Simulation Based Finite and Large Sample Tests in Multivariate Regressions [a]

Jean-Marie Dufour [b]
Université de Montréal

Lynda Khalaf [c]
Université Laval

First version: November 1996
Revised: August 1998
This version: January 2000
Compiled: December 8, 2000, 11:04am

[a] This work was supported by the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds FCAR (Government of Québec). The authors thank Emanuela Cardia, Marcel Dagenais, John Galbraith, Eric Ghysels, James McKinnon, Christophe Muller, Olivier Torrès, Michael Veall, two anonymous referees and the Editor Richard Smith for several useful comments. Earlier versions of this paper were presented at the North American Meetings of the Econometric Society, the Third International Conference on Computing and Finance (Hoover Institution, Stanford University), the Annual Meetings of the American Statistical Association, the Canadian Econometric Study Group, the Canadian Economics Association, and at Ohio State University (Economics).

[b] C.R.D.E, CIRANO and Département de sciences économiques, Université de Montréal. Mailing address: C.R.D.E., Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. Tel.: (514) 343 2400. FAX: (514) 343 5831. e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.crde.umontreal.ca/personnel/dufour.html.

[c] GREEN, Université Laval, Pavillon J.-A.-De Sève, St. Foy, Québec, Canada, G1K 7P4. TEL: (418) 656 2131-2409; FAX: (418) 656 7412. email: lynda.khalaf@ecn.ulaval.ca

ABSTRACT

In the context of multivariate linear regression (MLR) models, it is well known that commonly employed asymptotic test criteria are seriously biased towards overrejection. In this paper, we propose a general method for constructing exact tests of possibly nonlinear hypotheses on the coefficients of MLR systems. For the case of uniform linear hypotheses, we present exact distributional invariance results concerning several standard test criteria. These include Wilks' likelihood ratio (LR) criterion as well as trace and maximum root criteria. The normality assumption is not necessary for most of the results to hold. Implications for inference are two-fold. First, invariance to nuisance parameters entails that the technique of *Monte Carlo tests* can be applied on all these statistics to obtain exact tests of uniform linear hypotheses. Second, the invariance property of the latter statistic is exploited to derive general nuisance-parameter-free bounds on the distribution of the LR statistic for arbitrary hypotheses. Even though it may be difficult to compute these bounds analytically, they can easily be simulated, hence yielding *exact bounds Monte Carlo tests*. Illustrative simulation experiments show that the bounds are sufficiently tight to provide conclusive results with a high probability. Our findings illustrate the value of the bounds as a tool to be used in conjunction with more traditional simulation-based test methods (*e.g.*, the parametric bootstrap) which may be applied when the bounds are not conclusive.

Key words: multivariate linear regression; seemingly unrelated regressions; uniform linear hypothesis; Monte Carlo test; bounds test; nonlinear hypothesis; finite sample test; exact test; bootstrap.

Journal of Economic Literature classification: C3; C12; C33; C15 O4; O5.

RÉSUMÉ

Dans le contexte des modèles de régression multivariés (MLR), il est bien connu que les tests asymptotiques usuels tendent rejeter trop souvent les hypothèses considérées. Dans cet article, nous proposons une méthode générale qui permet de construire des tests exacts pour des hypothèses possiblement non linéaires sur les coefficients de tels modèles. Pour le cas des hypothèses uniformes linéaires, nous présentons des résultats sur la distribution exacte de plusieurs statistiques de test usuelles. Ces dernières incluent le critère du quotient de vraisemblance (Wilks), de même que les critères de la trace et de la racine maximale. L'hypothèse de normalité des erreurs n'est pas requise pour la plupart des résultats présentés. Ceux-ci ont deux types de conséquences pour l'inférence statistique. Premièrement, l'invariance par rapport aux paramètres de nuisance signifie que l'on peut appliquer la technique des *tests de Monte Carlo* afin de construire des tests exacts pour les hypothèses uniformes linéaires. Deuxièmement, nous montrons comment exploiter cette propriété afin d'obtenir des bornes sans paramètres de nuisance sur la distribution des statistiques de quotient de vraisemblance pour des hypothèses générales. Même si les bornes ne sont pas faciles à calculer par des moyens analytiques, on peut les simuler aisément et ainsi effectuer des tests de Monte Carlo à bornes. Nous présentons une expérience de simulation qui montre que ces bornes sont suffisamment serrées pour fournir des résultats concluants avec une forte probabilité. Nos résultats démontre la valeur de ces bornes comme instrument à utiliser conjointement avec des méthodes d'inférence simulée plus traditionnelles (telles que le bootstrap paramétrique) que l'on peut appliquer lorsque le test à borne n'est pas concluant.

Mots clés: modèle de régression multivarié; régressions empilées; hypothèse linéaire uniforme; test de Monte Carlo; test à borne; hypothèse non linéaire; test à distance finie; test exact; bootstrap.

Classification du Journal of Economic Literature: C3; C12; C33; C15 O4; O5.

Contents

List of Definitions, Propositions and Theorems

List of Tables

1 Introduction

Testing the validity of restrictions on the coefficients of a multivariate linear regression (MLR) model is a common issue which arises in statistics and econometrics. A serious problem with the MLR model is the fact that, except for very special cases, the distributions of standard test criteria are either intractable or unknown, because of the presence of nuisance parameters. In general, only asymptotic approximations are operational. These however may be highly unreliable, especially in systems with large numbers of equations. In view of this, the development of finite-sample procedures appears to be particularly important.

Exact results are available in the literature only for specific test problems. Early references can be found in connection with multivariate analysis of variance (MANOVA). These include the likelihood ratio criterion [Wilks (1932), Bartlett (1947)], the Lawley-Hotelling trace criterion [Lawley (1938), Bartlett (1939), Hotelling (1947, 1951)], the Bartlett-Nanda-Pillai trace criterion [Bartlett (1939), Nanda (1950), Pillai (1955)] and the maximum root criterion [Roy (1953)]. The literature concerning the moments, Laplace transforms and exact densities of these statistics is vast; see, for example, Rao (1973, Chapter 8), Anderson (1984, chapters 8 and 13) and Kariya (1985). However, most of the existing exact results in this area are limited to a very specific class of hypotheses, namely the *uniform mixed linear* (UL) class [see Berndt and Savin (1977)]. Examples of UL hypotheses include: (i) the case where identical transformations of the regression coefficients (within or across equations) are set to given values, and (ii) the hypothesis that a single parameter equals zero. For some recent exact results on tests of UL hypotheses, see Stewart (1997). Note however not all linear hypotheses can be put in UL form. Further, except for even more restricted classes of UL hypotheses (for which tables are available), the existing results on general UL hypotheses are difficult to exploit and approximate distributions are usually suggested.

Thus far less restrictive testing problems have not apparently been considered from a finite sample perspective, with perhaps the exception of Hashimoto and Ohtani's (1990) exact test for general linear restrictions. This procedure is similar to Jayatissa's (1977) test for equality of regression coefficients in two linear regressions with unequal error variances. However, the authors recognize that, similarly with Jayatissa's procedure, this test involves complicated computations and has low power. Further, the test relies on a non-unique transformation of the OLS residuals. These observations suggest that this test has limited practical interest.

Asymptotic Wald, Lagrange multiplier and likelihood ratio tests are available and commonly employed in econometric applications of the MLR model; see Berndt and Savin (1977), Evans and Savin (1982), Breusch (1979), Gouriéroux, Monfort and Renault (1993, 1995) and Stewart (1995, 1997). It has been shown, however, that in finite samples, these asymptotic criteria are seriously biased towards overrejection when the number of equations relative to the sample size is large (even moderately). Well known examples include Laitinen (1978), Meisner (1979), Bera, Byron and Jarque (1981) and Theil and Fiebig (1985) on testing homogeneity and symmetry in demand systems. Further evidence for the case of multivariate tests in capital asset pricing models (CAPM) is also available; see Stambaugh (1982), Jobson and Korkie (1982), Amsler and Schmidt (1985) and MacKinlay (1987). These and other references are discussed by Stewart (1997).

It is clear that standard asymptotic approximations are quite unsatisfactory in this context. Attempts to improve these include, in particular: (i) Bartlett-type corrections, and (ii) bootstrap methods. Bartlett corrections involve rescaling the test statistic by a suitable constant obtained such that the mean of the scaled statistic equals that of the approximating distribution to a given order [Bartlett (1937), Lawley (1956), Rothenberg (1984), Barndorff-Nielsen and Blaesild (1986)]. Formulae explicitly directed towards systems of equations are given in Attfield (1995). Overall, the correction factors require cumulants and joint cumulants of first and second order derivatives of the log-likelihood function, and, outside a small class of problems, are complicated to implement. Furthermore, simulation studies [*e.g.*, Ohtani and Toyoda (1985), Frydenberg and Jensen (1989), Hollas (1991), Rocke (1989), Wong (1989, 1991) and Gonzalo and Pitarakis (1994)] suggest that in many instances Bartlett adjustments do not work well. A simpler correction factor is proposed by Italianer (1985), but the procedure is rather heuristic and has little theoretical background.

The use of bootstrap methods for MLR models has been discussed by several authors, *e.g.* Williams (1986), Rocke (1989), Rayner (1990*a*, 1990*b*), Eakin, McMillen and Buono (1990), Affleck-Graves and McDonald (1990), Martin (1990), Atkinson and Wilson (1992), and Rilstone and Veall (1996). Although long recognized as a useful alternative to standard asymptotic methods, the bootstrap only has an asymptotic justification when the null distribution of the test statistic involves nuisance parameters, hence the finite sample properties of bootstrap tests remain to be established. For general discussion of bootstrap methods, the reader may consult Hall (1992), Efron and Tibshirani (1993) and Shao and Tu (1995); on econometric applications, see Jeong and Maddala (1993), Vinod (1993) and Davidson and MacKinnon (1999*a*, 1999*b*, 1999*c*). In a different vein, randomized tests have been suggested in the MLR literature for a number of special test problems and are referred to under the name of *Monte Carlo tests*; see Theil, Shonkwiler and Taylor (1985), Theil, Taylor and Shonkwiler (1986), Taylor, Shonkwiler and Theil (1986) and Theil and Fiebig (1985). However, these authors do not supply a distributional theory, either exact or asymptotic.

Further results relevant to MLR-based test problems may also be found in the econometric literature on seemingly unrelated regressions (SURE) and simultaneous equations. Indeed, the MLR model can be interpreted as a SURE model with identical regressors across equations. Alternatively, the SURE model may be nested within an MLR framework, imposing exclusion constraints. Hypotheses maintaining the SURE exclusion restrictions are clearly not compatible with the UL format, which precludes the application of existing exact procedures. With regards to simultaneous equations, recall that an unrestricted reduced form is an MLR model. In this context, given the relationship between the structural and reduced-form parameters, one usually meets non-linear hypotheses. Since hardly any practical exact procedure appears to be available for the nonlinear case, testing procedures in simultaneous equations remain for the most part asymptotic. In fact, the existing literature on exact tests in both SURE and simultaneous equations models [see, for example, Dufour and Khalaf (1998*a*, 1998*b*), and Dufour (1997)] is very limited. Furthermore, simulation evidence (reported in the latter references) indicates that the asymptotic tests may have serious shortcomings in such models, and standard size correction techniques are not appropriate.

In this paper, we propose a general exact method for testing arbitrary - possibly non-linear - hypotheses on the coefficients of a standard MLR. We first prove a number of finite sample results dealing with the UL case. While the normality assumption underlies the motivation for the statistics we consider, this is not necessary for most of the results obtained. More precisely, an important feature of the MLR model is the fact that several test criteria derived under the Gaussian assumption [including the likelihood ratio (LR), the Lawley-Hotelling and Bartlett-Nanda-Pillai trace criteria, and Roy's maximum root criterion] are all functions of the eigenvalues of a characteristic determinantal equation which involves the restricted and unrestricted residual sum-of-squares matrices. Further, for UL hypotheses, we show these eigenvalues have a distribution that does not depend on nuisance parameters under the null hypothesis, as soon as the error distribution is parametrically specified up to an unknown linear transformation (or covariance matrix, when second moments exist). This invariance property does not appear to have been pointed out in the earlier literature on inference in the MLR model, especially for non-Gaussian settings.

Second, even though the entailed (nuisance-parameter-free) null distributions of the test statistics are typically non-standard, we observe that finite-sample (randomized) tests of UL hypotheses may then easily be obtained by applying the technique of Monte Carlo (MC) tests [originally proposed by Dwass (1957) and Barnard (1963)] to the test statistics considered]. MC tests may be interpreted as parametric bootstrap tests applied to statistics whose null distribution does not involve nuisance parameters, with however the central additional observation that the randomized test procedure so obtained can easily be performed in such a way that the test exactly has the desired size (for a given, possibly small number of MC simulations); for further discussion, see Jöckel (1986), Hall (1992), Dufour (1995), Dufour and Kiviet (1996, 1998), Kiviet and Dufour (1997), and Dufour, Farhat, Gardiol and Khalaf (1998).

Thirdly, for the problem of testing general possibly nonlinear hypotheses, we use the above invariance results to construct nuisance-parameter-free bounds on the null distribution of the LR criterion. A very remarkable feature of these bounds is the fact that they hold *without imposing any regularity condition* on the form of the null hypothesis, something even the most general asymptotic theories do not typically achieve. The bounds proposed are motivated by the propositions in Dufour (1997) relating to likelihood based inference in MLR settings: using an argument similar to the one in Dufour (1989) for a univariate regression, we show that LR statistics have null distributions which are boundedly pivotal, *i.e.* they admit nuisance-parameter-free bounds. Here we extend this result, *e.g.* by allowing for non-Gaussian models, and outline a general procedure to construct typically tighter bounds. Note however that the bound implicit in Dufour (1997)'s demonstrations may be obtained as a special - although non-optimal - case of the bounds presented here.

To be more specific, the bounds test procedure for general restrictions can be described as follows. First, we introduce a UL hypothesis which is a special case of the restrictions to be tested. Then we argue that the LR criterion associated with the suggested UL hypothesis provides the desired bound. The result follows from two considerations. First, since the UL constraints in question were constructed as a special case of the tested hypothesis, it is evident that the LR statistic for the UL hypothesis (UL-LR) is larger than the LR test statistic of interest, and thus the UL-LR distribution yields an upper bound (and conservative critical points) applicable to the LR statistic. Second, the pivotal property which characterizes the UL-LR statistic (established in our paper) guarantees invariance with respect to nuisance parameters. The null distributions so obtained associated are non-standard, so it may be difficult to compute analytically the corresponding conservative p-values. However, the bounding UL-LR statistics can be easily simulated, hence yielding *exact bounds MC tests*.

We conduct a simulation experiment to assess the performance of the bound. We find that the bounds are sufficiently tight to yield conclusive results with a high probability. These findings illustrate the value of the bounds test as a tool to be used in conjunction (possibly) with more traditional methods (*e.g.,* the parametric bootstrap) and not necessarily as an alternative to these methods. Finally, we refer the reader to Dufour and Khalaf (1998*a*, 1998*b*) for extensions to the SURE and simultaneous equations models.

The paper is organized as follows. Section 2 describes the notation and definitions. Section 3 discusses the distributional results pertaining to uniform linear hypotheses. Section 4 discusses the testing of general hypotheses in the MLR model and establishes bounds on the significance points for these statistics. Simulation results are reported in Section 5, and Section 6 concludes.

2 Framework

The MLR model can be expressed as follows:

$$Y = XB + U \tag{2.1}$$

where $Y = [Y_1, \ \ldots \ , \ Y_p]$ is an $n \times p$ matrix of observations on p dependent variables, X is an $n \times K$ full-column rank matrix of fixed regressors, $B = [b_1, \ \ldots \ , \ b_p]$ is a $K \times p$ matrix of unknown coefficients, and $U = [U_1, \ \ldots \ , \ U_p] = [\widetilde{U}_1, \ \ldots \ , \ \widetilde{U}_n]'$ is an $n \times p$ matrix of random disturbances. For further reference, let $b_j = (b_{0j}, \ b_{1j} \ \ldots \ , \ b_{sj})', j = 1, \ \ldots \ , \ p,$ where $s = K - 1$. We also assume that the rows $\widetilde{U}_i', \ i = 1, \ \ldots \ , \ n$, of U satisfy the following distributional assumption:

$$\widetilde{U}_i = JW_i \ , \ i = 1, \ \ldots \ , \ n \, , \tag{2.2}$$

where the vector $w = vec(W_1, \ \ldots \ , \ W_n)$ has a known distribution and J is an unknown, non-singular matrix. In this context, the covariance matrix of $\widetilde{U}_i$ is$\Sigma = JJ'$, where $\det(\Sigma) \neq 0$. For further reference, let $W = [W_1, \ \ldots \ , \ W_n]' = U\left(J^{-1}\right)'$. In particular, assumption (2.2) is satisfied when

$$W_i \overset{i.i.d.}{\sim} N(0, I_p) \, , \ \ i = 1, \ \ldots \ , \ n \, . \tag{2.3}$$

An alternative representation of the model is

$$y = (I_p \otimes X)b + u \tag{2.4}$$

where $y = vec(Y), \ b = vec(B)$, and $u = vec(U)$. The least squares estimate of B is

$$\widehat{B} = (X'X)^{-1}X'Y \tag{2.5}$$

and the corresponding residual matrix is

$$\widehat{U} = Y - X\widehat{B} = MY = MU \tag{2.6}$$

where $M = I - X(X'X)^{-1}X'$. In this model, it is well known that under (2.3) the maximum likelihood estimators (MLE) of the parameters reduce to $\widehat{B}$ and $\widehat{\Sigma} = \widehat{U}'\widehat{U}/n$. Thus the maximum of the log-likelihood function (MLF) over the unrestricted parameter space is

$$\max_{B,\Sigma} \ln(L) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln(|\widehat{\Sigma}|) - \frac{np}{2}. \tag{2.7}$$

3 Uniform linear hypotheses in the multivariate linear model

In this section, we establish an exact finite-sample distributional invariance result for several usual test statistics in the MLR model (2.1). This result obtains on assuming Gaussian or non-Gaussian errors, provided the latter have a distribution which is specified up to the unknown matrix J. Specifically, we show that, for a wide class of linear hypotheses, the null distributions of the test statistics are free of nuisance parameters.

The fundamental invariance property applies to the case where the constraints take the special UL form

$$H_0 : RBC = D \tag{3.1}$$

where R is a known $r \times K$ matrix of rank $r \leq K$, C is a known $p \times c$ matrix of rank $c \leq p$, and D is a known $r \times c$ matrix. We will first study the problem of testing

$$H_{01} : R\beta_j = \delta_j \ , \ j = 1, \ \ldots \ , \ p \ , \tag{3.2}$$

which corresponds to $C = I_p$. In this context, the most commonly used criteria are: the LR criterion [Wilks (1932), Bartlett (1947)], the Lawley-Hotelling (LH) trace criterion [Lawley (1938), Bartlett (1939), Hotelling (1947, 1951)], the Bartlett-Nanda-Pillai (BNP) trace criterion [Bartlett (1939), Nanda (1950), Pillai (1955)], and the maximum Root (MR) criterion [Roy (1953)].[1] All these test criteria are functions of the roots $m_1, m_2, \ \ldots \ , \ m_p$ of the equation

$$|\widehat{U}'\widehat{U} - m\widehat{U}_0'\widehat{U}_0| = 0 \tag{3.3}$$

where $\widehat{U}_0'\widehat{U}_0$ and $\widehat{U}'\widehat{U}$ are respectively the constrained and unconstrained sum of squared errors (SSE) matrices. For convenience, the roots are reordered so that $m_1 \geq \ldots \geq m_p$. In particular, we have:

$$LR = -n\ln(\mathbf{L}) \ , \ \mathbf{L} = |\widehat{U}'\widehat{U}|/|\widehat{U}_0'\widehat{U}_0| = \prod_{i=1}^{p} m_i \tag{3.4}$$

where $\mathbf{L}$ is the well known Wilks statistic, and

$$LH = \sum_{i=1}^{p}(1 - m_i)/m_i \ ,$$

[1] Note that the criteria LH and BNP can be interpreted as Wald and Lagrange multiplier test statistics, respectively. For details of the relationship, see Berndt and Savin (1977), Breusch (1979) or Stewart (1995).

$$
\begin{aligned}
BNP &= \sum_{i=1}^{p}(1-m_i) \ , \\
MR &= \max_{1\le i\le p}\,(1-m_i)/m_i \ .
\end{aligned}
$$

Now consider the following decomposition of the SSE matrix $\widehat{U}'\widehat{U}$:

$$
\begin{aligned}
\widehat{U}'\widehat{U} &= U'MU = J\left[U\left(J^{-1}\right)'\right]'M\left[U\left(J^{-1}\right)'\right]J' \\
&= JW'MWJ' \qquad (3.5)
\end{aligned}
$$

where the matrix $W = U(J^{-1})'$ defined by (2.2) has a distribution that does not involve nuisance parameters. In other words, $\widehat{U}'\widehat{U}$ depends on Σ only through J. Similarly, $\widehat{U}_0'\widehat{U}_0$ can be expressed as

$$
\widehat{U}_0'\widehat{U}_0 = JW'M_0WJ' \qquad (3.6)
$$

where $M_0 = M - X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'$. These observations yield the following basic distributional result.

Theorem 3.1 DISTRIBUTION OF DETERMINANTAL ROOTS. *Under* (2.1), (2.2) *and* H_{01}, *the vector* $(m_1,\ m_2,\ \ldots\ ,\ m_p)'$ *of the roots of* (3.3) *is distributed like the vector of the corresponding roots of*

$$
\left|W'M\ W - mW'M_0W\right| = 0 \qquad (3.7)
$$

where M is defined as in (2.6), *M_0 as in* (3.6), *$W = U\left(J^{-1}\right)'$ and the roots are put in descending order in both cases.*

PROOF: From (3.5) and (3.6), we have:

$$
\begin{aligned}
\widehat{U}'\widehat{U} &= JW'MWJ' \ , \\
\widehat{U}_0'\widehat{U}_0 &= JW'M_0WJ' \ .
\end{aligned}
$$

Consequently, the determinantal equation (3.3) can be expressed as

$$
\left|JW'MWJ' - mJW'M_0WJ'\right| = 0 \ ,
$$

hence

$$
|J|\left|W'MW - mW'M_0W\right|\left|J'\right| = 0
$$

and

$$
\left|W'M\ W - m\ W'M_0\ W\right| = 0 \ .
$$

Since the vector $w = vec(W_1\ ,\ ...\ ,\ W_N)$ has a completely specified distribution, the roots of equation (3.7) have distributions which does not involve *any* unknown parameter. Q.E.D.

The above result entails that the joint distribution of $(m_1,\ \ldots\ ,\ m_p)'$ does not depend on nuisance parameters. Hence the test criteria obtained as functions of the roots are pivotal under the null and have a completely specified distribution under assumption (2.2). Further, this distribution depends on the R matrix but not on the constants δ_j , $j = 1, \ldots, p$, in (3.2). On the basis of this theorem, the distribution of the Wilks' $\mathbf{L}$ criterion can be readily established.

Corollary 3.2 DISTRIBUTION OF WILKS' STATISTIC. *Under the assumptions of Theorem* **3.1***, Wilks'* $\mathbf{L}$ *statistic for testing* H_{01} *is distributed like the product of the roots of* $|W'MW - mW'M_0W| = 0$.

It may be useful, for simulation purposes, to restate Corollary **3.2** as follows.

Corollary 3.3 DISTRIBUTION OF WILKS' STATISTIC AS RATIO. *Under the assumptions of Theorem* **3.1***, Wilks'* $\mathbf{L}$ *statistic for testing* H_{01} *is distributed like* $|W'MW| / |W'M_0W|$.

We now turn to the general UL hypothesis (3.1). In this case, the model may be reparametrized as follows:

$$Y_c = XB_c + U_c \tag{3.8}$$

where $Y_c = YC$, $B_c = BC$ and $U_c = UC$ with covariance $C'\Sigma C$. The corresponding null hypothesis takes the form $RB_c = D$. The proof then proceeds as for Theorem **3.1**. We emphasize again that the above results do not require the normality assumption.

Eventually, when the normality hypothesis (2.3) holds, the distribution of the Wilks criterion is well known and involves the product of p independent *beta* variables with degrees of freedom that depend on the sample size, the number of restrictions and the number of parameters involved in these restrictions. The reader may consult Anderson (1984) and Rao (1973). For completeness sake, we restate this result in Appendix A. Note that Theorem **3.1** is not explicitly stated by Anderson (1984) or Rao (1973), although it can be derived by looking at their demonstrations.[2]

For non-Gaussian errors [*i.e.* when W_i follows a known distribution which differs from the $N(0, I_p)$ distribution], the null distribution of Wilks' statistic may not be analytically tractable. However, the above invariance results can be used to obtain Monte Carlo tests that are applicable given the distributional assumption (2.2). Such procedures were originally suggested by Dwass (1957) and Barnard (1963). In Appendix B, we briefly outline the methodology involved as it applies to the present context; for a more detailed discussion, see Dufour (1995), Dufour and Kiviet (1996, 1998), Kiviet and Dufour (1997), and Dufour et al. (1998).

To conclude, observe that even in the Gaussian case, it may be more convenient to obtain critical points by simulation. Indeed, it is clear that the null distribution as characterized by Anderson or Rao is not so suitable, in general, for analytical computations (except for specific cases reviewed in Appendix A). Finally, recall that not all linear hypotheses can be expressed as in H_0; we discuss other types of hypotheses in the following section.

[2]The distributions derived in Anderson (1984) and Rao (1973) establish the pivotal characateristic of Wilks' criterion in Gaussian contexts.

4 General hypotheses in the multivariate linear model

In this section, we study the problem of testing general hypotheses on the coefficients of the MLR model. Exact bounds on the null distributions of the LR statistic are derived, which extend the results in Dufour (1989) to the multi-equation context. The bounds are based on the distributional results of the previous section and can be easily simulated. Formally, in the context of (2.4) consider the general hypothesis

$$H_0^* : R^* b \in \Delta_0 \tag{4.1}$$

where R^* is a $q^* \times (pK)$ matrix of rank q^*, and Δ_0 is a non-empty subset of $\mathbb{R}^{q*}$. This characterization of the hypothesis includes cross-equation linear restrictions and allows for nonlinear as well as inequality constraints. The relevant LR statistic is:

$$LR^* = n \ln(\Lambda^*) \ , \ \ \Lambda^* = |\widehat{\Sigma}_0^*|/|\widehat{\Sigma}| \tag{4.2}$$

where $\widehat{\Sigma}_0^*$ and $\widehat{\Sigma}$ are the MLE of Σ imposing and ignoring H_0^*. In general, the null distribution of LR^* depends on nuisance parameters [see Breusch (1980) in connection with the general linear case]. Here we show that LR^* is a boundedly pivotal statistic under the null hypothesis, *i.e.* its distribution can be bounded in a non-trivial way by a nuisance-parameter-free function. To do this, we shall extend the methodology proposed in Dufour (1989) in the context of single equation linear models. Furthermore, we exploit the invariance result which we established above in the UL hypothesis case. The method of proof we present next is likelihood based, in the sense that we explicitly use the Gaussian log-likelihood function. However, as will become clear from our analysis, it is trivial to rewrite proofs and results in the Least Squares framework.

Consider the MLR model (2.4) and let $L(H_U)$ denote the unrestricted MLF. In the Gaussian model, $L(H_U)$ is expressed by (2.7). Further, suppose we can find another set of UL restrictions $H_0^{**} : \widetilde{R}BC = D$ such that $H_0^{**} \subseteq H_0^*$. Now define $L(H_0^*)$, $L(H_0^{**})$ to be the MLF under H_0^* and H_0^{**} respectively. Under assumption (2.3),

$$\begin{aligned} L(H_0^*) &= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln(|\widehat{\Sigma}_0^*|) - \frac{np}{2}, & (4.3)\\ L(H_0^{**}) &= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln(|\widehat{\Sigma}_0^{**}|) - \frac{np}{2}, & (4.4)\end{aligned}$$

where $\widehat{\Sigma}_0^{**}$ is the MLE under H_0^{**}. Then it is straightforward to see that

$$L(H_0^{**}) \le L(H_0^*) \le L(H_U) \ . \tag{4.5}$$

Using (4.3), (4.4) and (4.5), we see that

$$\Lambda^* \le \Lambda^{**} \tag{4.6}$$

where

$$\Lambda^{**} = |\widehat{\Sigma}_0^{**}|/|\widehat{\Sigma}| \ . \tag{4.7}$$

It follows that $P[\Lambda^* \ge x] \le P[\Lambda^{**} \ge x]$, $\forall x$, where $P[\Lambda^{**} \ \ge \ x]$, as demonstrated in Section 3, is nuisance-parameter free and may be used to obtain exact procedures in finite samples on applying

Monte Carlo test methods (see Appendix B).

At this point, it is worth noting that normality [hypothesis (2.3)] by no way constitutes a necessary assumption in this case. Indeed, inequality (4.6) follows from the properties of least squares estimation irrespective of the true density function. Furthermore, the critical values of the bounding statistic may still be determined using the MC test method under the general assumption (2.2). For further reference, we call the MC test based on the conservative bound a *bounds Monte Carlo* (BMC) test. We now state our main result for model (2.4) given the distributional assumption (2.2).

Theorem 4.1 BOUNDS FOR GENERAL LR STATISTICS. *Consider the MLR model* (2.4) *with* (2.2)*. Let* Λ^* *be the statistic defined by* (4.2) *for testing* $R^*b \in \Delta_0$*, where* R^* *is a* $q^* \times (pK)$ *full column rank matrix and* Δ_0 *is a non-empty subset of* $\mathbb{R}^{q^*}$*. Further, consider restrictions of the form* $\widetilde{R}BC = D$ *that satisfy* $R^*b \in \Delta_0$*, where* $\widetilde{R}$ *and* C *are* $r \times K$ *and* $p \times c$ *matrices such that* $r = rank(\widetilde{R})$ *and* $c = rank(C)$*. Let* Λ^{**} *be the inverse of Wilks criterion for testing the latter restrictions. Then under the null hypothesis,* $P[\Lambda^* \geq \lambda^{**}(\alpha)] \leq \alpha$*, for all* $0 \leq \alpha \leq 1$ *, where* $\lambda^{**}(\alpha)$ *is determined such that* $P[\Lambda^{**} \geq \lambda^{**}(\alpha)] = \alpha$*.*

For completeness, we proceed next to state our main conclusion for the Gaussian model. Let $\Psi_\alpha(\cdot)$ be such that

$$P[\Psi(v_1, v_2, v_3) \ \geq \Psi_\alpha(v_1, v_2, v_3)] = \alpha \ , \ 0 \leq \alpha \leq 1 \, , \tag{4.8}$$

where $\Psi(v_1, v_2, v_3)$ is distributed like the product of the inverse of v_2 independent *beta* variables with parameters $(\frac{1}{2}(v_1 - v_2 + i), \ \frac{v_3}{2})$, $i = 1, \ \ldots \ , v_2$. Then, we can prove the following theorem.

Theorem 4.2 BOUNDS FOR GENERAL LR STATISTICS: GAUSSIAN MODEL. *Consider the MLR model* (2.4) *with* (2.2) *and* (2.3)*. Let* Λ^* *be the statistic defined by* (4.2) *for testing* $R^*b \in \Delta_0$*, where* R^* *is a* $q^* \times pK$ *with rank* q^* *and* Δ_0 *is a non-empty subset of* $\mathbb{R}^{q*}$*. Further, consider restrictions of the form* $\widetilde{R}BC = D$ *that satisfy* $R^*b \in \Delta_0$*, where* $\widetilde{R}$ *and* C *are* $r \times K$ *and* $p \times c$ *matrices such that* $r = rank(\widetilde{R})$ *and* $c = rank(C)$*. Then, under the null hypothesis, for all* $0 \leq \alpha \leq 1$, $P[\Lambda^* \geq \Psi_\alpha(n - K, p, \widetilde{q})] \leq \alpha$ *, where* $\widetilde{q} = \min(r, c)$ *and* $\Psi_\alpha(\cdot)$ *is defined by* (4.8)*.*

Indeed, under (2.3), Λ^{**} is distributed like $\Psi(n - K, \ p, \ \widetilde{q})$; see Appendix A. Using (4.6) and (4.8), we have

$$P[\Lambda^* \geq \Psi_\alpha(n - K, \ p, \ \widetilde{q})] \leq \alpha \ , \ 0 \leq \alpha \leq 1 \, . \tag{4.9}$$

Consequently, the critical value Q_α defined by

$$Q_\alpha = \Psi_\alpha(n - K, \ p, \ \widetilde{q}) \tag{4.10}$$

is conservative at level α. Of course, one should seek the smallest critical bound possible. This would mean expressing $\widetilde{R}$ so that $\widetilde{q}$ is as small as possible.

Clearly, the above results hold when the hypothesis is linear of the form $R^*b = \delta_0$. It is worth mentioning at this stage that exact bounds have been proposed in the literature for a specific test problem, namely testing the efficiency hypothesis in the Capital Asset Pricing model (CAPM) contexts; see Stewart (1997) and Shanken (1986). Dufour and Khalaf (1998a) reconsider this example from the finance literature and show that both bounds obtain as a special case of Theorem **4.2**.

To conclude, note that Theorems **4.1** and **4.2** have further implications on LR-based hypothesis tests. The fact that the null distribution of the LR statistic can be bounded (in a non trivial way) implies that alternative simulation-based test techniques may be used to obtain valid p-values based on the statistic in (4.2). See Dufour (1997) for further discussion of the boundedly-pivotal test property and its implications on the potential usefulness of standard size correction techniques.

Eventually, when the BMC p-value is not conclusive, alternative MC and/or bootstrap type methods may be considered. However, we emphasize the fact that the BMC procedure can be implemented in complementarity with such methods. Indeed, if the BMC p-value is less than or equal than α, then it follows from Theorem **4.1** that the exact p-value will most certainly reject the null at level α. Our point is that the bounds are very easy to simulate, since they are based on UL-LR criteria; to see this, refer to Corollaries **3.2** - **3.3**. In contrast, alternative simulation based size corrections procedures including the bootstrap require realizations of the test statistic at hand. It is well known that general-restrictions-LR criteria typically require numerical iterative procedures (even under certain non-UL linear constraints). In view of this, it is advantageous to construct a BMC p-value first, to avoid costly constrained maximizations and the associated numerical problems.

5 Simulation study

This section reports an investigation, by simulation, of the performance of the various proposed statistics under UL constraints as well as more general contexts.

5.1 Design

We considered the following designs.

D1. **MLR system, within-equation UL constraints**
Model: (2.1) with $K = p + 1$;
$H_0^{D1} : (0, 1, ..., 1)B = 0$;
$p = 5, 7, 8$; $n = 20, 25, 40, 50, 100$.

D2. **MLR system, cross-equation UL constraints**
Model: (2.1) ;
H_0^{D2} : (3.1) ;
$p = 11, 12, 13$; $K = 12, 13$; $r = 12, 13$; $c = 11, 12, 13$.

D3. **MLR system, cross-equation constraints**
Model: (2.1) ;
$H_0^{D3} : b_{jj} = b_{11}$, $j = 2$, ... , p and $b_{kj} = 0$, $j \neq k$, j, $k = 1$, . . . , p ;
$p = 3, 5$; $n = 25$.

D4. **MLR system, non-linear constraints**
Model: (2.1) with $K = 2$;
$H_0^{D4} : b_{0j} = \gamma(1 - b_{1j})$, $j = 1$, ... , p , γ unknown;
$p = 40$; $n = 60$.

Experiments D1 and D2 illustrate the UL case. D1 is modelled after the study in Attfield (1995) whose purpose was to demonstrate the effectiveness of Bartlett adjustments. However, the example analyzed there was restricted to a two-equations model. This experiment may be viewed as an illustration of homogeneity tests in demand systems. D2 studies the size of Rao's F test when (A.1) is valid only asymptotically; in the subsequent tables, the latter test is denoted F_{asy}^{RAO}. Experiments D3 and D4 consider more general restrictions and are designed to assess the performance of the bounds procedure. Experiment D3 focuses on general linear restrictions, including exclusion and cross-equation equality constraints. Experiment D4 is modelled after multivariate CAPM tests [see Stewart (1997)]. We considered 40 equations with 60 observations following the empirical example analyzed in Stewart (1997).

For each model, a constant regressor was included and the other regressors were independently drawn (once) from a normal distribution; the errors were independently generated as *i.i.d.* $N(0, \Sigma)$ with $\Sigma = GG'$ and the elements of G drawn (once) from a normal distribution. The regression coefficients are reported in Table 1. The power of the tests in (D1, $n = 25$, $p = 8$), and D3 were investigated by simulating the model with the same parameter values except for b_{11}.

The statistics examined are the relevant LR criteria defined by (3.4) and (4.2). To derive the LR statistic in D4, the constrained MLE was numerically computed according to Shanken (1986). For the purpose of the power comparisons conducted in D3 and D4, we performed: (i) the standard asymptotic LR test (size corrected when needed, using an independent simulation), and (ii) the parametric bootstrap test to which we refer as the "Local" Monte Carlo (LMC) test. The latter procedure is based on simulations that use a restricted ML estimator. The subscripts asy, BMC, LMC and PMC refer respectively to the standard asymptotic tests, the MC bounds tests, the LMC test and the pivotal statistics-based Monte Carlo test. The BMC test performed in D3 is based on the LR statistic which corresponds to the UL constraints setting all coefficients except the intercepts to specific values. In the case of D4, the BMC test corresponds to the following UL restrictions: $b_{0j} = \gamma(1 - b_{1j})$, $j = 1$, ... , p, γ known. In D1 we have also considered the Bartlett-corrected LR test [Attfield (1995, Section 3.3)] which we denote LR_c. The MC tests were applied with 19 and 99 replications. We computed empirical rejection frequencies, based on a nominal size of 5% and 1000 replications. All the experiments were conducted using Gauss-386i VM version 3.1. Note here that the number of simulated samples used for the MC tests has no effect on size, but it may affect power.

5.2 Results and discussion

The results of experiments D1-D3 are summarized in Tables 2 to 6. The results of experiment D4 are as follows. The observed size of the asymptotic test was 89.5%. In contrast, the LMC and BMC tests show empirical type I error rates (.047 and .038) compatible with their nominal 5% level. Our results show the following.

5.2.1 Test sizes

First, it is evident that the asymptotic tests overreject substantially. Although this problem is well documented, observe that in some cases empirical sizes ranged from 75% to 100%. Second, the

TABLE 1. COEFFICIENTS FOR THE SIMULATION EXPERIMENTS

D1.	$b_{kj} = \begin{cases} .1\,, & j = 1\,,\ \ldots\,,\ I[p/2] \\ .2, & j = I[p/2]+1\,,\ \ldots\,,\ p \end{cases}\ ,\ k = 1\,,\ \ldots\,,\ p-1$ $b_{pj} = \sum_{k=1}^{p-1} b_{kj},\ j = 1\,,\ \ldots\,,\ p\,,$ $b_{0j} = \begin{cases} 1.2\,, & j = 1\,,\ \ldots\,,\ I[p/2] \\ 1.8, & j = I[p/2]+1\,,\ \ldots\,,\ p \end{cases}$
D2.	The elements of the matrices R, B, C were selected (once) independently from the $N(0,\ 1)$ distribution
D3.	$p = 3$: $B = \begin{bmatrix} 1.2 & .8 & -1.1 \\ .1 & 0 & 0 \\ 0 & .1 & 0 \\ 0 & 0 & .1 \end{bmatrix}$ $p = 5$: $B = \begin{bmatrix} 1.2 & .8 & -1.1 & 1.9 & -.2 \\ .1 & 0 & 0 & 0 & 0 \\ 0 & .1 & 0 & 0 & 0 \\ 0 & 0 & .1 & 0 & 0 \\ 0 & 0 & 0 & .1 & 0 \\ 0 & 0 & 0 & 0 & .1 \end{bmatrix}$
D4.	$\gamma = .009$ and $b_{1j},\ j = 1\,,\ \ldots\,,\ p$, drawn (once) as $NID(0, .16)$

TABLE 2. EMPIRICAL LEVELS OF VARIOUS TESTS: EXPERIMENT D1

	$p=5$			$p=7$			$p=8$		
n	LR_{asy}	LR_c	LR_{PMC}	LR_{asy}	LR_c	LR_{PMC}	LR_{asy}	LR_c	LR_{PMC}
20	.295	.100	.050	.599	.250	.042	.760	.404	.051
25	.174	.075	.045	.384	.145	.036	.492	.190	.045
40	.130	.066	.052	.191	.068	.045	.230	.087	.049
50	.097	.058	.049	.138	.066	.041	.191	.073	.054
100	.070	.052	.050	.078	.051	.049	.096	.052	.053

TABLE 3. TEST POWERS: EXPERIMENT D1

$n = 25, p = 8; H_0 : b_{11} = .1$

b_{11}	.2	.4	.8	1.0	1.4
LR_{asy}	.055	.176	.822	.965	1.0
LR_{PMC} $(N = 19)$	.054	.165	.688	.881	.991
LR_{PMC} $(N = 99)$	.056	.173	.799	.950	.999

TABLE 4. EMPIRICAL LEVELS OF VARIOUS TESTS: EXPERIMENT D2

(p, K, r, c)	LR_{asy}	F_{asy}^{RAO}	LR_{PMC}
13, 12, 12, 13	1.00	.198	.047
11, 12, 12, 11	1.00	.096	.054
12, 12, 12, 12	1.00	.114	.048
12, 13, 13, 12	1.00	.225	.038

TABLE 5. EMPIRICAL LEVELS OF VARIOUS TESTS: EXPERIMENT D3

$p = 3$			$p = 5$		
LR_{asy}	LR_{LMC}	LR_{BMC}	LR_{asy}	LR_{LMC}	LR_{BMC}
.122	.055	.036	.310	.044	.029

TABLE 6. TEST POWERS: EXPERIMENT D3

$H_0 : b_{11} = .1$

	$N = 19$					$N = 99$				
	$p = 3$									
b_{11}	.3	.5	.7	.9	1.0	.3	.5	.7	.9	1.0
LR_{asy}	.140	.522	.918	.995	1.0	.140	.522	.918	.995	1.0
LR_{LMC}	.137	.468	.849	.987	.991	.135	.539	.912	.995	1.0
LR_{BMC}	.095	.404	.799	.963	.987	.099	.441	.861	.986	.999
	$p = 5$									
b_{11}	.3	.5	.7	.9	1.1	.3	.5	.7	.9	1.1
LR_{asy}	.128	.515	.904	.995	1.0	.128	.515	.904	.995	1.0
LR_{LMC}	.138	.467	.937	.967	1.0	.137	.537	.904	.994	1.0
LR_{BMC}	.120	.427	.792	.958	.995	.110	.484	.877	.990	1.0

Bartlett correction, though providing some improvement, does not control the size in larger systems. From the results of D2, we can see that the asymptotic F test - when applicable - performs better than the standard χ^2 test, but size correction is still needed. The size of the PMC test corresponds closely to 5%. The levels of the BMC test are adequate in all experiments.

5.2.2 Test powers

Experiment D1 reveals that the PMC tests have good power (see Table 3) even with N as low as 19. With $N = 99$, we do not observe any significant power loss for tests having comparable size, although the power study focuses on the eight-equations model with just 25 observations. LMC tests provide substantial improvement over conventional asymptotics: the procedure corrects test sizes with no substantial power loss. A striking observation in the case of D3 is that the conservative bound provides conclusive results with high probability. Increasing the number of equations does not have a great effect on the relative performance of all MC methods proposed. An interesting experiment that bears on this problem is reported in Cribari-Neto and Zarkos (1997) in connection with MLR-based bootstrap tests for homogeneity and symmetry of demand. These authors find that the standard bootstrap achieves size control at the expense of important power losses.

Although the LMC test appears superior, which is somewhat suspected given that the bound is conservative by construction, this experiment shows that the bound has relatively good power. Indeed, we emphasize that LMC and BMC tests should be viewed as complementary rather than alternative procedures. As argued above, the bounds procedure is computationally inexpensive and exact. In addition, whenever the bounds test reject, inference may be made without further appeal to LMC tests. In this regard, our results illustrate the usefulness of the proposed bounds.

6 Conclusion

In this paper we have shown that the LR test on the coefficients of the MLR model is boundedly pivotal under the null hypothesis. The bounds we have derived for general, possibly non-linear hypotheses are exact in finite samples and may easily be implemented by simulation. The basic results were stated in terms of arbitrary hypotheses in MLR contexts. No regularity condition is imposed on the form of restrictions tested, which can be highly nonlinear and may not satisfy the conditions usually required for deriving an asymptotic theory.

For the special case of uniform linear hypotheses, which include many types of restrictions important in practice, we have also focused on special cases, namely uniform and general linear hypotheses. In fact, in the uniform linear case, we have shown that the LR statistic is pivotal even if the normality hypothesis is not imposed. This result has provided the foundations for the construction of the proposed general bounds. We have reported the results of a Monte Carlo experiment that covered uniform linear, cross-equation and non-linear restrictions. We have found that standard asymptotic tests exhibit serious errors in level, particularly in larger systems; usual size correction techniques (*e.g.* the Bartlett adjustment) may not be fully successful. In contrast, the bounds tests we have proposed displayed excellent properties.

Finally, even though the finite-sample validity of the proposed Monte Carlo test procedures only holds under parametric distributional assumptions on model disturbances, it is straightforward to see that such tests will be asymptotically valid (in the usual sense as the sample size goes to infinity) under much weaker distributional assumptions as soon as two conditions are met: (1) the assumptions used to derive an asymptotic distribution include as special case the parametric distributional assumptions used in performing the Monte Carlo tests (*e.g.*, a Gaussian assumption); (2) the asymptotic distribution of the test statistic does not involve unknown nuisance parameters (*e.g.*, it is a chi-square distribution with a known number of degrees of freedom). So there is typically nothing to lose (and potentially much to gain in terms of finite-sample reliability) in applying a finite-sample procedure of the type proposed here as opposed to only an asymptotic approximation. For further discussion of this sort of *generic asymptotic validity* of a finite-sample test, the reader may consult Dufour and Kiviet (1998).

A Appendix: Wilks' and Hotelling's null distributions

We restate here known finite sample distributional results [see Anderson (1984) or Rao (1973)] pertaining to the LR criteria for testing uniform linear hypotheses in the context of the MLR model (2.1) under (2.3). The first result characterizes the exact distribution of Wilks' statistic.

Theorem A.1 DISTRIBUTION OF WILKS' STATISTIC UNDER GAUSSIAN MODELS. *Under* (2.1), (2.2), (2.3) *and* (3.2), *Wilks'* $\mathbf{L}$ *statistic for testing* H_{01} *is distributed like the product of* p *independent beta variables with parameters* $(\frac{1}{2}(n - r_X - p + i)\,,\, \frac{r}{2})\,,\ i = 1,\, \ldots\,,\, p$, *where* r_X *is the rank of the regressor matrix and* r *is the rank of the matrix* R.

This result has formally been derived for the case where the constraints take the special form (3.2), although it is easy to see that it also holds under (3.1). For certain values of r and c and normal errors, the null distribution of the Wilks criterion reduces to the F distribution. For instance, if $\min(r, c) \le 2$, then

$$\left(\frac{\rho\tau - 2\lambda}{rc}\right)\frac{1 - \mathbf{L}^{1/\tau}}{\mathbf{L}^{1/\tau}} \sim F(rc\,,\, \rho\tau - 2\lambda) \tag{A.1}$$

where

$$\rho = \frac{n - K(c - r + 1)}{2}\ ,\ \lambda = \frac{rc - 2}{4}$$

and

$$\tau = \begin{cases} [(r^2c^2 - 4)/(r^2 + c^2 - 5)]^{1/2} & ,\, if\ \ r^2 + c^2 - 5 > 0 \\ 1 & ,\, otherwise \end{cases} .$$

Further, the special case $r = 1$ leads to the Hotelling's T^2 criterion which is a monotonic function of $\mathbf{L}$. If $r > 2$ and $c > 2$, then the distributional result (A.1) holds asymptotically [Rao (1973, Chapter 8)]. Stewart (1997) provides an extensive discussion of these special F tests; see also Shukur and Edgerton (1994).

B Appendix: Monte Carlo tests

MC test procedures were originally suggested by Dwass (1957) and Barnard (1963). In the following, we briefly outline the methodology involved as it applies to the present context; for a more detailed discussion, see Dufour (1995).

Consider first the UL test case. We focus on the statistic $\Lambda = \mathbf{L}^{-1}$, where $\mathbf{L}$ is the Wilks criterion, as defined in (3.4). Let Λ_0 denote the observed test statistic. By Monte Carlo methods and for a given number N of replications, generate $\Lambda_j,\ j = 1\ ,\ \ldots\ ,\ N$ independent realizations of the statistic in question, under the null hypothesis. This may be conveniently implemented using Corollaries **3.2-3.3**. While the level of the test is controlled irrespective of the number of replications, the statistic typically performs better in terms of power the larger the number of replications. Rank $\Lambda_j,\ j = 0\ ,\ \ldots\ ,\ N$ in non-decreasing order and obtain the MC p-value $\widehat{p}_N(\Lambda_0\,)$ where

$$\widehat{p}_N(x) = \frac{N\widehat{G}_N(x) + 1}{N + 1}\ , \tag{B.2}$$

with

$$\widehat{G}_N(x) = \frac{1}{N}\sum_{i=1}^{N} I_{[0,\infty]}\left(\Lambda_i - x\right), \quad I_A(z) = \begin{cases} 1, & if\ z \in A \\ 0, & if\ z \notin A \end{cases} \quad . \tag{B.3}$$

Then the test's critical region corresponds to

$$\widehat{p}_N(\Lambda_0\ ) \leq \alpha\ ,\ 0 < \alpha < 1. \tag{B.4}$$

In the pivotal statistic case, the latter critical region is provably exact, *i.e.* $P[\widehat{p}_N(\Lambda_0\ ) \leq \alpha] \leq \alpha$ with $P[\widehat{p}_N(\Lambda_0\ ) \leq \alpha] = \alpha$ when there is an integer k such that $\alpha = k/(N+1)$ we have . Thus $\widehat{p}_N(\Lambda_0\ )$ provides an exact $p-$value. For example, for $\alpha = 0.05$, the number of replications can be as low as $N = 19$, although of course one could use a larger number [*e.g.*, $N = 49, 99,\ 299,\ 999$]. Clearly, the fact that a small number of replications is sufficient to achieve the desired level does not entail that a larger number of replications is not preferable: raising the value of N will typically increase power and decrease the sensitivity of inference to the randomization inherent to any MC procedure.[3] However (and somewhat surprisingly), our simulation results suggest that increasing the number of replications only has a small effect on power at least for the cases considered.

We now turn to the case of the Λ^* statistic defined by (4.2) for testing (4.1). Denote by θ the vector of relevant nuisance parameters. From the observed data, compute: (i) the test statistic which we will denote Λ_0^*, and (ii) a restricted consistent estimator $\widehat{\theta}_n^0$ of θ [*i.e.*, an estimator $\widehat{\theta}_n^0$ of θ estimator such that the data generating process associated with $\theta = \widehat{\theta}_n^0$ satisfies H_0, and $\widehat{\theta}_n^0 \xrightarrow{p} \theta$ as $n \to \infty$ under H_0]. Using $\widehat{\theta}_n^0$, generate N simulated samples and, from them, N simulated values of the test statistic: $\Lambda_j^*,\ j = 0\ ,\ \ldots\ ,\ N$. Then compute $\widehat{p}_N(\Lambda_0^*|\widehat{\theta}_n^0)$, where $\widehat{p}_N(x|\overline{\theta})$ refers to $\widehat{p}_N(x)$ based on realizations of Λ^* generated given $\theta = \overline{\theta}$ and $\widehat{p}_N(x)$ is defined in (B.2), replacing $\Lambda_j,\ j = 0\ ,\ \ldots\ ,\ N$ by Λ_j^* in (B.3). A MC test may be based on the critical region

$$\widehat{p}_N(T_0|\widehat{\theta}_n^0) \leq \alpha\ ,\ \alpha \leq 0 \leq 1\ .$$

This yields a parametric bootstrap or, in our notation an LMC *p*-value. Using the results from Dufour (1995) on LMC tests, we have that under H_0,

$$\lim_{n\to\infty}\left\{P[\widehat{p}_N(\Lambda_0^*|\widehat{\theta}_n^0) \leq \alpha] - P[\widehat{p}_N(\Lambda_0^*|\theta) \leq \alpha]\right\} = 0\ , \tag{B.5}$$

which means that the LMC test has the correct level asymptotically (as $n \to \infty$). The latter limiting result takes the number of simulated samples explicitly into account, *i.e.* does not depend on $N \to \infty$. Finally, To obtain a BMC test, implement the PMC procedure based on realizations of the bounding statistic. These realizations may be obtained applying Corollaries **3.2-3.3**, where M_0 is chosen conformably with H_0^{**} which should be constructed as outlined in Section 4.

[3]Under quite general conditions, using a randomized procedure based on a finite number of replications N induces a power loss relative to the corresponding non-randomized procedure based on the analytical calculation of the relevant critical values (which is typically infeasible). Power increases (often monotonically) with the number of replications and converges to the power of the non-randomized procedure as $N \to \infty$. For further discussion, see Dwass (1957), Birnbaum (1974), Jöckel (1986) and Dufour (1995).

References

Affleck-Graves, J. and McDonald, B. (1990), 'Multivariate tests of asset pricing: The comparative power of alternative statistics', *Journal of Financial and Quantitative Analysis* **25**, 163–185.

Amsler, C. E. and Schmidt, P. (1985), 'A Monte Carlo investigation of the accuracy of multivariate CAPM tests', *Journal of Financial Economics* **14**, 359–375.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, second edn, John Wiley & Sons, New York.

Atkinson, S. E. and Wilson, P. W. (1992), 'The bias of bootstrapped versus conventional standard errors in the general linear and SUR models', *Econometric Theory* **8**, 258– 275.

Attfield, C. L. F. (1995), 'A Bartlett adjustment to the likelihood ratio test for a system of equations', *Journal of Econometrics* **66**, 207–223.

Barnard, G. A. (1963), 'Comment on 'The spectral analysis of point processes' by M. S. Bartlett', *Journal of the Royal Statistical Society, Series B* **25**, 294.

Barndorff-Nielsen, O. E. and Blaesild, P. (1986), 'A note on the calculation of Bartlett adjustments', *Journal of the Royal Statistical Society, Series B* **48**, 353–358.

Bartlett, M. S. (1937), 'Properties of sufficiency and statistical tests', *Proceedings of the Royal Society of London A* **160**, 268– 282.

Bartlett, M. S. (1939), 'A note on tests of significance in multivariate analysis', *Proceedings of the Cambridge Philosophical Society* **35**, 180–185.

Bartlett, M. S. (1947), 'Multivariate analysis', *Journal of the Royal Statistical Society (Supplement)* **9**, 176–197.

Bera, A. K., Byron, R. P. and Jarque, C. M. (1981), 'Further evidence on asymptotic tests for homogeneity and symmetry in large demand systems', *Economics Letters* **8**, 101–105.

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

Birnbaum, Z. W. (1974), Computers and unconventional test-statistics, *in* F. Proschan and R. J. Serfling, eds, 'Reliability and Biometry', SIAM, Philadelphia, PA, pp. 441–458.

Breusch, T. S. (1979), 'Conflict among criteria for testing hypotheses: Extensions and comments', *Econometrica* **47**, 203–207.

Breusch, T. S. (1980), 'Useful invariance results for generalized regression models', *Journal of Econometrics* **13**, 327–340.

Cribari-Neto, F. and Zarkos, S. G. (1997), 'Finite-sample adjustments for homogeneity and symmetry tests in systems of demand equations: A Monte Carlo evaluation', *Computational Economics* **10**, 337–351.

Davidson, R. and MacKinnon, J. G. (1999*a*), 'Bootstrap testing in non-linear models', *International Economic Review* **forthcoming**.

Davidson, R. and MacKinnon, J. G. (1999*b*), 'Bootstrap tests: How many bootstraps', *Econometric Reviews* **forthcoming**.

Davidson, R. and MacKinnon, J. G. (1999*c*), 'The size distortion of bootstrap tests', *Econometric Theory* **15**, 361–376.

Dufour, J.-M. (1989), 'Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions', *Econometrica* **57**, 335–355.

Dufour, J.-M. (1995), Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics, Technical report, C.R.D.E., Université de Montréal.

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

Dufour, J.-M., Farhat, A., Gardiol, L. and Khalaf, L. (1998), 'Simulation-based finite sample normality tests in linear regressions', *The Econometrics Journal* **1**, 154–173.

Dufour, J.-M. and Khalaf, L. (1998*a*), Simulation based finite and large sample inference methods in multivariate regressions and seemingly unrelated regressions, Technical report, C.R.D.E., Université de Montréal. 36 pages.

Dufour, J.-M. and Khalaf, L. (1998*b*), Simulation-based finite and large sample inference methods in simultaneous equations, Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M. and Kiviet, J. F. (1996), 'Exact tests for structural change in first-order dynamic models', *Journal of Econometrics* **70**, 39–68.

Dufour, J.-M. and Kiviet, J. F. (1998), 'Exact inference methods for first-order autoregressive distributed lag models', *Econometrica* **66**, 79–104.

Dwass, M. (1957), 'Modified randomization tests for nonparametric hypotheses', *Annals of Mathematical Statistics* **28**, 181–187.

Eakin, B. K., McMillen, D. P. and Buono, M. J. (1990), 'Constructing confidence intervals using bootstrap: An application to a multiproduct cost function', *Review of Economics and Statistics* **72**, 339–344.

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Vol. 57 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, New York.

Evans, G. B. A. and Savin, N. E. (1982), ‘Conflict among the criteria revisited: The W, LR and LM tests’, *Econometrica* **50**, 737–748.

Frydenberg, M. and Jensen, J. L. (1989), ‘Is the ‘improved likelihood’ really improved in the discrete case?’, *Biometrika* **76**, 655–661.

Gonzalo, J. and Pitarakis, J.-Y. (1994), Cointegration analysis in large systems, Technical report, Department of Economics, Boston university.

Gouriéroux, C., Monfort, A. and Renault, E. (1993), ‘Tests sur le noyau, l’image et le rang de la matrice des coefficients d’un modèle linéaire multivarié’, *Annales d’Économie et de Statistique* **11**, 81–111.

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.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.

Hashimoto, N. and Ohtani, K. (1990), ‘An exact test for linear restrictions in seemingly unrelated regressions with the same regressors’, *Economics Letters* **32**, 243–346.

Hollas, E. (1991), ‘Performance of Bartlett adjustment for certain likelihood ratio tests’, *Communications in Statistics, Simulation and Computation* **20 (2&3)**, 449–462.

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.

Hotelling, H. (1951), A generalized T-test and measure of multivariate dispersion, *in* J. Neyman, ed., ‘Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability’, The University of California Press, Berkeley, CA, pp. 23–42.

Italianer, A. (1985), ‘A small-sample correction for the likelihood ratio test’, *Economics Letters* **19**, 315–317.

Jayatissa, W. A. (1977), ‘Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal’, *Econometrica* **45**, 1291–1292.

Jeong, J. and Maddala, G. S. (1993), A perspective on application of bootstrap methods in econometrics, *in* (Maddala, Rao and Vinod 1993), pp. 573–610.

Jobson, J. and Korkie, B. (1982), ‘Potential performance and tests of portfolio efficiency’, *Journal of Financial Economics* **10**, 433–466.

Jöckel, K.-H. (1986), ‘Finite sample properties and asymptotic efficiency of Monte Carlo tests’, *The Annals of Statistics* **14**, 336–347.

Kariya, T. (1985), *Testing in the Multivariate General Linear Model*, number 22 *in* 'Economic Research Series, The Institute of Economic Research, Hitotsubashi University, Japan', Kinokuniya Company Ltd., Tokyo.

Kiviet, J. F. and Dufour, J.-M. (1997), 'Exact tests in single equation autoregressive distributed lag models', *Journal of Econometrics* **80**, 325–353.

Laitinen, K. (1978), 'Why is demand homogeneity so often rejected?', *Economics Letters* **1**, 187–191.

Lawley, D. N. (1938), 'A generalization of Fisher's z-test', *Biometrika* **30**, 180–187.

Lawley, D. N. (1956), 'A general method for approximating the distribution of the likelihood ratio statistic', *Biometrika* **43**, 295–303.

MacKinlay, A. C. (1987), 'On multivariate tests of the Capital Asset Pricing Model', *Journal of Financial Economics* **18**, 341–372.

Maddala, G. S., Rao, C. R. and Vinod, H. D., eds (1993), *Handbook of Statistics 11: Econometrics*, North-Holland, Amsterdam.

Martin, M. A. (1990), 'On bootstrap iteration for coverage correction in confidence intervals', *Journal of the American Statistical Association* **85**, 1105–1118.

Meisner, J. F. (1979), 'The sad fate of the asymptotic Slutsky symmetry test for large systems', *Economics Letters* **2**, 231–233.

Nanda, D. N. (1950), 'Distribution of the sum of roots of a determinantal equation under a certain condition', *Annals of Mathematical Statistics* **21**, 432–439.

Ohtani, K. and Toyoda, T. (1985), 'A Monte Carlo study of the Wald, LM and LR tests in a heteroscedastic linear model', *Communications in Statistics, Simulation and Computation* **14**, 735–746.

Pillai, K. C. S. (1955), 'Some new test criteria in multivariate analysis', *Annals of Mathematical Statistics* **43**, 117–121.

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

Rayner, R. K. (1990*a*), 'Bartlett's correction and the bootstrap in normal linear regression models', *Economics Letters* **33**, 255–258.

Rayner, R. K. (1990*b*), 'Bootstrapping p-values and power in the First-Order autoregression: A Monte Carlo study', *Journal of Business and Economic Statistics* **8**, 251–263.

Rilstone, P. and Veall, M. (1996), 'Using bootstrapped confidence intervals for improved inferences with seemingly unrelated regression equations', *Econometric Theory* **12**, 569–580.

Rocke, D. M. (1989), 'Bootstrap Bartlett adjustment in seemingly unrelated regressions', *Journal of the American Statistical Association* **84**, 598–601.

Rothenberg, T. J. (1984), 'Hypothesis testing in linear models when the error covariance matrix is nonscalar', *Econometrica* **52**, 827–842.

Roy, S. N. (1953), 'On a heuristic method of test construction and its use in multivariate analysis', *Annals of Mathematical Statistics* **24**, 220–238.

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

Shao, S. and Tu, D. (1995), *The Jackknife and Bootstrap*, Springer-Verlag, New York.

Shukur, G. and Edgerton, D. (1994), The small sample properties of the RESET test as applied to systems of equations, Technical report, Department of Statistics and Department of Economics, Lund university, Sweden.

Stambaugh, R. F. (1982), 'On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis', *Journal of Financial Economics* **10**, 237–268.

Stewart, K. G. (1995), 'The functional equivalence of the W, LR and LM statistics', *Economics Letters* **49**, 109–112.

Stewart, K. G. (1997), 'Exact testing in multivariate regression', *Econometric Reviews* **16**, 321–352.

Taylor, T. G., Shonkwiler, J. S. and Theil, H. (1986), 'Monte Carlo and bootstrap testing of demand homogeneity', *Economics Letters* **20**, 55–57.

Theil, H. and Fiebig, D. G. (1985), Small sample and large equation systems, *in* E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds, 'Handbook of Statistics 5: Time Series in the Time Domain', North-Holland, Amsterdam, pp. 451–480.

Theil, H., Shonkwiler, J. S. and Taylor, T. G. (1985), 'A Monte Carlo test of Slutsky symmetry', *Economics Letters* **19**, 331–332.

Theil, H., Taylor, T. and Shonkwiler, J. S. (1986), Monte Carlo testing in systems of equations, *in* D. J. Slottje and G. F. Rhodes Jr., eds, 'Advances in Econometrics, Volume 5, Innovations in Quantitative Economics: Essays in Honor of Robert L. Basmann', Vol. 5, JAI Press, Greenwich, Conn., pp. 227–239.

Vinod, H. D. (1993), Bootstrap methods: Applications in econometrics, *in* (Maddala et al. 1993), pp. 629–661.

Wilks, S. S. (1932), 'Certain generalizations in the analysis of variance', *Biometrika* **24**, 471–494.

Williams, M. A. (1986), 'An economic application of bootstrap statistical methods: Addyston pipe revisited', *American Economic Review* **30**, 52–58.

Wong, M. Y. (1989), 'Likelihood estimation of a simple linear regression model when both variables have error', *Biometrika* **76**, 141–148.

Wong, M. Y. (1991), 'Bartlett adjustment to the likelihood ratio statistic for testing several slopes', *Biometrika* **78**, 221–224.