Identification robust inference in structural multivariate factor models with rank restrictions *

Marie-Claude Beaulieu † Université Laval

Jean-Marie Dufour ‡ McGill University

Lynda Khalaf § Carleton University

February 2009

* This work was supported by the Institut de Finance Mathématique de Montréal [IFM2], the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Research Chair Program, the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fond Banque Royale en Finance and the Fonds FQRSC and NATECH (Government of Québec).

† Centre Interuniversitaire de Recherche sur les Politiques Économique et de l'Emploi (CIRPÉE), Département de finance et assurance, Pavillon Palasis-Prince, local 3632, Université Laval, Québec, Canada G1K 7P4. TEL: (418) 656-2131-2926, FAX: (418) 656-2624. email: Marie-Claude.Beaulieu@fas.ulaval.ca

‡ Centre interuniversitaire de recherche en économie quantitative (CIREQ), Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Canada Research Chair holder université de Montréal and William Dow Professor of Economics, Department of Economics, McGill University. Mailing address: Leacock Building, Rm 519, 855 Sherbrooke Street West, Montréal, Québec, Canada H3A 2T7. TEL: (514) 398 8879; FAX: (514) 398 4938; e-mail: jean-marie.dufour@mcgill.ca.

§ Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles (GREEN) Université Laval, Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Economics Department, Carleton University. Mailing address: Economics Department, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel (613) 520-2600-8697; FAX: (613)-520-3906. email: Lynda_Khalaf@carleton.ca.

ABSTRACT

We propose identification robust inference methods for structural multivariate factor models with rank restrictions. Such models involve nonlinear reduced rank restrictions whose identification may raise serious non-regularities leading to the failure of standard asymptotics. First, we prove several invariance and nuisance-parameter reduction results for commonly used eigenvalue and minimum root based statistics; we also derive useful scale-invariance results for heteroskedasticity-autocorrelation robust multivariate Wald-type criteria. Second, we derive confidence set estimates for structural parameters based on inverting minimum-distance type pivotal statistics. We provide analytical solutions to the latter problem which hold exactly (or asymptotically) imposing (or relaxing) Gaussian fundamentals. Simulation-based counterparts are also suggested for non-Gaussian hypotheses. Results are not restricted to the intertemporal *i.i.d.* setting. The statistics we invert include Hotelling's $\mathbf{T}^2$ criterion, which is widely used in multivariate analysis for *test* purposes. Our proposed confidence sets have much more informational content than Hotelling-type tests, and extend their relevance beyond reduced form specifications. Our approach further provides multivariate extensions of the classical Fieller problem, and may be viewed as a generalization of Dufour and Taamouti's (Econometrica, 2005) quadrics-based set estimation method beyond the linear limited information simultaneous equations setting. Third, we provide a formal definition of a statistically non-informative factor and prove necessary conditions linking the presence of such factors to unbounded set estimation outcomes. We also document the perverse effects of adding such factors on J-type minimum-root-based model tests. Fourth, we provide a unified analytical treatment of point estimation. With reduced rank constraints, normalizations raise uniqueness issues that may not matter in some contexts, yet in many econometric or financial structural models, normalizations are motivated by underlying theory. Fifth, proposed inference methods are applied to a multi-factor Capital Asset Pricing type model with unobservable risk-free rates and an Arbitrage Pricing Theory based model with Fama-French factors. Results reveal dramatic differences between the standard Wald-type confidence set estimates and our proposed identification robust ones and illustrate the severe implications of redundant factors. In particular, we find that the term structure variables and the momentum factor are statistically non-informative in many sub-periods. This causes tests for model fit to spuriously pass the underlying pricing restrictions, though associated confidence sets are much too wide. These results document the serious pitfalls of usual asset pricing tests and illustrate the worth of our proposed confidence set based analysis.

1. Introduction

Structural multivariate econometric models commonly involve reduced rank [**RR**] restrictions. Well known applications include: (i) RR regression and dimensionality test problems [see Anderson (1951, 1999, 2006)][1], (ii) simultaneous equations [see Kleibergen and van Dijk (1998), Hoogerheide, Kaashoek and van Dijk (2007) and Anderson (2006) on the link between simultaneous equations and RR regressions][2], and (iii) financial econometric works related to the Arbitrage Pricing Theory (**APT**) [Black (1972), Ross (1976)][3] and to factor models. In this context, this paper has both methodological and empirical motivations.

From a methodological perspective, RR restrictions raise serious statistical challenges. Even with linear multivariate models, discrepancies between standard asymptotic and finite sample distributions can be very severe and are usually attributed to the curse of dimensionality; see *e.g.* Dufour and Khalaf (2002) and the references therein. Nonlinear rank restrictions pose further identification non-regularities that can lead to the failure of standard asymptotics; the above cited literature on instrumental regressions provides prominent illustrations of this problem. We propose inference methods immune to both dimensionality and identification difficulties.

From an empirical perspective, we focus on rank-based inference methods relevant to asset pricing factor models. In general multivariate regression based financial models, finite sample motivated testing is important because tests that are only approximate and/or do not account for non-normality can lead to unreliable empirical interpretations of standard financial models; see *e.g.* Shanken (1996), Campbell et al. (1997), Dufour, Khalaf and Beaulieu (2003, 2008) and Beaulieu, Dufour and Khalaf (2005, 2007, 2008*a*). The few finite sample rank-based asset pricing methods [see *e.g.* Zhou (1991, 1995), Beaulieu, Dufour and Khalaf (2008*c*), Shanken (1985, 1986) and Shanken and Zhou (2007)] have focused on testing rather than on set inference.[4]

On the above theoretical and empirical issues, the paper has five main contributions. First, we prove several invariance and nuisance-parameter reduction results for commonly

[1] See also *e.g.* Rao (1973, Chapter 8), Fujikoshi (1974), Calinsky and Lejeune (1998), Gouriéroux, Monfort and Renault (1995), Johansen (1995), Geweke (1996), Reinsel and Velu (1998), Cragg and Donald (1997), Kleibergen (1999), Kleibergen and Paap (1998) and the references therein.

[2] See also *e.g.* Dufour (1997, 2003), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz and Nelson (1998), Dufour and Jasiak (2001), Kleibergen (2002, 2005), Stock, Wright and Yogo (2002), Moreira (2003), Dufour and Taamouti (2005, 2007) and Andrews, Moreira and Stock (2006).

[3] See also Gibbons (1982), Barone-Adesi (1985), Shanken (1986), Zhou (1991, 1995), Bekker, Dobbelstein and Wansbeek (1996), Costa, Gardini and Paruolo (1997), Campbell, Lo and MacKinlay (1997, Chapter 6), Velu and Zhou (1999) and Barone-Adesi, Gagliardini and Urga (2004*a*, 2004*b*).

[4] In Beaulieu, Dufour and Khalaf (2008*b*), we propose a confidence set estimate approach, yet the weakly identified parameter [the zero-beta rate, in this case] is scalar. In the present paper our results cover parameter vectors, which expands the class of applicable financial models.

used eigenvalue and minimum root based statistics. Depending on the restrictions considered, results hold: (i) given a wide class of multivariate mixtures of normals [that includes *e.g.* the skewed normal or Student-t as special cases], or (ii) assuming that the error distribution is parametrically specified up to an unknown scale transformation [that does not even require the existence of moments]. The latter case includes the distribution family studied by Dufour and Khalaf (2002); in the present paper however, results are not restricted to the intertemporal *i.i.d.* framework. Indeed, we also obtain useful scale-invariance results for heteroskedasticity-autocorrelation robust (**HAR**) criteria [as considered in *e.g.* Ravikumar, Ray and Savin (2000), Ray and Savin (2008)].

Second, for the model's parameters of interest, we derive confidence set (**CS**) estimates for structural parameters based on inverting minimum-distance type pivotal statistics. The statistics we invert include Hotelling's $\mathbf{T}^2$ criterion [Hotelling (1947)]. In multivariate analysis, Hotelling's statistic is mostly used for *test* purposes, and its popularity stems from its least-squares (**LS**) foundations which lead to exact **F**-based null distributions in Gaussian settings. We provide analytical solutions to the test inversion problem which hold exactly (or asymptotically) imposing (or relaxing) Gaussian error distributions. Simulation-based counterparts are also suggested for non-Gaussian parametric hypotheses. Our CSs provides much more information than Hotelling-type tests, and extend their relevance beyond reduced form specifications.

Our test inversion approach can be viewed as an extension of the classical inference procedure proposed by Fieller (1954) [see also Zerbe, Laska, Meisner and Kushner (1982), Dufour (1997), Beaulieu et al. (2008*b*), Bolduc, Khalaf and Yelou (2008)] to the multivariate setting. As with standard χ^2-type methods including *e.g.* the *delta* method, our generalized Fieller approach relies on LS. Yet both methods exploit LS theory in fundamentally different ways. In contrast to the former which excludes parameter discontinuity regions out and which by construction yields bounded confidence intervals, our inverted test does not require parameter identification and allows for unbounded solutions. Fieller-type methods are rare in multivariate settings and may involve numerical complications [see e.g. Bennett (1959, 1961)]. Our methodology may also be viewed as a generalization of the Dufour and Taamouti (2005, 2007) quadrics-based set estimation method beyond the linear limited information simultaneous equations setting.

Third, we provide a formal definition of a statistically non-informative factor and prove necessary conditions linking the presence of such factors to unbounded set estimation outcomes. We also document the perverse effects of adding such factors on J-type minimum-root-based model tests. Although unbounded set estimators are not always uninformative, our results call for caution in relying on J-tests alone as measures of model fit. This warning has obvious implications for the asset pricing models that motivate our empirical analysis.

Fourth, we provide a unified analytical treatment of point estimation. With RR constraints, normalizations raise uniqueness issues that may not matter in some contexts, yet in many econometric or financial structural models, normalizations are motivated by underlying theory. This holds particularly true for the asset pricing models we consider empirically.

Fifth, our results are applied to a multi-factor Capital Asset Pricing model and to an APT-based factor model with Fama and French [Fama and French (1992, 1993)] factors. A simulation study modelled on this case and allowing for multivariate GARCH reveals serious distortions with usual asymptotic tests; in contrast, our proposed inference methods perform well, even when GARCH effects are not accounted for. Empirically, we find dramatic differences between standard inference methods and our identification robust procedures. In particular, on comparing our results to the linear tests of the Fama-French model performed in Beaulieu, Dufour and Khalaf (2008a), we illustrate the severe implications of over-conditioning, following our formal definition of a statistically redundant factor. Indeed, we find that the term structure variables and the momentum factor are uninformative in many sub-periods. This causes tests for model fit to spuriously pass the underlying restrictions, in the following sense. The tests are statistically sound and fail to reject the pricing relations; unfortunately, associated confidence sets for key model parameters are unbounded. We argue that reliance on usual χ^2 tests may have delayed awareness on problems stemming from such factors in empirical finance. Our results support the Lewellen, Nagel and Shanken (2006) approach that favours confidence set based inference. Since to the best of our knowledge, identification-robust set estimates for APT-type regressions are as yet unavailable, the results of this paper provide valuable tools in asset pricing analysis.

The paper is organized as follows. Section 2 sets the notation and framework. Sections 3 and 4 present the estimators and tests considered. In section 5, we present our test inversion strategy. Non-linear tests are analyzed in section 4. Applications are discussed in section 7; the simulation study is presented in section 7.1 and our empirical analysis is discussed in section 7.2. Section 8 concludes followed by a technical appendix.

2. Framework and notation

We consider structural models which impose rank restrictions on the following reduced form:

$$Y = XB + U, \quad U = WJ' \Leftrightarrow Y_t = B'X_t + U_t, \quad U_t = JW_t, \ \ t = 1, \ \dots \ , T \tag{2.1}$$

where Y is a $T \times n$ matrix of observations on n endogenous variables, X is a $T \times k$ full-column rank matrix of exogenous variables, Y_t' and X_t' are, respectively, the t-th row of Y and X so that Y_t and X_t provide the t-th observation on the dependant variables and regressors, J is unknown, non-singular and possibly random, U_t' is the t-th row of U, W is a $T \times n$ matrix of random errors, W_t' is the t-th row of W, and the joint distribution of $W_1, \ldots, W_T$ is either fully specified, or specified up to a nuisance parameter κ. Finite sample results in this paper assume we can condition on X for statistical analysis. The structural models we consider can thus be defined via the following restrictions on (2.1):

$$\mathcal{H}_1\left(\bar{\theta}\right) \;:\; (1, \bar{\theta}')B = 0, \text{ for some unknown vector } \bar{\theta}, \tag{2.2}$$

$$\mathsf{H}_1\left(\bar{\delta}\right) \;:\; B(1, \bar{\delta}')' = 0, \text{ for some unknown vector } \bar{\delta}, \tag{2.3}$$

$$\mathcal{H}_2\left(\bar{\zeta}, \bar{\mu}\right) \;:\; (1, \bar{\zeta}')B = \bar{\mu}\iota_n', \text{ for some unknown vector } (\bar{\zeta}', \bar{\mu})', \tag{2.4}$$

$$\mathsf{H}_2\left(\bar{\phi}, \bar{\eta}\right) \;:\; B(1, \bar{\phi}')' = \bar{\eta}\iota_k, \text{ for some unknown vector } (\bar{\phi}', \bar{\eta})', \tag{2.5}$$

where ι_j refers to a j-dimensional vector of ones, $\bar{\theta}$ and $\bar{\zeta}$ are $q \times 1$, $\bar{\delta}$ and $\bar{\phi}$ are $m \times 1$ with $q = k - 1$, $m = n - 1$ and $\bar{\mu}$ and $\bar{\eta}$ are scalars. Given the underlying RR restrictions on the matrix B, typically, (2.2) and (2.4) assume that $k \leq n$ and (2.3) and (2.5) that $k \geq n$.[5] We focus on estimating the vectors $\bar{\theta}$, $\bar{\delta}$, $(\bar{\zeta}', \bar{\mu})'$ or $(\bar{\phi}', \bar{\eta})'$, respectively. As is illustrated in section 7, (2.2) - (2.5) capture several empirically relevant multivariate structural econometric models, including simultaneous equations.

For some though not all inferential problems considered, we will assume the following mixture distributional setting:

$$W = VZ \tag{2.6}$$

where V is $T \times T$, unknown and possibly random (in which case it is independent of Z), and Z is a $T \times n$ matrix of *i.i.d.* n-dimensional standard normal variables *i.e.* if we denote the t-th row of Z as Z_t', then

$$Z_t \overset{i.i.d.}{\sim} N(0, I_n). \tag{2.7}$$

The error structure in (2.1) is not restricted to the intertemporal *i.i.d.* setting. Time-dependence may be fit via appropriate specifications for the joint distribution of W_t, $t = 1, \ldots, T$. For instance, a multivariate GARCH structure [see Bauwens, Laurent and Rombouts (2006) for a recent survey] may be considered if T is sufficiently large relative to n. In particular, the following parsimonious parametric GARCH structure

[5] If $\bar{\mu}$ or $\bar{\eta}$ are strictly positive (or negative), then (2.4) and (2.5) may be viewed as inequality restrictions; results below illustrate that projection methods provide an interesting solution to moment-inequality based estimation.

suits our modelling purposes, since we propose procedures that are invariant to J:

$$W_t = \mathbf{G}_t^{1/2} Z_t, \quad \mathbf{G}_t = (1-\kappa_1-\kappa_2) I_n + \kappa_1 W_{t-1} W_{t-1}' + \kappa_2 \mathbf{G}_{t-1} \tag{2.8}$$

where Z_t are uncorrelated n-dimensional standard normal variables, so the conditional variance of JW_t is given by Σ_t with

$$\begin{aligned} \Sigma_t &= J\mathbf{G}_t J' = (1-\kappa_1-\kappa_2) JJ' + \kappa_1 J W_{t-1} W_{t-1}' J' + \kappa_2 J \mathbf{G}_{t-1} J' \\ &= (1-\kappa_1-\kappa_2) JJ' + \kappa_1 J W_{t-1} W_{t-1}' J' + \kappa_2 \Sigma_{t-1} \end{aligned} \tag{2.9}$$

which corresponds to the process considered by Engle, Sheppard and Sheppard (2008) and is a special case of the multivariate GARCH proposed by Engle and Kroner (1995). Assumption (2.6) is sufficiently general and includes various n-dimensional elliptically contoured distributions and skew-elliptical distributions. Special cases of (2.6) include the normal distribution

$$W_t = Z_t \overset{i.i.d}{\sim} N(0, I_n) \tag{2.10}$$

and the multivariate Student-t distribution with κ degrees-of freedom [denoted as $\mathtt{t}(\kappa)$].

Throughout the paper, we maintain the following notation. The unrestricted LS estimator of B and corresponding sum of squared errors are:

$$\hat{B} = (X'X)^{-1} X'Y\ , \quad S = \hat{U}'\hat{U}, \quad \hat{U} = Y - X\hat{B}. \tag{2.11}$$

In addition, we assume that X includes a constant regressor, so that

$$X = \left[\begin{array}{cc} \iota_T & \mathbf{X} \end{array}\right] \tag{2.12}$$

where $\mathbf{X}$ is $T \times q$. $s_j[i]$ denotes a j-dimensional selection vector with all elements equal to zero except for the i-th element which equals 1. $\mathcal{D}(d_1, \ \ldots \ , \ d_m)$ refers to an m-dimensional diagonal matrix with diagonal elements $d_1, \ \ldots \ , \ d_m$. DIAG(A) refers to a column vector from the diagonal of a matrix A. For any $N \times K$ matrix A, vec(A) returns an $NK \times 1$ vector, with the columns of A stacked on top of each other; $\mathcal{M}[A] = I - A(A'A)^{-1}A'$ for any full column rank matrix A. We refer to a $1-\alpha$ level CS for a parameter ϑ as $\mathrm{CS}_\alpha(\vartheta)$. For presentation ease, we partition the coefficient matrix as follows:

$$B = \left[\begin{array}{c} a' \\ b \end{array}\right], \quad a = (a_1, ..., a_n)'\,, \quad b = \left[\begin{array}{ccc} b_1 & ... & b_n \end{array}\right] = \left[\begin{array}{c} \beta_2' \\ \vdots \\ \beta_k' \end{array}\right] \tag{2.13}$$

where a is the vector of n intercepts, and b is $q \times n$. Conformably, we partition $\hat{B}$ as follows

$$\hat{B} = \begin{bmatrix} \hat{a}' \\ \hat{b} \end{bmatrix}, \quad \hat{b} = \begin{bmatrix} \hat{\beta}_2' \\ \vdots \\ \hat{\beta}_k' \end{bmatrix}. \tag{2.14}$$

We define a re-scaled residuals matrix $\bar{U}$ such that

$$\hat{U} = \mathcal{M}\,[X]\,Y = \mathcal{M}\,[X]\,WJ' = \bar{U}J', \quad \bar{U} = \mathcal{M}\,[X]\,W \tag{2.15}$$

or alternatively, if we denote by $\hat{U}_t'$ and $\bar{U}_t'$ the t-th row of $\hat{U}$ and $\bar{U}$, respectively, then $\hat{U}_t = J\bar{U}_t$. We also use the following matrix partitions:

$$Y = [\ y \quad \mathbf{Y}\ ] \tag{2.16}$$

where y is the first column of Y, and

$$(X'X)^{-1} = \begin{bmatrix} x^{11} & x^{12} \\ x^{21} & x^{22} \end{bmatrix} \tag{2.17}$$

where x^{11} is a scalar, $x^{21} = x^{12\prime}$ is $q \times 1$ and x^{22} is $q \times q$. Furthermore, we let

$$\vartheta = \text{vec}(B'), \quad \hat{\vartheta} = \text{vec}(\hat{B}'). \tag{2.18}$$

It is well known that $\hat{\vartheta}$ coincides with the GMM estimator and that a robust estimator for the associated asymptotic variance covariance matrix [Campbell et al. (1997, Chapter 5, section 5.6) and Ravikumar et al. (2000)] may be obtained as

$$\mathbf{V}_T = \frac{1}{T}\left[\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right]\mathbf{S}_T\left[\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right], \tag{2.19}$$

$$\mathbf{S}_T = \mathit{\Gamma}_{0,t} + \sum_{j=1}^{p}\left(\frac{p-j}{p}\right)\left[\mathbf{\Gamma}_{j,T} + \mathbf{\Gamma}_{j,T}'\right], \tag{2.20}$$

$$\mathbf{\Gamma}_{j,T} = \frac{1}{T}\sum_{t=j+1}^{T}\left(X_t \otimes \hat{U}_t\right)\left(X_{t-j} \otimes \hat{U}_{t-j}\right)' \tag{2.21}$$

for $j = 0, ..., p$ where p is a given lag length.

Finally, to derive and motivate our non-linear set estimation methods, we consider various linear constraints on (2.1), in particular:

$$\mathbf{H}\,(C, G, D) : CBG = D \text{ for known } C,\ G \text{ and } D \tag{2.22}$$

where C is $c \times k$ with rank c, $0 \leq c \leq k$, G is $n \times g$, with rank g, $0 \leq g \leq n$ and D the following special cases

$$\begin{aligned}
\mathbf{H}(C, I_n, D) &: CB = D, \text{ for known } C \text{ and } D, & (2.23)\\
\mathbf{H}(I_k, G, D) &: BG = D\,, \text{ for known } G \text{ and } D, & (2.24)\\
\mathbf{H}((1, \theta'), I_n, 0) &: (1, \theta')B = 0, \text{ for known } \theta, & (2.25)\\
\mathbf{H}(I_k, (1, \delta')', 0) &: B(1, \delta')' = 0, \text{ for known } \delta, & (2.26)\\
\mathbf{H}((1, \zeta'), I_n, \mu\iota_n') &: (1, \zeta')B = \mu\iota_n', \text{ for known } (\zeta, \mu)\,, & (2.27)\\
\mathbf{H}(I_k, (1, \phi')', \eta\iota_k) &: B(1, \phi')' = \eta\iota_k, \text{ for known } (\phi, \eta)\,, & (2.28)
\end{aligned}$$

where θ and ζ are $q \times 1$, δ and ϕ are $m \times 1$ and μ and η are scalars. Observe that (2.25) - (2.28) are the linear counterparts of (2.2) - (2.5).

3. Eigenvalue based and Hotelling-Type s tatistics

Consider hypothesis $\mathbf{H}(C, G, D)$ in (2.22). In this case, least squares estimators of B and corresponding sum of squared errors [denoted $\hat{B}_0$ and $\hat{S}_0$] are provided by Berndt and Savin (1977):

$$\hat{B}_0 = B_0(C, G, D)\,, \quad \hat{S}_0 = \hat{U}_0'\hat{U}_0 = S_0(C, G, D)\,, \quad \hat{U}_0 = Y - X\hat{B}_0$$

where for any given C, G and D, the functions $B_0(C, G, D)$ and $S_0(C, G, D)$ obtain as

$$B_0(C, G, D) = \hat{B} - (X'X)^{-1} C'[C(X'X)^{-1}C']^{-1}(C\hat{B}G - D)\tilde{S}^{-1}G'\hat{S}, \quad (3.1)$$

$$S_0(C, G, D) = \hat{S} + \hat{S}G\tilde{S}^{-1}(C\hat{B}G - D)'[C(X'X)^{-1}C']^{-1}(C\hat{B}G - D)\tilde{S}^{-1}G'\hat{S}, \quad (3.2)$$

$$\tilde{S} = G'\hat{S}G. \quad (3.3)$$

Commonly used statistics including the LR and Wald criteria [see Berndt and Savin (1977), Gouriéroux et al. (1995), Dufour and Khalaf (2002) and the references therein] to test $\mathbf{H}(C, G, D)$ can be expressed as

$$\mathcal{L}(C, G, D) = T\ln\left(|\hat{S}_0|/|\hat{S}|\right) = -T\sum_{i=1}^{l}\ln\left(1 - \lambda_i(C, G, D)\right) \quad (3.4)$$

$$\mathcal{W}(C, G, D) = T\,\mathrm{tr}\left(\hat{S}^{-1}[\hat{S}_0 - \hat{S}]\right) = T\sum_{i=1}^{l}\frac{\lambda_i(C, G, D)}{1 - \lambda_i(C, G, D)} \quad (3.5)$$

where $l = \min\{c,\, g\}$ and $\lambda_1(C, G, D) \geq \cdots \geq \lambda_n(C, G, D)$ are the eigenvalues of

$$\hat{\Delta}\,(C, G, D) = \hat{S}_0{}^{-1}[\hat{S}_0 - \hat{S}].$$

Clearly, $\lambda_i(C, G, D)$, $i = 1, ..., l$ coincide with the roots of $\tilde{S}_0^{-1}[\tilde{S}_0 - \tilde{S}]$ where

$$\tilde{S}_0 = \tilde{S} + (C\hat{B}G - D)'[C(X'X)^{-1}C']^{-1}(C\hat{B}G - D). \tag{3.6}$$

Solving for eigenvalues in question thus requires considering the determinantal equation

$$\left|(\tilde{S}_0 - \tilde{S}) - \lambda\tilde{S}_0\right| = 0. \tag{3.7}$$

Theorem 3.1 *In the context of* (2.1) *and under the null hypothesis* $\mathbf{H}\,(C, G, D)$ *in* (2.22)*, the vector of the roots of* (3.7) *is distributed like the vector of the roots of*

$$|\Gamma'W'\,(\mathcal{M}_0[X, C])\,W\Gamma - \lambda\Gamma'W'\,(\mathcal{M}[X] + \mathcal{M}_0[X, C])\,W\Gamma| = 0 \tag{3.8}$$

where Γ *is the orthogonal* $n \times g$ *matrix which includes the eigenvectors associated with the non-zero eigenvalues of* $J'GG'J$ *and*

$$\mathcal{M}_0[X, C] = X(X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C(X'X)^{-1}X'. \tag{3.9}$$

Furthermore, under assumption (2.6)*, the distribution in question follows that of the roots of*

$$|\mathcal{Z}'V'\,(\mathcal{M}_0[X, C])\,V\mathcal{Z} - \lambda\mathcal{Z}'V'\,(\mathcal{M}[X] + \mathcal{M}_0[X, C])\,V\mathcal{Z}| = 0 \tag{3.10}$$

where $\mathcal{Z}$ *is a* $T \times g$ *matrix of i.i.d. g-dimensional standard normal variables, and is thus invariant to* B *and* J*. For the special case where* $G = I_n$*, i.e. hypothesis* $\mathbf{H}\,(C, I_n, D)$ *in* (2.23)*, the distribution in question follows that of the roots of*

$$|W'\,(\mathcal{M}_0[X, C])\,W - \lambda W'\,(\mathcal{M}[X] + \mathcal{M}_0[X, C])\,W| = 0 \tag{3.11}$$

so invariance to B *and* J *holds imposing or ignoring assumption* (2.6)*.*

In general, under assumption (2.6), the distribution of the roots will depend on C but not on D, and depends on G only through its rank. Simulation-based p-values can be obtained using the Monte Carlo test method [see e.g. Dufour and Khalaf (2002) and Dufour (2006)] if the variates underlying V are simulable. Given normal errors and if

$\min(c,g) \leq 2$ [see Rao (1973, Chapter 8)] then

$$\left(\frac{\varkappa_1\varkappa_3 - 2\varkappa_2}{cg}\right)\frac{1-\left(|\hat{S}|/|\hat{S}_0|\right)^{1/\varkappa_3}}{\left(|\hat{S}|/|\hat{S}_0|\right)^{1/\varkappa_3}} \sim F(cg\,,\,\varkappa_1\varkappa_3 - 2\varkappa_2) \tag{3.12}$$

where $\varkappa_1 = T - k - ((g-c+1)/2)$, $\varkappa_2 = (cg-2)/4$, $\varkappa_3 = [(c^2g^2-4)/(c^2+g^2-5)]^{1/2}$, if $c^2+g^2-5 > 0$ or $\varkappa_3 = 1$ otherwise. The latter result holds as a reliable [relative to the standard χ^2] approximation when $\min(c,g) > 2$. The following approximation is also recommended [see McKeon (1974)]: $\mathrm{tr}\left(\hat{S}^{-1}[\hat{S}_0 - \hat{S}]\right)/\varkappa_6 \sim F(cg\,,\,\varkappa_5)$, where $\varkappa_4 = (T-k+c-g-1)(T-k-1)/[(T-k-g-3)(T-k-g)]$, $\varkappa_5 = 4 + (gc+2)/(\varkappa_4 - 1)$, $\varkappa_6 = (cg)(\varkappa_5 - 2)/[\varkappa_5(T-k-g-1)]$ and is exact [exactly coincides with (3.12)] for $\min(c,g) = 1$.

Theorem 3.2 *In the context of* (2.1) *and the null hypothesis* $\mathbf{H}(C, I_n, D)$ *in* (2.23)*, the LR criterion obtains as* $T\ln\left(\left|I_c + \bar{\Lambda}(C, I_n, D)\right|\right)$ *where*

$$\bar{\Lambda}(C, I_n, D) = [C(X'X)^{-1}C']^{-1}(C\hat{B} - D)\hat{S}^{-1}(C\hat{B} - D)'. \tag{3.13}$$

Alternatively, under the null hypothesis $\mathbf{H}(I_k, G, D)$ *in* (2.24)*, the LR criterion obtains as* $T\ln\left(\left|I_g + \bar{\Omega}(I_k, G, D)\right|\right)$ *where*

$$\bar{\Omega}(I_k, G, D) = [G'Y'\mathcal{M}[X]YG]^{-1}(\hat{B}G - D)'(X'X)(\hat{B}G - D). \tag{3.14}$$

Results in the next sections use Theorem (**3.2**) and the following special cases.

Under the null hypotheses $\mathbf{H}((1,\theta'), I_n, 0)$ in (2.25) and $\mathbf{H}((1,\zeta'), I_n, \mu\iota_n')$ (2.27), and applying (3.13), the LR criteria obtain (respectively) as

$$\mathcal{L}_\Lambda(\theta) = T\ln\left(1 + \Lambda(\theta)\right), \quad \mathcal{L}_{\Lambda_*}(\zeta,\mu) = T\ln\left(1 + \Lambda_*(\zeta,\mu)\right) \tag{3.15}$$

where considering the function $\bar{\Lambda}(.)$ in (3.13)

$$\Lambda(\theta) = \bar{\Lambda}((1,\theta'), I_n, 0) = \frac{(1,\theta')\hat{B}\hat{S}^{-1}\hat{B}'(1,\theta')'}{(1,\theta')(X'X)^{-1}(1,\theta')'}, \tag{3.16}$$

$$\Lambda_*(\zeta,\mu) = \bar{\Lambda}((1,\zeta'), I_n, \mu\iota_n') = \frac{\left((1,\zeta')\hat{B} - \mu\iota_n'\right)\hat{S}^{-1}\left(\hat{B}'(1,\zeta')' - \mu\iota_n\right)}{(1,\zeta')(X'X)^{-1}(1,\zeta')'}. \tag{3.17}$$

The associated Wald statistics obtain respectively as

$$\mathcal{W}_\Lambda(\theta) = T\Lambda(\theta), \quad \mathcal{W}_{\Lambda_*}(\zeta,\mu) = T\Lambda_*(\zeta,\mu). \tag{3.18}$$

Furthermore, the special case

$$\mathbf{H}\left(s_k[i]', I_n, 0\right) : s_k[i]'B = 0, \quad i \in \{1, ..., k\} \tag{3.19}$$

leads to the well known Hotelling statistic

$$\varLambda_i = \frac{s_k[i]'\hat{B}\hat{S}^{-1}\hat{B}'s_k[i]}{s_k[i]'(X'X)^{-1}s_k[i]}. \tag{3.20}$$

When errors are normal as in (2.10), then [from (3.12)] under (2.25), (2.27) and (3.19) respectively, where $\tau_n = T - k - n + 1$

$$\varLambda\left(\theta\right)\frac{\tau_n}{n} \sim F\left(n, \tau_n\right), \quad \varLambda_*\left(\zeta, \mu\right)\frac{\tau_n}{n} \sim F\left(n, \tau_n\right), \quad \varLambda_i\frac{\tau_n}{n} \sim F\left(n, \tau_n\right). \tag{3.21}$$

Alternatively, under the null hypotheses $\mathbf{H}\left(I_k, (1, \delta')', 0\right)$ in (2.26) and $\mathbf{H}\left(I_k, (1, \phi')', \eta\iota_k\right)$ in (2.28), and applying (3.14), the LR criteria obtain (respectively) as:

$$\mathcal{L}_\Omega(\delta) = T\ln\left(\left|1 + \varOmega\left(\delta\right)\right|\right), \quad \mathcal{L}_{\Omega_*}(\phi, \eta) = T\ln\left(\left|1 + \varOmega_*\left(\phi, \eta\right)\right|\right) \tag{3.22}$$

where considering the function $\bar{\varOmega}\left(.\right)$ in (3.14)

$$\varOmega\left(\delta\right) = \bar{\varOmega}\left(I_k, (1, \delta')', 0\right) = \frac{(1, \delta')Y'X(X'X)^{-1}X'Y(1, \delta')'}{(1, \delta')Y'\mathcal{M}[X]Y(1, \delta')'}, \tag{3.23}$$

$$\varOmega_*\left(\phi, \eta\right) = \bar{\varOmega}\left(I_k, (1, \phi')', \eta\iota_k\right) = \frac{\left((1, \phi')\hat{B}' - \eta\iota_k'\right)(X'X)\left(\hat{B}(1, \phi')' - \eta\iota_k\right)}{(1, \phi')Y'\mathcal{M}[X]Y(1, \phi')'}. \tag{3.24}$$

The associated Wald statistics obtain respectively as

$$\mathcal{W}_\Omega(\delta) = T\varOmega\left(\delta\right), \quad \mathcal{W}_{\Omega_*}(\phi, \eta) = T\varOmega_*\left(\phi, \eta\right). \tag{3.25}$$

In addition, the special case

$$\mathbf{H}\left(I_k, s_n[i], 0\right) : Bs_n[i] = 0, \quad i \in \{1, ..., n\} \tag{3.26}$$

leads to the usual univariate F-statistic

$$\varOmega_i = \frac{s_n[i]'Y'X(X'X)^{-1}X'Ys_n[i]}{s_n[i]'Y'\mathcal{M}[X]Ys_n[i]} \tag{3.27}$$

associated with the joint significance of all regressors in the ith equation of (2.1). When errors are normal as in (2.10), then [from (3.12)] under (2.26), (2.28) and (3.26) respectively, where $\tau_k = T - k$

$$\varOmega(\delta)\frac{\tau_k}{k} \sim F(k,\tau_k), \quad \varOmega_*(\phi,\eta)\frac{\tau_k}{k} \sim F(k,\tau_k), \quad \varOmega_i\frac{\tau_k}{k} \sim F(k,\tau_k). \tag{3.28}$$

In addition to the above statistics, a Wald-type criterion for the linear hypothesis $\mathbf{R}\vartheta = 0$ where $\mathbf{R}$ is an $r \times (kn)$ matrix can be derived from the robust variance/covariance estimator of $\hat{\vartheta}$ [refer to (2.19)-(2.21)] as

$$\mathcal{W}_p = T\,\hat{\vartheta}'\mathbf{R}'\left[\mathbf{R}\left(\left(\frac{X'X}{T}\right)^{-1}\otimes I_n\right)\mathbf{S}_T\left(\left(\frac{X'X}{T}\right)^{-1}\otimes I_n\right)\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\vartheta}, \tag{3.29}$$

where $\mathbf{S}_T$ is given by (2.20) and p refers to the number of lags underlying (2.20).[6] For the case where $\mathbf{R} = C \otimes I_n$ which corresponds to $\mathbf{H}(C, I_n, 0)$ [(2.23) with $D = 0$], the following invariance result holds.

Theorem 3.3 *In the context of* (2.1) *and the uniform linear hypothesis* $\mathbf{H}(C, I_n, 0)$ [(2.23) *with* $D = 0$], *the robust Wald statistic* (3.29) *is distributed, under the null hypothesis following*

$$\tilde{\mathcal{W}}_p = \tilde{\vartheta}'(C' \otimes I_n)\left\{(C \otimes I_n)\hat{\mathbf{V}}_T(C' \otimes I_n)\right\}^{-1}(C \otimes I_n)\tilde{\vartheta} \tag{3.30}$$

where $\tilde{\vartheta} = \mathrm{vec}\big(\tilde{B}'\big) = \mathrm{vec}\big[W'X(X'X)^{-1}\big], \bar{U}_t'$ *the* t*-th row of* $\bar{U}$, *and* $\bar{U}$ *is defined in* (2.15).

$$\hat{\mathbf{V}}_T = \frac{1}{T}\left[\left(\frac{X'X}{T}\right)^{-1}\otimes I_n\right]\hat{\mathbf{S}}_T\left[\left(\frac{X'X}{T}\right)^{-1}\otimes I_n\right] \tag{3.31}$$

$$\hat{\mathbf{S}}_T = \hat{\varGamma}_{0,T} + \sum_{j=1}^{p}\left(\frac{p-j}{p}\right)\left[\hat{\mathbf{\Gamma}}_{j,T} + \hat{\mathbf{\Gamma}}'_{j,T}\right], \tag{3.32}$$

$$\hat{\mathbf{\Gamma}}_{j,T} = \frac{1}{T}\sum_{t=j+1}^{T}\left(X_t \otimes \bar{U}_t\right)\left(X_{t-j} \otimes \bar{U}_{t-j}\right)'. \tag{3.33}$$

Since both $\tilde{\vartheta}$ and $\hat{\mathbf{V}}_T$ have distributions which do not depend on B or J, B and J are completely evacuated from the null distribution of the statistic $\mathcal{W}_p$. This expression allows to derive convenient simulation-based test procedures for non-i.i.d. time series specifications for vec(W), including *e.g.* multivariate GARCH.

[6] The non-zero constraint case derives straightforwardly from our analysis. We consider zero-restrictions for presentation clarity.

4. Minimum distance estimation

Consider the structural models defined by hypotheses $\mathcal{H}_1\left(\bar{\theta}\right)$, $\mathsf{H}_1\left(\bar{\delta}\right)$, $\mathcal{H}_2\left(\bar{\zeta},\bar{\mu}\right)$ and $\mathsf{H}_2\left(\bar{\phi},\bar{\eta}\right)$ in (2.2) - (2.5). We first provide analytical expressions for associated test statistics, using fundamental results from the RR regression literature [references are cited in section 1]; note that analytical solutions are not readily available in the case of (2.4).

The LR statistics associated with $\mathcal{H}_1\left(\bar{\theta}\right)$, $\mathsf{H}_1\left(\bar{\delta}\right)$, $\mathcal{H}_2\left(\bar{\zeta},\bar{\mu}\right)$ and $\mathsf{H}_2\left(\bar{\phi},\bar{\eta}\right)$ are (respectively):

$$\begin{aligned} \mathcal{L}_{\Lambda}(\hat{\vartheta}) &= \min_{\theta}\mathcal{L}_{\Lambda}(\theta), \quad \mathcal{L}_{\Lambda_*}(\hat{\zeta},\hat{\mu}) = \min_{\zeta,\mu}\mathcal{L}_{\Lambda_*}(\zeta,\mu), &(4.1)\\ \mathcal{L}_{\Omega}(\hat{\delta}) &= \min_{\delta}\mathcal{L}_{\Omega}(\delta), \quad \mathcal{L}_{\Omega_*}(\hat{\phi},\hat{\eta}) = \min_{\phi,\eta}\mathcal{L}_{\Omega_*}(\phi,\eta) &(4.2)\end{aligned}$$

where $\mathcal{L}_{\Lambda}(\theta)$, $\mathcal{L}_{\Lambda_*}(\zeta,\mu)$, $\mathcal{L}_{\Omega}(\delta), \mathcal{L}_{\Omega_*}(\phi,\eta)$ are as defined in (3.15) and (3.22), and $\hat{\theta}$, $\hat{\delta}$, $(\hat{\zeta},\hat{\mu})$ and $(\hat{\phi},\hat{\eta})$ are the corresponding extremum estimators. Conformably, the Wald statistics are (respectively):

$$\begin{aligned} \mathcal{W}_{\Lambda}(\hat{\theta}) &= \min_{\theta}\mathcal{W}_{\Lambda}(\theta), \quad \mathcal{W}_{\Lambda_*}(\hat{\zeta},\hat{\mu}) = \min_{\zeta,\mu}\mathcal{W}_{\Lambda_*}(\zeta,\mu)\\ \mathcal{W}_{\Omega}(\hat{\delta}) &= \min_{\delta}\mathcal{W}_{\Omega}(\delta), \quad \mathcal{W}_{\Omega_*}(\hat{\phi},\hat{\eta}) = \min_{\phi,\eta}\mathcal{W}_{\Omega_*}(\phi,\eta)\end{aligned}$$

where $\mathcal{W}_{\Lambda}(\theta)$, $\mathcal{W}_{\Lambda_*}(\zeta,\mu)$, $\mathcal{W}_{\Omega}(\delta)$ and $\mathcal{W}_{\Omega_*}(\phi,\eta)$ are as defined in (3.18) and (3.25). These definitions imply that in order to to derive the extremum estimators for the four problems under consideration, it suffices to minimize $\Lambda\left(\theta\right)$, $\Lambda_*\left(\zeta,\mu\right)$, $\Omega\left(\delta\right)$ and $\Omega_*\left(\phi,\eta\right)$, respectively. From the references cited above, the following result obtains:

$$\min_{\theta}\Lambda\left(\theta\right) = \min_{\delta}\Omega\left(\delta\right) = \hat{\rho}/(1-\hat{\rho}) \equiv \hat{\gamma} \tag{4.3}$$

where $\hat{\rho}$ is the minimum non-zero root of

$$\hat{\Delta} = (X'X)^{-1}X'Y(Y'Y)^{-1}Y'X \tag{4.4}$$

leading to

$$\mathcal{L}_{\Lambda}(\hat{\theta}) = \mathcal{L}_{\Omega}(\hat{\delta}) = -T\ln(1-\hat{\rho}), \quad \mathcal{W}_{\Lambda}(\hat{\theta}) = \mathcal{W}_{\Omega}(\hat{\delta}) = T\hat{\gamma}. \tag{4.5}$$

Observe that $\hat{\gamma}$ coincides with the minimum root of both determinantal equations

$$\begin{aligned} \left|\hat{B}\hat{S}^{-1}\hat{B}' - \gamma(X'X)^{-1}\right| &= 0 &(4.6)\\ \left|\hat{B}'\left(X'X\right)\hat{B} - \gamma\hat{S}\right| &= 0. &(4.7)\end{aligned}$$

Theorem 4.1 *In the context of* (2.1) *and the non-linear hypothesis* $\mathcal{H}_2\left(\bar{\zeta}, \bar{\mu}\right)$ *in* (2.4), *the LR criterion obtains as*

$$\mathcal{L}_{\Lambda_*}(\hat{\zeta}, \hat{\mu}) = T \ln\left(1 + \hat{\nu}\right) \Leftrightarrow \hat{\nu} = \min_{\zeta, \mu} \Lambda_*(\zeta, \mu),$$

where $\Lambda_*(\zeta, \mu)$ *is defined in* (3.18) *and* $\hat{\nu}$ *is the minimum root of*

$$\left| \hat{B}\left(\hat{S}^{-1} - \hat{S}^{-1}\iota_n\left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1}\iota_n'\hat{S}^{-1}\right)\hat{B}' - \nu(X'X)^{-1} \right| = 0. \tag{4.8}$$

Alternatively, the LR criterion associated with $\mathsf{H}_2\left(\bar{\phi}, \bar{\eta}\right)$ *in* (2.5) *obtains as*

$$\mathcal{L}_{\Omega_*}(\hat{\phi}, \hat{\eta}) = T \ln\left(1 + \hat{\sigma}\right) \Leftrightarrow \hat{\sigma} \min_{\phi, \eta} \Omega_*(\phi, \eta)$$

where $\Omega_*(\phi, \eta)$ *is defined in* (3.25) *and* $\hat{\sigma}$ *is the minimum root of*

$$\left| \hat{B}'\left(X'X - (X'X)\iota_k\left(\iota_k'(X'X)\iota_k\right)^{-1}\iota_k'(X'X)\right)\hat{B} - \sigma\hat{S} \right| = 0. \tag{4.9}$$

Unique point estimators for $\bar{\theta}$, $\bar{\delta}$, $(\bar{\zeta}, \bar{\mu})$ and $(\bar{\phi}, \bar{\eta})$ can also be obtained analytically. We provide, in Theorem **4.2** below, simple expressions which exploit a general solution [provided in the Appendix] for the following vector equation in an m-dimensional vector π given an $(m+1) \times (m+1)$ matrix Σ, of the form:

$$\Sigma(1, \pi')' = 0, \qquad |\Sigma| = 0, \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{4.10}$$

where Σ_{11} is a scalar, $\Sigma_{12} = \Sigma_{21}'$ is $1 \times m$ and Σ_{22} is $m \times m$ and is invertible. Indeed,

$$\pi = -\Sigma_{22}^{-1}\Sigma_{21} \tag{4.11}$$

provides a unique solution to this system. Such an approach has been used in econometrics in *e.g.* limited information simultaneous equations or cointegration models; see Dhrymes (1974, Chapter 7), Davidson and MacKinnon (2004, Chapter 12), or Johansen (1995). To the best of our knowledge, a unified treatment of all point estimation problems given the specific normalizations we adopt here is not readily available, particularly in the case of (2.4) and (2.5). Actually, even in the case of (2.2), iterative estimation methods are often still applied [see e.g. Barone-Adesi, Gagliardini and Urga (2004*b*)].

Theorem 4.2 *In the context of* (2.1), *the MLE of* $\bar{\theta}$, *as defined by* $\mathcal{H}_1\left(\bar{\theta}\right)$ [*hypothesis*

(2.2)], *and of* $\bar{\delta}$*, as defined by* $\mathsf{H}_1\left(\bar{\delta}\right)$ [*hypothesis* (2.3)]*, are* (*respectively*)*:*

$$\hat{\theta} = \left(\hat{b}\hat{S}^{-1}\hat{b}' - \hat{\gamma}x^{22}\right)^{-1}\left(\hat{b}\hat{S}^{-1}\hat{a} - \hat{\gamma}x^{21}\right) \quad (4.12)$$

$$\hat{\delta} = \left(\mathbf{Y}'X(X'X)^{-1}X'\mathbf{Y} - \hat{\gamma}\mathbf{Y}'\mathcal{M}[X]\mathbf{Y}\right)^{-1}\left(\mathbf{Y}'X(X'X)^{-1}X'y - \hat{\gamma}\mathbf{Y}'\mathcal{M}[X]y\right) \quad (4.13)$$

where $\hat{\gamma}$ *is defined by* (4.3)*. Furthermore, the MLE of* $\bar{\zeta}$ *and* $\bar{\mu}$*, as defined by* $\mathcal{H}_2\left(\bar{\zeta},\bar{\mu}\right)$ [*hypothesis* (2.4)]*, and of* $\bar{\phi}$ *and* $\bar{\eta}$*, as defined by* $\mathsf{H}_2\left(\bar{\phi},\bar{\eta}\right)$ [*hypothesis* (2.5)]*, are* (*respectively*)*:*

$$\hat{\zeta} = \left(\hat{b}\left(\hat{S}^{-1} - \hat{S}^{-1}\iota_n\left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1}\iota_n'\hat{S}^{-1}\right)\hat{b}' - \hat{\nu}x^{22}\right)^{-1} \left(\hat{b}\left(\hat{S}^{-1} - \hat{S}^{-1}\iota_n\left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1}\iota_n'\hat{S}^{-1}\right)\hat{a} - \hat{\nu}x^{21}\right)$$

$$\hat{\phi} = \left(\mathbf{Y}'X\left(I - \iota_k\left(\iota_k'(X'X)\iota_k\right)^{-1}\iota_k'\right)X'\mathbf{Y} - \hat{\sigma}\mathbf{Y}'\mathcal{M}[X]\mathbf{Y}\right)^{-1} \left(\mathbf{Y}'X\left(I - \iota_k\left(\iota_k'(X'X)\iota_k\right)^{-1}\iota_k'\right)X'y - \hat{\sigma}\mathbf{Y}'\mathcal{M}[X]y\right)$$

where $\hat{\nu}$ *and* $\hat{\sigma}$ *are* (*respectively*) *the minimum roots of* (4.8) *and* (4.9)*. Conformably, the MLE of* μ *and* η *obtain, respectively, as*

$$\hat{\mu} = \frac{(1,\hat{\zeta}')\hat{B}\hat{S}^{-1}\iota_n}{\iota_n'\hat{S}^{-1}\iota_n}, \quad \hat{\eta} = \frac{(1,\hat{\phi}')\hat{B}'(X'X)\iota_k}{\iota_k'(X'X)\iota_k}. \quad (4.14)$$

Given the above derived estimates of $\bar{\theta}$, $\bar{\delta}$, $(\bar{\zeta},\bar{\mu})$ and $(\bar{\phi},\bar{\eta})$, constrained estimates [imposing, in turn, $\mathcal{H}_1\left(\bar{\theta}\right)$, $\mathsf{H}_1\left(\bar{\delta}\right)$, $\mathcal{H}_2\left(\bar{\zeta},\bar{\mu}\right)$ and $\mathsf{H}_2\left(\bar{\phi},\bar{\eta}\right)$, in (2.2) - (2.5)] for the regression parameters [denoted $\hat{B}_{\Lambda}(\hat{\theta})$, $\hat{B}_{\Lambda_*}(\hat{\zeta},\hat{\mu})$, $\hat{B}_{\Omega}(\hat{\delta})$, $\hat{B}_{\Omega_*}(\hat{\phi},\hat{\eta})$] may be obtained by plugging $\hat{\theta}$, $\hat{\delta}$, $(\hat{\zeta},\hat{\mu})$ or $(\hat{\phi},\hat{\eta})$ (respectively) into the function (3.1):

$$\hat{B}_{\Lambda}(\hat{\theta}) = B_0\left((1,\hat{\theta}'),I_n,0\right), \quad \hat{B}_{\Lambda_*}(\hat{\zeta},\hat{\mu}) = B_0\left((1,\hat{\zeta}'),I_n,\hat{\mu}\iota_n'\right), \quad (4.15)$$

$$\hat{B}_{\Omega}(\hat{\delta}) = B_0\left(I_k,(1,\hat{\delta}')',0\right), \quad \hat{B}_{\Omega_*}(\hat{\phi},\hat{\eta}) = B_0\left(I_k,(1,\hat{\phi}')',\hat{\eta}\iota_k\right). \quad (4.16)$$

Finally, observe that minimizing the special case of (3.29) with $C = (1,\theta')$ over θ yields a continuous updated GMM test statistic for hypothesis (2.25). Given the mounting evidence which documents the poor performance of this test with usual χ^2 cut-off points [see *e.g.* Ray and Savin (2008), Dufour, Khalaf and Beaulieu (2008), and Gungor and Luger (2008)], we consider simulation-based alternatives in the following sections.

5. Inverting Hotelling-type tests

For the structrual models defined, in turn, by hypotheses $\mathcal{H}_1\left(\bar{\theta}\right)$, $\mathsf{H}_1\left(\bar{\delta}\right)$, $\mathcal{H}_2\left(\bar{\zeta},\bar{\mu}\right)$ and $\mathsf{H}_2\left(\bar{\phi},\bar{\eta}\right)$ in (2.2) - (2.5), this section discusses identification-robust CS estimation. CSs are obtained, respectively, by inverting the tests in (3.21) and (3.28). We derive an analytical solution to these problems as follows. We formulate each problem as an inequation of the form

$$(1,\pi')A(1,\pi')' \leq 0 \tag{5.1}$$

where π is the $m \times 1$ vector of unknown parameters and A is an $(m+1) \times (m+1)$ data dependent matrix. The solution for this inequality [which builds on the mathematics of quadrics as introduced by Dufour and Taamouti (2005, 2007)] and projection based solutions for each component of π and any linear transformation of latter of the form $\omega'\pi$ where ω is a non-zero $m \times 1$ vector, is provided in the Appendix. Our results require partitioning A as follows, where A_{11} is a scalar, A_{22} is $m \times m$, and $A_{12} = A_{21}'$ is $1 \times m$:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \tag{5.2}$$

So we proceed next to show that our problems of interest can all be cast as in (5.1).

Inverting the test defined by (3.16) and (3.21) leads to the following inequality in θ:

$$\frac{(1,\theta')\hat{B}\hat{S}^{-1}\hat{B}'(1,\theta')'}{(1,\theta')(X'X)^{-1}(1,\theta')'}\frac{\tau_n}{n} \leq f_{n,\tau_n,\alpha}$$

where $f_{n,\tau_n,\alpha}$ denotes the α-level cut off point from the $F\left(n,\tau_n\right)$ distribution. This falls into the (5.1) form with $\pi = \theta$ and

$$A = \hat{B}\hat{S}^{-1}\hat{B}' - (X'X)^{-1}f_{n,\tau_n,\alpha}\left(n/\tau_n\right). \tag{5.3}$$

Using the partitionings (2.14) and (2.17), $A_{11} = \hat{a}'\hat{S}^{-1}\hat{a} - \left(f_{n,\tau_n,\alpha}\left(n/\tau_n\right)\right)x^{11}$, $A_{12} = A_{21}' = \hat{a}'\hat{S}^{-1}\hat{b}' - \left(f_{n,\tau_n,\alpha}\left(n/\tau_n\right)\right)x^{12}$ and

$$A_{22} = \hat{b}\hat{S}^{-1}\hat{b}' - \left(f_{n,\tau_n,\alpha}\left(n/\tau_n\right)\right)x^{22}. \tag{5.4}$$

The diagonal elements of A_{22} can be linked to a series of joint tests on the significance of each factor in (2.1). As may be checked from Dufour and Taamouti (2005, 2007), an unbounded solution to the inversion problem based on (5.1) would occur if A_{22} is not positive definite. Theorem **5.1** below characterizes the latter situation as it applies to the problem at hand. The latter characterization also holds when inverting the test defined in (3.17) and (3.21). In this case, we are led to the following inequality in ζ and μ:

$$\frac{\left((1,\zeta')\hat{B}-\mu\iota_n'\right)\hat{S}^{-1}\left(\hat{B}'(1,\zeta')'-\mu\iota_n\right)}{(1,\zeta')(X'X)^{-1}(1,\zeta')'}\frac{\tau_n}{n}\leq f_{n,\tau_n,\alpha}$$

which takes the (5.1) form with $\pi=(\zeta',\mu)'$ and

$$A=\left[\begin{array}{cc}\hat{B}\hat{S}^{-1}\hat{B}'-(X'X)^{-1}f_{n,\tau_n,\alpha}\frac{n}{\tau_n} & -\hat{B}\hat{S}^{-1}\iota_n\\ -\iota_n'\hat{S}^{-1}\hat{B}' & \iota_n'\hat{S}^{-1}\iota_n\end{array}\right]. \tag{5.5}$$

Theorem 5.1 *In the context of* (2.1), *if*

$$(\tau_n/n)\,\Lambda_i<f_{n,\tau_n,\alpha},\quad i\in\{2,...,k\} \tag{5.6}$$

where Λ_i are the Hotelling statistics defined in (3.20)*, then the CS for $\bar{\theta}$ [as defined in $\mathcal{H}_1\left(\bar{\theta}\right)$, in* (2.2)*] which inverts the statistic* (3.16) *is unbounded.*

Inverting the test defined by (3.23) and (3.28) leads to the following inequality in δ:

$$\frac{(1,\delta')Y'X(X'X)^{-1}X'Y\ (1,\delta')'}{(1,\delta')Y'\mathcal{M}[X]Y(1,\delta')'}\frac{\tau_k}{k}\leq f_{k,\tau_k,\alpha}$$

where $f_{k,\tau_k,\alpha}$ denotes the α-level cut off point from the $F\left(k,\tau_k\right)$ distribution. This falls into the (5.1) form with $\pi=\delta$ and

$$A=Y'X(X'X)^{-1}X'Y-Y'\mathcal{M}[X]Y f_{k,\tau_k,\alpha}\frac{k}{\tau_k}. \tag{5.7}$$

Using the partitioning (2.16), $A_{11}=y'X(X'X)^{-1}X'y-\left(f_{k,\tau_k,\alpha}\left(k/\tau_k\right)\right)y'\mathcal{M}[X]y$, $A_{12}=A_{21}'=y'X(X'X)^{-1}X'y-\left(f_{k,\tau_k,\alpha}\left(k/\tau_k\right)\right)y'\mathcal{M}[X]\mathbf{Y}$ and

$$A_{22}=\mathbf{Y}'X(X'X)^{-1}X'\mathbf{Y}-\left(f_{k,\tau_k,\alpha}\left(k/\tau_k\right)\right)\mathbf{Y}'\mathcal{M}[X]\mathbf{Y}. \tag{5.8}$$

The diagonal elements of A_{22} can be linked to a series of tests which assess the significance of each univariate regression for all components of $\mathbf{Y}$. We characterize in Theorem **5.2** a necessary [although not sufficient] condition which allows to check whether (5.8) is positive definite. The same condition also holds when inverting the test defined in (3.24) and (3.21). In this case, we are led to the following inequality in ϕ and η:

$$\frac{\left((1,\phi')\hat{B}'-\eta\iota_k'\right)(X'X)\left(\hat{B}(1,\phi')'-\eta\iota_k\right)}{(1,\phi')Y'\mathcal{M}[X]Y(1,\phi')'}\frac{\tau_k}{k}\leq f_{k,\tau_k,\alpha}$$

which takes the (5.1) form with $\pi = (\phi', \eta)'$ and

$$A = \begin{bmatrix} Y'X(X'X)^{-1}X'Y - Y'\mathcal{M}[X]Yf_{k,\tau_k,\alpha}\frac{k}{\tau_k} & -\hat{B}'\iota_k \\ -\iota_k'\hat{B} & \iota_k'(X'X)^{-1}\iota_k \end{bmatrix}. \tag{5.9}$$

Theorem 5.2 *In the context of* (2.1), *if*

$$(\tau_k/k)\Omega_i < f_{k,\tau_k,\alpha}, \quad i \in \{2, ..., n\} \tag{5.10}$$

where Ω_i *are the Hotelling statistics defined in* (3.27), *then the CS for* $\bar{\delta}$ [*as defined in* $\mathsf{H}_1\left(\bar{\delta}\right)$ *in* (2.3)] *which inverts the statistics* (3.23) *is unbounded.*

The conditions in Theorem **5.1** and **5.2** also apply (as may be inferred from the proofs provided in the Appendix) for inference on $(\bar{\zeta}', \bar{\mu})'$ [defined by $\mathcal{H}_2\left(\bar{\zeta}, \bar{\mu}\right)$ in (2.4)] and on $(\bar{\phi}', \bar{\eta})'$ [defined by $\mathsf{H}_2\left(\bar{\phi}, \bar{\eta}\right)$ in (2.5)]. These results illustrate the implications of "redundant" regressors on the precision of RR regression estimation. We proceed next to characterize the conditions underlying empty CSs for the problems considered. Such characterizations lead to LR-based bound J-type test procedures.

Theorem 5.3 *In the context of* (2.1) *and* (2.10), *let* $\text{CS}_\alpha(\bar{\theta})$, $\text{CS}_\alpha((\bar{\zeta}', \bar{\mu})')$, $\text{CS}_\alpha(\bar{\delta})$ *and* $\text{CS}_\alpha((\bar{\phi}', \bar{\eta})')$ *denote, respectively, the CSs for* $\bar{\theta}$ [*defined by* $\mathcal{H}_1\left(\bar{\theta}\right)$ *in* (2.2)], *the CS for* $\bar{\zeta}$ *and* $\bar{\mu}$ [*defined by* $\mathsf{H}_2\left(\bar{\phi}, \bar{\eta}\right)$ *in* (2.4)], *the CS for* $\bar{\delta}$ [*defined by* $\mathsf{H}_1\left(\bar{\delta}\right)$ *in* (2.3)], *and the CS for* $\bar{\phi}$ *and* $\bar{\eta}$ [*defined by* $\mathsf{H}_2\left(\bar{\phi}, \bar{\eta}\right)$ *in* (2.5)], *which invert* (*respectively*) *the statistics* (3.16) *and* (3.17) *given the distributions in* (3.21), *and the statistics* (3.23) *and* (3.24) *given the distributions in* (3.28). *Then*

$$\begin{aligned} \text{CS}_\alpha(\bar{\theta}) &= \varnothing \Leftrightarrow (\tau_n/n)\,\hat{\gamma} \geq f_{n,\tau_n,\alpha}, \quad \text{CS}_\alpha((\bar{\zeta}', \bar{\mu})') = \varnothing \Leftrightarrow (\tau_n/n)\,\hat{\nu} \geq f_{n,\tau_n,\alpha}, &(5.11) \\ \text{CS}_\alpha(\bar{\delta}) &= \varnothing \Leftrightarrow (\tau_k/k)\,\hat{\gamma} \geq f_{k,\tau_k,\alpha}, \quad \text{CS}_\alpha((\bar{\phi}', \bar{\eta})') = \varnothing \Leftrightarrow (\tau_k/k)\,\hat{\sigma} \geq f_{k,\tau_k,\alpha}, &(5.12) \end{aligned}$$

where $\hat{\gamma}$ *is defined by* (4.3), $\hat{\nu}$ *and* $\hat{\sigma}$ *are the minimum roots of* (4.8) *and* (4.9) *respectively.*

For completion, we consider next the case of inverting the Hotelling test associated with the hypothesis $s_k[i]'B = \bar{\varphi}', \quad i \in \{1, ..., k\}$ where $\bar{\varphi}$ is a $n \times 1$ vector of unknown parameters, which focuses on the coefficients of a single regressor in all equations of the system. To the best of our knowledge, an explicit discussion of this problem has escaped notice. The problem amounts to solving the following inequality

$$\frac{(\hat{\beta}_i - \varphi)'\hat{S}^{-1}(\hat{\beta}_i - \varphi)}{s_k[i]'(X'X)^{-1}s_k[i]}\frac{\tau_n}{n} \leq f_{n,\tau_n,\alpha} \tag{5.13}$$

which takes the quadric form (A.3) with $A_{22} = \hat{S}^{-1}$, $A_{12} = -\hat{\beta}_i{}'\hat{S}^{-1}$, $A_{11} =$

$-n\left(s_k[i]'(X'X)^{-1}s_k[i]\right)/\left(\tau_n\right)$. Then projection based CSs for the components of $\bar{\varphi}$ proceeds as in the Appendix. It is straightforward to see that inverting the LR test for a hypothesis of the form $Bs_n[i] = \bar{\psi}, \quad i \in \{1, ..., n\}$, where $\bar{\psi}$ is a $k \times 1$ vector of unknown parameters leads to inverting the usual univariate linear regression based F-test; on this problem, refer to e.g. Savin (1984) and the references therein.

The above results rely heavily on the fact that the same **F**-critical point is used to test all values of the parameters under considerations. This occurs with Gaussian errors, but not necessarily if errors are not Gaussian. Formally, the CSs as derived remain approximately valid as long as the **F**-distributions in (3.21) and (3.28) provide a reliable approximation to the null distribution of the Hotelling tests inverted [refer to section 3, Rao (1973, Chapter 8) or McKeon (1974)]. Furthermore, Theorem **3.1** provides distributional results which can be exploited to obtain exact CSs allowing for possibly non-Gaussian disturbances. The problem can be tackled by inverting an α-level test performed by simulation as a Monte Carlo (**MC**) test [see Dufour and Khalaf (2002) and Dufour (2006)], which will require numerical methods. Suppose for example that W_t are *i.i.d.* following an n-dimensional $\mathtt{t}(\kappa)$ distribution. Then the null distributions of the Hotelling statistics can be simulated using the pivotal expressions: (i) (3.11) in the case of (2.25) or (2.27), and (ii) (3.10) in the case of (2.26) and (2.28). This requires, respectively, draws from: (i) an n-dimensional $\mathtt{t}(\kappa)$ distribution, or (ii) a univariate $\mathtt{t}(\kappa)$ distribution. In the latter case, the resulting distributions are the same for all values of δ, ϕ or η under test. In the former case, expression (3.11) implies that the distributions in question will not depend on μ, yet θ and ζ intervene via the terms $\mathcal{M}[X, (1, \theta')]$ and $\mathcal{M}[X, (1, \zeta')]$. The only nuisance parameter that needs to be dealt with is κ; maximized Monte Carlo (**MMC**) strategies [where the simulation based p-value is maximized over a relevant choice set for κ; see Dufour (2006) or Beaulieu, Dufour and Khalaf (2007)] may be followed to obtain exact p-values where numerical costs are minimal since κ is a scalar.

We thus see that if error distributions that are simulable given a parsimonious number of nuisance parameters are considered, then exact CSs can be obtained numerically with relative ease, depending on the dimension of θ, ζ, δ or ϕ. Furthermore, a simulation-based version of the robust Hotelling test based on (3.29) with $C = (1, \theta')$ can also be inverted using Theorem **3.3** when a parsimonious process is considered to model time dependence. In view of our empirical analysis, consider the multivariate GARCH from (2.8) as an illustrative case. Using the pivotal expression (3.30) and draws form a multivariate standard normal variate, the null distribution of the robust Hotelling statistic can easily be simulated for any θ under test, given κ_1 and κ_2. Again, maximized Monte Carlo strategies where the simulation based p-value is maximized over a relevant choice set for κ_1 and κ_2 may be applied with relative ease. In this regard, we concur with Engle et al. (2008) on the merits of parsimonious multivariate GARCH modelling.

Checking whether the CSs are empty also follows straightforwardly, by referring the minimum Hotelling statistics to their simulated cut-off points. For this exercise, the simulation in the case of (2.25) or (2.27) can be run by fixing θ and ζ to the value which minimizes the associated statistic, here $\hat{\theta}$ and $\hat{\zeta}$ respectively [refer to Theorem **4.2**]; if the associated p-value exceeds the desired level, then one can be sure that the confidence set is not empty. To obtain decisive inference in this regard, and because of dependence on θ and ζ, the p-value should be maximized with respect to these parameters, which may be numerically less expensive that the construction of the confidence set (mainly because, depending on the dimension of the problem, global minimizing software may be more efficient than e.g. grid-based test inversion). We study exact distributions of minimum Hotelling statistics in the next Section.

6. Distributions of minimum-distance criteria

In this section, we study the distribution of the test statistics associated with $\mathcal{H}_1\left(\bar{\theta}\right)$ and $\mathsf{H}_1\left(\bar{\delta}\right)$ [hypotheses (2.2), (2.3)]. In particular, we show that these distributions depend on B and J only through a lower-dimensional function of B and J, even when the null hypotheses do not hold. We also narrow down the number of free parameters under each null hypothesis and write the statistics as tractable functions of these parameters, and of the error term W. This is achieved via invariance-based reductions, as shown below.

The Gaussian log-likelihood function for model (2.1) is

$$\boldsymbol{L}(Y, B\,, \varSigma) = -\frac{T}{2}[n(2\pi) + \ln(|\varSigma|)] - \frac{1}{2}\mathrm{tr}[\varSigma^{-1}(Y - XB)'(Y - XB)].$$

Setting $\bar{\varSigma}(B) \equiv \frac{1}{T}(Y - XB)'(Y - XB)$, for any given value of B, $\boldsymbol{L}(Y, B\,, \varSigma)$ is maximized by taking $\varSigma = \bar{\varSigma}(B)$ yielding the concentrated log-likelihood

$$\bar{\boldsymbol{L}}(Y, B\,, \varSigma) = -\frac{nT}{2}[(2\pi) + 1] - \frac{T}{2}\ln(|\bar{\varSigma}(B)|)\,. \tag{6.14}$$

The Gaussian ML estimator of B thus minimizes $|\bar{\varSigma}(B)|$ with respect to B. In this context, the unrestricted ML estimator of B and associated sum-of-squared residuals are $\hat{B}$ and $\hat{S}$ [as defined in (2.11)]; furthermore, given a restriction function of the form $\mathbf{H}(C, G, D)$ [for any C, G and D as in (2.22)], the restricted estimator of B and associated sum of squares are $B_0\left(C, G, D\right)$ and $S_0\left(C, G, D\right)$, as defined in (3.1)-(3.2); for presentation clarity, $B_0\left(.\right)$ and $S_0\left(.\right)$ may be viewed as functions of C, G, D and the data.

Theorem 6.1 *The LR statistics* $\mathcal{L}_{\Lambda}(\theta)$ *and* $\mathcal{L}_{\Lambda}(\hat{\theta}) = \min_{\theta}\mathcal{L}_{\Lambda}(\theta)$ *associated* [*as in* (3.16)] *with* $\mathcal{H}_1(\bar{\theta})$ (2.2) *are invariant to replacing* Y *by*

$$Y_{\top} = YM \tag{6.15}$$

where M *is an arbitrary nonsingular* $n \times n$ *matrix. The LR statistics* $\mathcal{L}_{\Omega}(\delta)$ *and* $\mathcal{L}_{\Omega}(\hat{\delta}) = \min_{\delta} \mathcal{L}_{\Omega}(\delta)$ *associated* [*as in* (3.23)] *with* $\mathsf{H}_1(\bar{\delta})$ (2.3) *are invariant to replacing* X *by*

$$X_{\perp} = XK \tag{6.16}$$

where K *is an arbitrary nonsingular* $k \times k$ *matrix.*

Given the specific normalization underlying $\mathcal{H}_1(\bar{\theta})$ and $\mathsf{H}_1(\bar{\delta})$ (2.2) and (2.3), the associated likelihood ratio statistics and the point estimators of $\bar{\theta}$ and of $\bar{\delta}$ are thus unaffected by the considered transformations. This may also be checked by referring to the analytical solutions provided above for these special cases. Indeed, on replacing Y by $Y_{\top}$ and X by $X_{\perp}$, in turn, into (4.6) and (4.7), we see that the roots of $\hat{\Delta}$ [in 4.4] are unaffected; furthermore, on replacing Y by $Y_{\top}$ into the determinantal equation (A.17) that underlies the estimator $\hat{\theta}$ in (4.12), we can check that the latter is also unchanged by this mapping; the same observation may be checked regarding the estimator $\hat{\delta}$ in (4.13), by replacing X by $X_{\perp}$ into (A.18). We can now establish the following theorem which characterizes the distribution of these test statistics.

Theorem 6.2 *Under* (2.1), *the distributions of the statistics* $\mathcal{L}_{\Lambda}(\theta)$ *and* $\mathcal{L}_{\Lambda}(\hat{\theta})$ *depend on the model parameters* (B, J) *only through* $\bar{B} = B\,(J')^{-1}$:

$$\begin{aligned} \mathcal{L}_{\Lambda}(\theta) &= T\,\ln\left(|\hat{U}_{\Lambda}'(\theta)\hat{U}_{\Lambda}(\theta)|/|W'\mathcal{M}[X]W|\right), \quad \mathcal{L}_{\Lambda}(\hat{\theta}) = \inf_{\theta} \mathcal{L}_{\Lambda}(\theta)\,, & (6.17)\\ \hat{U}_{\Lambda}(\theta) &= \bar{M}_{\Lambda}(\theta)(X\bar{B} + W) = \bar{M}_{\Lambda}(\theta)\left[\iota_T(a' + \theta' b)\right](J')^{-1} + \bar{M}_{\Lambda}(\theta)W, & (6.18)\end{aligned}$$

where, conforming with the definition (3.9),

$$\begin{aligned} \bar{M}_{\Lambda}(\theta) &= \mathcal{M}[X] + \mathcal{M}_0[X, (1, \theta')],\\ \mathcal{M}_0[X, (1, \theta')] &= X(X'X)^{-1}(1, \theta')'[(1, \theta')(X'X)^{-1}(1, \theta')']^{-1}(1, \theta')(X'X)^{-1}X'.\end{aligned}$$

If, furthermore, the null hypothesis $\mathcal{H}_1(\bar{\theta})$ (2.2) *holds i.e. if there exists an unknown vector* $\bar{\theta}$ *such that* $a = b\bar{\theta}$, *then*

$$\hat{U}_{\Lambda}(\theta) = \bar{M}_{\Lambda}(\theta)\left[\iota_T(\bar{\theta} - \theta)' b\right](J')^{-1} + \bar{M}_{\Lambda}(\theta)W. \tag{6.19}$$

Further information on the structure of the nuisance parameters can be drawn from considering the *singular value decomposition* of $\bar{B}$. Let $\bar{r}$ be the rank of $\bar{B}$. Since $\bar{B}$ is a

$k \times n$ matrix, and we can write (see Harville (1997, Theorem 21.12.1)):

$$\bar{B} = \bar{P} \begin{bmatrix} \bar{D} & 0 \\ 0 & 0 \end{bmatrix} \bar{Q}', \quad \bar{D} = \text{diag}\left(\lambda_1^{1/2}, \ldots, \lambda_{\bar{r}}^{1/2}\right), \tag{6.20}$$

where $\lambda_1, \ldots, \lambda_{\bar{r}}$ are the $\bar{r}$ non-zero eigenvalues of $\bar{B}'\bar{B}$, $\bar{Q} = \left(\bar{Q}_1, \bar{Q}_2\right)$ is an orthogonal $n \times n$ matrix $[\bar{Q}'\bar{Q} = I_n]$ and $\bar{P} = \left(\bar{P}_1, \bar{P}_2\right)$ is an orthogonal $k \times k$ matrix $[\bar{P}'\bar{P} = I_k]$, $\bar{Q}_1$ and $\bar{P}_1$ have $\bar{r}$ columns, $\bar{P}_1'\bar{P}_2 = 0$ and

$$\bar{Q}'\,\bar{B}'\bar{B}\bar{Q} = \begin{bmatrix} \bar{D}^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad \bar{P}'\bar{B}\bar{Q} = \begin{bmatrix} \bar{D} & 0 \\ 0 & 0 \end{bmatrix}, \quad \bar{P}_1 = \bar{B}\bar{Q}_1\bar{D}^{-1}.$$

If $k \leq n$, we can set $\bar{P}_1 = \bar{P}$ when $\bar{r} = k$, so in this case $\bar{Q}'\,\bar{B}'\bar{B}\bar{Q} = \begin{bmatrix} \bar{D}^2 & 0 \end{bmatrix}$, $\bar{P}'\bar{B}\bar{Q} = \begin{bmatrix} \bar{D} & 0 \end{bmatrix}$, $\bar{B} = \bar{P}\begin{bmatrix} \bar{D} & 0 \end{bmatrix}\bar{Q}'$. Using Lemma **6.1** and Theorem **6.2**, $\mathcal{L}_\Lambda(\theta)$ and $\mathcal{L}_\Lambda(\hat{\theta})$ may then be re-expressed as:

$$\begin{aligned} \mathcal{L}_\Lambda(\theta) &= T\,\ln\left(|\tilde{U}_\Lambda'(\theta)\tilde{U}_\Lambda(\theta)|/|\bar{Q}'W'\mathcal{M}[X]W\bar{Q}|\right), \quad \mathcal{L}_\Lambda(\hat{\theta}) = \inf_\theta \mathcal{L}_\Lambda(\theta), \\ \tilde{U}_\Lambda(\theta) &= \hat{U}_\Lambda(\theta)\bar{Q} = \bar{M}_\Lambda(\theta)\left(X\begin{bmatrix} \bar{P}\bar{D} & 0 \end{bmatrix} + W\bar{Q}\right). \end{aligned}$$

When W follows a Gaussian assumption, the rows of $W\bar{Q}$ are *i.i.d.* $\mathrm{N}(0, I_n)$, so that the statistics $\mathcal{L}_\Lambda(\theta)$ and $\mathcal{L}_\Lambda(\hat{\theta})$ follow distributions which depend on B, J' only through $\bar{P}\bar{D}$, and the orthogonality of $\bar{P}$ leads to further nuisance parameter reductions. Under $\mathcal{H}_1(\bar{\theta})$, nuisance parameters may be tracked down conveniently, using the singular value decomposition of $b_J = (J')^{-1}\,b'$. Indeed, we can write

$$b_J = b_*\mathbf{b}^{1/2}b_\top$$

where $\mathbf{b}$ is a q-dimensional diagonal matrix which includes the non-zero eigenvalues of $b_J b_J'$, b_* is the $n \times q$ matrix which includes the corresponding eigenvectors so $b_*'b_* = I_q$ and $b_\top$ is a q-dimensional orthogonal matrix such that $b_\top b_\top' = I_q$. Furthermore, we can find an $n \times n$ matrix $\bar{b} = \begin{bmatrix} b_* & \bar{b}_* \end{bmatrix}$ such that $\bar{b}'\bar{b} = \bar{b}\bar{b}' = I_n$ so $b_*'\bar{b} = \begin{bmatrix} I_q & 0 \end{bmatrix}$. If we define

$$b_\theta = (\bar{\theta} - \theta)'b_\top'\mathbf{b}^{1/2}\begin{bmatrix} I_q & 0 \end{bmatrix},$$

$\mathcal{L}_\Lambda(\theta)$ and $\mathcal{L}_\Lambda(\hat{\theta})$ may then be re-expressed as:

$$\begin{aligned} &\mathcal{L}_\Lambda(\theta) = T\,\ln\left(|\tilde{U}_\Lambda'(\theta)\tilde{U}_\Lambda(\theta)|/|\bar{b}'W'\mathcal{M}[X]W\bar{b}|\right), \quad \mathcal{L}_\Lambda(\hat{\theta}) = \inf_\theta \mathcal{L}_\Lambda(\theta), \\ &\tilde{U}_\Lambda(\theta) = \hat{U}_\Lambda(\theta)\bar{b} = \bar{M}_\Lambda(\theta)\iota_T(\bar{\theta} - \theta)'b_J'\bar{b} + \bar{M}_\Lambda(\theta)W\bar{b} = \bar{M}_\Lambda(\theta)\iota_T b_\theta + \bar{M}_\Lambda(\theta)W\bar{b}. \end{aligned}$$

When W follows a Gaussian assumption, the rows of $W\bar{b}$ are *i.i.d.* $\mathrm{N}(0, I_n)$, so that the statistics $\mathcal{L}_{\Lambda}(\theta)$ and $\mathcal{L}_{\Lambda}(\hat{\theta})$ follow distributions which depend only on q free parameters [from b_{θ}].

Let us now turn to the case of $\mathsf{H}_1(\bar{\delta})$. In this case, it useful to study the distribution of

$$\Omega_G = \min_G \left[|\tilde{S}_0(G)|/|\tilde{S}(G)| \right]$$

where G is an $n \times g$ matrix of rank g, and conforming with (3.3) and (3.6)

$$\tilde{S}(G) = G'\hat{S}G, \quad \tilde{S}_0(G) = \tilde{S}(G) + G'\hat{B}(X'X)\hat{B}G.$$

As in the case of Theorem **3.1**, consider the singular value decomposition of $J'G$:

$$J'G = \Gamma \Psi^{1/2} \Xi$$

where Ψ is a g-dimensional diagonal matrix which includes the non-zero eigenvalues of $J'GG'J$, Γ is the $n \times g$ matrix which includes the corresponding eigenvectors so $\Gamma'\Gamma = I_g$ and Ξ is the g-dimensional matrix $\Xi = \Psi^{-1/2}\Gamma'J'G$ so that $\Xi\Xi' = \Xi'\Xi = I_g$. So

$$YG = \left[\hat{U}_{\Omega}(\Gamma) \right] \Psi^{1/2}\Xi, \quad \hat{U}_{\Omega}(\Gamma) = XBG\Xi'\Psi^{-1/2} + W\Gamma$$

$$|\tilde{S}_0(G)|/|\tilde{S}(G)| = \frac{\left| \hat{U}_{\Omega}(\Gamma)'\hat{U}_{\Omega}(\Gamma) \right|}{|\Gamma'W'P[X]W\Gamma|}.$$

When $G = (1, \delta')'$ is an n-dimensional vector, then Γ is an orthonormal vector and $\Psi^{-1/2}\Xi$ is a scalar which we denote $\bar{c}$, so we define

$$B(1, \delta')'\Xi'\Psi^{-1/2} = \bar{c}B(1, \delta')' \tag{6.21}$$

which leads to the following distributional result.

Theorem 6.3 *Under* (2.1), *the distributions of the statistics* $\mathcal{L}_{\Omega}(\delta)$ *and* $\mathcal{L}_{\Omega}(\hat{\delta})$ *depend on the model parameters* (B, J) *only through*

$$\mathcal{L}_{\Omega}(\theta) = T \ln \left(|\hat{U}_{\Omega}'(\delta)\hat{U}_{\Omega}(\delta)|/|\Gamma'W'\mathcal{M}[X]W\Gamma| \right), \quad \mathcal{L}_{\Lambda}(\hat{\theta}) = \inf_{\theta} \mathcal{L}_{\Lambda}(\theta), \tag{6.22}$$

where Γ *is the orthonormal* $n \times 1$ *vector which includes the eigenvector associated with the non-zero eigenvalue of* $J'(1, \delta')'(1, \delta')J$,

$$\hat{U}_{\Omega}(\delta) = X\bar{c}B(1, \delta')' + W\Gamma$$

and $\bar{c}$ *is the scalar from* (6.21). *If, furthermore, the null hypothesis* $\mathsf{H}_1(\bar{\delta})$ (2.2) *holds, i.e.*

if there exists an unknown vector $\bar{\delta}$ *such that* $B(1, \bar{\delta}')' = 0$, *then*

$$\hat{U}_{\Omega}(\delta) = X\,đ(\delta - \bar{\delta}) + W\Gamma$$

where $đ$ *is the* $k \times (n-1)$ *submatrix of* $\bar{c}B$ *which excludes its first column.*

Indeed, when $G = (1, \delta')'$ is an n-dimensional vector, then

$$BG\Xi'\Psi^{-1/2} = \bar{a} + đ\delta$$

where $\bar{a}$ is the first column of $\bar{c}B$ and $đ$ is the remaining $k \times (n-1)$ submatrix of $\bar{c}B$. Under the null hypothesis, there exists a vector $\bar{\delta}$ such that $B(1, \bar{\delta}')' = 0$, so

$$B(1, \delta')'\Xi'\Psi^{-1/2} = đ(\delta - \bar{\delta}).$$

Furthermore, if $k \geq n$, then using the singular value decomposition of $đ$, we can write

$$đ = d_* \mathbf{đ}^{1/2} d_\perp$$

where $\mathbf{đ}$ is an $(n-1)$-dimensional diagonal matrix which includes the square roots of non-zero eigenvalues of $đđ'$, d_* is the $k \times (n-1)$ matrix which includes the corresponding eigenvectors so $d_*' d_* = I_{n-1}$ and $d_\perp$ is an $(n-1)$-dimensional orthogonal matrix such that $d_\perp d_\perp' = I_{n-1}$. We can thus find a $k \times k$ matrix $\bar{\mathbf{d}} = [\ d_* \quad \bar{d}_*\ ]$ such that $\bar{\mathbf{d}}'\bar{\mathbf{d}} = \bar{\mathbf{d}}\bar{\mathbf{d}}' = I_k$ and $\bar{\mathbf{d}}' d_* = [\ I_{n-1} \quad 0\ ]'$. If we define

$$b_\delta = \bar{\mathbf{d}} \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}' \mathbf{đ}^{1/2} d_\perp (\delta - \bar{\delta})$$

and in view of the invariance of the statistics and test problem to post-multiplying X by a $k \times k$ invertible matrix, then

$$\hat{U}_{\Omega}(\delta) = X\bar{\mathbf{d}}\bar{\mathbf{d}}'\,đ(\delta - \bar{\delta}) + W\Gamma = Xb_\delta + W\Gamma.$$

When W follows a Gaussian assumption, the elements of $W\Gamma$ are *i.i.d.* $\mathrm{N}(0, 1)$, so that the statistics $\mathcal{L}_{\Omega}(\delta)$ and $\mathcal{L}_{\Omega}(\hat{\delta})$ follow distributions which depend only on $n-1$ free parameters [from b_δ].

The above results on nuisance parameter reductions can also be useful for the mixture of normals hypothesis (2.6). Indeed, the above arguments easily imply that the number of nuisance parameters in this case corresponds to the Gaussian one, plus the parameters which define the distribution of the variates in the V matrix. This result is interesting since (2.6) covers many empirically relevant distributions used in practice, including the Skewed normal or Student-t. Nevertheless, our results reveal null distributions that depend on nuisance parameters possibly in a complex way. So the test inversion approach we followed above may provide tractable finite sample motivated tests of $\mathcal{H}_1(\bar{\theta})$ and $\mathsf{H}_1(\bar{\delta})$.

7. Applications

Hypotheses of the uniform linear class (2.22) are common in statistics, econometrics and finance [see Dufour and Khalaf (2002) and Beaulieu, Dufour and Khalaf (2005, 2007, 2008*a*) for examples and further references]. In particular, note that (2.26) and (2.3) relate to the linear simultaneous equations settings, which leads to the analytical setting of Dufour and Taamouti (2005, 2007). For this problem, Theorem **5.2** implies that if the usual F-test for the (joint) significance of instruments is not significant in at least one equation from the first stage multivariate regression, then the CS for the coefficient of the right-hand-side endogenous variable will be unbounded. In addition, Theorem **5.3** implies that an empty CS corresponds to a significant bounds-based J-test using the usual LIML root. Alternatively, (2.5) corresponds to the dimensionality hypothesis (as first formulated in Rao (1973, Chapter 8)). Applications of (2.2) and (2.4) include various important asset pricing models which form the basis of our empirical analysis.

Let r_i, $i = 1, \ldots, n$, be a vector of T returns on n securities or portfolios for the period $t = 1, \ldots, T$, and $\tilde{r} = \begin{bmatrix} \tilde{r}_1 & \ldots & \tilde{r}_q \end{bmatrix}$ a $T \times q$ matrix of observations on q factors or benchmarks. Factor models used to assess various forms of the APT rely on regressions such as

$$r_i = a_i \iota_T + \tilde{r} b_i + u_i, \qquad i = 1, \ldots, n. \tag{7.1}$$

Here we focus on restrictions [see Campbell et al. (1997, Chapter 6)] such as:

$$a_i = \bar{\zeta}' b_i + \bar{\mu}, \text{ for some unknown } \bar{\zeta} \text{ and } \bar{\mu}. \tag{7.2}$$

Clearly, (7.1) takes the form (2.1) where $Y = \begin{bmatrix} r_1 & \ldots & r_n \end{bmatrix}$ and $X = \begin{bmatrix} \iota_T & \tilde{r} \end{bmatrix}$ and (7.2) generally implies a restriction of the form (2.4). Models which include a market factor may also lead to constraints of the form (2.2). For example, suppose that $\tilde{r}_1$ corresponds to a vector of returns on a market benchmark; then an alternative model consistent with (7.1) would take the form

$$r_i - \tilde{r}_1 = a_i \iota_T + \tilde{r} d_i + u_i, \qquad i = 1, \ldots, n, \tag{7.3}$$

$$a_i = \bar{\theta}' d_i, \qquad \text{for some unknown } \bar{\theta}. \tag{7.4}$$

For instance, the fundamental single factor CAPM due to Black (1972) (see e.g. Zhou (1991)) may be assessed as in (7.3) with $\tilde{r} = [\tilde{r}_1]$. The multi-beta CAPM extension of Black's model (see e.g. Shanken (1986) and Velu and Zhou (1999)) may be expressed in the (7.1) form imposing $a_i = \bar{\theta}(1-\iota_q' b_i)$, $i = 1, \ldots, n$. This model can easily be expressed into our framework by a suitable projection.

7.1. Simulation Study

We consider a simulation study to assess the properties of our proposed CSs in non-*i.i.d.* contexts associated with a single factor APT, namely (7.3) with $\tilde{r} = [\tilde{r}_1]$, a vector of returns on a market benchmark with multivariate GARCH errors. (Beaulieu et al. (2008*b*)) used this design to illustrate the superiority of the LR-type statistic with or without GARCH corrections, relative to the robust Wald test. This is scalar parameter problem which is easy to interpret, so we revisit the study to assess whether the parsimonious GARCH structure that allows to evacuate J out affects the properties of QMLE relative to HAR-type statistics.

The design relies on real value weighted monthly returns from January 1926 to December 1995, obtained from the University of Chicago's Center for Research in Security Prices (CRSP) on 12 portfolios of New York Stock Exchange (NYSE) firms grouped by standard two-digit industrial classification (SIC); for details see Beaulieu et al. (2007). The market return corresponds to the value-weighted NYSE returns, from CRSP. The real risk-free rate is proxied by the one month Treasury bill rate also from CRSP net of inflation. The designs with $T = 60, 120$ and 828 correspond respectively to the 1991-95, 1990-95 subperiods, and to the full sample, respectively. The regressor matrix uses the market regressor form each sample, and a constant. W_t is drawn as in (2.8) with $(\phi_1, \phi_2) = (.15, .80)$. The matrix J is also calibrated to its QMLE counterpart [denoted $\hat{J}$ in Table 1] from each sample; we also use $J = I_{12}$ in the size study, to illustrate the superiority of the scale-invariant criteria. We study the size and power of tests associated with a linear version of (7.4):

$$a_i = \theta' d_i, \text{ for a known scalar } \theta. \tag{7.5}$$

This corresponds to assessing the coverage properties of our proposed CSs for $\bar{\theta}$. To draw the samples imposing (7.5), θ is calibrated to its QMLE counterpart from each sample. Power is analyzed maintaining (7.4) for θ equal to the null value + $\mathsf{step} \times \widehat{\sigma}_i^{\min}$, where step ranges from .25 to 2.5 and $\widehat{\sigma}_i^{\min} = [\min\{\widehat{\sigma}_i^2\}]^{1/2}$, $\widehat{\sigma}_i^2$ are the diagonal terms of $\hat{J}\hat{J}'$. This formulation accounts for the scale in line with the design data.

Tests with unknown nuisance parameters are MMC tests [as described in section 5], where maximizations are run over the nuisance parameter range $(\phi_1, \phi_2) \in \{(.05,\ .90)$; $(.10,.85)$; $(.15,.80)$; $(.20,.75)$; $(.25,.70)$; $(.30,.65)$; $(.35,.60)$; $(.40,.55)$; $(.45,.50)$; $(.50;.45)\}$. To assess the power losses arising from maximizing p-values, we also study the MC test that treats ϕ_1, ϕ_2 as known. The tests in Tables 1-2 pertain to the following: (i) the

Table 1. Size Monte Carlo Study

$n = 12$	$T = 60$,		$T = 120$		$T = 828$	
	J		J		J	
Test	I_{12}	$\hat{J}$	I_{12}	$\hat{J}$	I_{12}	$\hat{J}$
$\hat{\theta}\pm$ Asy-SE($\hat{\theta}$)	.639	.136	.637	.089	.563	.046
$\mathcal{LR}_{\Lambda}(\theta)$, MC, $\phi_1 = \phi_2 = 0$	.045	.045	.048	.048	.046	.046
$\mathcal{W}_p$, $\chi^2(12)$	.959	.959	.683	.683	.116	.116
$\mathcal{W}_p$, MC, ϕ_1, ϕ_2 known	.038	.038	.044	.044	.055	.055
$\mathcal{W}_p$, MMC, ϕ_1, ϕ_2 unknown	.030	.030	.037	.037	.037	.037
$\mathcal{LR}_{\Lambda}(\theta)$, MC, ϕ_1, ϕ_2 known	.053	.053	.050	.050	.044	.044
$\mathcal{LR}_{\Lambda}(\theta)$, MMC, ϕ_1, ϕ_2 unknown	.041	.041	.028	.028	.030	.030

Note – Numbers reported are empirical rejection rates for a 5% nominal size for tests of (7.5); inverting these tests provides a confidence set for $\bar{\theta}$ in (7.4). The tests pertain to the following: (i) the LR statistic (3.16) ignoring GARCH [$\mathcal{LR}_{\Lambda}(\theta)$, MC, $\phi_1 = \phi_2 = 0$], imposing GARCH yet treating ϕ_1, ϕ_2 as known [$\mathcal{LR}_{\Lambda}(\theta)$, MC, ϕ_1, ϕ_2 known] and the MMC case where ϕ_1, ϕ_2 are nuisance parameters [$\mathcal{LR}_{\Lambda}(\theta)$, MMC, ϕ_1, ϕ_2 unknown]; (ii) the HAR statistic (3.29) for 12 lags, with asymptotic cut-off points, and simulation-based versions imposing GARCH with known parameters [$\mathcal{W}_p$, MC, ϕ_1, ϕ_2 known] and the nuisance parameter case [$\mathcal{W}_p$, MMC, ϕ_1, ϕ_2 unknown]; and (iii) the asymptotic Wald-type confidence interval for $\bar{\theta}$, using the QMLE information-matrix-based asymptotic standard errors. Samples are generated with conditional variance as in (2.9) using $\hat{J}$ or I_{12} for J.

LR statistic (3.16) ignoring GARCH [$\mathcal{LR}_{\Lambda}(\theta)$, MC, $\phi_1 = \phi_2 = 0$][7], imposing GARCH yet treating ϕ_1, ϕ_2 as known [$\mathcal{LR}_{\Lambda}(\theta)$, MC, ϕ_1, ϕ_2 known] and the MMC case where the GARCH parameters are nuisance parameters [$\mathcal{LR}_{\Lambda}(\theta)$, MMC, ϕ_1, ϕ_2 unknown]; (ii) the HAR statistic (3.29) with $p = 11$ [12 lags], with asymptotic cut-off points, and simulation-based versions imposing GARCH with known parameters [$\mathcal{W}_p$, MC, ϕ_1, ϕ_2 known] and the nuisance parameter case [$\mathcal{W}_p$, MMC, ϕ_1, ϕ_2 unknown]; and (iii) the asymptotic Wald-type confidence interval for $\bar{\theta}$, using the QMLE information-matrix-based asymptotic standard errors [see Appendix .3] and a normal limiting distribution. Formally, the 5% confidence interval in question is first built; if this set does not cover the true θ, then the test is considered significant at the 5% level. This test is studied in line with our objective, namely the consequences of ignoring GARCH. MC tests are applied with 99 simulated samples, and each simulation is run with 1000 replications. We report empirical rejections for test with a 5% nominal size.

Results can be summarized as follows. The asymptotic *i.i.d.* or robust procedures perform poorly in finite samples. Interestingly, although we impose GARCH errors, tests based on the *i.i.d.* asymptotic intervals seem even more stable than the robust ones, for samples generated with $\hat{J}$. These results are not representative, since the asymptotic

[7] This corresponds to the F-based test.

Table 2. Power Monte Carlo Study

$n = 12$	$T = 60$		$T = 120$		$T = 828$	
Test	Step	Power	Step	Power	Step	Power
$\mathcal{LR}_\Lambda(\theta)$, MC, $\phi_1 = \phi_2 = 0$	.50	.344	.35	.252	.25	.392
$\mathcal{W}_p$, MC, ϕ_1, ϕ_2 known		.236		.204		.366
$\mathcal{W}_p$, MMC, ϕ_1, ϕ_2 unknown		.193		.160		.308
$\mathcal{LR}_\Lambda(\theta)$, MC, ϕ_1, ϕ_2 known		.354		.267		.385
$\mathcal{LR}_\Lambda(\theta)$, MMC, ϕ_1, ϕ_2 unknown		.303		.214		.317
$\mathcal{LR}_\Lambda(\theta)$, MC, $\phi_1 = \phi_2 = 0$	.75	.716	.50	.573	.35	.726
$\mathcal{W}_p$, MC, ϕ_1, ϕ_2 known		.538		.399		.703
$\mathcal{W}_p$, MMC, ϕ_1, ϕ_2 unknown		.469		.340		.637
$\mathcal{LR}_\Lambda(\theta)$, MC, ϕ_1, ϕ_2 known		.719		.577		.724
$\mathcal{LR}_\Lambda(\theta)$, MMC, ϕ_1, ϕ_2 unknown		.671		.511		.670
$\mathcal{LR}_\Lambda(\theta)$, MC, $\phi_1 = \phi_2 = 0$	1.0	.946	.60	.766	.40	.860
$\mathcal{W}_p$, MC, ϕ_1, ϕ_2 known		.760		.549		.840
$\mathcal{W}_p$, MMC, ϕ_1, ϕ_2 unknown		.720		.491		.792
$\mathcal{LR}_\Lambda(\theta)$, MC, ϕ_1, ϕ_2 known		.943		.759		.870
$\mathcal{LR}_\Lambda(\theta)$, MMC, ϕ_1, ϕ_2 unknown		.920		.712		.833
$\mathcal{LR}_\Lambda(\theta)$, MC, $\phi_1 = \phi_2 = 0$	1.5	.999	.75	.927	.50	.978
$\mathcal{W}_p$, MC, ϕ_1, ϕ_2 known		.942		.757		.970
$\mathcal{LR}_\Lambda(\theta)$, MC, ϕ_1, ϕ_2 known		.920		.693		.955
$\mathcal{LR}_\Lambda(\theta)$, MC, ϕ_1, ϕ_2 known		.999		.925		.976
$\mathcal{LR}_\Lambda(\theta)$, MMC, ϕ_1, ϕ_2 unknown		.988		.909		.972

Note – Numbers reported are empirical rejection rates for 5% tests of (7.5). Power is analyzed maintaining (7.4) for θ equal to the null value + $\texttt{step} \times \widehat{\sigma}_i^{\min}$, where $\widehat{\sigma}_i^{\min} = [\min\{\widehat{\sigma}_i^2\}]^{1/2}$, $\widehat{\sigma}_i^2$ are the diagonal terms of $\hat{J}\hat{J}'$. The samples are generated with conditional variance as in (2.9) using $\hat{J}$ for J.

interval is not invariant to J. Indeed, the case with $J = I_{12}$ reveals size important size distortions. The LR and robust MC tests achieve size control, even when GARCH is not accounted for. The power study [which focuses on the size correct procedures in view of the very poor performance on the asymptotic tests] reflect this feature. Indeed, the LR uncorrected for GARCH performs best; this results motivates our reliance on F-based LR statistics in our empirical study. Even when GARCH corrections are performed [via the MMC p-values], the corrected LR statistic performs generally better than the robust one, at least in smaller samples. Overall, the MMC correction is not utterly costly power-wise. On balance, given the numerical computations involved, these results illustrate the worth of our analytical F-based motivation for relying on our proposed analytical test inversion formula.

7.2. Empirical Analysis

We focus on the factor model (7.3) and the multi-beta CAPM using Fama and French [Fama and French (1992, 1993)] factors: (1) the excess return on the market, defined as the value-weighted return on all NYSE, AMEX, and NASDAQ stocks (from CRSP), denoted **MRKT**; (2) the average return on three small portfolios minus the average return on three big portfolios, denoted **SMB** (Small Minus Big); (3) the average return on two value portfolios minus the average return on two growth portfolios, denoted **HML** (High Minus Low); (4) the difference between the monthly long term government bond return (from Ibbotson Associates) and the one-month Treasury bill rate measured at the end of the previous month, denoted **TERM**; (5) the difference between the return on a market portfolio of long term corporate bonds (the Composite portfolio on the corporate bond module of Ibbotson Associates) and the long term government bond return, denoted **DEF**; and (6) the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios, denoted **MOM** (Momentum).

We use Fama and French's data base. We produce results for monthly returns of 25 value weighted and equally weighted portfolios from 1961-2000. The portfolios which are constructed at the end of June, are the intersections of five portfolios formed on size (market equity) and five portfolios formed on the ratio of book equity to market equity. The size breakpoints for year s are the New York Stock Exchange (NYSE) market equity quintiles at the end of June of year s. The ratio of book equity to market equity for June of year s is the book equity for the last fiscal year end in $s-1$ divided by market equity for December of year $s-1$. The ratio of book equity to market equity are NYSE quintiles. The portfolios for July of year s to June of year $s+1$ include all NYSE, AMEX, NASDAQ stocks for which we have market equity data for December of year $s-1$ and June of year s, and (positive) book equity data for $s-1$. Fama and French benchmark factors, SMB and HML, are constructed from six size/book-to-market benchmark portfolios that do not include hold ranges and do not incur transaction costs. The portfolios for these factors are rebalanced quarterly using two independent sorts, on size (market equity, ME) and book-to-market (the ratio of book equity to market equity, BE/ME). The size breakpoint (which determines the buy range for the small and big portfolios) is the median NYSE market equity. The BE/ME breakpoints (which determine the buy range for the growth, neutral, and value portfolios) are the 30th and 70th NYSE percentiles. For the construction of the MOM factor, six value-weighted portfolios formed on size and prior (2-12) returns are used. The portfolios, which are formed monthly, are the intersections of two portfolios formed on size (market equity, ME) and three portfolios formed on prior (2-12) return. The monthly size breakpoint is the median NYSE market equity. The monthly prior (2-12) return breakpoints are the 30th and 70th NYSE percentiles.

Results can be summarized as follows. Black's multifactor model is rejected at the 5% level (the CS for $\theta = \varnothing$) everywhere expect for 1961-65 with value-weighted portfolios where $\mathrm{CS}_{.05}(\theta) = [-.005, -.001]$. Results for the APT model are reported in the following Tables.

As may be checked from Table 3, the TERM and DEF factors are not jointly significant using our Hotelling test. This implies that an unbounded CSs for θ will occur if these two factors are included in model. Indeed, we find that for all models considered which include TERM and DEF [solely or in addition to any other or all other factors], the CSs for the θ parameter is the real line. This result is quite striking and illustrates the serious consequences of adding redundant factors to the asset pricing model. Of course, relying on a usual Wald-type procedure would not have uncovered this serious shortcoming. We thus analyze a model with MRKT, SMB, HML and MOM.

Overall, our results lend support to the APT only in the 1991-95 subperiod with equally weighted data and the 1976-80 subperiod with value weighted data. The model is rejected in 1976-80 subperiod with equally value data, and extremely wide CSs obtains for all other subperiods with both equally and value weighted data. Most importantly, our CSs differ dramatically and arbitrarily from the standard Wald-type ones.

Indeed, on comparing the identification robust CSs to their Wald-based standard counterparts from Table 4, we note striking and arbitrary differences. For instance, several Wald-type intervals though quite precise, do cover zero, whereas their identification robust counterparts though very diffuse, exclude the zero value. This occurs in particular: (i) with equally weighted data, in the case of the market factor in the 1966-70 subperiod and the momentum factor in the 1986-90 subperiod, and (ii) with value-weighted data, in the case of the market factor in the 1966-70 and 1996-2000 subperiod and the momentum factor in the 1986-90 and 1991-95 subperiods. Such sharp contrasts are worth noting. Furthermore, in the few cases where the identification robust sets are bounded, the latter may also still conflict with regular Wald-type intervals regarding covering the zero value;

Table 3. Hotelling-type tests on individual factors

EW	MRKT		SMB		HML		TERM		DEF		MOM	
	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$
61-65	353.2	.001	255.4	.001	229.3	.001	62.75	.027	50.41	.143	68.67	.009
66-70	308.1	.001	301.2	.001	190.8	.001	39.16	.450	57.16	.054	94.58	.001
71-75	327.2	.001	282.4	.001	245.3	.001	51.37	.133	54.00	.091	45.77	.219
76-80	299.8	.001	255.6	.001	204.2	.001	44.85	.257	51.72	.114	109.0	.001
81-85	266.8	.001	236.3	.001	174.3	.001	59.60	.037	46.85	.200	77.86	.003
86-90	386.7	.001	276.8	.001	213.8	.001	39.05	.441	37.57	.495	88.28	.002
91-95	308.2	.001	261.5	.001	217.0	.001	27.26	.851	26.81	.868	81.54	.001
96-00	232.8	.001	255.8	.001	203.8	.001	50.14	.134	64.32	.026	75.69	.003
VW	MRKT		SMB		HML		TERM		DEF		MOM	
	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$	LR	$p_{\mathcal{N}}$
61-65	442.8	.001	294.4	.001	232.5	.001	54.81	.073	41.36	.354	44.93	.244
66-70	374.8	.001	321.0	.001	228.8	.001	49.18	.180	47.14	.207	86.35	.001
71-75	382.8	.001	326.3	.001	311.7	.001	44.21	.284	43.68	.296	50.04	.136
76-80	317.8	.001	293.4	.001	250.0	.001	38.51	.456	43.38	.307	92.21	.001
81-85	322.8	.001	322.6	.001	228.4	.001	44.82	.274	32.73	.638	43.09	.290
86-90	425.8	.001	278.5	.001	254.8	.001	28.09	.824	25.06	.894	51.31	.117
91-95	355.9	.001	327.8	.001	286.0	.001	29.91	.772	33.37	.634	54.42	.082
96-00	278.8	.001	353.6	.001	261.1	.001	45.27	.231	55.91	.065	72.72	.006

Notes – The model corresponds to the unrestricted system in (7.3) where $\tilde{r}$ includes the returns on MRKT, SMB, HML, TERM, DEF and MOM. LR refers to the statistic $T \ln\left(1 + \Lambda_i\right)$,defined in (3.20) associated with the joint significance of each factor in all equations; $p_{\mathcal{N}}$ denotes its F-based p-value. EW and VW refer to equally weighted and value weighted portfolios respectively. Tests and CSs are conducted at the 5% level.

this occurs in the case of the SMB factor in the 1991-95 subperiod with equally weighted data.

Table 5 reports the general model test results. Although the χ^2 based p-values lead to rejecting the model (at the 5% level) in 4 subperiods with equally weighted data, and 6 subperiods with value weighted data, relying on the parametric bootstrap leads to 3 model rejections with equally weighted data, and only two rejections with value weighted data; the bound-based p-values lead to only one rejection at the 5% level with equally weighted data [the 1976-80 subperiod], whereas for value weighted data, the bounds test is not significant at 5% for all subperiods. As a robustness check, we compute the bounds maximized Monte Carlo test as described in section 5 which allows one to check whether the CS for θ is empty assuming a $\mathtt{t}(\kappa)$ distribution. The choice set for κ was selected as in Beaulieu et al. (2007) by inverting a goodness-of-fit test (at the 2.5% level) for the hypothesized distribution; this requires a level correction if an exact J-type test is

Table 4. CSs for key parameters
Model with Fama-French Factors

EW	MRKT	SMB	HML	MOM
61-65	$\mathbb{R}$ ([.001,.015])	$\mathbb{R}$ ([−.003,−.001])	$\mathbb{R}$ ([−.003,−.001])	$\mathbb{R}$ ([−.001,.009])
66-70	$\{\leq -.026\} \cup \{\geq .012\}$ ([−11.240,14.456])	$\mathbb{R}$ ([−.074,.083])	$\mathbb{R}$ ([−.521,.667])	$\mathbb{R}$ ([−2.864,2.229])
71-75	$\mathbb{R}$ ([−.093,.014])	$\mathbb{R}$ ([−.008,.004])	$\mathbb{R}$ ([−.020,.002])	$\mathbb{R}$ ([.005,.323])
76-80	$\varnothing$ ([−.156,.042])	$\varnothing$ ([−.006,.002])	$\varnothing$ ([−.004,.006])	$\varnothing$ ([.024,.079])
81-85	$\mathbb{R}$ ([−2.994,2.247])	$\mathbb{R}$ ([−1.657,1.206])	$\mathbb{R}$ ([−.516,.704])	$\mathbb{R}$ ([−12.103,8.872])
86-90	$\mathbb{R}$ ([−.060,.024])	$\mathbb{R}$ ([−.016,−.001])	$\mathbb{R}$ ([−.047,.001])	$\{\leq -.011\} \cup \{\geq .054\}$ ([−.347,.020])
91-95	$[.035, .047]$ ([.027,.053])	$[.002, .005]$ ([−.000,.006])	$[-.015, -.011]$ ([−.017,−.009])	$[.062, .086]$ ([.049,.095])
96-00	$\mathbb{R}$ ([.042,.116])	$\mathbb{R}$ ([−.001,.010])	$\mathbb{R}$ ([−.004,.005])	$\mathbb{R}$ ([.030,.089])

VW	MRKT	SMB	HML	MOM
61-65	$\mathbb{R}$ ([−.529,.018])	$\mathbb{R}$ ([−.013,.004])	$\mathbb{R}$ ([−.021,.005])	$\mathbb{R}$ ([−.226,.038])
66-70	$\{\leq -.099\} \cup \{\geq .131\}$ ([−2.087,.548])	$\mathbb{R}$ ([−.017,.017])	$\mathbb{R}$ ([−.025,.037])	$\mathbb{R}$ ([−.133,.473])
71-75	$\mathbb{R}$ ([.065,.273])	$\mathbb{R}$ ([−.024,−.004])	$\mathbb{R}$ ([.001,.018])	$\mathbb{R}$ ([.040,.292])
76-80	$[-.933, -.067]$ ([−.209,−.048])	$[.002, .075]$ ([.000,.015])	$[-.072, -.004]$ ([−.019,−.003])	$[.063, .839]$ ([.045,.191])
81-85	$\mathbb{R}$ ([.085,.273])	$\mathbb{R}$ ([−.000,.006])	$\mathbb{R}$ ([−.023,−.006])	$\mathbb{R}$ ([.051,.230])
86-90	$\mathbb{R}$ ([−.131,−.022])	$\mathbb{R}$ ([−.009,.003])	$\mathbb{R}$ ([−.014,−.001])	$\{\leq -.027\} \cup \{\geq .138\}$ ([−.330,−.072])
91-95	$\mathbb{R}$ ([−.086,.033])	$\mathbb{R}$ ([−.002,.009])	$\mathbb{R}$ ([−.018,.002])	$\{\leq -.137\} \cup \{\geq .075\}$ ([−.039,.681])
96-00	$\{\leq -.026\} \cup \{\geq .051\}$ ([−.084,.710])	$\mathbb{R}$ ([−.004,.008])	$\mathbb{R}$ ([−.004,.020])	$\mathbb{R}$ ([−.067,.405])

Notes – The model corresponds to (7.3) where $\tilde{r}$ includes the returns on MRKT, SMB, HML and MOM. LR presents the statistic defined in (4.5) associated with hypothesis (2.2). CSs [with standard Wald-type confidence intervals in parentheses] for the components of θ as defined in (7.3) corresponding to each factor are reported. EW and VW refer to equally weighted and value weighted portfolios respectively. Tests and CSs are conducted at the 5% level.

Table 5. Model specification tests
Model with Fama-French Factors

EW	J-Test				VW	J-Test			
	LR	p_∞	$p_{\mathcal{N}}$	p_{t}		LR	p_∞	$p_{\mathcal{N}}$	p_{t}
61-65	22.33	.500 (.771,.772)	.943	.951	61-65	39.78	.016 (.135,.157)	.392	.401
66-70	51.39	.001 (.017,.018)	.101	.116	66-70	37.14	.032 (.183,.190)	.509	.505
71-75	35.73	.044 (.197,.210)	.537	.560	71-75	31.85	.103 (.340,.353)	.695	.706
76-80	61.61	.000 (.004,.004)	.032	.033	76-80	55.92	.000 (.010,.012)	.067	.076
81-85	34.47	.059 (.253,.260)	.585	.605	81-85	40.78	.012 (.099,.101)	.354	.357
86-90	28.84	.185 (.486,.500)	.778	.789	86-90	22.48	.492 (.767,.777)	.943	.947
91-95	57.68	.000 (.003,.006)	.054	.058	91-95	49.05	.001 (.027,.038)	.156	.175
96-00	30.53	.135 (.410,.428)	.767	.772	96-00	36.68	.035 (.189,.204)	.547	.579

Notes – The model corresponds to (7.3) where $\tilde{r}$ includes the returns on MRKT, SMB, HML and MOM. LR presents the statistic defined in (4.5) associated with hypothesis (2.2); p_∞ gives its standard χ^2 based p-value and its Gaussian-based and Student-t based parametric bootstrap counterpart in parentheses in the following order (Gaussian-based, Student-t based), $p_{\mathcal{N}}$ is the F-based bound p-value from Theorem **5.3**, and p_{t} is its bounds maximized Monte Carlo counterpart assuming a $\mathsf{t}(\kappa)$ distribution as described in section 5. The Student-t based parametric bootstrap is a maximized Monte Carlo p-value, obtained by calculating a parametric bootstrap p-value for each κ using the QMLE constrained estimated regression coefficients to generate the simulated samples, then taking the maximum of these bootstrap p-values over κ. EW and VW refer to equally weighted and value weighted portfolios respectively. Tests and CSs are conducted at the 5% level.

desired; we suggest to subtract 2.5% from the significance level the test p-value is referred to. We also derive an maximized MC version of the parametric bootstrap assuming Student-t based errors. Specifically, we calculate a parametric bootstrap p-value for each κ using the QMLE constrained estimated regression coefficients to generate the simulated samples, then we take the maximum of these bootstrap p-values over κ. MC p-values rely on 999 replications. Results as reported in Table 5 are globally qualitatively similar to the Gaussian approximated ones. To sum up, observe that relying on a simple test for the specification would have led to accepting most models under test; our CS approach which includes a specification and an identification built-in check provides a much more complete statistical analysis. These results concur with the emerging literature on redundant factors [see *e.g.* Kan and Zhang (1999*a*, 1999*b*)], on tight factor structures and statistical pitfalls of asset pricing tests [see Lewellen and Nagel (2006), and Lewellen et al. (2006)], and on the importance of joint (across-portfolios) tests [see Shanken and Weinstein (2006) and

Beaulieu et al. (2007)].

8. Conclusion

This paper proposes identification robust inference methods for structural multivariate factor models with rank restrictions. We derive confidence set estimates for structural parameters based on inverting minimum-distance Hotelling-type pivotal statistics. We provide analytical solutions to the latter problem which hold exactly (or asymptotically) imposing (or relaxing) Gaussian error distributions. Our methodology relies on useful invariance results which we prove allowing for non-*i.i.d.* set-ups. Our proposed CSs have much more informational content than usual Hotelling tests and have various useful applications in statistics, econometrics and finance. Our approach further provides multivariate extensions of the classical Fieller problem. We apply our methodology to an asset pricing model with unobservable risk-free rates and an APT-type model using Fama-French factors. A simulation study modelled on this case and allowing for multivariate GARCH reveals serious distortions with traditional asymptotic tests; in contrast, our proposed inference methods perform well, even when GARCH effects are not accounted for. Empirical results illustrate, among others, severe problems with redundant factors.

Appendix

.1. Eigenvalue based equations and inequations

Equation (4.10) may be re-expressed as

$$\Sigma_{11} + \Sigma_{12}\pi = 0, \tag{A.1}$$

$$\Sigma_{21} + \Sigma_{22}\pi = 0, \tag{A.2}$$

and solving (A.2) for π leads to (4.11). Substituting $\hat{\pi}$ into (A.1) yields $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 0$. Assuming that Σ_{11} is non-singular, on recalling that $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is a scalar and using the formulae for the determinant of partitioned matrices

$$|\Sigma| = |\Sigma_{22}| \left|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right| = |\Sigma_{22}| \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

we thus see that if $\hat{\pi}$ satisfies (A.2) then it satisfies (A.1).

Inequality (5.1) may be re-expressed as

$$\pi' A_{22}\pi + 2A_{12}\pi + A_{11} \leq 0 \tag{A.3}$$

which leads to set-up of Dufour and Taamouti (2005, 2007) so projections based CSs for any linear transformation of π of the form $\omega'\pi$ can be obtained as follows. Let $\tilde{A} = -A_{22}^{-1}A_{12}'$, $\tilde{D} = A_{12}A_{22}^{-1}A_{12} - A_{11}$. If all the eigenvalues of A_{22} [as defined in (5.2)] are positive so A_{22} is positive definite then:

$$\begin{aligned} \mathrm{CS}_\alpha(\omega'\pi) &= \left[\omega'\tilde{A} - \sqrt{\tilde{D}\left(\omega' A_{22}^{-1}\omega\right)}, \omega'\tilde{A} + \sqrt{\tilde{D}\left(\omega' A_{22}^{-1}\omega\right)}\right], \quad if \quad \tilde{D} \geq 0 \quad (A.4) \\ \mathrm{CS}_\alpha(\omega'\pi) &= \varnothing, \quad if \quad < \tilde{D}0. \quad (A.5) \end{aligned}$$

If A_{22} is non-singular and has one negative eigenvalue then: (i) if $\omega' A_{22}^{-1}\omega < 0$ and $\tilde{D} < 0$:

$$\mathrm{CS}_\alpha(\omega'\pi) = \left]-\infty, \omega'\tilde{A} - \sqrt{\tilde{D}\left(\omega' A_{22}^{-1}\omega\right)}\right] \cup \left[\omega'\tilde{A} + \sqrt{\tilde{D}\left(\omega' A_{22}^{-1}\omega\right)}, +\infty\right[; \qquad (A.6)$$

(ii) if $\omega' A_{22}^{-1}\omega > 0$ or if $\omega' A_{22}^{-1}\omega \leq 0$ and $\tilde{D} \geq 0$ then:

$$\mathrm{CS}_\alpha(\omega'\pi) = \mathbb{R}; \qquad (A.7)$$

(iii) if $\omega' A_{22}^{-1}\omega = 0$ and $\tilde{D} < 0$ then:

$$\mathrm{CS}_\alpha(\omega'\pi) = \mathbb{R}\backslash\left\{\omega'\tilde{A}\right\}. \qquad (A.8)$$

The projection is given by (A.7) if A_{22} is non-singular and has at least two negative eigenvalues.

.2. Proofs of Theorems

Proof of Theorem **3.1** Consider the following decomposition of $\tilde{S}$ and $\tilde{S}_0$:

$$\begin{aligned} \tilde{S} &= G'JW'\mathcal{M}[X]WJ'G \\ C\hat{B}G - D &= C(X'X)^{-1}X'[XB + WJ']G - D = CBG - D + C(X'X)^{-1}X'WJ'G \end{aligned}$$

so under the null hypothesis $C\hat{B}G - D = C(X'X)^{-1}X'WJ'G$ and

$$\tilde{S}_0 - \tilde{S} = G'JW'\mathcal{M}_0[X, C]WJ'G, \quad \tilde{S}_0 = G'JW'\left(\mathcal{M}[X] + \mathcal{M}_0[X, C]\right)WJ'G.$$

Since J is invertible and G has full column rank, then using the singular value decomposition of $J'G$, we have

$$WJ'G = W\Gamma\Psi^{1/2}\Xi$$

where Ψ is a g-dimensional diagonal matrix which includes the non-zero eigenvalues of $J'GG'J$, Γ is the $n \times g$ matrix which includes the corresponding eigenvectors so $\Gamma'\Gamma = I_g$ and Ξ is the g-dimensional matrix $\Xi = \Psi^{-1/2}\Gamma'J'G$ so that $\Xi\Xi' = I_g$. Replacing the latter expressions in (3.7), we have

$$|G'JW'\left(\mathcal{M}_0[X,C]\right)WJ'G - \lambda G'JW'\left(\mathcal{M}[X] + \mathcal{M}_0[X,C]\right)WJ'G| = 0, \tag{A.9}$$

which leads to (3.8). In particular, under assumption (2.6), (3.8) reduces to (3.10) where $\mathcal{Z} = Z\Gamma$ and in view of (2.10), the rows of $Z\Gamma$ are $\overset{i.i.d.}{\sim} N(0, I_g)$. It follows that the null distribution of all test statistics which depend on the data via the roots of (3.7) are invariant to B and J. When $G = I_n$, (A.9) takes the form

$$|JW'\left(\mathcal{M}_0[X,C]\right)WJ' - \lambda JW'\left(\mathcal{M}[X] + \mathcal{M}_0[X,C]\right)WJ'| = 0 \tag{A.10}$$

which leads to (3.8) so B and J are evacuated out. □

Proof of Theorem **3.2** Given (2.23)

$$\begin{aligned}|\hat{S}_0|/|\hat{S}| &= \left|I_n + \hat{S}^{-1}(C\hat{B} - D)'[C(X'X)^{-1}C']^{-1}(C\hat{B} - D)\right| \\ &= \left|I_c + [C(X'X)^{-1}C']^{-1}(C\hat{B} - D)\hat{S}^{-1}(C\hat{B} - D)'\right|\end{aligned}$$

using a well known result on determinants, which leads to (3.13).[8] The expression for the statistic in (3.14) associated with (2.24) obtains from (3.6) setting $C = I_k$. □

Proof of Theorem **3.3** In this model, $\hat{U}_t = J\bar{U}_t$. It follows that, for $j = 0, ..., p$ where p is a given lag length

$$\begin{aligned}\left(X_t \otimes \hat{U}_t\right)\left(X_{t-j} \otimes \hat{U}_{t-j}\right)' &= \left(X_t \otimes J\bar{U}_t\right)\left(X_{t-j} \otimes J\bar{U}_{t-j}\right)' \\ &= \left(X_t \otimes J\bar{U}_t\right)\left(X'_{t-j} \otimes \bar{U}'_{t-j}J'\right) = X_tX'_{t-j} \otimes J\bar{U}_t\bar{U}'_{t-j}J' \\ &= \left(I_k \otimes J\right)\left(X_tX'_{t-j} \otimes \bar{U}_t\bar{U}'_{t-j}\right)\left(I_k \otimes J'\right) \\ &= \left(I_k \otimes J\right)\left(X_t \otimes \bar{U}_t\right)\left(X_{t-j} \otimes \bar{U}_{t-j}\right)'\left(I_k \otimes J'\right).\end{aligned}$$

So $\mathbf{\Gamma}_{j,T} = \left(I_k \otimes J\right)\hat{\mathbf{\Gamma}}_{j,T}\left(I_k \otimes J'\right)$ and

$$\mathbf{\Gamma}_{j,T} + \mathbf{\Gamma}'_{j,T} = \left(I_k \otimes J\right)\hat{\mathbf{\Gamma}}_{j,T}\left(I_k \otimes J'\right) + \left(I_k \otimes J\right)\hat{\mathbf{\Gamma}}'_{j,T}\left(I_k \otimes J'\right)$$

[8] For any $n \times m$ matrix S and any $m \times n$ matrix U, $|I_n + SU| = |I_m + US|$; see e.g. Harville (1997, section 18.1, p. 416).

$$
\begin{aligned}
&= (I_k \otimes J)\left[\hat{\mathbf{\Gamma}}_{j,T} + \hat{\mathbf{\Gamma}}'_{j,T}\right](I_k \otimes J')\,. \\
\mathbf{S}_T &= (I_k \otimes J)\,\hat{\Gamma}_{0,T}\,(I_k \otimes J') + \sum_{j=1}^{p}\left(\frac{p-j}{p}\right)(I_k \otimes J)\left[\hat{\mathbf{\Gamma}}_{j,T} + \hat{\mathbf{\Gamma}}'_{j,T}\right](I_k \otimes J') \\
&= (I_k \otimes J)\,\hat{\mathbf{S}}_T\,(I_k \otimes J')\,.
\end{aligned}
$$

The robust Wald statistic may then be rewritten as

$$
\begin{aligned}
\mathcal{W}_p &= T\,\hat{\vartheta}'\mathbf{R}'\left[\mathbf{R}\left(\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right)\mathbf{S}_T\left(\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right)\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\vartheta}, \\
&= T\,\hat{\vartheta}'\mathbf{R}'\left[\mathbf{R}\left(\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right)(I_k \otimes J)\,\hat{\mathbf{S}}_T\,(I_k \otimes J')\left(\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right)\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\vartheta} \\
&= T\,\hat{\vartheta}'\mathbf{R}'\left[\mathbf{R}\left(\left(\frac{X'X}{T}\right)^{-1} \otimes J\right)\hat{\mathbf{S}}_T\left(\left(\frac{X'X}{T}\right)^{-1} \otimes J'\right)\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\vartheta} \\
&= T\,\hat{\vartheta}'\mathbf{R}'\left[\mathbf{R}\,(I_k \otimes J)\left(\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right)\hat{\mathbf{S}}_T\left(\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right)(I_k \otimes J')\,\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\vartheta} \\
&= T\,\hat{\vartheta}'\mathbf{R}'\left[\mathbf{R}\,(I_k \otimes J)\left(\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right)\hat{\mathbf{S}}_T\left(\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right)(I_k \otimes J')\,\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\vartheta} \\
&= \hat{\vartheta}'\mathbf{R}'\left[\mathbf{R}\,(I_k \otimes J)\,\hat{\mathbf{V}}_T\,(I_k \otimes J')\,\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\vartheta}.
\end{aligned}
$$

When $\mathbf{R} = C \otimes I_n$, we can write[9]

$$
\mathbf{R}\vartheta = (C \otimes I_n)\mathrm{vec}(B') = \mathrm{vec}(B'C')\,, \quad \mathbf{R}\hat{\vartheta} = \mathrm{vec}(\hat{B}'C')\,, \tag{A.12}
$$

so that $\mathbf{R}\vartheta = 0 \Leftrightarrow B'C' = 0 \Leftrightarrow CB = 0$, and

$$
\begin{aligned}
\mathcal{W}_p &= \hat{\vartheta}'(C \otimes I_n)'\left[(C \otimes I_n)\,(I_k \otimes J)\,\hat{\mathbf{V}}_T\,(I_k \otimes J')\,(C \otimes I_n)'\right]^{-1}(C \otimes I_n)\hat{\vartheta} \\
&= \hat{\vartheta}'\,(C' \otimes I_n)\left[(C \otimes I_n)\,(I_k \otimes J)\,\hat{\mathbf{V}}_T\,(I_k \otimes J')\,(C' \otimes I_n)\right]^{-1}(C \otimes I_n)\,\hat{\vartheta} \\
&= \mathrm{vec}(\hat{B}'C')'\left[(C \otimes I_n)\,(I_k \otimes J)\,\hat{\mathbf{V}}_T\,(I_k \otimes J')\,(C' \otimes I_n)\right]^{-1}\mathrm{vec}(\hat{B}'C') \\
&= \mathrm{vec}(\hat{B}'C')'\left[(C \otimes J)\,\hat{\mathbf{V}}_T\,(C' \otimes J')\right]^{-1}\mathrm{vec}(\hat{B}'C').
\end{aligned}
$$

[9] We use the identity

$$
\mathrm{vec}(\mathsf{ABC}) = (\mathsf{C}' \otimes \mathsf{A})\mathrm{vec}(\mathsf{B}) \tag{A.11}
$$

where A, B and C are any three matrices such that the matrix product ABC is defined.

Under the null hypothesis $CB=0$, we have:

$$C\hat{B}=C(X'X)^{-1}X'\left[XB+WJ'\right]=C(X'X)^{-1}X'WJ'=C\tilde{B}J'$$

where $\tilde{B}=(X'X)^{-1}X'W$ is an $k\times T$ full column rank matrix, hence[10]

$$\begin{aligned}\mathrm{vec}(\hat{B}'C') &= \mathrm{vec}[J\tilde{B}'C']=(C\otimes J)\mathrm{vec}\big(\tilde{B}'\big)\\ &= (C\otimes I_n)\,(I_k\otimes J)\,\mathrm{vec}\big(\tilde{B}'\big)=(I_r\otimes J)\,(C\otimes I_n)\,\tilde{\vartheta}\end{aligned}$$

where $\tilde{\vartheta}=\mathrm{vec}\big(\tilde{B}'\big)=\mathrm{vec}\big[W'X(X'X)^{-1}\big]$ has a distribution which does not depend on B or J. The Wald statistic can then be rewritten as:

$$\begin{aligned}\mathcal{W}_p &= \tilde{\vartheta}'\,(C'\otimes I_n)\,(I_r\otimes J')\left\{(C\otimes\ J)\,\hat{\mathbf{V}}_T\,(C'\otimes J')\right\}^{-1}(I_r\otimes J)\,(C\otimes I_n)\,\tilde{\vartheta}\\ &= \tilde{\vartheta}'\,(C'\otimes I_n)\,(I_r\otimes J')\left\{(I_r\otimes J)\,(C\otimes\ I_n)\,\hat{\mathbf{V}}_T\,(C'\otimes I_n)\,(I_r\otimes J')\right\}^{-1}\\ &\quad (I_r\otimes J)\,(C\otimes I_n)\,\tilde{\vartheta}\\ &= \tilde{\vartheta}'\,(C'\otimes I_n)\,(I_r\otimes J')\,(I_r\otimes J')^{-1}\left\{(C\otimes\ I_n)\,\hat{\mathbf{V}}_T\,(C'\otimes I_n)\right\}^{-1}\\ &\quad (I_r\otimes J)^{-1}\,(I_r\otimes J)\,(C\otimes I_n)\,\tilde{\vartheta}\\ &= \tilde{\vartheta}'\,(C'\otimes I_n)\left\{(C\otimes\ I_n)\hat{\mathbf{V}}_T(C'\otimes I_n)\right\}^{-1}(C\otimes I_n)\,\tilde{\vartheta}\,.\end{aligned}$$

□

Proof of Theorem **4.1**

$$\begin{aligned}\mathit{\Omega}_*\,(\phi,\eta) &= \frac{(1,\phi')\hat{B}'(X'X)\hat{B}(1,\phi')'-2\eta(1,\phi')\hat{B}'(X'X)\iota_k+\eta^2\iota_k'(X'X)\iota_k}{(1,\phi')Y'\mathcal{M}[X]Y(1,\phi')'} \quad (A.13)\\ \frac{\partial\mathit{\Omega}_*\,(\phi,\eta)}{\partial\eta} &= \frac{-2(1,\phi')\hat{B}'(X'X)\iota_k+2\eta\iota_k'(X'X)\iota_k}{(1,\phi')Y'\mathcal{M}[X]Y(1,\phi')'}\end{aligned}$$

and the (non-zero) value of η which sets the latter partial derivative to zero is

$$\tilde{\eta}\,(\phi)=\frac{(1,\phi')\hat{B}'(X'X)\iota_k}{\iota_k'(X'X)\iota_k}. \quad (A.14)$$

Substituting $\tilde{\eta}\,(\phi)$ in (A.13) leads to

$$\mathit{\Omega}_*\,(\phi,\tilde{\eta}\,(\phi))=\frac{(1,\phi')\hat{B}'\left[X'X-(X'X)\iota_k\,(\iota_k'(X'X)\iota_k)^{-1}\,\iota_k'(X'X)\right]\hat{B}(1,\phi')'}{(1,\phi')\,(Y'\mathcal{M}[X]Y)\,(1,\phi')'}$$

[10] On using again (A.11).

so $\min_\phi \Omega_* (\phi, \eta) = \hat{\sigma}$. This is exactly the same solution obtained (using another method of proof) by Fujikoshi (1974); note that this author did not provide analytical formulae for point estimates. In the same vein,

$$\Lambda_* (\zeta, \mu) = \frac{(1, \zeta')\hat{B}\hat{S}^{-1}\hat{B}'(1, \zeta')' - 2\mu(1, \zeta')\hat{B}\hat{S}^{-1}\iota_n + \mu^2\iota_n'\hat{S}^{-1}\iota_n}{(1, \zeta')(X'X)^{-1}(1, \zeta')'} \tag{A.15}$$

$$\frac{\partial\Lambda_* (\zeta, \mu)}{\partial\mu} = \frac{-2(1, \zeta')\hat{B}\hat{S}^{-1}\iota_n + 2\mu\iota_n'\hat{S}^{-1}\iota_n}{(1, \zeta')(X'X)^{-1}(1, \zeta')'}$$

and the (non-zero) value of μ which sets the latter partial derivative to zero is

$$\tilde{\mu} (\zeta) = \frac{(1, \zeta')\hat{B}\hat{S}^{-1}\iota_n}{\iota_n'\hat{S}^{-1}\iota_n}. \tag{A.16}$$

Substituting $\tilde{\mu} (\zeta)$ in (A.15) leads to

$$\begin{aligned}\Lambda_* (\zeta, \tilde{\mu} (\zeta)) &= \frac{(1, \zeta')\hat{B}\hat{S}^{-1}\hat{B}'(1, \zeta')' - (1, \zeta')\hat{B}\hat{S}^{-1}\iota_n \left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1} \iota_n'\hat{S}^{-1}\hat{B}'(1, \zeta')'}{(1, \zeta')(X'X)^{-1}(1, \zeta')'} \\ &= \frac{(1, \zeta')\hat{B} \left[\hat{S}^{-1} - \hat{S}^{-1}\iota_n \left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1} \iota_n'\hat{S}^{-1}\right] \hat{B}'(1, \zeta')'}{(1, \zeta')(X'X)^{-1}(1, \zeta')'}\end{aligned}$$

so $\min_{\zeta,\mu} \Lambda_* (\zeta, \mu) = \hat{\nu}$. □

Proof of Theorem **4.2** $\min_\theta \Lambda (\theta)$ is the minimum root [denoted $\hat{\gamma}$] of the determinantal equation (4.6) so the minimization problem can be cast as an equation of the (4.10) where $\pi = \hat{\theta}$, where $\hat{\theta}$ is the point estimate of θ and

$$\Sigma = \hat{B}\hat{S}^{-1}\hat{B}' - \hat{\gamma}(X'X)^{-1}, \tag{A.17}$$

$\Sigma_{11} = \hat{a}'\hat{S}^{-1}\hat{a} - \hat{\gamma}x^{11}$, $\Sigma_{12} = \Sigma_{21}' = \hat{a}'\hat{S}^{-1}\hat{b}' - \hat{\gamma}x^{12}$, $\Sigma_{22} = \hat{b}\hat{S}^{-1}\hat{b}' - \hat{\gamma}x^{22}$, using the partitioning (2.14). So $\hat{\theta}$ obtains as in (4.11). $\min_\delta \Omega (\delta)$ is the minimum root [which coincides with $\hat{\gamma}$] of the determinantal equation (4.7), so again the minimization problem takes the (4.10) form with $\pi = \hat{\delta}$

$$\Sigma = Y'X(X'X)^{-1}X'Y - \hat{\gamma}Y'\mathcal{M}[X]Y \tag{A.18}$$

$\Sigma_{11} = y'X(X'X)^{-1}X'y - \hat{\gamma}y'\mathcal{M}[X]y$, $\Sigma_{12} = \Sigma_{21}' = y'X(X'X)^{-1}X'\mathbf{Y} - \hat{\gamma}y'\mathcal{M}[X]\mathbf{Y}$, $\Sigma_{22} = \mathbf{Y}'X(X'X)^{-1}X'\mathbf{Y} - \hat{\gamma}\mathbf{Y}'\mathcal{M}[X]\mathbf{Y}$, using the partitioning (2.16). So the point estimate of $\bar{\delta}$ denoted $\hat{\delta}$ obtain as in (4.11). $\min_{\zeta,\mu} \Lambda_* (\zeta, \mu)$ requires [refer to the proof of Theorem

4.1] to solve a system of the (4.10) form with $\pi = \zeta$, and

$$\begin{aligned}
\Sigma &= \hat{B}\left[\hat{S}^{-1} - \hat{S}^{-1}\iota_n\left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1}\iota_n'\hat{S}^{-1}\right]\hat{B}' - \hat{\nu}(X'X)^{-1} \\
\Sigma_{11} &= \hat{a}'\left[\hat{S}^{-1} - \hat{S}^{-1}\iota_n\left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1}\iota_n'\hat{S}^{-1}\right]\hat{a} - \hat{\nu}x^{11} \\
\Sigma_{12} &= \Sigma_{21}' = \hat{a}'\left[\hat{S}^{-1} - \hat{S}^{-1}\iota_n\left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1}\iota_n'\hat{S}^{-1}\right]\hat{b}' - \hat{\nu}x^{12} \\
\Sigma_{22} &= \hat{b}\left[\hat{S}^{-1} - \hat{S}^{-1}\iota_n\left(\iota_n'\hat{S}^{-1}\iota_n\right)^{-1}\iota_n'\hat{S}^{-1}\right]\hat{b}' - \hat{\nu}x^{22}
\end{aligned}$$

using the partitionings (2.14) and (2.17). So a point estimate for $\bar{\zeta}$ [denoted $\hat{\zeta}$] obtains as in (4.11) and an point estimate for $\bar{\mu}$ thus follows using (A.16) leading to (4.14). This is exactly the same solution obtained (using another method of proof) by Shanken and Zhou (2007). $\min_{\phi,\eta} \Omega_* (\phi, \eta)$ requires [refer to the proof of Theorem **4.1**] to solve a system of the (4.10) form with $\pi = \phi$, and

$$\begin{aligned}
\Sigma &= Y'X\left(I_k - \iota_k\left(\iota_k'(X'X)\iota_k\right)^{-1}\iota_k'\right)X'Y - \hat{\sigma}Y'\mathcal{M}[X]Y \\
\Sigma_{11} &= y'X\left(I - \iota_k\left(\iota_k'(X'X)\iota_k\right)^{-1}\iota_k'\right)X'y - \hat{\sigma}y'\mathcal{M}[X]y \\
\Sigma_{12} &= \Sigma_{21}' = y'X\left(I - \iota_k\left(\iota_k'(X'X)\iota_k\right)^{-1}\iota_k'\right)X'\mathbf{Y} - \hat{\sigma}y'\mathcal{M}[X]\mathbf{Y} \\
\Sigma_{22} &= \mathbf{Y}'X\left(I - \iota_k\left(\iota_k'(X'X)\iota_k\right)^{-1}\iota_k'\right)X'\mathbf{Y} - \hat{\sigma}\mathbf{Y}'\mathcal{M}[X]\mathbf{Y}
\end{aligned}$$

using the partitioning (2.16). So a point estimate for $\bar{\phi}$ [denoted $\hat{\phi}$] obtains as in (4.11) and an point estimate for $\bar{\eta}$ thus follows using (A.14) leading to (4.14). □

Proof of Theorem **5.1** Equations (A.4) - - (A.8) applied with A as defined in (5.3) imply that an unbounded solution to the problem of inverting the test defined by (3.16) and (3.21) would occur if A_{22} [refer to the partitioning in (5.2) and (5.4)] is not positive definite. In this case, the diagonal term of A_{22} is given by $\text{DIAG}(A_{22}) = (\; F_2 \;\; \ldots \;\; F_k \;)'$ where

$$F_i = s_k[i]'\hat{B}\hat{S}^{-1}\hat{B}'s_k[i] - s_k[i]'(X'X)^{-1}s_k[i]\frac{nf_{n,\tau_n,\alpha}}{\tau_n}.$$

Clearly, if any of the Hotelling tests based on Λ_i, $i \in \{2, ..., k\}$ [as in (3.20) and using the distribution in (3.21)] is not significant at level α, then by the definition of Λ_i and F_i, $\Lambda_i(\tau_n)/n < f_{n,\tau_n,\alpha} \Leftrightarrow F_i < 0$, in which case A_{22} cannot be positive definite. On comparing (5.3) and (5.5) we see that $\Lambda_i(\tau_{n-1})/(n-1) \geq f_{n-1,\tau_{n-1},\alpha}$, $i \in \{2, ..., k\}$ holds for the problem of inverting the test defined by (3.17) and (3.21) as a necessary but not sufficient condition to obtain bounded CSs. □

PROOF OF THEOREM **5.2** Equations (A.4) - - (A.8) applied with A as defined in (5.7) imply that an unbounded solution to the problem of inverting the test defined by (3.23) and (3.28) would occur if A_{22} [refer to the partitioning in (5.2) and (5.8)] is not positive definite. In this case, the diagonal term of A_{22} is given by $\text{DIAG}(A_{22}) = (\ F_2 \ \ \dots \ \ F_n\ )'$ where

$$F_i = s_n[i]'\mathbf{Y}'X(X'X)^{-1}X'\mathbf{Y}s_n[i] - s_n[i]'\mathbf{Y}'\mathcal{M}[X]\mathbf{Y}s_n[i]\frac{kf_{k,\tau_k,\alpha}}{\tau_k}.$$

Clearly, if any of the Hotelling tests based on Ω_i, $i \in \{2, ..., n\}$ [as in (3.27) and using the distribution in (3.28)] is not significant at level α, then by the definition of Ω_i and F_i, $\Omega_i(\tau_k)/k < f_{k,\tau_k,\alpha} \Leftrightarrow F_i < 0$, in which case A_{22} cannot be positive definite. On comparing (5.7) and (5.9) we see that $\Omega_i(\tau_{k-1})/(k-1) \geq f_{k-1,\tau_{k-1},\alpha}$, $i \in \{2, ..., n\}$ holds for the problem of inverting the test defined by (3.24) and (3.28) as a necessary but not sufficient condition to obtain bounded CSs. □

PROOF OF THEOREM **5.3** By the definition of the test inversion procedures underlying the CSs under consideration and using (3.21)-(3.28) we have:

$$\begin{aligned}
\text{CS}_\alpha(\bar{\theta}) &= \varnothing \Leftrightarrow \min_{\theta} \Lambda(\theta) \geq f_{n,\tau_n,\alpha}, \quad \text{CS}_\alpha((\bar{\zeta}', \bar{\mu})') = \varnothing \Leftrightarrow \min_{\zeta,\mu} \Lambda_*(\zeta, \mu) \geq f_{n-1,\tau_{n-1},\alpha} \\
\text{CS}_\alpha(\bar{\delta}) &= \varnothing \Leftrightarrow \min_{\delta} \Omega(\delta) \geq f_{k,\tau_k,\alpha}, \quad \text{CS}_\alpha((\bar{\phi}', \bar{\eta})') = \varnothing \Leftrightarrow \min_{\phi,\eta} \Omega_*(\phi, \eta) \geq f_{k-1,\tau_{k-1},\alpha}.
\end{aligned}$$

Then (5.11) and (5.12) follow from the minimization results in (4.3) and Theorem **4.1**. □

PROOF OF THEOREM **6.1** Suppose that Y is replaced by $Y_\top$, and define $\bar{\Sigma}_\top(B_\top) \equiv \frac{1}{T}(Y_\top - XB_\top)'(Y_\top - XB_\top)$ where $B_\top = BM$. Using (6.15), we have $\bar{\Sigma}_\top(B_\top) = M'\bar{\Sigma}(B)M$ so that

$$\left|\bar{\Sigma}_\top(B_\top)\right| = |M'M|\,|\bar{\Sigma}(B)|$$

for any M and any $B_\top = BM$. This implies that $|\bar{\Sigma}_\top(B_\top)|$ is minimized by $\hat{B}_\top = \hat{B}M$, the corresponding sum of squared residuals obtain as $\hat{S}_\top = M'\hat{S}M$ and $|\bar{\Sigma}_\top(\hat{B}_\top)| = |M'M||\bar{\Sigma}(\hat{B})|$. Similarly, suppose that X is replaced by $X_\perp$, and define $\bar{\Sigma}_\perp(B_\perp) \equiv \frac{1}{T}(Y - X_\perp B_\perp)'(Y - XB_\perp)$ where $B_\perp = K^{-1}B$. Using (6.16), we have $\bar{\Sigma}_\perp(B_\perp) = \bar{\Sigma}(B)$ so that

$$\left|\bar{\Sigma}_\perp(B_\perp)\right| = |\bar{\Sigma}(B)|$$

for any K and any $B_\perp = K^{-1}B$. This implies that $|\bar{\Sigma}_\perp(B_\perp)|$ is minimized by $\hat{B}_\perp = K^{-1}\hat{B}$, the corresponding sum of squared residuals obtain as $\hat{S}_\perp = \hat{S}$ and $|\bar{\Sigma}_\perp(\hat{B}_\perp)| = |\bar{\Sigma}(\hat{B})|$. Now observe that for any full row-rank rank matrix C, $CB = 0 \Longleftrightarrow CBM = 0 \Longleftrightarrow$

$CB_\top = 0$, and for any full column-rank matrix $G, BG = 0 \Longleftrightarrow K^{-1}BG = 0 \Longleftrightarrow B_\perp G = 0$. It follows that the restricted estimators of B under $\mathbf{H}(C, I_n, 0)$ and $\mathbf{H}(I_k, G, 0)$ are transformed in the same way, respectively:

$$\hat{B}_{0\top} = \hat{B}_0 M, \quad \hat{B}_{0\perp} = K^{-1}\hat{B}_0$$

where $\hat{B}_{0\top}$ is the constrained [imposing $\mathbf{H}(C, I_n, 0)$] estimator from the transformed model [following transformation (6.15)] and $\hat{B}_{0\perp}$ is the constrained [imposing $\mathbf{H}(I_k, G, 0)$] estimator from the transformed model [following transformation (6.16)]. Indeed, let $B_{0\top}(C, I_n, 0)$ and $S_{0\top}(C, I_n, 0)$ correspond to $B_0(C, G, 0)$ and $S_0(C, G, 0)$ with $G = I_n$ where Y is replaced by $Y_\top$. Similarly, let $B_{0\perp}(I_k, G, 0)$ and $S_{0\perp}(I_k, G, 0)$ correspond to $B_0(C, G, 0)$ and $S_0(C, G, 0)$ with $C = I_k$ where X is replaced by $X_\perp$. From (3.1)-(3.2), we see that

$$\begin{aligned} B_{0\top}(C, I_n, 0) &= B_0(C, I_n, 0)\, M, \quad S_{0\perp}(C, I_n, 0) = M' S_0(C, I_n, 0)\, M, \\ B_{0\perp}(I_k, G, 0) &= K^{-1} B_0(I_k, G, 0), \quad S_{0\perp}(I_k, G, 0) = S_0(I_k, G, 0), \end{aligned}$$

for any full row-rank rank matrix C and for any full column-rank matrix G. So conforming with (3.13)-(3.14), we have:

$$\bar{\Lambda}_\top(C, I_n, 0) = [C(X_\perp' X_\perp)^{-1}C']^{-1}C\hat{B}_\top \hat{S}_\top^{-1}\hat{B}'C' = \bar{\Lambda}(C, I_n, 0) \quad \text{(A.19)}$$

$$\bar{\Omega}_\perp(I_k, G, 0) = [G'Y_\top' \mathcal{M}[X] Y_\top G]^{-1} G'\hat{B}_\perp' X'X\hat{B}_\perp G = \bar{\Omega}(I_k, G, 0) \quad \text{(A.20)}$$

so the variance ratios are unaffected by each of the transformations (6.15) or (6.16), respectively. It also follows that

$$\inf_C \bar{\Lambda}_\top(C, I_n, 0) = \inf_C \bar{\Lambda}(C, I_n, 0), \quad \inf_G \bar{\Omega}_\perp(I_k, G, 0) = \inf_G \bar{\Omega}(I_k, G, 0).$$

This holds true a-fortiori for the special cases $\bar{\Lambda}_\top((1, \theta'), I_n, 0)$ and $\bar{\Omega}_\perp(I_k, (1, \delta')', 0)$ where $\bar{\Lambda}_\top((1, \theta'), I_n, 0) = \bar{\Lambda}((1, \theta'), I_n, 0)$, for any θ and $\bar{\Omega}_\perp(I_k, (1, \delta')', 0) = \bar{\Omega}(I_k, (1, \delta')', 0)$, for any δ, so

$$\inf_\theta \bar{\Lambda}_\top((1, \theta'), I_n, 0) = \inf_\theta \bar{\Lambda}((1, \theta'), I_n, 0), \quad \inf_\delta \bar{\Omega}(I_k, (1, \delta')', 0) = \inf_\delta \bar{\Omega}(I_k, (1, \delta')', 0).$$

□

PROOF OF THEOREM **6.2** Consider a transformation of the form $Y_\top = Y(J')^{-1}$. Using

(2.1) we then have:

$$Y_{\top} = (XB + WJ')\,(J')^{-1} = XB\,(J')^{-1} + W = X\bar{B} + W$$

In view of the invariance established in Lemma **6.1**, the statistics $\mathcal{L}_{\Lambda}(\theta)$ and $\mathcal{L}_{\Lambda}(\hat{\theta})$ can be viewed as functions of $Y_{\top}$, which entails that they depend on the model parameters (B, J) only through the lower dimensional vector $\bar{B}$. Furthermore, under the null hypothesis, the nuisance parameter only involves θ and $b\,(J')^{-1}$ where b is the $q \times n$ matrix defined in (2.13). Now the distribution of $\mathcal{L}_{\Lambda}(\theta)$ can be explicitly characterized by using the derivations from Theorem **3.2** as follows:

$$\mathcal{L}_{\Lambda}(\theta) = T\,\ln\left(|\hat{U}_{\Lambda}'(\theta)\hat{U}_{\Lambda}(\theta)|/|W'\mathcal{M}[X]W|\right),$$

where $\hat{U}_{\Lambda}(\theta) = \bar{M}_{\Lambda}(\theta)(X\bar{B} + W) = \bar{M}_{\Lambda}(\theta)\,[\iota_T a' + \mathbf{X}b] + \bar{M}_{\Lambda}(\theta)W$ (using the partitioning of X as in (2.12) and of B as in (2.13))

$$\begin{aligned}
\bar{M}_{\Lambda}(\theta) &= \mathcal{M}[X] + \mathcal{M}_0[X,(1,\theta')] \\
\mathcal{M}_0[X,(1,\theta')] &= \mathcal{M}[X] + X(X'X)^{-1}(1,\theta')'[(1,\theta')(X'X)^{-1}(1,\theta')']^{-1}(1,\theta')(X'X)^{-1}X'
\end{aligned}$$

conforming with the definition (3.9) and

$$\begin{aligned}
\hat{U}_{\Lambda}(\theta) &= \bar{M}_{\Lambda}(\theta)\,[\iota_T a' + \mathbf{X}b - \iota_T\theta' b + \iota_T\theta' b]\,(J')^{-1} + \bar{M}_{\Lambda}(\theta)W, \\
&= \bar{M}_{\Lambda}(\theta)\,[\iota_T(a' + \theta' b) + (\mathbf{X} - \iota_T\theta')\,b]\,(J')^{-1} + \bar{M}_{\Lambda}(\theta)W, \\
&= \bar{M}_{\Lambda}(\theta)\,[\iota_T(a' + \theta' b)]\,(J')^{-1} + \bar{M}_{\Lambda}(\theta)\,(\mathbf{X} - \iota_T\theta')\,b\,(J')^{-1} + \bar{M}_{\Lambda}(\theta)W \\
&= \bar{M}_{\Lambda}(\theta)\,[\iota_T(a' + \theta' b)]\,(J')^{-1} + \bar{M}_{\Lambda}(\theta)W
\end{aligned}$$

since $\bar{M}_{\Lambda}(\theta)\,(\mathbf{X} - \iota_T\theta') = 0$. If, furthermore, the null hypothesis $\mathcal{H}_1(\bar{\theta})$ (2.2) holds, which implies that there exists some unknown scalar say $\bar{\theta}$ such that $a = b\bar{\theta}$, then

$$\hat{U}_{\Lambda}(\theta) = \bar{M}_{\Lambda}(\theta)\,\left[\iota_T(\bar{\theta} - \theta)' b\right]\,(J')^{-1} + \bar{M}_{\Lambda}(\theta)W$$

and the distribution of $\mathcal{L}(\hat{\theta})$ depends on (B, J') only through $\bar{\theta}$ and $b\,(J')^{-1}$. The rest follows form the definition of $\mathcal{L}_{\Lambda}(\hat{\theta}) = \inf_{\theta}\mathcal{L}_{\Lambda}(\theta)$ and Theorem **6.1**. □

.3. Information matrix and Standard Errors

We derive the formulae for the regular standard errors of the estimates for the model underlying our empirical analysis [(2.1)- (2.2)]. Standard errors for the estimates of θ and μ under (2.4) are provided in Campbell et al. (1997, Chapter 6); see also Barone-Adesi et al. (2004b). It is straightforward to show (by first computing the partial derivatives of the Gaussian likelihood, denoted $\boldsymbol{L}(\theta, \beta_2, ..., \beta_k, J)$, with respect to $\theta, \beta_2, ..., \beta_k$ and then taking their negative expectation) that the information sub-matrix relevant to θ and B takes the form

$$\ddot{I}\left(\theta, \beta_2, ..., \beta_k\right) = \begin{bmatrix} \ddot{I}_{11} & \ddot{I}_{12} \\ \ddot{I}_{21} & \ddot{I}_{22} \end{bmatrix}, \quad \ddot{I}^{-1}\left(\theta, \beta_2, ..., \beta_k\right) = \begin{bmatrix} \ddot{I}^{11} & \ddot{I}^{12} \\ \ddot{I}^{21} & \ddot{I}^{22} \end{bmatrix} \tag{A.21}$$

where $\ddot{I}_{11}$ is $q \times q$, $\ddot{I}_{22}$ is $nq \times nq$ and $\ddot{I}_{21} = \ddot{I}_{12}'$ is $nq \times q$ such that

$$\begin{aligned} \ddot{I}_{11} &= E\left(\frac{-\partial^2 \boldsymbol{L}(\theta, \beta_2, ..., \beta_k, J)}{\partial\theta\partial\theta'}\right) = Tb\Upsilon^{-1}b' \\ \ddot{I}_{22} &= \left[E\left(\frac{-\partial^2 \boldsymbol{L}(\theta, \beta_2, ..., \beta_k, J)}{\partial\beta_i\partial\beta_j'}\right)_{i,j=2,...,k}\right] = \left(\mathbf{X} - i_T\theta'\right)'\left(\mathbf{X} - i_T\theta'\right) \otimes \Upsilon^{-1} \\ \ddot{I}_{21} &= \left[E\left(\frac{-\partial^2 \boldsymbol{L}(\theta, \beta_2, ..., \beta_k, J)}{\partial\beta_i\partial\theta'}\right)_{i=2,...,k}\right] = T(\overline{\mathbf{X}} - \theta) \otimes \Upsilon^{-1}b' \end{aligned}$$

where $\Upsilon = JJ'$ and $\overline{\mathbf{X}}$ is the $q \times 1$ vector of the (time series) column means of $\mathbf{X}$. Using the formulae for partitioned matrix inversion, and the notation in (A.21)

$$\begin{aligned} \ddot{I}^{11} &= \left(Tb\Upsilon^{-1}b'\right)^{-1}\left(1 - T(\overline{\mathbf{X}} - \theta)'Q_{\mathbf{X}}^{-1}\left(\theta\right)\left(\overline{\mathbf{X}} - \theta\right)\right)^{-1} \\ \ddot{I}^{21} &= \left(\left[-Q_{\mathbf{X}}^{-1}\left(\theta\right)\left(\overline{\mathbf{X}} - \theta\right)\right] \otimes b'\right)\ddot{I}^{11} \\ \ddot{I}^{22} &= \left[Q_{\mathbf{X}}^{-1}\left(\theta\right)/T \otimes \Upsilon\right] + \left(\left[Q_{\mathbf{X}}^{-1}\left(\theta\right)\left(\overline{\mathbf{X}} - \theta\right)\right] \otimes b'\right)\ddot{I}^{11}\left(\left[(\overline{\mathbf{X}} - \theta)'Q_{\mathbf{X}}^{-1}\left(\theta\right)\right] \otimes b\right). \end{aligned}$$

where $Q_{\mathbf{X}}\left(\theta\right) = \left(\mathbf{X} - i_T\theta'\right)'\left(\mathbf{X} - i_T\theta'\right)/T$. Variance/covariance matrices for the estimators of b and θ can thus be obtained analytically using the latter expressions. The delta-method can be applied to derive the variance/covariance matrix denoted $[\ddot{A}]$ of the (constrained) estimate of the intercept vector: $\ddot{A} = \ddot{a}'\ddot{I}^{-1}\left(\theta, \beta_2, ..., \beta_k\right)\ddot{a}, \quad \ddot{a} = \begin{bmatrix} b \\ (\theta \otimes I_n) \end{bmatrix}.$

References

Anderson, T. W. (1951), 'Estimating linear restrictions on regression coefficients for multivariate normal distributions', *Annals of Mathematical Statistics* **22**, 327–351.

Anderson, T. W. (1999), 'Asymptotic distribution of the reduced rank regression estimator under general conditions', *The Annals of Statistics* **27**, 1141–1154.

Anderson, T. W. (2006), 'Reduced rank regression for blocks of simultaneous equations', *Journal of Econometrics* **135**, 55–76.

Andrews, D. W. K., Moreira, M. J. and Stock, J. H. (2006), 'Optimal two-sided invariant similar tests for instrumental variables regression', *Econometrica* **74**, 715–752.

Barone-Adesi, G. (1985), 'Arbitrage equilibrium with skewed asset returns', *Journal of Financial and Quantitative Analysis* **20**(3), 299–313.

Barone-Adesi, G., Gagliardini, P. and Urga, G. (2004*a*), A test of the homogeneity hypothesis on asset pricing models, *in* E. Jurczenko and B. Maillett, eds, 'Multi-Moment CAPM and Related Topics', Springer-Verlag, Berlin.

Barone-Adesi, G., Gagliardini, P. and Urga, G. (2004*b*), 'Testing asset pricing models with coskewness', *Journal of Business and Economic Statistics* **22**, 474–485.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006), 'Multivariate GARCH models: a survey', *Journal of Applied Econometrics* **21**, 79–109.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2005), Exact multivariate tests of asset pricing models with stable asymmetric distributions, *in* M. Breton and H. Ben Ameur, eds, 'Numerical Methods in Finance', Kluwer, Dordrecht, The Netherlands.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2007), 'Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: An exact simulation-based approach', *Journal of Business and Economic Statistics* **25**, 398–410.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2008*a*), Asset-pricing anomalies and spanning: Multivariate multifactor tests with heavy tailed distributions, Technical report, Mc Gill University, Université Laval and Carleton University.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2008*b*), Finite-sample identification-robust inference for unobservable zero-beta rates and portfolio efficiency with non-Gaussian distributions, Technical report, Mc Gill University, Université Laval and Carleton University.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2008*c*), 'Finite sample multivariate tests of asset pricing models with coskewness', *Computational Statistics and Data Analysis* **forthcoming**.

Bekker, P., Dobbelstein, P. and Wansbeek, T. (1996), 'The APT model as reduced-rank regression', *Journal of Business and Economic Statistics* **14**, 199–202.

Bennett, B. M. (1959), 'On a multivariate version of Fieller's theorem', *Journal of the Royal Statistical Society, Series B* **21**, 59–62.

Bennett, B. M. (1961), 'Confidence limits for multiple ratios', *Journal of the Royal Statistical Society, Series B* **23**, 108–112.

Berndt, E. R. and Savin, N. E. (1977), 'Conflict among criteria for testing hypotheses in the multivariate linear regression model', *Econometrica* **45**, 1263–1277.

Black, F. (1972), 'Capital market equilibrium with restricted borrowing', *Journal of Business* **45**, 444–454.

Bolduc, D., Khalaf, L. and Yelou, C. (2008), 'Identification robust confidence sets methods for inference on parameter ratios with application to discrete choice models', *Computational Statistics and Data Analysis* **forthcoming**.

Calinsky, T. and Lejeune, M. (1998), 'Dimensionality in manova tested by a closed testing procedure', *Journal of Multivariate Analysis* **65**, 181–194.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

Costa, M., Gardini, A. and Paruolo, P. (1997), 'A reduced rank regression approach to tests of asset pricing', *Oxford Bulletin of Economics and Statistics* **59**, 163–181.

Cragg, J. G. and Donald, S. G. (1997), 'Inferring the rank of a matrix', *Journal of Econometrics* **76**, 223–250.

Davidson, R. and MacKinnon, J. G. (2004), *Econometric Theory and Methods*, Oxford University Press, New York.

Dhrymes, P. J. (1974), *Econometrics: Statistical Foundations and Applications*, Springer-Verlag, New York.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M. (2006), 'Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **133**, 443–478.

Dufour, J.-M. and Jasiak, J. (2001), 'Finite sample limited information inference methods for structural equations and models with generated regressors', *International Economic Review* **42**, 815–843.

Dufour, J.-M. and Khalaf, L. (2002), 'Simulation based finite and large sample tests in multivariate regressions', *Journal of Econometrics* **111**(2), 303–322.

Dufour, J.-M., Khalaf, L. and Beaulieu, M.-C. (2003), 'Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models', *Oxford Bulletin of Economics and Statistics* **65**, 891–906.

Dufour, J.-M., Khalaf, L. and Beaulieu, M.-C. (2008), 'Multivariate residual-based finite-sample tests for serial dependence and GARCH with applications to asset pricing models', *Journal of Applied Econometrics* **forthcoming**.

Dufour, J.-M. and Taamouti, M. (2005), 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica* **73**, 1351–1365.

Dufour, J.-M. and Taamouti, M. (2007), 'Further results on projection-based inference in IV regressions with weak, collinear or missing instruments', *Journal of Econometrics* **139**, 133–153.

Engle, R. F. and Kroner, K. F. (1995), 'Multivariate generalized ARCH', *Econometric Theory* **11**, 122–150.

Engle, R. F., Sheppard, K. and Sheppard, N. (2008), Fitting and testing vast dimensional time-varying covariance, Technical report, Nuffield College Economics Working Papers 2008-W11.

Fama, E. F. and French, K. R. (1992), 'The cross-section of expected stock returns', *Journal of Finance* **47**, 427–465.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**, 3–56.

Fieller, E. C. (1954), 'Some problems in interval estimation', *Journal of the Royal Statistical Society, Series B* **16(2)**, 175–185.

Fujikoshi, Y. (1974), 'The likelihood ratio tests for the dimensionality of regression coefficients', *Journal of Multivariate Analysis* **4**, 327–340.

Geweke, J. (1996), 'Bayesian reduced rank regression in econometrics', *Journal of Econometrics* **75**, 121–146.

Gibbons, M. R. (1982), 'Multivariate tests of financial models: A new approach', *Journal of Financial Economics* **10**, 3–27.

Gouriéroux, C., Monfort, A. and Renault, E. (1995), Inference in factor models, *in* G. S. Maddala, P. C. B. Phillips and T. N. Srinivasan, eds, 'Advances in Econometrics and Quantitative Economics', Blackwell, Oxford, U.K., chapter 13, pp. 311–353.

Gungor, S. and Luger, R. (2008), Exact distribution-free tests of mean-variance efficiency, Technical report, Department of Economics, Emory University.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Hoogerheide, L. F., Kaashoek, J. F. and van Dijk, H. K. (2007), 'On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: An application of flexible sampling methods using neural networks', *Journal of Econometrics* **139**(4), 154–180.

Hotelling, H. (1947), *Multivariate Quality Control Illustrated by the Air Testing of Sample Bomb Sights,Techniques of Statistical Analysis, Ch. II*, McGraw-Hill, New York.

Johansen, S. (1995), *Likelihood-based Inference in Cointegrated Vector Autoregressive Model*, second edn, Oxford University Press, New York.

Kan, R. and Zhang, C. (1999*a*), 'GMM tests of stochastic discount factor models with useless factors', *Journal of Financial Economics* **54**(1), 103–127.

Kan, R. and Zhang, C. (1999*b*), 'Two-pass tests of asset pricing models with useless factors', *Journal of Finance* **54**, 204–235.

Kleibergen, F. (1999), Reduced rank regression using GMM, *in* L. Mátyás, ed., 'Generalized Method of Moments Estimation', Cambridge University Press, Cambridge, U.K., chapter 7, pp. 171–210.

Kleibergen, F. (2002), 'Pivotal statistics for testing structural parameters in instrumental variables regression', *Econometrica* **70**, 1781–1803.

Kleibergen, F. (2005), 'Testing parameters in gmm without assuming that they are identified', *Econometrica* **73**, 1103–1123.

Kleibergen, F. and Paap, R. (1998), 'Generalized reduced rank tests using the singular value decomposition', *Journal of Econometrics* **133**, 97–126.

Kleibergen, F. and van Dijk, H. K. (1998), 'Bayesian simultaneous equations analysis using reduced rank structures', *Econometric Theory* **14**, 699–744.

Lewellen, J., Nagel, S. and Shanken, J. (2006), A skeptical appraisal of asset-pricing tests, Technical report, NBER Working Papers 12360.

Lewellen, J. W. and Nagel, S. (2006), 'The conditional CAPM does not explain asset-pricing anomalies', *Journal of Financial Economics* **82**, 289–314.

McKeon, J. J. (1974), 'F approximation to the distribution of hotelling's t^2', *Journal of Business and Economic Statistics* **15**, 74–81.

Moreira, M. J. (2003), 'A conditional likelihood ratio test for structural models', *Econometrica* **71**(4), 1027–1048.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, second edn, John Wiley & Sons, New York.

Ravikumar, B., Ray, S. and Savin, E. (2000), 'Robust wald tests in sur systems with adding-up restrictions', *Econometrica* **68**, 715–720.

Ray, S. and Savin, E. (2008), 'The performance of heteroskedasticity and autocorrelation robust tests: a Monte Carlo study with an application to the three-factor Fama-French asset-pricing model', *Journal of Applied Econometrics* **23**, 91–109.

Reinsel, G. and Velu, R. (1998), *Multivariate Reduced-Rank Regression, Theory and Applications*, Springer-Verlag, New York.

Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *Journal of Economic Theory* **13**, 341–360.

Savin, N. E. (1984), Multiple hypothesis testing, *in* Z. Griliches and M. D. Intrilligator, eds, 'Handbook of Econometrics, Volume 2', North-Holland, Amsterdam, chapter 14, pp. 827–879.

Shanken, J. (1985), 'Multivariate tests of the zero-beta CAPM', *Journal of Financial Economics* **14**, 325–348.

Shanken, J. (1986), 'Testing portfolio efficiency when the zero-beta rate is unknown: A note', *Journal of Finance* **41**, 269–276.

Shanken, J. (1996), Statistical methods in tests of portfolio efficiency: A synthesis, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics 14: Statistical Methods in Finance', North-Holland, Amsterdam, pp. 693–711.

Shanken, J. and Weinstein, M. I. (2006), 'Economic forces and the stock market revisited', *Journal of Empirical Finance* **13**, 129–144.

Shanken, J. and Zhou, G. (2007), 'Estimating and testing beta pricing models: Alternative methods and their performance in simulations', *Journal of Financial Economics* **84**, 40–86.

Staiger, D. and Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**(3), 557–586.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business and Economic Statistics* **20**(4), 518–529.

Velu, R. and Zhou, G. (1999), 'Testing multi-beta asset pricing models', *Journal of Empirical Finance* **6**, 219–241.

Wang, J. and Zivot, E. (1998), 'Inference on structural parameters in instrumental variables regression with weak instruments', *Econometrica* **66**(6), 1389–1404.

Zerbe, G. O., Laska, E., Meisner, M. and Kushner, H. B. (1982), 'On multivariate confidence regions and simultaneous confidence limits for ratios', *Communications in Statistics, Theory and Methods* **11**, 2401–2425.

Zhou, G. (1991), 'Small sample tests of portfolio efficiency', *Journal of Financial Economics* **30**, 165–191.

Zhou, G. (1995), 'Small sample rank tests with applications to asset pricing', *Journal of Empirical Finance* **2**, 71–93.

Zivot, E., Startz, R. and Nelson, C. R. (1998), 'Valid confidence intervals and inference in the presence of weak instruments', *International Economic Review* **39**, 1119–1144.