

GMM Overidentification Test with First Order Underidentification*

Prosper DOVONON[†] and Éric RENAULT[‡]

August 24, 2009

Abstract

This paper revisits the asymptotic theory of GMM when the moment conditions identify a unique parameter true value θ^0 but the rank condition of the Jacobian matrix at θ^0 fails. The possibility in case of nonlinear moment restrictions of such simultaneous global identification but first order under-identification has already been pointed out by Sargan (1983). The contribution of this paper is to provide a general asymptotic theory when one can maintain an assumption of second order identification. While this issue has been addressed in a maximum likelihood context by Lee and Chesher (1986) and Rotnitzky, Cox, Bottai and Robbins (2000), we set the focus on the asymptotic behaviour of the GMM overidentification test statistic J_T . We show that with H moment conditions, when the Jacobian matrix of the moment conditions evaluated at θ^0 is of rank $p - 1$, where p is the number of parameters, the asymptotic distribution of J_T is a half-half mixture of χ_{H-p}^2 and $\chi_{H-(p-1)}^2$ instead of the standard χ_{H-p}^2 . In other words, the distribution of J_T for large sample sizes T has fatter tails, leading to over-rejection of the null of valid moments when using standard critical values. The practical significance of this oversize problem is illustrated by Monte Carlo experiments in the context of a test for common GARCH features proposed by Engle and Kozicki (1993).

Keywords: Nonstandard asymptotics, GMM, GMM overidentification test, identification, first order identification, second order identification, common GARCH features.

*We would like to thank Manuel Arellano, Valentina Corradi, Sílvia Gonçalves and Enrique Sentana for many helpful discussions.

[†]Barclays Wealth, London UK. Email: prosper.dovonon@barclayswealth.com.

[‡]University of North Carolina at Chapel Hill, CIRANO and CIREQ. Email: renauld@email.unc.edu.

1 Introduction

Generalized Method of Moments (GMM) provides a computationally convenient method for inference on the structural parameters of economic models. The method has been applied in many areas of economics but it was in empirical finance that the power of the method was first illustrated. Hansen (1982) introduced GMM and presented its fundamental statistical theory, Hansen and Hodrick (1980) and Hansen and Singleton (1982) showed the potential of the GMM approach to testing economic theories through their empirical analyzes of, respectively, foreign exchange markets and asset pricing. In such contexts, the cornerstone of the GMM inference is a set of conditional moment restrictions. More generally, GMM is well suited for the test of an economic theory every time the theory can be encapsulated in the postulated unpredictability of some error term $u(Y_t, \theta)$ given as a known function of p unknown parameters $\theta \in \Theta \subset \mathbb{R}^p$ and a vector of observed random variables Y_t . Then, the testability of the theory of interest is akin to the testability of a set of conditional moment restrictions that takes the form

$$E_t(u(Y_{t+1}, \theta^0)) = 0 \tag{1}$$

where the operator $E_t(\cdot)$ stands for the conditional expectation given the available information at time t . Moreover, under the null hypothesis that the theory summarized by the restrictions (1) is true, these restrictions are supposed to uniquely identify the true unknown value θ^0 of the parameters. The GMM way to proceed is to consider a set of H instruments z_t assumed to belong to the available information at time t and to summarize the testable implications of (1) by the implied unconditional moment restrictions

$$E(z_t \otimes u(Y_{t+1}, \theta^0)) = 0 \tag{2}$$

The recent weak-instruments literature (see e.g. Stock and Wright (2000)) has stressed that the standard asymptotic theory of GMM inference may be misleading due to the insufficient correlation between the instruments z_t and the *local explanatory variables* $\partial u(Y_{t+1}, \theta^0)/\partial \theta'$. Many asset pricing applications of GMM are focused on the study of a pricing kernel as provided by some financial theory. This pricing kernel will be typically either a linear function of the parameters of interest, as in linear-beta pricing models, or a log-linear one as in most of the equilibrium based pricing models where parameters of interest are preference parameters. In all these examples, the weak instrument problem simply relates to some lack of predictability of pricing factors from some lagged variables.

We rather focus in this paper on the predictability of some polynomial functions like conditional variance, skewness, kurtosis or any conditional higher order moments. The resulting non-linearity with respect to the parameters of interest is then likely to cause a different kind of weak identification issue. To see this, let us informally assume that, for some given exponent $n > 1$, the error term in (1) is

$$u(Y_{t+1}, \theta) = v^n(Y_{t+1}, \alpha) - \beta \tag{3}$$

While the vector of unknown parameters is $\theta = (\alpha', \beta')$, the main focus of our interest is the existence

of some true unknown value α^0 making $v^n(Y_{t+1}, \alpha^0)$ unpredictable. This predictability issue is at stake to understand the time variability of risk premiums ($n = 2$), of skewness compensation ($n = 3$), etc. When addressing the issue of multivariate dynamic modeling of higher order moments, the researcher will typically try to capture the commonality of these conditional moment dynamics through some common factors. The validity of the conditional moment restrictions (1) in the context (3) is precisely tantamount to the possibility of reduced rank dynamic modeling through some common factors (see e.g. Doz and Renault (2006) and references therein). The problem with the standard GMM inference in this context is that the focus on the correlation between the instruments z_t and the variables $\frac{\partial u(Y_{t+1}, \theta_0)}{\partial \theta'}$ is indeed misleading. In particular, the genuine *local explanatory variables* are now $\frac{\partial v(Y_{t+1}, \alpha^0)}{\partial \alpha'}$ rather than $\frac{\partial u(Y_{t+1}, \theta^0)}{\partial \alpha'} = nv^{n-1}(Y_{t+1}, \alpha^0) \frac{\partial v(Y_{t+1}, \alpha^0)}{\partial \alpha'}$. The difference is important because intuitively (see Section 5 for a formal proof), in the context of a data generating process with common factors, the unpredictability of $v^n(Y_{t+1}, \alpha^0)$ is likely to cause a lack of predictability of $v^{n-1}(Y_{t+1}, \alpha^0)$ and in turn a weak correlation between the instruments z_t and $\frac{\partial u(Y_{t+1}, \theta^0)}{\partial \theta'}$. In other words, we face a kind of weak identification problem not because the phenomenon of interest is weakly identified (the instruments z_t and some polynomial function of the genuine *local explanatory variables* $\frac{\partial v(Y_{t+1}, \alpha_0)}{\partial \alpha'}$ are strongly correlated) but because the standard GMM asymptotic theory does not set the focus on the right object.

The main contribution of this paper is to derive the asymptotic distribution of the Hansen's (1982) J_T test statistic for overidentification in Equation (2) while the standard theory does not apply due to a rank deficiency in the covariance matrix between instruments z_t and $\frac{\partial u(Y_{t+1}, \theta^0)}{\partial \theta'}$. Following the intuition above, the correct derivation of the asymptotic distribution of the test statistic under the null (2) rests upon a non-singularity assumption about a polynomial function of parameters built from the covariance matrix between the instruments and some polynomial functions of the genuine local explanatory variables. This polynomial function is the relevant local approximation of the moment conditions as produced by their higher order Taylor expansions. In this paper, we put a special emphasis on predictability of conditional variances ($n = 2$) and then, the asymptotic theory rests upon a second order identification condition (identification through the second derivative of moment conditions) when first order identification fails. The generalization of this approach to higher orders ($n = 3, 4, \dots$) would be conceptually straightforward, possibly at the price of tedious matricial derivation formulas. The key conclusion is that the common use as a critical value of a quantile of chi-square with a number of degrees of freedom equal to the dimension of $z_t \otimes u(Y_{t+1}, \theta)$, say H (assuming to simplify a real valued error term $u(Y_{t+1}, \theta)$), minus the dimension p of the vector θ of unknown parameters may lead to severe over-rejection. The intuition is the following. Since the full informational content of the moment restrictions (2) is displayed only when considering higher order derivatives (order 2 or more), higher order Taylor expansions may give a negligible weight to parameter uncertainty by considering higher order powers of $(\hat{\theta}_T - \theta^0)$ (where $\hat{\theta}_T$ is a GMM estimator). Therefore, we must see the asymptotic distribution of the J_T statistic under the null as a mixture of chi-square distributions

with degrees of freedom $(H - p_i), i = 1, 2, \dots, I$ with $p_I < p_{I-1} < \dots < p_1 \leq p$ instead of $(H - p)$. Hence the over-rejection implied by the use of chi-square $(H - p)$ to compute critical values.

The valid asymptotic theory is actually more involved since the lack of first order identification has also an impact on the rate of convergence of GMM estimators. In this respect, our asymptotic theory generalizes the work of Sargan (1983) who had been the first, in the context of instrumental variables regression, to note that in case of non-linearity with respect to the parameters, global identification might come with first order lack of identification. Like Sargan (1983), we derive our asymptotic results by assuming that there is a set of parameters with respect to which the first derivative of the moment conditions is always of full rank and a set of remaining parameters with respect to which the first derivative is null at the true value of the parameters. We refer to the first set of parameters as those identified at first order while the other ones will only be second order identified. Our framework generalizes Sargan (1983) in particular because we allow for any number of first order non-identified parameters. We find that not all the components of the GMM estimator have the same rate of convergence. The components that are only second order identified may converge only at rate $T^{1/4}$ while the square-root convergence is warranted for first order identified parameters, although the limit distribution is not normal. Our main contribution is to show that the asymptotic distribution of the J_T statistic is still based on chi-square distributions but through mixtures of them, with possibly different number of degrees of freedom larger than the standard $(H - p)$. The reason is that, even parameters which are only second order identified may be consistently estimated at rate $T^{1/2}$ in some parts of the sample space. In this case, parameter uncertainty becomes negligible when squared in second order Taylor expansions and the corresponding dimensions should no longer be subtracted from H in the computation of degrees of freedom.

A similar partition of the sample space according to different rates of convergence has been put forward by Rotnitzky, Cox, Bottai and Robins (2000) in the context of likelihood-based inference with singular information matrix. In this latter context too, Lee and Chesher (1986) had already noted that valid inference may resort to higher order Taylor expansions. Since likelihood-based inference can always be nested in a GMM framework by focusing on first order conditions, our setting encompasses the former ones. However, our main issue of interest on testing for overidentification is not addressed by likelihood-based inference. Moreover, the standard weak identification literature does not provide a solution. As stressed by Antoine and Renault (2009), when all parameters are identified but some rates of convergence may be as slow as $T^{1/4}$, neither the standard GMM asymptotics nor the weak instruments asymptotics (Kleibergen (2005)) allow to characterize the relevant asymptotic distributions.

We provide a numerical illustration in the context of factor models of multivariate conditional heteroskedasticity as previously studied by Diebold and Nerlove (1989), Engle and Susmel (1993), King, Sentana and Wadwhani (1994), Fiorentini, Sentana and Shephard (2004) and Doz and Renault (2006) among others. In this context, as first enhanced by Engle and Kozicki (1993), testing for a

factor structure may go through testing for common features, that is for the existence of portfolio returns whose conditional variance is time-invariant. Therefore, this issue fits well within our general framework of conditional variance predictability. This leads us to point out that the factor structure may invalidate the common chi-square critical value for the corresponding overidentification test. The correct mixture of chi-square distributions may have significantly larger tails, leading the standard test to severe over-rejection, especially in large samples.

The paper is organized as follows. In Section 2, we introduce the first order and the second order identification concepts and show how they have already been at stake in the selectivity bias literature. In Section 3, we discuss the impact of lack of first order identification on rates of convergence of extremum estimators by generalizing the approach of Sargan (1983) and Rotnitzky, Cox, Bottai and Robins (2000) as well. The implied non-standard asymptotic behaviour under the null of the J_T test statistic for overidentification is studied in Section 4. In Section 5, we apply this theory to the device of a test for common GARCH features. A Monte Carlo study in Section 6 compares the properties of our new testing approach with the test proposed by Engle and Kozicki (1993). The main proofs are gathered in an Appendix.

Throughout the paper $\|\cdot\|$ denotes not only the usual Euclidean norm but also a matrix norm $\|A\| = (\text{Trace}(AA'))^{1/2}$. 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 First order underidentification and second order identification

2.1 General framework

We consider a general minimum distance estimation problem of an unknown vector θ of p parameters given as solution of H estimating equations

$$\rho(\theta) = 0. \tag{4}$$

These estimating equations are assumed to identify the true unknown value θ^0 of θ by to the following assumptions

Assumption 1 (Global Identification). $\rho(\theta) = \{\rho_h(\theta)\}_{1 \leq h \leq H}$ is a continuous function defined on a compact parameter space $\Theta \subset \mathbb{R}^p$ such that for all θ in Θ , $\rho(\theta) = 0 \Leftrightarrow \theta = \theta^0$.

Assumption 1 is maintained for the sake of expositional simplicity even though it could be easily relaxed by only assuming that θ_0 is a well-separated minimum of norm of $\rho(\theta)$ (see Van der Vaart (1998, p. 46)).

For the purpose of minimum distance estimation, a data set of size T will give us some sample counterparts of the estimating equations. More precisely, with time series notations, we consider that with a sample size T , corresponding to observations at dates $t = 1, 2, \dots, T$ and for any possible value θ of the parameters, we have at our disposal a H -dimensional sample-based vector $\bar{\phi}_T(\theta) =$

$\{\bar{\phi}_{h,T}(\theta)\}_{1 \leq h \leq H}$. In most cases, minimum distance estimation is akin to GMM estimation because $\bar{\phi}_T(\theta)$ is obtained as a sample mean

$$\bar{\phi}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \phi_t(\theta). \quad (5)$$

In any case, we define a minimum distance estimator for a given sequence of weighting matrices.

Definition 2.1. *A minimum distance estimator $\hat{\theta}_T$ of θ is defined as solution of*

$$\min_{\theta \in \Theta} \bar{\phi}_T'(\theta) \Omega_T \bar{\phi}_T(\theta),$$

where Ω_T is a sequence of symmetric positive definite matrices which converges when T goes to infinity to Ω , a positive definite matrix.

The asymptotic properties of a minimum distance estimator are classically deduced from the asymptotic behaviour of the sample counterpart $\bar{\phi}_T(\theta)$ of the estimating equations.

Assumption 2 (Well-behaved moments). (a) $\bar{\phi}_T(\theta)$ converges in probability to $\rho(\theta)$, uniformly in $\theta \in \Theta$; (b) $\sqrt{T} \bar{\phi}_T(\theta^0)$ converges in distribution to a normal distribution with mean 0 and non-singular variance matrix $\Sigma(\theta^0)$.

It is well-known (see e.g. Amemiya (1989)) that Assumption 2.a implies that any minimum distance estimator $\hat{\theta}_T$ is weakly consistent for θ^0 . The asymptotic distribution of $\hat{\theta}_T$ is then usually deduced from a Taylor expansion of the first order conditions

$$\frac{\partial \bar{\phi}_T'}{\partial \theta}(\hat{\theta}_T) \Omega_T \sqrt{T} \bar{\phi}_T(\hat{\theta}_T) = 0. \quad (6)$$

Of course, this kind of approach is based on the maintained assumption below.

Assumption 3 (Differentiability of estimating equations). $\bar{\phi}_T(\theta)$ and $\rho(\theta)$ are continuously differentiable on the interior $\overset{\circ}{\Theta}$ of Θ , $\theta^0 \in \overset{\circ}{\Theta}$ and $\partial \bar{\phi}_T(\theta) / \partial \theta'$ converges to $\partial \rho(\theta) / \partial \theta'$, uniformly on $\theta \in \overset{\circ}{\Theta}$.

2.2 Local identification

The classical local condition for identification is that the matrix

$$\frac{\partial \rho}{\partial \theta'}(\theta^0)$$

is of rank p . This allows to interpret the first order condition (6) as asymptotically picking p independent linear combinations of the overidentifying estimating equations. This plays a crucial role in the standard asymptotic distribution theory of GMM because it allows to see $\sqrt{T}(\hat{\theta}_T - \theta^0)$ as asymptotically equivalent to a linear function of the Gaussian vector $\sqrt{T} \bar{\phi}_T(\theta^0)$.

The purpose of this paper is to relax the standard first order condition for local identification. For that purpose, we need to ensure the validity of second order Taylor expansions by maintaining the following assumption.

Assumption 4 (Higher order regularity of the estimating equations). (a) For all v in the null space of $\partial\rho(\theta^0)/\partial\theta'$, $\sqrt{T}(\partial\bar{\phi}_T(\theta^0)/\partial\theta')v = O_P(1)$; (b) $\bar{\phi}_T(\theta)$ and $\rho(\theta)$ are twice continuously differentiable on the interior $\mathring{\Theta}$ of Θ and for all $h = 1, 2, \dots, H$, $\partial^2\bar{\phi}_{h,T}(\theta)/\partial\theta\partial\theta'$ converges to $\partial^2\rho_h(\theta)/\partial\theta\partial\theta'$, uniformly on $\theta \in \mathring{\Theta}$.

Like Kleibergen (2005) we need in particular to complete the central limit theorem for the moment conditions by a similar assumption about the limit behaviour of the Jacobian matrix of these moment conditions.

Our new local identification assumption is then

Assumption 5 (Second order identification). For all u in the range of $\frac{\partial\rho'}{\partial\theta}(\theta^0)$ and all v in the null space of $\frac{\partial\rho}{\partial\theta'}(\theta^0)$, we have

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

The standard first order identification condition amounts to assume that the null space of $\frac{\partial\rho}{\partial\theta'}(\theta^0)$ is reduced to the null vector. In this case, the range of $\frac{\partial\rho'}{\partial\theta}(\theta^0)$ is \mathbb{R}^p and for all $u \in \mathbb{R}^p$,

$$\left(\frac{\partial\rho}{\partial\theta'}(\theta^0)u = 0 \right) \Rightarrow u = 0.$$

In other words, our local identification Assumption 5 contains the standard assumption as a particular case.

The following toy example gives the intuition of the relevance of this more general local identification assumption.

Example 2.1. (A toy example)

Assume we observe two stationary and ergodic time series, x_t and y_t , $t = 1, 2, \dots, T$ of real random variables with well-defined moments $E(x_t^a y_t^b)$, $a = 1, \dots, 6$, $b = 1, 2$. We want to characterize two parameters θ_1 and θ_2 solution of the three following overidentified moment conditions

$$\rho_1(\theta) = E((y_t - \theta_1 x_t)x_t) = 0 \tag{7}$$

$$\rho_2(\theta) = E((y_t - \theta_1 x_t)x_t^2) = 0 \tag{8}$$

$$\rho_3(\theta) = E((y_t - \theta_2 x_t)^2 x_t) = 0. \tag{9}$$

These conditions have been discussed by several authors within the framework of the market model for asset pricing (Sharpe and Lintner's CAPM). In this context, y_t and x_t stand for net asset returns, in excess of the risk free rate: x_t is the net market return while y_t is a net return on another risky asset whose market beta coincides, under the maintained assumption of the validity of CAPM, with the parameter θ_1 defined by (7). Then, the overidentifying moment restriction (8) will be fulfilled in particular if the affine regression of the individual asset return on the market return coincides with

the conditional expectation (linear market model). Mackinlay and Richardson (1991) have stressed the importance of the validity of the joint system of three conditions (7), (8), and (9) with $\theta_1 = \theta_2$ for the validity of the common test of CAPM (or equivalently mean-variance efficiency of the market return) based on normality of returns. Otherwise, the contemporaneous cross sectional conditional heteroskedasticity of idiosyncratic risk implied by the violation of (9) with $\theta_1 = \theta_2$ invalidates the standard test and requires an adjusted weighting matrix for a correct GMM-based test. Bakshi, Kapadia and Madan (2003) and Engle and Mistry (2007) have also used this assumption of contemporaneous homoskedasticity of idiosyncratic risk in order to put forward a decomposition of individual skewness between systematic and idiosyncratic sources. Of course, these discussions are more practically relevant when considering a bunch of individual risky assets with an individual market beta for each of them. However, for the purpose of this toy example, one can focus without loss of generality on one given risky asset besides the market return. A generalization to a multivariate y_t would be easy.

The key remark is actually that, for some plausible data generating processes (DGPs), the joint overidentified system of equations (7), (8), and (9) will identify a unique true unknown value $\theta^0 = (\theta_1^0, \theta_2^0)'$ such that $\theta_1^0 = \theta_2^0$. To see this, let us rewrite (9) as:

$$\theta_2^2 E(x_t^3) - 2\theta_2 E(x_t^2 y_t) + E(x_t y_t^2) = 0. \quad (10)$$

By (8), $E(x_t^3) = 0 \Rightarrow E(x_t^2 y_t) = 0$ and then θ_2 is not properly defined by (10) or equivalently (9). We must then maintain the assumption $E(x_t^3) \neq 0$ and consider (9) as a second degree equation to define the parameter θ_2 . Note that in case of a nonnegative variable x_t , the Cauchy-Schwarz inequality would imply that equation (10) has either no solution or only one solution when this inequality is an equality, that is

$$\{E(x_t^2 y_t)\}^2 = E(x_t^3) E(x_t y_t^2). \quad (11)$$

Of course, we do not want to maintain an assumption of nonnegativity for a net return like x_t but, in the spirit of the aforementioned financial literature, it makes sense to assume that the true unknown values of the two parameters coincide and then, from (8),

$$\theta_2 = \theta_1 = \frac{E(x_t^2 y_t)}{E(x_t^3)}. \quad (12)$$

In other words, the aforementioned literature implicitly focuses on DGPs conformable to (11). Then:

$$\frac{\partial \rho}{\partial \theta'}(\theta^0) = \begin{pmatrix} -E(x_t^2) & 0 \\ -E(x_t^3) & 0 \\ 0 & -2E((y_t - \theta_2^0 x_t)x_t^2) \end{pmatrix}$$

with, by (12), $E((y_t - \theta_2^0 x_t)x_t^2) = 0$. Thus the standard first order condition for identification is violated and the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$ is spanned by the vector $(0, 1)'$.

However,

$$\frac{\partial^2 \rho_1}{\partial \theta \partial \theta'}(\theta^0) = \frac{\partial^2 \rho_2}{\partial \theta \partial \theta'}(\theta^0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\frac{\partial^2 \rho_3}{\partial \theta \partial \theta'}(\theta^0) = \begin{pmatrix} 0 & 0 \\ 0 & 2E(x_t^3) \end{pmatrix}.$$

In other words, the second order identification assumption states that for all $u = (u_1, 0)'$, and $v = (0, v_2)'$,

$$\left(u_1 \begin{pmatrix} -E(x_t^2) \\ -E(x_t^3) \\ 0 \end{pmatrix} + 2v_2^2 \begin{pmatrix} 0 \\ 0 \\ E(x_t^3) \end{pmatrix} = 0 \right) \Rightarrow (u = v = 0).$$

In other words, Assumption 5 is fulfilled if and only if $E(x_t^3) \neq 0$.

Note that, since in this example the estimating equations are quadratic with respect to the unknown parameters, second order identification is tantamount to global identification.

The key feature of the local identification Assumption 5 is, by contrast with the standard first order identification, to introduce quadratic equations. Lemma 2.1 below sets the focus on this quadratic identification condition. This lemma will be crucial in the next section to see why only a rate of convergence $T^{1/4}$ may be warranted instead of $T^{1/2}$ as it is in the case of first order identification. Note that the focus of interest of this paper is only the impact of these non-standard rates of convergence on the GMM overidentification test. We do not address the likely lack of power issue resulting from the above singularity for an econometrician who would like to test the moment condition (9) for the purpose of a skewness decomposition a la Bakshi, Kapadia and Madan (2003). Of course, as it is often the case, the singularity problem in the toy example above could also be fixed by an alternative parameterization which would recognize that, when condition (11) is fulfilled, $\theta_1 = \theta_2$ is the only parameter of interest. However, for the purpose of financial interpretation, it may be convenient to disentangle the possible violation of conditions (8) and (9). Moreover, in many circumstances, the well-suited reparameterization is not as obvious as in this toy example.

Lemma 2.1. *Let P denote the orthogonal projection matrix on the range of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$. Let $M = I - P$ be the orthogonal projection matrix on the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$. Then Assumption 5 is equivalent to each of the following conditions.*

(a) *For all v in the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$,*

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

(b) *There exists a positive number γ such that, for any v in the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$,*

$$\left\| M \left(v' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) v \right)_{1 \leq h \leq H} \right\| \geq \gamma \|v\|^2.$$

Example 2.2. (Testing for selectivity bias)

Consider the two-equation selectivity bias model examined by Lee and Chesher (1986)

$$y_i = x_i' \beta + u_i, \quad i = 1, \dots, N, \tag{13}$$

$$y_i^* = z_i' \gamma - \varepsilon_i, \quad i = 1, \dots, N, \quad (14)$$

in which $x_i (k_1 \times 1)$ and $z_i (k_2 \times 1)$ are values of exogenous variables and the vectors (u_i, ε_i) , $i = 1, \dots, N$, are mutually independently normally distributed $\mathcal{N}(0, 0, \sigma^2, 1, \xi)$, where ξ stands for the correlation coefficient. The variate y_i^* is not observed but defines the binary indicator I_i , related to y_i^* , by

$$\begin{aligned} I_i &= 1 && \text{if and only if } y_i^* \geq 0, \\ &= 0 && \text{if and only if } y_i^* < 0. \end{aligned}$$

The variable y_i is observed only when $y_i^* \geq 0$, i-e when $I_i = 1$. To test whether there is selectivity bias, we examine the hypothesis $H_0 : \xi = 0$. The vector of unknown parameters is $\theta = (\beta', \gamma', \sigma^2, \xi)'$ and it is globally identified by the estimating equations of maximum likelihood

$$\rho(\theta) = E \left(\frac{\partial \ln L}{\partial \theta}(\theta) \right),$$

where

$$\begin{aligned} \ln L &= \sum_{i=1}^N \left\{ (1 - I_i) \ln(1 - \Phi(z_i' \gamma)) - \frac{1}{2} I_i \ln(2\pi\sigma^2) - (1/2\sigma^2) I_i (y_i - x_i' \beta)^2 \right. \\ &\quad \left. + I_i \ln \Phi \left((z_i' \gamma - (\xi/\sigma)(y_i - x_i' \beta)) / \sqrt{1 - \xi^2} \right) \right\}, \end{aligned}$$

where Φ is the standard normal distribution function and ϕ is the standard normal density function.

The lack of first order identification then corresponds to the singularity of the Fisher information matrix since

$$\frac{\partial \rho}{\partial \theta'}(\theta^0) = E \left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}(\theta^0) \right).$$

Lee and Chesher (1986) show that it is the case under the null H_0 when $\phi(z_i' \gamma^0) / \Phi(z_i' \gamma^0)$ is a linear function of x_i . This may occur not only if z_i is a constant while x_i contains the constant variable but more generally when z_i contains only dummy variables and x_i includes the same set of dummy variables and their interaction terms. The intuition of this lack of first order identification is related to the failure of the Heckman two-stage estimation of Equation (13) as noted by Melino (1982). Since, for two-stage estimation, this equation is re-written, when $I_i = 1$,

$$y_i = x_i' \beta - \sigma \xi \frac{\phi(z_i' \hat{\gamma})}{\Phi(z_i' \hat{\gamma})} + \eta_i, \quad (15)$$

it does not allow to identify $\sigma \xi$ in the aforementioned case. However Melino (1982) shows that the score test for $\xi = 0$ is asymptotically equivalent to the t -statistic associated with the coefficient $\sigma \xi$ in (15). The score test is then invalid, as expected in case of lack of first order identification. In such a case, Lee and Chesher (1986) show that a properly devised likelihood ratio test of H_0 amounts to test for skewness of the disturbance $I_i u_i$. If $\xi \neq 0$, the disturbance $I_i u_i$ is not symmetric. Lee and Chesher (1986) note that $E(I_i u_i^3)$ is actually proportional to the third order directional derivative of the log-likelihood computed in the direction of the null space of the Fisher information matrix. When $v \neq 0$ is in the null space of

$$\frac{\partial \rho}{\partial \theta'}(\theta^0) = E \left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}(\theta^0) \right),$$

we must have with the Lee and Chesher (1986) notations (see p. 136)

$$\frac{\partial^2 \rho}{\partial \mu^2}(\theta^0) = \frac{\partial^3 \ln L}{\partial \mu^3}(\theta^0) \neq 0$$

if μ denotes the coefficient of v in the directional derivative.

This condition, which is key for the validity of their likelihood ratio test, is exactly our second order identification condition in a case where there is only one dimension of lack of first order identification.

3 Rates of convergence

Following Stock and Wright (2000), it is convenient to derive rates of convergence of minimum distance estimators in presence of weak identification from a functional central limit theorem about the empirical process of moment conditions. More precisely, we reinforce Assumption 2 by the following.

Assumption 6. $\sqrt{T}(\bar{\phi}_T(\theta) - \rho(\theta))$ converges weakly with respect to the sup norm towards a Gaussian stochastic process with mean zero on Θ .

The crucial role of Assumption 6 is to provide a simple and general characterization of the rate of convergence of $\rho(\hat{\theta}_T)$.

Proposition 3.1. *If $\hat{\theta}_T$ is a minimum distance estimator conformable to Definition 2.1, we have, under Assumption 1 and 6*

$$\left\| \rho(\hat{\theta}_T) \right\| = O_P \left(T^{-1/2} \right).$$

The proof of Proposition 3.1 is given in Antoine and Renault (2009) as an extension of the Stock and Wright (2000) result. As announced, the second order identification assumption will allow to deduce from Proposition 3.1 the minimum rate of convergence of $\hat{\theta}_T$.

Proposition 3.2. *Under Assumptions 1 to 6*

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

and for any α given in the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$,

$$\left| \alpha' \hat{\theta}_T - \alpha' \theta^0 \right| = O_P \left(T^{-1/2} \right).$$

The intuition of Proposition 3.2 is quite clear. If we could replace the p unknown parameters θ by only $r = \text{Rank} \frac{\partial \rho'}{\partial \theta}(\theta^0)$ independent linear combinations $\alpha' \theta$ which are all in the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$, we would be back to the standard asymptotic theory of GMM under first order identification and we would get \sqrt{T} -asymptotically normal estimators. Unfortunately, this is not feasible and due to the lack of first order identification ($r < p$), we only ensure convergence at a slower rate $T^{1/4}$. Even though existing faster rates in some specific directions may not be feasible, because we do not know

in practice the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$, their characterization is important for the asymptotic theory of the GMM overidentification test, which is the main focus of interest of this paper. For this purpose, it is important to realize that even when some linear combinations $\alpha'\theta$ are slowly estimated (because α does not belong to the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$), $\alpha'\hat{\theta}_T - \alpha'\theta^0$ may have a faster rate of convergence, actually \sqrt{T} , in some specific parts of the sample space.

To see this, it is worth revisiting an argument put forward by Rotnitzky, Cox, Bottai and Robins (2000). Their focus of interest is the maximum likelihood with singular information matrix which, as explained in Example 2.2, can be seen as a particular case of our setting. They first give (see their Section 3) some heuristics to state their result in a one-dimensional parametric model with zero Fisher information at the parameter's value. More generally, let us consider the following example.

Example 3.1. (*One-dimensional parameter*)

Let $\theta \in \mathbb{R}$ be the parameter of interest. The first order lack of identification means

$$\frac{\partial \rho}{\partial \theta}(\theta^0) = 0.$$

The GMM objective function is $Q_T(\theta) = T\bar{\phi}'_T(\theta)\Omega_T\bar{\phi}_T(\theta)$. Similarly to the expansion of the log-likelihood considered by Rotnitzky et al (2000), let us consider the following Taylor expansion of Q_T .

$$\begin{aligned} Q_T(\theta) &= Q_T(\theta^0) + \frac{\partial Q_T}{\partial \theta}(\theta^0)(\theta - \theta^0) + \frac{1}{2} \frac{\partial^2 Q_T}{\partial \theta^2}(\theta^0)(\theta - \theta^0)^2 \\ &\quad + \frac{1}{6} \frac{\partial^3 Q_T}{\partial \theta^3}(\theta^0)(\theta - \theta^0)^3 + \frac{1}{24} \frac{\partial^4 Q_T}{\partial \theta^4}(\theta^0)(\theta - \theta^0)^4 + O((\theta - \theta^0)^5). \end{aligned}$$

Since by Assumptions 2 and 4, $\sqrt{T}\bar{\phi}_T(\theta^0)$ and $\sqrt{T}\frac{\partial \bar{\phi}_T}{\partial \theta}(\theta^0)$ are $O_P(1)$, the fact that $\|\hat{\theta}_T - \theta^0\| = O_P(T^{-1/4})$ implies that the only possibly non-negligible terms in the above expansion computed at $\hat{\theta}_T$ are

$$Q_T(\hat{\theta}) \sim Q_T(\theta^0) + \left(\frac{\partial^2 \bar{\phi}'_T}{\partial \theta^2}(\theta^0)\Omega_T\sqrt{T}\bar{\phi}_T(\theta^0) \right) \sqrt{T}(\hat{\theta}_T - \theta^0)^2 + \frac{1}{4} \left(\frac{\partial^2 \bar{\phi}'_T}{\partial \theta^2}(\theta^0)\Omega_T\frac{\partial^2 \bar{\phi}_T}{\partial \theta^2}(\theta^0) \right) T(\hat{\theta}_T - \theta^0)^4.$$

In other words, $\hat{\theta}_T$ is asymptotically equivalent to $\theta^0 + x_T/T^{1/4}$ where x_T minimizes

$$Q_T^*(x) = Q_T(\theta^0) + a_T x^2 + b_T x^4$$

with

$$a_T = \frac{\partial^2 \bar{\phi}'_T}{\partial \theta^2}(\theta^0)\Omega_T\sqrt{T}\bar{\phi}_T(\theta^0) \quad \text{and} \quad b_T = \frac{1}{4} \frac{\partial^2 \bar{\phi}'_T}{\partial \theta^2}(\theta^0)\Omega_T\frac{\partial^2 \bar{\phi}_T}{\partial \theta^2}(\theta^0).$$

By the second order identification assumption, b_T is positive for large T . Following Rotnitzky et al (2000, p. 250), the intuition is then the following.

If $a_T < 0$, the minimum is reached at $x_T = \pm \sqrt{-\frac{a_T}{b_T}}$. Thus, we need that in the above expansion, the two terms are of the same order of magnitude. Hence $T^{1/4}(\hat{\theta}_T - \theta^0)$ should be asymptotically non-degenerate.

By contrast, when $a_T > 0$, the minimum is reached at $x_T = 0$ and this allows a faster rate of convergence making the second term negligible. In this part of the sample space, $(\hat{\theta}_T - \theta^0)$ will be $O_P(T^{-1/2})$.

A precise statement would require a partition of the sample space which does not depend on the sample size (as the condition $a_T > 0$) but goes through weak convergence. Moreover, in order to extend this inequality condition to a multiple parameter setting, we will simplify the notation by assuming that the limit weighting matrix $\Omega = \text{plim} \Omega_T$ is the identity matrix. Note that we can maintain this assumption without loss of generality since it is always possible to rescale the moments $\bar{\phi}_T(\theta)$ as $\Omega_T^{1/2} \bar{\phi}_T(\theta)$. This rescaling is immaterial for the validity of Assumptions 1 to 6.

Proposition 3.3. *Assume $\Omega = Id_H$ and N is a $(p, p - r)$ -matrix whose columns are a basis of the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$. We consider the symmetric random matrix Z_T of size $(p - r)$ whose coefficients (i, j) , $i, j = 1, \dots, p - r$ are*

$$a'_{ij} M \sqrt{T} \bar{\phi}_T(\theta^0)$$

with $a_{ij} = (a_{ij}^h)_{1 \leq h \leq H}$ and $(a_{ij}^h)_{1 \leq i, j \leq p-r}$ is the matrix

$$N' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) N, \quad h = 1, \dots, H.$$

Let Z denote the distribution limit of Z_T , $(Z \geq 0)$ the event “ Z is positive semidefinite” and $\overline{(Z \geq 0)}$ its complement.

Under Assumptions 1 to 6, for any subsequence of $T^{1/4}(\hat{\theta}_T - \theta^0)$ which converges towards V , we have

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

Note that $\text{Vec}(Z)$ is by definition a zero-mean Gaussian distribution linear function of the limit distribution $\mathcal{N}(0, \Sigma(\theta^0))$ of $\sqrt{T} \bar{\phi}_T(\theta^0)$. It is in particular important to realize that Z is positive definite if and only if $\text{Vec}(Z)$ fulfills $(p - r)$ multilinear inequalities corresponding to the positivity of the $(p - r)$ leading principal minors of the matrix Z (see e.g Horn and Johnson (1985, p. 404)). Therefore, the probability q of the event $(Z \geq 0)$ is strictly positive. In particular, $q = 1/2$ if $p - r = 1$. In the case $\dim \theta = 1$ of Example 3.1 above, we have $r = 0$, $p = 1$ and

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

Then, Z corresponds to the (non degenerate) zero-mean univariate normal asymptotic distribution of a_T in the example. Proposition 3.3 confirms the message of Example 3.1: the order of convergence of $\hat{\theta}_T$ is $T^{1/4}$ or more depending on the sign of Z . More generally, the message of Proposition 3.3 is twofold. First, in the part of the sample space where Z is positive semi-definite, all the components of $\hat{\theta}_T$ converge at a rate faster than $T^{1/4}$. By contrast, in general, only the directions in the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$ get the fast rate of convergence $T^{1/2}$ all over the sample space by Proposition 3.2. Proposition

3.3 tells us that $T^{1/4}(\hat{\theta}_T - \theta^0)$ must sometimes have a non-zero limit in the part of the sample space where Z is not positive semi-definite. This classification of rates of convergence for GMM estimators in the case of lack of first order identification has clearly been pointed out by Sargan (1983) in a particular case of instrumental variables estimation. It is also tightly related to the result of Rotnitzky et al. (2000) in the particular case of maximum likelihood estimation.

4 Overidentification test

In this section, we study the asymptotic behaviour of the GMM overidentification test statistic

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

when the Jacobian matrix of the moment function at the true parameter value is rank deficient. We will however maintain, as in the previous sections, the second order identification condition in Assumption 5. J_T is the minimum value of the GMM objective function expressed by the optimal weighting matrix defined as a consistent estimate of the inverse of the moment conditions' long run variance, $\Sigma(\theta^0) \equiv \lim_{T \rightarrow \infty} Var\left(\sqrt{T}\bar{\phi}_T(\theta^0)\right)$. This specific choice of weighting matrix ensures the required normalization of the moment functions that makes J_T behave in large samples as a chi-square random variable with $H - p$ degrees of freedom (Hansen (1982)) when the moment conditions are true and the null space of the Jacobian matrix of the moment conditions is reduced to the null vector (standard first order identification condition).

More generally, Assumption 5 covers also situations where the null space is not of null dimension. The main result of this section is a characterization of the asymptotic distribution of J_T when the null space of the Jacobian has a dimension larger than or equal to one. While the characterization is only partial when the dimension is larger than one, we give a full characterization when the dimension is exactly one. Similarly to the previous section, we assume without loss of generality that $\Sigma(\theta^0) = Id_H$. This condition is immaterial regarding the conclusion of the result below. In particular, upon a re-normalization of the moment conditions, it is always satisfied. On the other hand, it considerably simplifies the presentation as we can set Ω_T to Id_H in the definition of J_T .

Theorem 4.1. *Assume $\Sigma(\theta^0) = Id_H$. Under Assumptions 1-6, the overidentification J-test statistic, $J_T = T\bar{\phi}'_T(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T)$ associated to the estimating equations*

$$\rho(\theta) = 0$$

is, with probability $q \geq \text{Prob}(Z \geq 0) > 0$, asymptotically distributed as χ^2_{H-r} , where Z is defined as in Proposition 3.3,

$$H = \dim \rho(\theta), \text{ and } r = \text{Rank}(\partial\rho(\theta^0)/\partial\theta').$$

In particular, if $r = p - 1$, $q = 1/2$ and J_T is asymptotically distributed as the mixture

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

Theorem 4.1 states a rather non-standard behaviour for J_T . If the moment functions do not have a Jacobian matrix of full rank, there is a probability $q > 0$ that J_T behaves asymptotically as a chi-square with a tail thicker than in the usual case. By contrast, when the Jacobian is of full rank, $Z = 0$ almost surely and therefore, J_T has a probability one to behave as a chi-square of $H - p$ degrees of freedom. This is the result in the standard case.

The classification of the rates of convergence of the GMM estimator as highlighted in the previous section is the main cause of this non-standard asymptotic distribution of J_T as stated by Theorem 4.1. The r independent directions in which the GMM estimator has the standard root-T rate of convergence lead to subtract r degrees of freedom to the dimension H of moment conditions as it is always the case with root-T consistent estimators. The key is the way to accommodate the $(p - r)$ remaining dimensions of estimation. Since these directions only show up in second order terms of Taylor expansions, they play no role when they are estimated at a rate faster than $T^{1/4}$, which, by virtue of Proposition 3.3, is the case in the part of the sample space where $(Z \geq 0)$. By contrast, if the complement of the event $(Z \geq 0)$ occurs instead, some of the first-order underidentified directions are estimated exactly at the rate $T^{1/4}$ and thus may be locally non negligible in second order expansions of J_T . This makes the full characterization of J_T more difficult when $p - r > 1$ since higher order expansions involve the product of such directions. When $p - r = 1$, the behaviour of the GMM estimator in the direction of the null space can be characterized clearly enough to help deduce the full asymptotic distribution of J_T which is a half-half mixture of chi-squares. When $(Z \geq 0)$, we have a chi-square with a number of degrees of freedom $(H - r)$ while we recover the standard $(H - p)$ in the complement $(Z < 0)$.

The bottom line is the occurrence of some mixture components with asymptotic chi-square distributions with more degrees of freedom than the standard $(H - p)$. The key consequence is that by using the standard critical value, one would be led to over-rejection in large sample.

5 Application to the test for common GARCH features

The conditionally heteroskedastic factor model introduced by Diebold and Nerlove (1989) (see also Fiorentini, Sentana and Shephard (2004) and Doz and Renault (2006)) allows a parsimonious structural representation of multivariate volatility. This model decomposes a finite set of asset returns in a systematic part and an idiosyncratic part. The idiosyncratic parts are supposed to have a constant conditional variance while the well documented conditional heteroskedasticity in asset returns is supported by the common systematic factors. Considering a bivariate vector, Y_{t+1} , of returns $y_{1,t+1}$, $y_{2,t+1}$ of two assets at time $t + 1$, a conditionally heteroskedastic factor representation of Y_{t+1} is given by

$$Y_{t+1} = \lambda f_{t+1} + U_{t+1}, \tag{16}$$

where f_{t+1} is the latent common conditionally heteroskedastic factor. λ is a bivariate non-random vector of factor loadings and U_{t+1} is the bivariate random vector of idiosyncratic shocks. The decom-

position in (16) is completed by the following restrictions with J_t denoting the increasing filtration containing the information available at the date t :

$$\begin{aligned} E(f_{t+1}|J_t) &= 0, & E(U_{t+1}|J_t) &= 0, & Var(f_{t+1}|J_t) &= \sigma_t^2, \\ E(\sigma_t^2) &= 1, & Var(U_{t+1}|J_t) &= \Omega, & E(f_{t+1}U_{t+1}|J_t) &= 0. \end{aligned} \quad (17)$$

In addition, one may restrict Ω to be positive definite and $Var(\sigma_t^2) > 0$. These further restrictions imply that any other single heteroskedastic factor decomposition of Y_{t+1} has factor loadings proportional to λ . (See Doz and Renault (2006).) It is worth noting that the representation (16) and (17) considers for simplicity that $E(Y_{t+1}|J_t) = 0$.

Assuming that both $y_{1,t+1}$ and $y_{2,t+1}$ display an evidence of conditional heteroskedasticity¹, it is reasonable to wonder if these two return processes share a common pattern of conditional heteroskedasticity. In other words, if they have a common GARCH feature. The factor representation in (16) and (17) is a valid set up to answer such a question as it offers some testable implications for common GARCH features.

In particular, any portfolio of $y_{1,t+1}$ and $y_{2,t+1}$ with weights orthogonal to λ , the so-called zero-beta portfolio with respect to the risk factor f_{t+1} has a constant conditional variance. Since both return processes have evidence of conditional heteroskedasticity, $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$ such that, up to a certain normalization, any zero-beta portfolio return can be written

$$u_{\theta,t+1} = y_{2,t+1} - \theta y_{1,t+1},$$

where $\theta = \lambda_2/\lambda_1$.

If $y_{1,t+1}$ and $y_{2,t+1}$ share their GARCH features, the representation in (16) and (17) holds and there exists a real valued parameter θ such that

$$Var(u_{\theta,t+1}|J_t) = c, \quad (18)$$

where c is a constant. The uniqueness of θ is guaranteed by the positive definiteness of Ω and the variability of σ_t^2 . In consequence, (18) represents a conditional moment restriction which can be tested using a set of suitable instruments, z_t extracted from the increasing filtration J_t . The resulting unconditional moment condition is given by

$$E(z_t(u_{\theta,t+1}^2 - c)) = 0. \quad (19)$$

Our goal is to test the commonality in the GARCH features by applying the GMM overidentification test to the moment condition model in (19).

Since two free parameters are involved in this moment condition model, we need z_t to contain more than two instruments to be able to test for overidentification. One may choose as instrument besides the constant 1 the lagged square returns and their product which all happen to belong to J_t .

¹Note that this can be investigated by the Lagrange multiplier test for autoregressive conditional heteroskedasticity (ARCH) effects proposed by Engle (1982).

Considering again the moment condition model (19), it appears that one can reduce the parameter dimension by eliminating the parameter c . Thanks to the constant instrument, $c = E(u_{\theta,t+1}^2)$. The moment condition for a given instrument z_t can then be written

$$E(z_t(u_{\theta,t+1}^2 - E(u_{\theta,t+1}^2))) = 0$$

or equivalently

$$E\{(z_t - E(z_t))(u_{\theta,t+1}^2 - E(u_{\theta,t+1}^2))\} = 0. \quad (20)$$

In the moment condition model (20), we center the instruments on purpose since the GMM overidentification test statistic deduced from (20) is similar if not identical to the test statistic of common ARCH features proposed by Engle and Kozicki (1993).

In the sample version, as to compute the GMM objective function, $E(z_t)$ and $E(u_{\theta,t+1}^2)$ are to be replaced by the usual uniform sample averages. Note that since the constant instrument is already exploited (in the aim of getting rid of the conditional variance c), it becomes redundant to use it again to fit (20) and one is better off using the other relevant instruments.

As suggested by our results from the previous section, the identification properties of the overidentifying moment conditions are fundamental in determining the asymptotic behaviour of the GMM overidentification test. The next result studies the global identification, the first and second order identification properties of the model in (20).

Theorem 5.1. *Let $Y_{t+1} = (y_{1t+1}, y_{2t+1})'$ satisfy (16) and (17). Let (z_t) be a H -dimensional process adapted to the filtration J_t and $\phi_t(\theta) = (z_t - E(z_t))(u_{\theta,t+1}^2 - E(u_{\theta,t+1}^2))$.*

If the vector $(z_t, \sigma_t^2)'$ is a stationary process such that $E(\|z_t\|) < \infty$ and $0 < \|Cov(z_t, \sigma_t^2)\| < \infty$, then

- (i) (Identification) there exists one and only one parameter value $\theta^0 \in \mathbb{R}$ satisfying the moment condition in (20),*
- (ii) (First order underidentification) $E(\partial\phi_t(\theta^0)/\partial\theta) = 0$,*
- (iii) (Second order identification) $E(\partial^2\phi_t(\theta^0)/\partial\theta^2) \neq 0$.*

The main message of Theorem 5.1 is from its point (ii). Even though the moment condition model in (20) globally identifies the parameter of interest, it suffers of lack of first order identification. The direct consequence of this is the non-applicability of the asymptotic distribution derived by Hansen (1982) for both the GMM estimator and the GMM overidentification test statistic, J_T . On the other hand, (iii) shows that the second order identification condition is satisfied. From our previous results, we can deduce that the GMM estimator of θ^0 has an unconditional rate of convergence of $T^{1/4}$ while J_T is asymptotically distributed as a half-half mixture of chi-squares of H and $H - 1$ degrees of freedom, respectively.

As we already mentioned, because the actual asymptotic distribution of J_T has a thicker tail than the one expected in the standard conditions (chi-square of $H - 1$ degrees of freedom), ignoring the first order lack of identification may lead to possibly severe over-rejections. We evaluate next the extent of this over-rejection.

Let α be the nominal level of the GMM overidentification test of the moment condition model (20) performed through the standard conditions and $c_{\alpha, H-1}$ the standard critical value of the test based on the hypothetical χ_{H-1}^2 as asymptotic distribution under the null. Let α_0 be the actual asymptotic size of the test. $c_{\alpha, H-1}$ and α_0 are defined by

$$\text{Prob}(\chi_{H-1}^2 > c_{\alpha, H-1}) = \alpha$$

and

$$\alpha_0 = \text{Prob}\left(\frac{1}{2}\chi_{H-1}^2 + \frac{1}{2}\chi_H^2 > c_{\alpha, H-1}\right).$$

The asymptotic relative rate of over-rejection is given by

$$100 \times (\alpha^{-1}\alpha_0 - 1)\%.$$

Table I below displays the theoretical relative over-rejection rate of the GMM overidentification test performed by ignoring the first order underidentification. Different number of instruments are considered as well as the nominal levels $\alpha = 0.05$ and $\alpha = 0.01$.

Similar facts can be documented for both levels. The amount of relative over-rejection rate is large for any number of included instruments even though this tends to decrease with larger number of instruments.

At its largest corresponding to 2 included instruments, the amount of over-rejection rate is almost 100% for a 0.05-nominal level test and about 130% for a 0.01-nominal level test. This amount hangs on 26.2% for a 0.05-nominal level test and 34.0% for a 0.01-nominal level test for the case where as many as 10 instruments are included.

Table I: Over-rejection rate of the common GARCH feature test at the nominal levels

$\alpha = 0.05$ and 0.01

Number of instruments H	Critical value		Exact asymptotic level		Relative over-rejection rate	
	$c_{\alpha, H-1}$		α_0		$\alpha^{-1}\alpha_0 - 1$	
Level: α	<i>0.05</i>	<i>0.01</i>	<i>0.05</i>	<i>0.01</i>	<i>0.05</i>	<i>0.01</i>
2	3.842	6.635	0.098	0.023	96.6%	131.0%
4	7.815	11.345	0.074	0.017	48.6%	95.0%
5	9.488	13.277	0.071	0.016	41.2%	55.0%
6	11.071	15.086	0.068	0.015	36.2%	48.0%
10	16.919	21.666	0.063	0.013	26.2%	34.0%

These asymptotic over-rejection rates are confirmed even in finite samples as we can see through the Monte Carlo experiments in the next section.

6 Monte Carlo evidence

The Monte Carlo experiments in this section aim to support the theoretical results that have been presented in the previous sections. We mainly give an illustration of the non-standard asymptotic behaviour of the test for common GARCH features as proposed in Section 5 and also confirm the slower rate of convergence of the GMM estimator, resulting from the mixture of rates $T^{1/2}$ and $T^{1/4}$. We simulate the bivariate return process

$$Y_{t+1} = \lambda f_{t+1} + U_{t+1},$$

where $\lambda = (1, 1)'$, f_{t+1} is a Gaussian *GARCH*(1, 1) process, i.e

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

$\varepsilon_{t+1} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ and is independent of the vector of idiosyncratic shocks $U_{t+1} \stackrel{iid}{\sim} \mathcal{N}(0, 0.5Id_2)$ (Id_2 is the identity matrix of size 2).

We consider two designs. The set of parameters values for Design *D1* is

$$\omega = 0.2, \quad \alpha = 0.4, \quad \text{and} \quad \beta = 0.4$$

and the set of parameters values for Design *D2* is

$$\omega = 0.2, \quad \alpha = 0.2, \quad \text{and} \quad \beta = 0.6.$$

In these two designs, the GARCH effect in the factor's dynamics has the same persistence. They differ only by their GARCH (and ARCH) persistences, respectively α (and β). The parameters values that we consider match those found in empirical applications for monthly returns and are also used by Fiorentini, Sentana and Shephard (2004) in their Monte Carlo experiments. Each design is replicated 5,000 times for each sample size T .

The sample sizes that we consider are 1,000, 2,000, 5,000, 10,000, 15,000, 20,000, 30,000, and 40,000. We include such large sample sizes in our experiments because of the slower rate of convergence of the GMM estimator. Since the unconditional rate of convergence of this estimator is $T^{1/4}$ and not \sqrt{T} as usual, we expect that the asymptotic behaviours of interest become perceptible for larger samples than those commonly used for such studies.

On each simulated sample, we evaluate the GMM estimator of the moment condition model in (20). To stay as close as possible to the similar testing procedure described by Engle and Kozicki (1993), we compute the efficient GMM estimator in one step by making the optimal weighting matrix parameter-dependent. This estimation procedure is known as the continuously updated GMM (see Hansen, Heaton and Yaron (1996)). We use a set of two instruments $z_t = (y_{1,t}^2, y_{2,t}^2)$. This is the minimum number of instruments allowing to overidentify the parameter to be estimated. We expect the test statistic to behave, as the sample size grows, as a half-half mixture of $\chi^2(1)$ and a $\chi^2(2)$. We

also expect the simulated variance of the estimator to blow to infinity as the sample size grows when scaled by the sample size, T , but stays reasonably stable when scaled by \sqrt{T} .

Table II displays the simulated 95th and 99th percentiles for J_T as well as theoretical 95th and 99th percentiles of $\chi^2(1)$, $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$ and $\chi^2(2)$ distributions. Clearly and for both designs, as T grows, the simulated percentiles of J_T get closer to those of $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$ as we expect from our theory.

Table II: Simulated critical values $c_{1-\alpha} : P(X > c_{1-\alpha}) = \alpha$. The theoretical $(1 - \alpha)$ -percentiles of the $\chi^2(1)$, $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$ and $\chi^2(2)$ distributions are displayed on the last three rows. The simulated percentiles of J_T are based on 5,000 replications.

	α	0.05		0.01			
		<i>Design</i>	<i>D1</i>	<i>D2</i>	<i>D1</i>	<i>D2</i>	
T	1,000		4.13	3.71	7.98	6.52	
	2,000		4.58	4.13	8.89	7.29	
	5,000		4.68	4.58	8.32	7.60	
	10,000		5.01	4.78	8.56	7.85	
	15,000		4.85	4.75	8.26	7.98	
	20,000		4.92	4.70	8.74	7.86	
	30,000		4.85	4.90	8.87	8.23	
	40,000		4.92	5.05	8.42	8.41	
		$\chi^2(1)$		3.84		6.63	
		$\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$		5.13		8.27	
	$\chi^2(2)$		5.99		9.21		

Table III : Simulated asymptotic order of magnitude of the GMM estimator.

T	1,000	2,000	5,000	10,000	15,000	20,000	30,000	40,000
<i>Design D1</i>								
$Var(\hat{\theta}_T)$	0.040	0.020	0.011	0.007	0.005	0.005	0.004	0.003
$T \times Var(\hat{\theta}_T)$	39.973	39.557	52.767	66.855	80.603	94.816	116.093	127.795
$\sqrt{T} \times Var(\hat{\theta}_T)$	1.264	0.885	0.746	0.669	0.658	0.670	0.670	0.639
<i>Design D2</i>								
$Var(\hat{\theta}_T)$	0.276	0.058	0.027	0.017	0.013	0.011	0.009	0.007
$T \times Var(\hat{\theta}_T)$	275.604	115.638	136.759	169.866	193.411	215.029	253.897	286.158
$\sqrt{T} \times Var(\hat{\theta}_T)$	8.715	2.586	1.934	1.699	1.579	1.521	1.466	1.431

Figure 1 presents a graphic view of Table II. It illustrates the tail behaviour of J_T and one can notice the departure from the percentiles of the $\chi^2(1)$ even in moderately large samples. One can also notice the convergence of J_T 's percentiles towards those of the $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$ distribution as the sample size grows. Figures 2 and 3 plot, for various sample sizes all the 99 simulated percentiles of

J_T . Each graph presents the percentiles for J_T as well as the theoretical percentiles of the $\chi^2(1)$, $\chi^2(2)$ and $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$ distributions. These graphs confirm that not only the 95th and 99th percentiles of J_T move toward those of the mixture as the sample size grows but actually all of the percentiles as predicted by the theory. This is true for both designs.

It is worth mentioning that the test statistic J_T is identical to the test statistic for common (G)ARCH features as proposed by Engle and Kozicki (1993) for our choice of instruments. In the light of the asymptotic distribution of J_T , their test would be oversized in our context since based on the standard critical value provided by the $\chi^2(1)$.

The results for the asymptotic order of magnitude of the GMM estimator are displayed by Table III. We can see that the GMM estimator, at a sample size of 1,000, still has a *wild* behaviour in terms of variance, particularly for Design *D2*. This can only be related to the lack of identification since one would not expect such a large gap between the simulated variance for $T = 1,000$ and $T = 2,000$ in a standard condition. This table also shows that when the simulated variance of $\hat{\theta}_T$ is scaled by \sqrt{T} , for large values of T , it stays quite stable in both designs while it skyrockets when scaled by the usual rate T . This is an evidence supporting the fact that $\hat{\theta}$ behaves asymptotically rather as an $O_P(T^{-1/4})$ random sequence than an $O_P(T^{-1/2})$ sequence as predicted by our theory.

Figure 1: Simulated $(1 - \alpha)$ -percentile of J_T and theoretical $(1 - \alpha)$ -percentile of the distributions $\chi^2(1)$, $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$ and $\chi^2(2)$. $\alpha = 0.05$ and 0.01 .

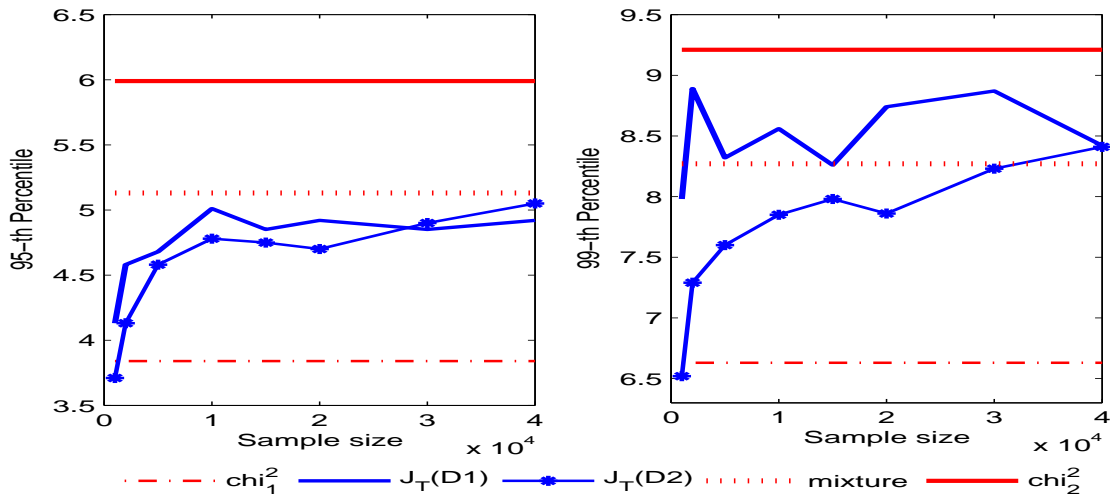


Figure 2: Simulated percentiles of J_T , Design $D1$ and the theoretical percentiles of the distributions $\chi^2(1)$, $\chi^2(2)$ and the mixture $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$.

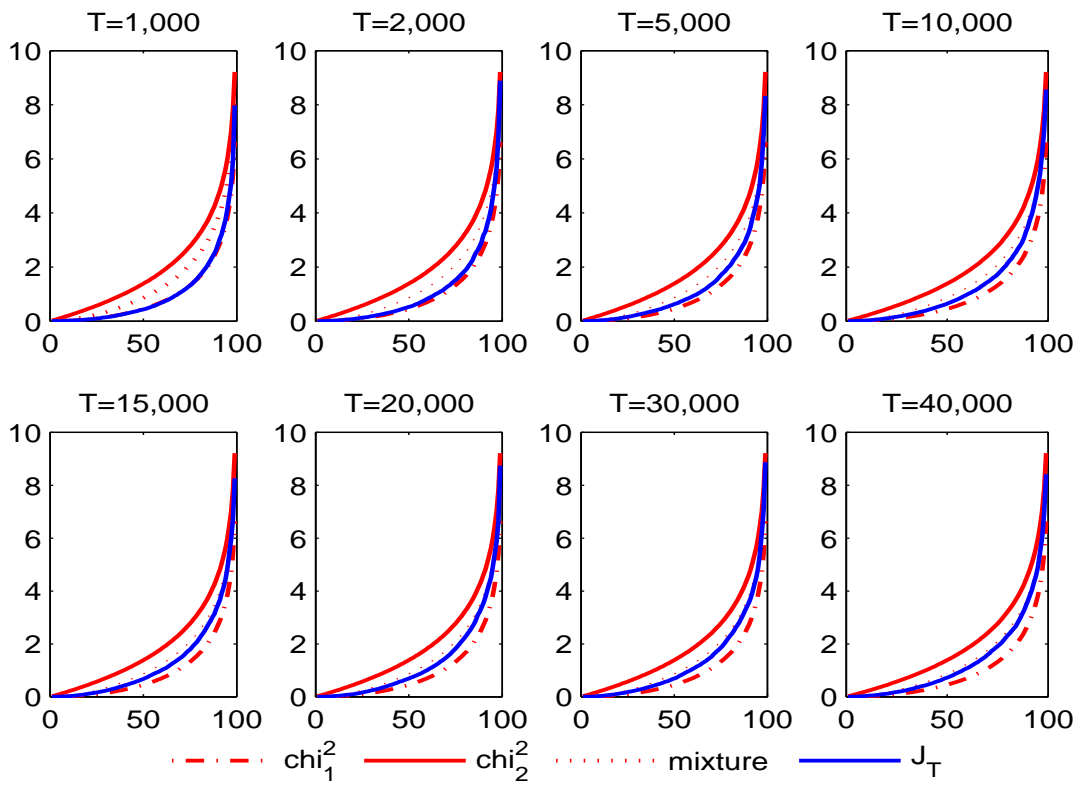
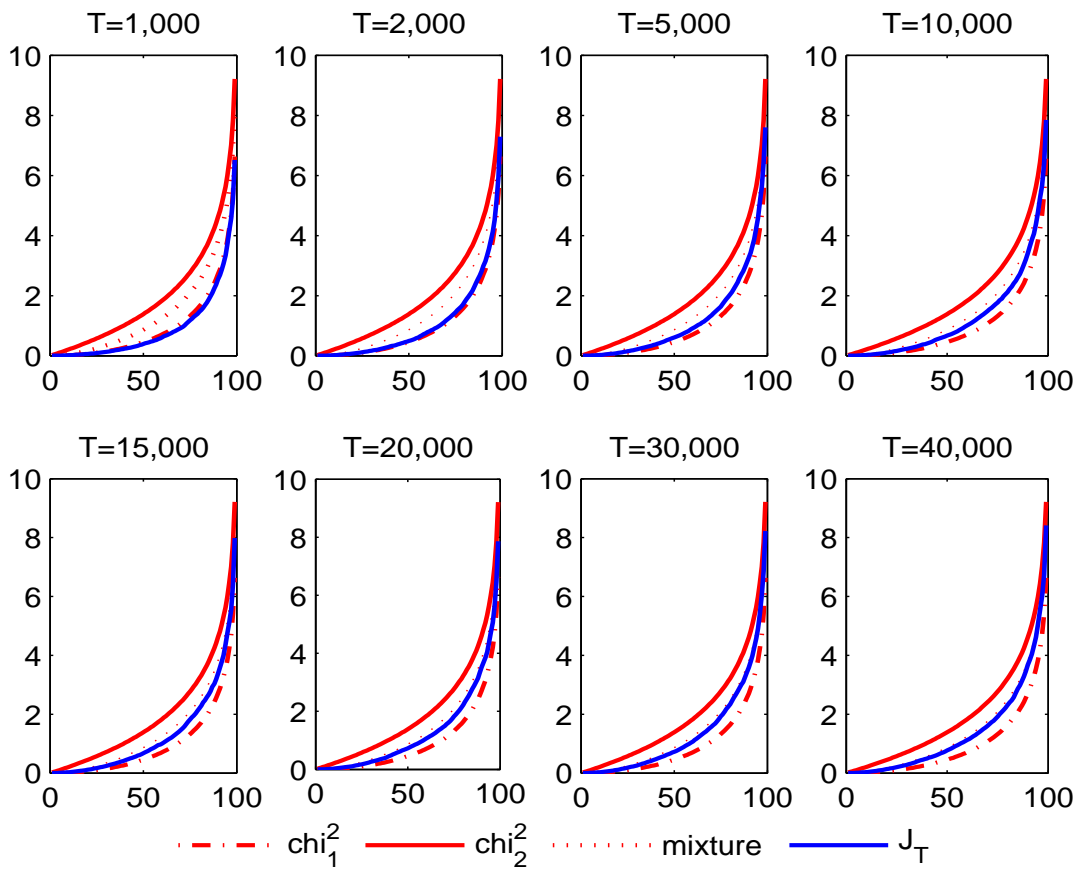


Figure 3: Simulated percentiles of J_T , Design $D2$ and the theoretical percentiles of the distributions $\chi^2(1)$, $\chi^2(2)$ and the mixture $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$.



7 Conclusion

This paper explores the asymptotic behaviour of the minimum distance estimators and the Hansen (1982) test for overidentifying moment restrictions statistic, J_T under non standard conditions dubbed first-order under-identification. While maintaining a second order identification condition, we derive the rate of convergence of the minimum distance estimators and the asymptotic distribution of J_T when the Jacobian matrix of the moment function evaluated at the true parameter value is not of full rank. We find that the linear combinations of the minimum distance estimator belonging to the range of the Jacobian matrix have the usual $O_P(T^{-1/2})$ asymptotic order of magnitude. Meanwhile, the linear combinations belonging to the null space of the Jacobian matrix converge more slowly as their unconditional asymptotic order of magnitude is $O_P(T^{-1/4})$. These results generalize for the minimum distance estimators (in particular for the GMM estimator) the findings by Sargan (1983) for the instrumental variables estimator. A further investigation also reveals that these latter linear combinations can actually go faster in some regions of the sample space. This non-standard behaviour affects the asymptotic distribution of J_T . Instead of a chi-square distribution, it is asymptotically distributed as a half-half mixture of two chi-squares distributions when the rank deficiency is equal to 1.

In the context of the conditionally heteroskedastic factor models, we propose an overidentification test for common GARCH features which has this mixture of chi-squares as asymptotic distribution. Our test statistic is identical to the one proposed by Engle and Kozicki (1993) for (G)ARCH features but both tests differ since the latter, by predicting the standard chi-square asymptotic distribution, is obviously oversized in our framework.

Appendix: Proofs

Proof of Lemma 2.1. Let us introduce

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

Let us assume that for all u in the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$ and all v in the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$,

$$\left(\frac{\partial \rho}{\partial \theta'}(\theta^0) u + \Delta(v) = 0 \right) \Rightarrow (u = v = 0)$$

and establish (a) by contradiction. For this, let $v_0 \neq 0$ in the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$ such that

$$M\Delta(v_0) = 0.$$

Clearly, the vector

$$P\Delta(v_0) = \Delta(v_0)$$

and as a result, there exists u_0 such that

$$\frac{\partial \rho}{\partial \theta'}(\theta^0) u_0 = \Delta(v_0).$$

Note that u_0 can be decomposed as $u_0 = u_0^* + u_0^{**}$, with u_0^* in the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$ and u_0^{**} in the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$. Thus

$$-\frac{\partial \rho}{\partial \theta'}(\theta^0) u_0^* + \Delta(v_0) = 0.$$

This contradicts Assumption 5 since $v_0 \neq 0$.

Next, we show that (a) implies (b). Let v be any vector in the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$. Since $v \mapsto \|M\Delta(v)\|$ is an homogeneous function of degree 2 with respect to v , we have

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

By considering

$$\gamma = \inf_{\|v\|=1, \frac{\partial \rho}{\partial \theta'}(\theta^0)v=0} \|M\Delta(v)\|,$$

we just have to show that $\gamma > 0$. By the compactness of the set $\left\{ v \in \mathbb{R}^p : \|v\| = 1, \frac{\partial \rho}{\partial \theta'}(\theta^0)v = 0 \right\}$ and the continuity of $v \mapsto \|M\Delta(v)\|$, $\gamma = \|M\Delta(v^*)\|$ for some v^* such that $\|v^*\| = 1$ and $\frac{\partial \rho}{\partial \theta'}(\theta^0)v^* = 0$. Therefore, thanks to (a), $\gamma > 0$ and for any v , $\|M\Delta(v)\| \geq \gamma \|v\|^2$.

To complete the proof, we just have to establish that Assumption 5 is implied by (b).

Let us consider u in the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$ and v in the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$, such that

$$\frac{\partial \rho}{\partial \theta'}(\theta^0) u + \Delta(v) = 0.$$

By pre-multiplying each side of this equation by M , we have $M\Delta(v) = 0$. Thus $\|M\Delta(v)\| = 0$. From (b), we deduce that $v = 0$. As a consequence, $\Delta(v) = 0$ and $\frac{\partial \rho}{\partial \theta'}(\theta^0) u = 0$. Since u belongs to the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$, this last equality implies that we also have $u = 0$ \square

Proof of Proposition 3.2. Let $r = \text{Rank} \frac{\partial \rho}{\partial \theta'}(\theta^0)$. Let R^1 and R^2 be two matrices of dimension (p, r) and $(p, p - r)$ respectively such that

- (i) The columns of R^1 are a basis of the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$.
- (ii) The columns of R^2 are a basis of the null space of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$.

Then $R = (R^1 | R^2)$ is a (p, p) -non-singular matrix and we will introduce the new parameterization

$$\eta = R^{-1}\theta \quad \text{and} \quad \eta^0 = R^{-1}\theta^0.$$

We denote $\hat{\eta}_T = R^{-1}\hat{\theta}_T$ and

$$\rho^*(\eta) = \rho(R\eta).$$

With obvious notations, we decompose

$$\theta = R\eta = R^1\eta_1 + R^2\eta_2.$$

Let us consider the second-order Taylor expansion

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

for some $\tilde{\eta}_T$ between $\hat{\eta}_T$ and η^0 . Note that, by a common abuse of notation, we omit to stress that $\tilde{\eta}_T$ actually depends on the component ρ_h^* of ρ^* .

By definition of R ,

$$\frac{\partial \rho^*}{\partial \eta'}(\eta^0) = \frac{\partial \rho}{\partial \theta'}(\theta^0)R = \left(\frac{\partial \rho}{\partial \theta'}(\theta^0)R^1, 0 \right)$$

and

$$\frac{\partial^2 \rho_h^*}{\partial \eta \partial \eta'}(\eta^0) = R' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0)R$$

Thus, we can rewrite the above expansion as

$$\rho(\hat{\theta}_T) = \frac{\partial \rho}{\partial \theta'}(\theta^0)R^1(\hat{\eta}_{1T} - \eta_1^0) + \frac{1}{2} \left((\hat{\eta}_T - \eta^0)' R' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\tilde{\theta}_T)R(\hat{\eta}_T - \eta^0) \right)_{1 \leq h \leq H}, \quad (\text{A.1})$$

with $\tilde{\theta}_T = R\tilde{\eta}_T$ between θ^0 and $\hat{\theta}_T$. Note that, since $\frac{\partial \rho}{\partial \theta'}(\theta^0)R^1$ is full column rank,

$$\left\| \frac{\partial \rho}{\partial \theta'}(\theta^0)R^1(\hat{\eta}_{1T} - \eta_1^0) \right\| \geq c \|\hat{\eta}_{1T} - \eta_1^0\|,$$

for some $c > 0$. Therefore, any term of the quadratic form in the RHS of (A.1) which involves $(\hat{\eta}_{1T} - \eta_1^0)$ is negligible in front of $\frac{\partial \rho}{\partial \theta'}(\theta^0)R^1(\hat{\eta}_{1T} - \eta_1^0)$. In other words, we can deduce from $\|\rho(\hat{\theta}_T)\| = O_P(T^{-1/2})$ that

$$\|z_T\| = O_P(T^{-1/2}),$$

where

$$z_T = \frac{\partial \rho}{\partial \theta'}(\theta^0)R^1(\hat{\eta}_{1T} - \eta_1^0) + \frac{1}{2} \left((\hat{\eta}_{2T} - \eta_2^0)' R^{2'} \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\tilde{\theta}_T)R^2(\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H}.$$

Moreover, we can decompose

$$z_T = z_T^* + z_T^{**}$$

with

$$z_T^* = \frac{\partial \rho}{\partial \theta'}(\theta^0)R^1(\hat{\eta}_{1T} - \eta_1^0) + \frac{1}{2} \left((\hat{\eta}_{2T} - \eta_2^0)' R^{2'} \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0)R^2(\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H}$$

and

$$z_T^{**} = \frac{1}{2} \left((\hat{\eta}_{2T} - \eta_2^0)' R^{2'} \left(\frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\tilde{\theta}_T) - \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) \right) R^2(\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H}.$$

Note that with M being the projection matrix introduced in Lemma 2.1,

$$Mz_T^* = \frac{1}{2}M \left((\hat{\eta}_{2T} - \eta_2^0)' R^{2'} \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0)R^2(\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H}$$

since by definition $M \frac{\partial \rho}{\partial \theta'}(\theta^0) = 0$. Hence, by applying Lemma 2.1,

$$\|z_T^*\| \geq \|Mz_T^*\| \geq \frac{\gamma}{2} \|R^2(\hat{\eta}_{2T} - \eta_2^0)\|^2. \quad (\text{A.2})$$

Since

$$\left\| \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\tilde{\theta}_T) - \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) \right\| = o_P(1),$$

we can deduce from (A.2) that

$$\|z_T^{**}\| = o_P(\|z_T^*\|).$$

Therefore, from $z_T = z_T^* + z_T^{**}$, we can deduce

$$\|z_T\| = O_P(T^{-1/2}) \Rightarrow \|z_T^*\| = O_P(T^{-1/2}).$$

Then (A.2) shows that

$$\|R^2(\hat{\eta}_{2T} - \eta_2^0)\| = O_P(T^{-1/4}).$$

Therefore,

$$a_T = \frac{\partial \rho}{\partial \theta'}(\theta^0) R^1(\hat{\eta}_{1T} - \eta_1^0) = z_T - \frac{1}{2} \left((\hat{\eta}_{2T} - \eta_2^0)' R^2' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\tilde{\theta}_T) R^2(\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H}$$

is the sum of two terms of order $O_P(T^{-1/2})$. Thus

$$\|a_T\| = O_P(T^{-1/2})$$

and since $\|a_T\| \geq c\|\hat{\eta}_{1T} - \eta_1^0\|$, we have

$$\|\hat{\eta}_{1T} - \eta_1^0\| = O_P(T^{-1/2}). \tag{A.3}$$

Then, for any α in the range of $\frac{\partial \rho'}{\partial \theta}(\theta^0)$, by definition, $\alpha = R^1 \beta$ for some β in \mathbb{R}^r , and then

$$\alpha'(\hat{\theta}_T - \theta^0) = \beta' R^1(\hat{\eta}_{1T} - \eta_1^0)$$

is also of order $O_P(T^{-1/2})$ \square

Lemma A.1. *Let $\{X_T : T \in \mathbb{N}\}$ be a sequence of real valued random variables converging in distribution towards X . Let F_T and F be the cumulative distribution functions of X_T and X , respectively.*

For any $x_0 \in \mathbb{R}$, we have

$$F(x_0-) \leq \liminf_{T \rightarrow \infty} F_T(x_0) \leq \limsup_{T \rightarrow \infty} F_T(x_0) \leq F(x_0),$$

with $F(x_0-) = \lim_{x \rightarrow x_0^-} F(x)$.

Proof of Lemma A.1. We only show that $F(x_0-) \leq \liminf_{T \rightarrow \infty} F_T(x_0)$. The last inequality can be established along similar lines while the second one is obvious.

Let us assume that $F(x_0-) > l \equiv \liminf_{T \rightarrow \infty} F_T(x_0)$. Since the sequence $\{F_T(x_0) : T \in \mathbb{N}\}$ lies between 0 and 1, $0 \leq l \leq 1$. Let $\epsilon = \frac{F(x_0-) - l}{3}$. By definition of $F(x_0-)$, there exists $\eta_0 < x_0$ such that

$$F(\eta_0) - F(x_0-) > -\epsilon.$$

Since the set of discontinuity points of F is countable, we can consider that F is continuous at η_0 . Therefore, $F_T(\eta_0) \rightarrow F(\eta_0)$ as $T \rightarrow \infty$. This implies that there exists T_0 such that, for all $T > T_0$,

$$F_T(\eta_0) - F(\eta_0) > -\epsilon.$$

Thus, for all $T > T_0$,

$$F_T(\eta_0) > -2\epsilon + F(x_0-).$$

Besides, by definition of l , there exists T_1 such that, for all $T \geq T_1$,

$$\inf_{n \geq T} F_n(x_0) - l \leq \epsilon/2.$$

As a consequence, there exists $T > \max(T_0, T_1)$ such that

$$F_T(x_0) \leq l + \epsilon$$

and

$$F_T(\eta_0) > -2\epsilon + F(x_0-).$$

Since $l + \epsilon = -2\epsilon + F(x_0-)$, we deduce that

$$F_T(\eta_0) > F_T(x_0).$$

The contradiction is set up by the fact that F_T is non-decreasing and $\eta_0 < x_0$ \square

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

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

Proof of Lemma A.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

Proof of Proposition 3.3. A second-order Taylor expansion similar to Proposition 3.2 gives

$$\begin{aligned} \sqrt{T} \bar{\phi}_T(\hat{\theta}_T) &= \sqrt{T} \bar{\phi}_T(\theta_0) + \frac{\partial \rho}{\partial \theta'}(\theta_0) R^1 \sqrt{T}(\hat{\eta}_{1T} - \eta_1^0) \\ &\quad + \frac{1}{2} \sqrt{T} \left((\hat{\eta}_{2T} - \eta_2^0)' R^{2'} \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) R^2 (\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H} + o_P(1) \end{aligned} \tag{A.4}$$

With $\Omega_T = \Omega = Id_H$, we can write the first order condition as

$$\frac{\partial \bar{\phi}_T'}{\partial \theta}(\hat{\theta}_T) \bar{\phi}_T(\hat{\theta}_T) = 0. \tag{A.5}$$

Plugging (A.4) into (A.5) and since $\frac{\partial \bar{\phi}_T'}{\partial \theta}(\hat{\theta}_T) = \frac{\partial \rho'}{\partial \theta}(\theta^0) + o_P(1)$, we have

$$\begin{aligned} \frac{\partial \rho'}{\partial \theta}(\theta_0) \left(\sqrt{T} \bar{\phi}_T(\theta_0) + \frac{\partial \rho}{\partial \theta'}(\theta_0) R^1 \sqrt{T}(\hat{\eta}_{1T} - \eta_1^0) \right. \\ \left. + \frac{1}{2} \sqrt{T} \left((\hat{\eta}_{2T} - \eta_2^0)' R^{2'} \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) R^2 (\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H} \right) = o_P(1). \end{aligned} \tag{A.6}$$

Moreover, with $X = \frac{\partial \rho}{\partial \theta'}(\theta^0) R^1$, the projection matrix P on the range of $\frac{\partial \rho}{\partial \theta'}(\theta^0)$ is

$$P = X(X'X)^{-1}X'.$$

From (A.6), we have

$$\begin{aligned} \sqrt{T}(\hat{\eta}_{1T} - \eta_1^0) &= -(X'X)^{-1}X' \left(\sqrt{T} \bar{\phi}_T(\theta_0) \right. \\ &\quad \left. + \frac{1}{2} \sqrt{T} \left((\hat{\eta}_{2T} - \eta_2^0)' R^{2'} \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) R^2 (\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H} \right) + o_P(1). \end{aligned} \tag{A.7}$$

Plugging (A.7) into (A.4), we get

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = M\sqrt{T}\bar{\phi}_T(\theta^0) + \frac{1}{2}M\sqrt{T} \left((\hat{\eta}_{2T} - \eta_2^0)' R^2 \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) R^2 (\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H} + o_P(1), \quad (\text{A.8})$$

with

$$M = Id_H - P.$$

It is worth comparing the minimum distance estimator $\hat{\theta}_T$ with the one we would have computed if we knew η_2^0 . This estimator would be

$$\tilde{\theta}_T = R(\tilde{\eta}'_{1T}, \eta_2^{0'})',$$

with $\tilde{\eta}_{1T}$ solution of

$$\arg \min_{\eta_1} \bar{\phi}_T^*(\eta_1, \eta_2^0) \bar{\phi}_T^*(\eta_1, \eta_2^0),$$

where $\bar{\phi}_T^*(\eta) = \bar{\phi}_T(R\eta)$.

By an argument similar to (A.8), we get

$$\sqrt{T}\bar{\phi}_T(\tilde{\theta}_T) = M\sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1). \quad (\text{A.9})$$

Therefore,

$$\begin{aligned} T\bar{\phi}'_T(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T) - T\bar{\phi}'_T(\tilde{\theta}_T)\bar{\phi}_T(\tilde{\theta}_T) &= \Delta' (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M\sqrt{T}\bar{\phi}_T(\theta^0) \\ &+ \frac{1}{4}\Delta' (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M\Delta (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) + o_P(1), \end{aligned} \quad (\text{A.10})$$

where $\Delta(v)$ is the H -dimensional vector

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

By definition,

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

Thus from (A.10),

$$\begin{aligned} \Delta' (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M\sqrt{T}\bar{\phi}_T(\theta^0) \\ + \frac{1}{4}\Delta' (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M\Delta (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) \leq o_P(1). \end{aligned} \quad (\text{A.11})$$

It is worth noting that

$$\Delta' (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M\sqrt{T}\bar{\phi}_T(\theta^0) = \left(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0) \right)' Z_T \left(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0) \right), \quad (\text{A.12})$$

where Z_T is a symmetric matrix defined in the statement of Proposition 3.3. Moreover we know that $T^{1/4}(\hat{\eta}_{2T} - \eta_2^0) = O_P(1)$. Thus, at least, for a subsequence, we can write

$$T^{1/4}(\hat{\eta}_{2T} - \eta_2^0) \xrightarrow{d} U.$$

For the sake of simplicity, we do not make explicit the notation for a subsequence. Thus

$$\Delta' \left(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0) \right) M\sqrt{T}\bar{\phi}_T(\theta^0) \xrightarrow{d} U' Z U.$$

From (A.11) and by Lemma A.2, we deduce that

$$\begin{aligned} \limsup_{T \rightarrow \infty} \text{Prob} \left(\Delta' (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M\sqrt{T}\bar{\phi}_T(\theta^0) + \right. \\ \left. + \frac{1}{4}\Delta' (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M\Delta (R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) \leq \epsilon \right) = 1, \end{aligned}$$

for all $\epsilon > 0$. Thus, by Lemma A.1, we have

$$\text{Prob} \left(U'ZU + \frac{1}{4}\Delta'(R^2U)M\Delta(R^2U) \leq \epsilon \right) = 1, \quad \forall \epsilon > 0.$$

Thus, by right continuity of cumulative distribution functions,

$$\text{Prob} \left(U'ZU + \frac{1}{4}\Delta'(R^2U)M\Delta(R^2U) \leq 0 \right) = 1.$$

We deduce in particular that if Z is positive semi-definite

$$\Delta'(R^2U)M\Delta(R^2U) \stackrel{a.s}{=} 0$$

and thus

$$\|M\Delta(R^2U)\| \stackrel{a.s}{=} 0$$

But, by Lemma 2.1,

$$\|M\Delta(R^2U)\| \geq \gamma\|R^2U\|^2.$$

Thus $R^2U \stackrel{a.s}{=} 0$.

By definition, $T^{1/4}(\hat{\theta}_T - \theta^0) = T^{1/4}R^1(\hat{\eta}_{1T} - \eta_1^0) + T^{1/4}R^2(\hat{\eta}_{2T} - \eta_2^0)$ with, by (A.3), $T^{1/4}R^1(\hat{\eta}_{1T} - \eta_1^0) \xrightarrow{P} 0$. Hence, when Z is positive semi-definite,

$$T^{1/4}(\hat{\theta}_T - \theta^0) \xrightarrow{d} R^2U \stackrel{a.s}{=} 0.$$

In other words, we have shown that

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

Conversely, let us assume that Z is not positive semi-definite. Then, we can find a vector e in the unit sphere of \mathbb{R}^{p-r} such that

$$\text{Prob}(e'Ze < 0) > 0.$$

Besides, the necessary second order condition for an interior solution for a minimization problem implies that

$$e' \left(\frac{\partial^2}{\partial \eta_2 \partial \eta_2'} (\bar{\phi}_T^{*'}(\eta) \bar{\phi}_T^*(\eta)) \Big|_{\eta = \hat{\eta}_T} \right) e \geq 0.$$

This can be written

$$e' \left(\tilde{Z}_T + G_T \right) e \geq 0, \tag{A.13}$$

where

$$\tilde{Z}_T = \left(\frac{\partial^2 \bar{\phi}_T^{*'}}{\partial \eta_{2i} \partial \eta_{2j}'} (\hat{\eta}_T) \sqrt{T} \bar{\phi}_T^* (\hat{\eta}_T) \right)_{1 \leq i, j \leq p-r}$$

and

$$G_T = \sqrt{T} \frac{\partial \bar{\phi}_T^{*'}}{\partial \eta_2} (\hat{\eta}_T) \frac{\partial \bar{\phi}_T^*}{\partial \eta_2'} (\hat{\eta}_T).$$

By a mean value expansion, we have

$$\frac{\partial \bar{\phi}_T^*}{\partial \eta_{2i}} (\hat{\eta}_T) = \frac{\partial^2 \bar{\phi}_T^*}{\partial \eta_{2i} \partial \eta_2'} (\bar{\eta})(\hat{\eta}_{2T} - \eta_2^0) + O_P(T^{-1/2}), \tag{A.14}$$

with $\bar{\eta} \in (\eta^0, \hat{\eta}_T)$ and $i = 1, \dots, p-r$. On the other hand, thanks to Equation (A.8), we have

$$\frac{\partial^2 \bar{\phi}_T^{*'}}{\partial \eta_{2i} \partial \eta_{2j}'} (\hat{\eta}_T) \bar{\phi}_T^* (\hat{\eta}_T) = \frac{\partial^2 \rho^{*'}}{\partial \eta_{2i} \partial \eta_{2j}'} (\eta^0) \left(M \bar{\phi}_T(\theta^0) + \frac{1}{2} M \Delta(R^2(\hat{\eta}_{2T} - \eta_2^0)) \right) + o_P(T^{-1/2}),$$

with $\rho^*(\eta) = \rho(R\eta)$. Noting that

$$\frac{\partial^2 \rho_h^*}{\partial \eta_2 \partial \eta_2'} (\eta^0) = R^{2'} \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'} (\theta^0) R^2 \equiv N' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'} (\theta^0) N, \quad h = 1, \dots, H,$$

we have

$$\frac{\partial^2 \bar{\phi}_T^{*'}}{\partial \eta_{2i} \partial \eta_{2j}}(\hat{\eta}_T) \sqrt{T} \bar{\phi}_T^*(\hat{\eta}_T) = a'_{ij} \sqrt{T} M \bar{\phi}_T(\theta^0) + \frac{1}{2} a'_{ij} M \Delta(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) + o_P(1).$$

Thus

$$\tilde{Z}_T = Z_T + \frac{1}{2} \left(a'_{ij} M \Delta(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) \right)_{1 \leq i, j \leq p-r} + o_P(1)$$

and

$$G_T = \left(T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)' \frac{\partial^2 \rho^{*'}}{\partial \eta_{2i} \partial \eta_2}(\eta^0) \frac{\partial^2 \rho^*}{\partial \eta_{2j} \partial \eta_2}(\eta^0) T^{1/4}(\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq i, j \leq p-r} + o_P(1).$$

From the inequality (A.13) and some successive applications of the Cauchy-Schwarz inequality, we have

$$-e' Z_T e \leq A \sqrt{T} \|\hat{\eta}_{2T} - \eta_2^0\|^2 + o_P(1), \quad (\text{A.15})$$

for some $A > 0$. Noting that $\|\hat{\eta}_{2T} - \eta_2^0\| \leq \|\hat{\eta}_T - \eta^0\|$ and recalling that $\eta = R^{-1}\theta$, we also have

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

for some $A > 0$, which may be different from the previous one. By Lemma A.2,

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

Then, from Lemma A.1, we have

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

Thus, by right continuity of cumulative distribution functions,

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

and consequently,

$$\text{Prob}(\|V\| > 0 | e' Z e < 0) = 1.$$

Hence,

$$\text{Prob}(e' Z e < 0) = \text{Prob}(\|V\| > 0, e' Z e < 0) \leq \text{Prob}(\|V\| > 0, \overline{(Z \geq 0)}).$$

We deduce that

$$\text{Prob}(\|V\| > 0, \overline{(Z \geq 0)}) > 0$$

and thus

$$\text{Prob}(\|V\| > 0, \overline{(Z \geq 0)}) < \text{Prob}(\overline{(Z \geq 0)})$$

that is

$$\text{Prob}(\|V\| > 0 | \overline{(Z \geq 0)}) < 1 \square$$

Proof of Theorem 4.1. From Equation (A.8),

$$\sqrt{T} \bar{\phi}_T(\hat{\theta}_T) = \sqrt{T} \bar{\phi}_T^*(\hat{\eta}_T) = M \sqrt{T} \bar{\phi}_T(\theta^0) + \frac{1}{2} M \Delta(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) + o_P(1).$$

(See the proof of Proposition 3.2 for the definition of $\hat{\eta}_T$, $\bar{\phi}_T^*(\cdot)$, R and R^2 .) Thus,

$$\begin{aligned} J_T &= T \bar{\phi}_T'(\theta^0) M \bar{\phi}_T(\theta^0) + \Delta'(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M \sqrt{T} \bar{\phi}_T(\theta^0) \\ &\quad + \frac{1}{4} \Delta'(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) M \Delta(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) + o_P(1). \end{aligned} \quad (\text{A.16})$$

Conditional on $(Z \geq 0)$, from Proposition 3.3, $T^{1/4}(\hat{\theta}_T - \theta^0) = o_P(1)$ and we also have $T^{1/4}(\hat{\eta}_T - \eta^0) = T^{1/4} R^{-1}(\hat{\theta}_T - \theta^0) = o_P(1)$. Therefore,

$$J_T = T \bar{\phi}_T'(\theta^0) M \bar{\phi}_T(\theta^0) + o_P(1).$$

Since M is an orthogonal projection matrix on a subspace of dimension $H - r$, $T\bar{\phi}'_T(\theta^0)M\bar{\phi}_T(\theta^0) \xrightarrow{d} \chi_{H-r}^2$. This establishes the first part of the Theorem. The positivity of q follows from the comment of Proposition 3.3 in the body of the text.

On the other hand, the first order condition for $\hat{\eta}_{2T}$ is

$$\frac{\partial \bar{\phi}'_T}{\partial \eta_2}(\hat{\eta}_T)\bar{\phi}_T^*(\hat{\eta}_T) = 0.$$

Using the same expansion as in (A.14), we can deduce that

$$\begin{aligned} T^{1/4}\frac{\partial \bar{\phi}'_T}{\partial \eta_2}(\hat{\eta}_T) &= \frac{\partial^2 \rho^*}{\partial \eta_2^2}(\eta_2^0)T^{1/4}(\hat{\eta}_{2T} - \eta_2^0) + o_P(1) \\ &= \left(R^2 \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0) R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0) \right)_{1 \leq h \leq H} + o_P(1). \end{aligned} \quad (\text{A.17})$$

By pre-multiplying (A.17) by $T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)$, we have

$$T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)T^{1/4}\frac{\partial \bar{\phi}'_T}{\partial \eta_2}(\hat{\eta}_T) = \Delta(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) + o_P(1).$$

The first order condition therefore implies that

$$\Delta'(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0))M\Delta(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) + 2\Delta'(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0))M\sqrt{T}\bar{\phi}_T(\theta^0) = o_P(1), \quad (\text{A.18})$$

Hence,

$$J_T = T\bar{\phi}'_T(\theta^0)M\bar{\phi}_T(\theta^0) + \frac{1}{2}\Delta'(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0))M\sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1).$$

Note that since $\dim \eta_2 = 1$, $\Delta'(R^2 T^{1/4}(\hat{\eta}_{2T} - \eta_2^0)) = GT^{1/2}(\hat{\eta}_{2T} - \eta_2^0)^2$, where $G = \Delta(R^2)$ and Equation (A.18) can be re-written as

$$\sqrt{T}(\hat{\eta}_{2T} - \eta_2^0)^2 \left(G'MG\sqrt{T}(\hat{\eta}_{2T} - \eta_2^0)^2 + 2G'M\sqrt{T}\bar{\phi}_T(\theta^0) \right) = o_P(1). \quad (\text{A.19})$$

Next we show that, conditional on $(Z < 0)$, there exists no subsequence of $\sqrt{T}\|\hat{\eta}_{2T} - \eta_2^0\|^2$ that converges in distribution to a random variable with an atom of probability at 0. For that, let a subsequence of $\sqrt{T}\|\hat{\eta}_{2T} - \eta_2^0\|^2$ converge in distribution to W . From, (A.15) and Lemma A.2,

$$\limsup_{T \rightarrow \infty} \text{Prob} \left(-Z_T - A\sqrt{T}\|\hat{\eta}_{2T} - \eta_2^0\|^2 \leq \epsilon \right) = 1, \quad \forall \epsilon > 0.$$

Note that since $\dim \eta_2 = 1$, $e'Z_T e = Z_T$. Therefore, from Lemma A.1,

$$\text{Prob}(-Z - AW \leq \epsilon) = 1, \quad \forall \epsilon > 0.$$

By right continuity of cumulative distribution functions, we have

$$\text{Prob}(Z + AW \geq 0) = 1.$$

Thus,

$$\text{Prob}(W > 0 | Z < 0) = 1.$$

This means that, conditional on $(Z < 0)$, W does not have an atom of probability at 0. As a consequence, $\sqrt{T}(\hat{\eta}_{2T} - \eta_2^0)^2$ does not converge to a random variable with an atom of probability at 0 along any subsequence. Therefore, (A.19) implies that

$$\sqrt{T}(\hat{\eta}_{2T} - \eta_2^0)^2 = -2\frac{G'M}{G'MG}\sqrt{T}\bar{\phi}_T(\theta^0) + o_P(1).$$

Thus

$$\begin{aligned} J_T &= T\bar{\phi}'_T(\theta^0)M\bar{\phi}_T(\theta^0) - T\frac{\bar{\phi}'_T(\theta^0)MGG'M\bar{\phi}_T(\theta^0)}{G'MG} + o_P(1) \\ &= T\bar{\phi}'_T(\theta^0)(Id_H - \mathcal{P})\bar{\phi}_T(\theta^0) + o_P(1), \end{aligned} \quad (\text{A.20})$$

where $\mathcal{P} = X(X'X)^{-1}X' + (MG)(MG)'/G'MG$, with $X = \frac{\partial p}{\partial \theta'}(\theta^0)$, is an orthogonal projection matrix on a space of dimension $p = r + 1$ (Note that $MG \neq 0$ by Assumption 5). This proves that

$$\text{Prob}(J_T \leq x | Z < 0) \rightarrow \text{Prob}(\chi_{H-p}^2 \leq x),$$

as $T \rightarrow \infty$, for all x . From the first part of the proof, we have

$$\text{Prob}(J_T \leq x | Z \geq 0) \rightarrow \text{Prob}(\chi_{H-(p-1)}^2 \leq x),$$

as $T \rightarrow \infty$, for all x . Thus,

$$\lim_{T \rightarrow \infty} \text{Prob}(J_T \leq x) = \text{Prob}(Z < 0) \text{Prob}(\chi_{H-p}^2 \leq x) + \text{Prob}(Z \geq 0) \text{Prob}(\chi_{H-(p-1)}^2 \leq x).$$

Since $Z \sim \mathcal{N}(0, G'MG)$, $\text{Prob}(Z < 0) = \text{Prob}(Z \geq 0) = 1/2$ and the expected result follows \square

Proof of Theorem 5.1. We have

$$E((z_t - E(z_t))(u_{\theta,t+1}^2 - Eu_{\theta,t+1}^2)) = 0 \Leftrightarrow E((z_t - E(z_t))u_{\theta,t+1}^2) = 0.$$

Since $E(y_{1,t+1}^2 | J_t) = \lambda_1^2 \sigma_t^2 + \Omega_{11}$, $E(y_{2,t+1}^2 | J_t) = \lambda_2^2 \sigma_t^2 + \Omega_{22}$ and $E(y_{1,t+1}y_{2,t+1} | J_t) = \lambda_1 \lambda_2 \sigma_t^2 + \Omega_{12}$, $E((z_t - E(z_t))u_{\theta,t+1}^2) = 0$ can be written

$$(\lambda_2 - \lambda_1 \theta)^2 E((z_t - E(z_t))\sigma_t^2) = 0$$

or equivalently

$$(\lambda_2 - \lambda_1 \theta)^2 \text{Cov}(z_t, \sigma_t^2) = 0$$

which in turn is equivalent to $\theta = \theta_0 = \lambda_2/\lambda_1$. This establishes the existence and the uniqueness of θ_0 as stated by (i).

Next, we show (ii). Clearly,

$$E(\partial \phi_t(\theta_0)/\partial \theta) = E((z_t - Ez_t)[-2y_{1,t+1}(y_{2,t+1} - \theta_0 y_{1,t+1}) + E(2y_{1,t+1}(y_{2,t+1} - \theta_0 y_{1,t+1}))]).$$

Since $y_{2,t+1} - \theta_0 y_{1,t+1} = U_{2,t+1} - \theta_0 U_{1,t+1}$ and $E(f_{t+1} U_{t+1} | J_t) = 0$,

$$E(y_{1,t+1}(y_{2,t+1} - \theta_0 y_{1,t+1}) | J_t) = \Omega_{12} - \theta_0 \Omega_{11} = E(y_{1,t+1}(y_{2,t+1} - \theta_0 y_{1,t+1})).$$

Therefore,

$$E(\partial \phi_t(\theta_0)/\partial \theta) = E((z_t - Ez_t) \times 0) = 0$$

Thus (ii).

On the other hand,

$$E(\partial^2 \phi_t(\theta_0)/\partial \theta^2) = -2E((z_t - Ez_t)(y_{1,t+1}^2 - E(y_{1,t+1}^2))).$$

Note that $E(y_{1,t+1}^2 | J_t) - E(y_{1,t+1}^2) = \lambda_1^2(\sigma_t^2 - 1)$. Thus,

$$E(\partial^2 \phi_t(\theta_0)/\partial \theta^2) = -2\lambda_1^2 E((z_t - Ez_t)\sigma_t^2) = -\lambda_1^2 \text{Cov}(z_t, \sigma_t^2) \neq 0.$$

This establishes (iii) \square

References

- [1] Andrews, D. W. K., 1994. "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity," *Econometrica*, 62, 43-72.
- [2] Antoine, B. and E. Renault, 2009. "Efficient GMM with Nearly-weak Instruments," *Econometrics Journal*, 12, S135-S171.
- [3] Arellano, M., L. P. Hansen and E. Sentana, 2009. "Underidentification?," *working paper*, CEMFI, <http://www.cemfi.es/~arellano/ahs0731.pdf>.
- [4] Bakshi, G., N. Kapadia and D. Madan, 2003. "Stock Return Characteristics, Skew Laws, and Differential Pricing of Individual Equity Options," *Review of Financial Studies*, 16, 101-143.
- [5] Choi, I. and P. C. B. Phillips, 1992. "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113-150.
- [6] 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.
- [7] Doz, C. and E. Renault, 2006. "Factor Volatility in Mean Models: a GMM approach," *Econometric Reviews*, 25, 275-309.
- [8] Dufour, J.-M. and P. Valéry, 2009. "GMM and Hypothesis Tests when Rank Conditions Fail: a Regularization Approach," *working paper*, CIREQ, <http://www.cireq.umontreal.ca/activites/papiers/08-09valery.pdf>.
- [9] Engle, R. F., 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation," *Econometrica*, 50, 987-1008.
- [10] Engle, R. F. and S. Kozicki, 1993. "Testing For Common Features," *Journal of Business and Economic Statistics*, 11(4), 369-395.
- [11] Engle, R. F. and A. Mistry, 2007. "Priced Risk and Asymmetric Volatility in the Cross-Section of Skewness," *working paper*, NYU, <http://pages.stern.nyu.edu/~rengle/>.
- [12] Engle, R. F. and R. Susmel, 1993. "Common Volatility in International Equity Markets," *Journal of Business and Economic Statistics*, 11, 167-176.
- [13] Fiorentini, G., E. Sentana and N. Shephard, 2004. "Likelihood-based Estimation of Generalised ARCH Structures," *Econometrica*, 72, 1481-1517.
- [14] Hansen, L. P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.
- [15] Hansen, L. P., J. Heaton and A. Yaron, 1996. "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-280.
- [16] Hansen, L. P. and R. J. Hodrick, 1980. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829-853.
- [17] Hansen, L. P. and K. J. Singleton, 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269-1286.
- [18] Hayashi, F., 2000. "Econometrics," *Princeton University Press*.
- [19] Horn, A. R. and C. R. Johnson 1985. "Matrix Analysis," *Cambridge University Press*.
- [20] King, M. A., E. Sentana and S. B. Wadhvani, 1994. "Volatility and Links Between National Stock Markets," *Econometrica*, 62, 901-933.
- [21] Kleibergen, F., 2005. "Testing Parameters in GMM Without Assuming that they are Identified," *Econometrica*, 73, 1103-1123.
- [22] Lee, L. F. and A. Chesher, 1986. "Specification Testing when Score Test Statistics are Identically Zero," *Journal of Econometrics*, 31, 121-149.

- [23] MacKinlay, A. C. and M. P. Richardson, 1991. "Using Generalized Method of Moments to Test Mean-Variance Efficiency ," *The Journal of Finance*, 46, 511-527.
- [24] Melino, A., 1982. "Testing for Sample Selection Bias," *Review of Economic Studies*, 49, 151-153.
- [25] Newey, K. W. and D. McFadden, 1994. "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, IV, Edited by R.F. Engle and D. L. McFadden, 2112-2245.
- [26] Newey, K. W. and R. J. Smith, 2004. "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.
- [27] Rotnitzky, A., D. R. Cox, M. Bottai and J. Robins 2000. "Likelihood-based Inference with Singular Information Matrix," *Bernoulli*, 6(2), 243-284.
- [28] Sargan, J. D., 1983. "Identification and lack of Identification," *Econometrica*, 51, 1605-1633.
- [29] Staiger, D. and J. H. Stock, 1997. "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.
- [30] Stock, J. H. and J. H. Wright, 2000. "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.
- [31] Van der Vaart, A. W., 1998. "Asymptotic Statistics," *Cambridge University Press*.