assumption that the true unknown value is an interior point.

ASSUMPTION 6: *The vector θ^0 belongs to the interior of $\Theta \subset \mathbb{R}^p = \mathbb{R}^{n-1}$.*

3.2. *Rates of Convergence*

Throughout the paper, inference about the unknown vector θ of common features will be based on the aforementioned feasible moment conditions:

$$E\big(\bar{\phi}_T(\theta)\big) = 0, \quad \theta \in \Theta \subset \mathbb{R}^p.$$

Then, along the lines of Chamberlain (1986), one could easily deduce from Proposition 2.1 that the partial information matrix for θ is zero. Therefore (see Chamberlain's Theorem 2), it must be impossible to build a regular $\sqrt{T}$-consistent estimator for θ. In our GMM context, the intuition of this result is quite clear. For any given positive definite matrix W of size H, one gets a consistent GMM estimator $\hat{\theta}_T$ as the solution of

$$(10) \qquad \hat{\theta}_T = \arg\min_{\theta \in \Theta} \bar{\phi}'_T(\theta) W_T \bar{\phi}_T(\theta),$$

where W_T is a sequence of positive definite random matrices such that $\mathrm{plim}(W_T) = W$. Then, due to Proposition 2.1, in the Taylor expansion of the first-order conditions of the minimization problem (10), the term of degree 1 (involving the Jacobian matrix of the moment conditions) will be negligible in front of the quadratic term. Thus, the latter term will determine the asymptotic order of magnitude of $\hat{\theta}_T - \theta^0$. Since the quadratic term will be of order $T^{1/2}$, we can only extract an order $T^{1/2}$ for $\|\hat{\theta}_T - \theta^0\|^2$. Hence we get the following result.

---

[3]It 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.

PROPOSITION 3.1: *Under Assumptions* 1–5, *if* $\hat{\theta}_T$ *is the GMM estimator as defined by Equation* (10), *then*

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

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 $T^{1/2}$. Proposition 3.2 below completes this statement as follows. On the one hand, it proves 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 $T^{1/4}(\hat{\theta}_T - \theta^0)$ may converge to zero with a positive probability. In other words, depending on which part of the sample space the infinite "observed" sample lies in, the rate of convergence may be either $T^{1/4}$ or faster, possibly $T^{1/2}$.

The relevant partition of the sample space will be defined by the asymptotic behavior of the sequence $Z_T(\theta^0)$ of symmetric random matrices of size p, with coefficients

$$Z_{ij,T}\left(\theta^0\right) = \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}\left(\theta^0\right) W \sqrt{T} \bar{\phi}_T\left(\theta^0\right), \quad i, j = 1, \ldots, p.$$

By Assumption 4, the sequence $Z_T(\theta^0)$ converges in distribution toward a random matrix $Z(X)$ with Gaussian coefficients

$$Z_{ij}(X) = \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}\left(\theta^0\right) W X; \quad X \sim N\left(0, \Sigma\left(\theta^0\right)\right).$$

We show in Appendix A of Dovonon and Renault (2013) that the random matrix $Z(X)$ can be interpreted as a random Fisher information matrix computed from the dominating part of the relevant Hessian matrix. As a matter of fact, Proposition 3.2 below shows that it is when the matrix $Z(X)$ is positive semidefinite (a random event that we denote by $(Z(X) \geq 0)$) that our statistical information about θ is the "richest" so that a rate of convergence faster than $T^{1/4}$ is warranted for any GMM estimator associated to the weighting matrix W. By contrast, when $Z(X)$ is not positive semidefinite (a random event that we denote by $\overline{(Z(X) \geq 0)}$), our statistical information about θ is really impaired by the lack of first-order identification, so that we cannot do better than the slow rate of convergence $T^{1/4}$. The precise result is as follows.

PROPOSITION 3.2: *If Assumptions* 1–6 *hold with* $K = n - 1$, *then the sequence* $(T^{1/4}(\hat{\theta}_T - \theta^0)', \text{Vec}'(Z_T(\theta^0)))'$ *has at least one subsequence that converges in distribution, and for any such subsequence with limit distribution* $(V', \text{Vec}'(Z(X)))'$, *we have* $0 < \text{Prob}(Z(X) \geq 0) \leq \frac{1}{2}$ *and*

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

Proposition 3.2 establishes that we have positive probability to observe each of the two possible patterns of convergence: (i) with positive probability $\pi \leq 1/2$, the rate of convergence of $\hat{\theta}_T$ is faster than $T^{1/4}$ and $T^{1/4}(\hat{\theta}_T - \theta^0)$ converges to 0 in probability; (ii) with probability $1 - \pi \geq 1/2$, the rate of convergence of $\hat{\theta}_T$ is $T^{1/4}$ but never faster and $T^{1/4}(\hat{\theta}_T - \theta^0)$ is asymptotically a nondegenerate random variable with a zero-probability mass at the value 0. At least some linear combinations of $\hat{\theta}_T$ cannot converge faster than the slowest possible rate $T^{1/4}$.

In the particular case of interest where $p = 1$ (two assets with one common feature), it is obvious that $\pi = 1/2$: for the real-valued and zero mean Gaussian variable $Z(X) = \frac{\partial^2 \rho'}{\partial \theta^2}(\theta^0) W X$, positive and negative values are equally likely. More generally, the heterogeneity of rates of convergence over the sample space is characterized by conditions that make the matrix $Z(X)$ positive semidefinite or not, that is, by sign restrictions on some multilinear functions of the Gaussian vector with zero mean X. We end up with a mixture of asymptotic distributions (with different rates) that are dependent on the location of the (asymptotically) observed sample, that is, a setting germane to the framework of pretest estimators.

This classification of rates of convergence extends the one described by Sargan (1983) in the context of linear instrumental variables (with nonlinearities 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 nonstandard asymptotic distribution of the J-test statistic for overidentification induced by the nonstandard 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 (10) 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 with respect to $(\hat{\theta}_T - \theta^0)$), 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 of Dovonon and Renault (2013)),

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

we have

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

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

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

where $X \sim N(0, \Sigma(\theta^0))$. Note that $X'WG \operatorname{Vec}(vv') = v'Z(X)v$ so that $J^W(v)$ is also given by

$$(11) \qquad J^W(v) = X'WX + v'Z(X)v + \frac{1}{4} \operatorname{Vec}'(vv') G'WG \operatorname{Vec}(vv'), \quad v \in \mathbb{R}^p.$$

By construction, for each $v \in \mathbb{R}^p$, $\hat{J}^W(v)$ converges in distribution toward $J^W(v)$. Lemma B.6 of Dovonon and Renault (2013, 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 toward the minimum of $J^W(v)$. This is formally stated in the following theorem.

THEOREM 3.1: *If Assumptions 1–5 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 toward $J^W = \min_{v \in \mathbb{R}^p} J^W(v)$.*

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 closed form expression. In usual cases, polynomials of degree 2 are often derived as the limiting process that yields the usual chi-squared 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 a 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 of simulating a large number of realizations of X and getting 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 (2013) 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 that is still tightly related to chi-squared, 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 closed form formulas for the distribution of J^W and not by efficiency considerations. Since the rates of convergence of any GMM estimator are heterogeneous depending on the occurrence of the event $(Z(X) \geq 0)$ (the occurrence of which itself depends on 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

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

where J_T is the minimum value of the GMM objective function using the so-called optimal weighting matrix, which is defined as a consistent estimate of the inverse of the moment conditions' long run variance, that is, $W^{-1} = \Sigma(\theta^0) \equiv \lim_{T \to \infty} \mathrm{Var}(\sqrt{T} \bar{\phi}_T(\theta^0))$. From Corollary 3.1, one may consider

$$(12) \qquad W_T^{-1} = \frac{1}{T} \sum_{t=1}^{T-1} (z_t - \bar{z}_T)(z_t - \bar{z}_T)' \big((\tilde{\theta}_T' Y_{t+1})^2 - \bar{c}_T(\tilde{\theta}_T)\big)^2,$$

where $\tilde{\theta}_T$ denotes any consistent estimator of θ^0, for instance, a GMM estimator obtained with identity matrix 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 J_T^W above, 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 no longer deserves

to be called optimal. However, it will allow us a direct comparison with two chi-squared distributions, namely $\chi^2(H)$ and $\chi^2(H - p)$. First, we have, by definition,

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

However, due to the nullity of the Jacobian matrix, the common asymptotic theory that states that J_T behaves in large samples as a chi-squared 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

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

Obviously

$$(13) \qquad J \geq L = \min_{u \in \mathbb{R}^{p^2}}\left\{X'WX + X'WGu + \frac{1}{4}u'G'WGu\right\}.$$

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, which is 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* (13) *(and $W = \Sigma^{-1}(\theta^0)$), we have the decomposition*

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

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

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*

$$L \leq J \leq J(0) \quad \text{with}$$

(i) $Z(X) \geq 0 \Rightarrow J = J(0)$ *and* (ii) $\forall c > 0, \operatorname{Prob}(L > c) < \operatorname{Prob}(J > c)$.

REMARK 3.2: The proof of Theorem 3.2 actually shows that $\forall c > 0$, $\mathrm{Prob}(L > c, Z(X) \geq 0) < \mathrm{Prob}(J > c, Z(X) \geq 0)$. 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, $\mathrm{Prob}((L > c) \cap B) \leq \mathrm{Prob}((J > c) \cap B)$. 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 3.3 below), for the overrejection 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 statistic 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 $\chi^2(H - q), q = 0, 1, \ldots, p$. Then it is not surprising that critical values computed from $\chi^2(H - p)$ lead to overrejection as formally proved in Remark 3.3 below.

REMARK 3.3: By application of the Portmanteau lemma (see, e.g., van der Vaart (1998, Lemma 2.2.(v), p. 6)), the convergence in distribution of J_T toward J implies that

$$\liminf_{T \to \infty} \mathrm{Prob}\big(J_T > \chi^2_{1-\alpha}(H - p)\big) \geq \mathrm{Prob}\big(J > \chi^2_{1-\alpha}(H - p)\big)$$

$$> \mathrm{Prob}\big(L > \chi^2_{1-\alpha}(H - p)\big) = \alpha.$$

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 $\chi^2(H - q), q = 0, 1, \ldots, p$, 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 always has a probability $1/2$. In other words, we can show the following corollary.

COROLLARY 3.2: *Under the same conditions as Theorem 3.2, if $p = 1$, the J-test statistic follows asymptotically under the null a fifty–fifty mixture of χ^2_H and χ^2_{H-1}. 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.

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 overreject 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 nonnegligible. 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 $\chi^2(H - p)$, 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.

REMARK 3.4: Of course, the price to pay for a conservative test based on $\chi^2(H)$ is some power loss. However, it is worth stressing that this power loss is not as much an issue here as in common settings. In general, failure to reject a null hypothesis (when it should be rejected) leads to omit some relevant variables and may result in biased estimators. By contrast, more often than not, the type 1 error (keeping some irrelevant variables) has only a price in terms of efficiency loss. Here, failure to reject the null (while it should be rejected) means that we overestimate the number of common features, that is, we underestimate the number of factors. In other words, some relatively minor sources of conditional heteroskedasticity might be omitted. However, the type 1 error, namely the overestimation of the number of factors, would be much more detrimental. It is not only a matter of efficiency loss, but even more importantly, a matter of identification. If some factors are irrelevant, they cannot be disentangled from homoskedastic noise (as captured by the matrix Ω in (1)) and they are hardly identified. This lack of identification will likely lead to very noisy estimators, not only for the weakly identified factor loadings, but also, by contamination, for all the model parameters.

Overall, the power issues, that are not explicitly addressed in this paper, are as follows.
  (i) We do claim that it is less detrimental to possibly underreject (for instance, by using the conservative critical values based on $\chi^2(H)$) than to overreject by using the common critical values based on $\chi^2(H - p)$.
  (ii) Of course, when the number of common features under test is small in front of the number of instruments, the difference between the two critical values becomes relatively small. For instance, from Corollary 3.2, one can compute the asymptotic null rejection probabilities of the J-test if an applied researcher (mistakenly) uses the $(1 - \alpha)$ quantile of a $\chi^2(1)$ distribution. With

nominal size 5% and $H = 2, 4, 6, 10$, they equal 9.8, 7.4, 6.8, and 6.3%, respectively. Therefore, in more overidentified situations, the overrejection becomes less detrimental.

(iii) Strictly speaking, the singularity issue that is the focus of interest of this paper does not occur under the alternative hypothesis. Therefore, the textbook treatment of power issues is still valid. First (see, e.g., Hall (2005, Theorem 5.2)), the J-test for overidentification is consistent under fixed alternatives and, second (see, e.g., Hall (2005, p. 156) or Andersen and Sorensen (1996)), in finite samples, the larger is the number of degrees of freedom, the smaller is the power. For this reason, a researcher may want to keep $(H - p)$ relatively small; that is typically the case where the risk of overrejection pointed out in this paper will be the largest.

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 CH factors. However, it is worth keeping in mind that the nonstandard behaviors 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 nonstandard behaviors are prevalent.

We simulate according to five designs (D1–D5) an asset return vector process Y_{t+1} that satisfies

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

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 generalized autoregressive conditionally heteroskedastic (GARCH(1, 1)) dynamic, that is,

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

where ε_t's are independent and identically distributed normal random variables with mean 0 and variance 1 ($\varepsilon_{t+1} \sim \text{NID}(0, 1)$). 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 is $U_{t+1} \sim \text{NID}(0, 0.5\text{Id}_2)$; 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 $\Lambda = \text{Id}_2$ and the idiosyncratic shocks $U_{t+1} \sim \text{NID}(0, 0.5\text{Id}_2)$.

Designs D3, D4, and D5 all generate trivariate vectors of three asset returns Y_{t+1} with one, two, and three GARCH factors, respectively. For all of them, the idiosyncratic shocks $U_{t+1} \sim \text{NID}(0, 0.5\text{Id}_3)$. 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 4 are fulfilled. (We refer to Bollerslev (1986) and Lindner (2009) for the conditions that guarantee these probabilistic properties for GARCH processes.)

Designs D1 and D4 satisfy 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 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 2.

The parameter 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 behaviors of interest to be confirmed for larger samples than those commonly relied on.

For each simulated sample, we evaluate the GMM estimator according to (10). The weighting matrix W_T is set to the inverse of $\hat{\Sigma}$ given by (12) 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 use $z_{2t} = (Y_{1t}^2, Y_{2t}^2, Y_{3t}^2)'$ to test for common GARCH factors for the trivariate simulated returns. While Engle and Susmel (1993) suggested 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: overrejection 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 fifty–fifty mix-
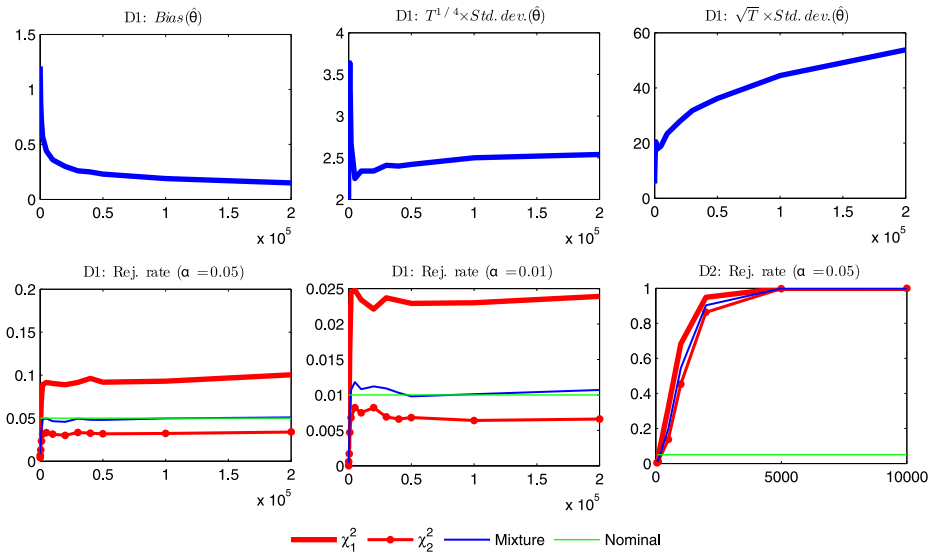
FIGURE 1.—Simulated bias and standard deviation of the GMM estimator $\hat{\theta}$ in design D1 (2 Assets—1 Factor: (2A-1F)) and simulated rejection rate of the J-test using critical values from χ_1^2, χ_2^2, and the mixture $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$. Nominal levels 5% and 1% for design D1 and 5% for D2 (2A-2F).

ture 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 overrejection 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 behavior of the J-test when the moment condition model is well specified but nonidentified, 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 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 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 fifty–fifty mixture of chi-squares derived as an asymptotic distribution provides a quite accurate approximation (an actual rejection rate between 4.5% and 5% for a nominal rate of 5%) for sample sizes greater than 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 are more than 170% of the nominal.

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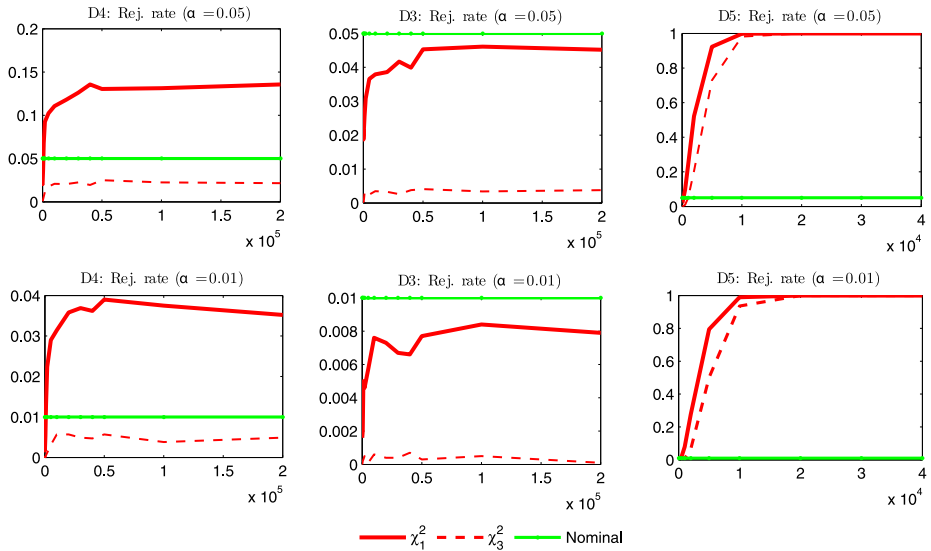
FIGURE 2.—Simulated rejection rate of the J-test using critical values from χ_1^2 and χ_3^2 for designs D3 (3A-1F), D4 (3A-2F), and D5 (3A-3F). Nominal levels 5% and 1%.

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 that accounts for first-order underidentification) would therefore underreject. This suggests that the results of Cragg and Donald (1996) and Staiger and Stock (1997), namely that the GMM overidentification test is undersized in the case of global identification failure, also hold in our context.

5. CONCLUSION

This paper sets the focus on an important albeit simple case of nonstandard behavior 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 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 with first-order identification failure is at stake with more complicated nonquadratic settings.

Extensions of these results to a GMM framework is worthwhile for at least two reasons. First, as exemplified in the present paper, nonstandard 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. Nonquadratic 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.

REFERENCES

ANDERSEN, T. G., AND B. E. SORENSEN (1996): "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business & Economic Statistics*, 14 (3), 328–352. [2581]

ANDREWS, D. W. K., AND X. CHENG (2012): "Estimation and Inference With Weak, Semi-Strong and Strong Identification," *Econometrica*, 80, 2153–2211. [2563]

ANDREWS, I., AND A. MIKUSHEVA (2012): "Maximum Likelihood Inference in Weakly Identified DSGE Models," Working Paper, MIT, available at http://economics.mit.edu/files/6593. [2563, 2564]

BILLINGSLEY, P. (1961): "The Lindeberg–Lévy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12, 788–792. [2572]

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. [2582]

CHAMBERLAIN, G. (1986): "Asymptotic Efficiency in Semi-Parametric Models With Censoring," *Journal of Econometrics*, 32, 189–218. [2573]

CRAGG, J. G., AND S. G. DONALD (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240. [2563]

——— (1996): "Testing Overidentifying Restrictions in Unidentified Models," Unpublished UBC Discussion Paper 96/20, UBC. [2563,2584]

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. [2562]

DOVONON, P., AND S. GONÇALVES (2013): "Bootstrapping the GMM Overidentification Test Under First Order Underidentification," Working Paper, Concordia University. [2577]

DOVONON, P., AND E. RENAULT (2009): "GMM Overidentification Test With First Order Underidentification," Working Paper, UNC, available at http://www.unc.edu/depts/econ/profs/renault/J_testDR20090824.pdf. [2570]

——— (2013): "Supplement to 'Testing for Common Conditionally Heteroskedastic Factors': Extensions and Proofs," *Econometrica Supplemental Material*, 81, http://www.econometricsociety.org/ecta/supmat/10082_proofs.pdf. [2564,2574-2576]

DOZ, C., AND E. RENAULT (2006): "Factor Stochastic Volatility in Mean Models: A GMM Approach," *Econometric Reviews*, 25, 275–309. [2562,2567]

ENGLE, R. F., AND S. KOZICKI (1993): "Testing for Common Features," *Journal of Business & Economic Statistics*, 11 (4), 369–395. [2561,2568,2570,2571,2580]

ENGLE, R. F., AND R. SUSMEL (1993): "Common Volatility in International Equity Markets," *Journal of Business & Economic Statistics*, 11, 167–176. [2582]

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. [2562]

FIORENTINI, G., E. SENTANA, AND N. SHEPHARD (2004): "Likelihood-Based Estimation of Latent Generalised ARCH Structures," *Econometrica*, 72, 1481–1517. [2562,2582]

HALL, A. R. (2005): *Generalized Method of Moments*. Advanced Texts in Econometrics. Oxford: Oxford University Press. [2581]

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [2561,2562,2567,2568,2571,2578]

HECQ, A., S. LAURENT, AND F. C. PALM (2012): "On the Univariate Representation of BEKK Models With Common Factors," Working Paper RM/12/018, Maastricht University. [2565]

KOUL, H. L. (2002): *Weighted Empirical Processes in Dynamic Nonlinear Models*. New York: Springer-Verlag. [2576]

LEE, L. F., AND A. CHESHER (1986): "Specification Testing When Score Test Statistics Are Identically Zero," *Journal of Econometrics*, 31, 121–149. [2563,2585]

LINDNER, A. M. (2009): "Stationarity, Mixing, Distributional Properties and Moments of GARCH(p, q)-Processes," in *Handbook of Financial Time Series* (ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch). New York: Springer-Verlag, 43–69. [2582]

MELINO, A. (1982): "Testing for Sample Selection Bias," *Review of Economic Studies*, 49, 151–153. [2563]

QU, Z. (2011): "Inference and Specification Testing in DSGE Models With Possible Weak Identification," Working Papers Series WP2011-058, Department of Economics, Boston University. [2563,2564]

ROTNITZKY, A., D. R. COX, M. BOTTAI, AND J. ROBINS (2000): "Likelihood-Based Inference With Singular Information Matrix," *Bernoulli*, 6 (2), 243–284. [2563,2575,2585]

SARGAN, J. D. (1983): "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633. [2562,2563,2575]

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [2562,2584]

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM With Weak Identification," *Econometrica*, 68, 1055–1096. [2562,2563,2572]

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press. [2579]

Dept. of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada, and CIRANO, and CIREQ; prosper. dovonon@concordia.ca

and

Dept. of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, U.S.A., and CIRANO, and CIREQ; Eric_Renault@brown.edu.