ELSEVIER

Journal of Econometrics 111 (2002) 303–322

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Simulation based finite and large sample tests in multivariate regressions

Jean-Marie Dufour[a,*], Lynda Khalaf[b]

[a] *Canada Research Chair, C.R.D.E-CIREQ, CIRANO and Département de Sciences Économiques, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7*

[b] *CIREQ and GREEN, Université Laval, Pavillon J.-A.-De Sève, St. Foy, Québec, Canada G1K 7P4*

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 perform well. 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.

JEL classification: C3; C12; C33; C15 O4; O5

Keywords: Multivariate linear regression; Seemingly unrelated regressions; Uniform linear hypothesis; Monte Carlo test; Bounds test; Nonlinear hypothesis; Finite sample test; Exact test; Bootstrap

* Corresponding author. Tel.: +514-343-2400; fax: +514-343-5831.

E-mail addresses: jean.marie.dufour@umontreal.ca (J.-M. Dufour), lynda.khalaf@ecn.ulaval.ca (L. Khalaf).

0304-4076/02/$ - see front matter
PII: S0304-4076(02)00108-2

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); 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. However, the authors recognize that this test involves complicated computations and has low power. Further, the test relies on a non-unique transformation of the OLS residuals. These observations suggests 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), and Breusch (1979). It has been shown, however, that in finite samples, these asymptotic criteria are seriously biased towards over-rejection when the number of equations relative to the sample size is large (even moderately). Well-known examples, including homogeneity and symmetry tests in demand systems and multivariate capital asset pricing tests, are discussed by Stewart (1997). Attempts to improve asymptotic approximations in this context 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. 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, simulations studies (e.g. Ohtani and Toyoda, 1985; Hollas, 1991; Rocke, 1989) suggests that in many instance Bartlett adjustments do not work well.

The use of bootstrap methods for MLR models has been discussed by several authors, e.g. Williams (1986), Rocke (1989), Affleck-Graves and McDonald (1990), and Atkinson and Wilson (1992). The bootstrap typically provides refinements for the precision of asymptotic tests; see, for example, Hall (1992), Efron and Tibshirani (1993), Shao and Tu (1995), and Davidson and MacKinnon (1999). When the null distribution of the test statistic involves nuisance parameters, level control is however not guaranteed by bootstrap arguments. The literature on MLR applications of the bootstrap provides examples where the method: (i) works well in finite samples (e.g. Rilstone and Veall, 1996), or (ii) fails to achieve size control (e.g. Dufour and Khalaf, 1998). 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 et al. (1985, 1986) and Taylor et al. (1986). However, these authors do not supply a distributional theory, either exact or asymptotic.

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 general 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. It is quite remarkable in view of the fact that the residuals themselves have distributions which do depend on the parameters of the disturbance covariance matrix (across equations), even under the null hypothesis.

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), Dufour and Kiviet (1996, 1998), Kiviet and Dufour (1997), and Dufour et al. (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 below) guarantees invariance with respect to nuisance parameters. The null distributions so obtained 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 on a set of linear and nonlinear parameter restrictions—for which no alternative finite-sample procedure appears to be available. The results indicate that the bounds proposed do not yield overly conservative tests and provide powers close to the ones of the other procedures considered (without the risk of being over-sized). These findings illustrate the usefulness of the bounds in this context—to be used possibly in conjunction 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, 2002) 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 $U = [U_1, \ldots, U_p] = [\tilde{U}_1, \ldots, \tilde{U}_n]'$ is an $n \times p$ matrix of random disturbances. For further reference, let $b_j = (b_{0j}, b_{1j}, \ldots, b_{K-1,j})'$, $j = 1, \ldots, p$. We also assume that the rows $\tilde{U}_i'$, $i = 1, \ldots, n$, of U satisfy the following distributional assumption:

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

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 $\tilde{U}_i$ is $\Sigma = JJ'$, where $\det(\Sigma) \neq 0$. For further reference, let $W = [W_1, \ldots, W_n]' = U(J^{-1})'$. In particular, assumption (2.2) is satisfied when

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

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 and the corresponding residual matrix are

$$\hat{B} = (X'X)^{-1}X'Y, \quad \hat{U} = Y - X\hat{B} = MY = MU, \tag{2.5}$$

where $M = I - X(X'X)^{-1}X'$. Note the distribution of $\hat{U}$ does not depend on B, although it is affected by the value of J (or $\Sigma = JJ'$). In this model, it is well known that under (2.3) the maximum likelihood estimators (MLE) of the parameters reduce to $\hat{B}$ and $\hat{\Sigma} = \hat{U}'\hat{U}/n$. The maximum of the log-likelihood function (MLF) over the unrestricted parameter space is

$$\max_{B,\Sigma} \ln(L) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln(|\hat{\Sigma}|) - \frac{np}{2}. \tag{2.6}$$

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). Further, for the cases where it applies, we observe that it can be used to obtain finite-sample MC tests. This result obtains irrespective of the disturbance distribution (whether it is Gaussian or non-Gaussian), provided the latter 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. Here it is important to note that, even though least squares residuals have distributions which do not depend on regression coefficients (B), it is easy to see that the latter *does involve* the unknown disturbance covariance matrix (or the J matrix). Thus we may expect that test statistics based on such residuals will depend on J as a nuisance parameter, and the invariance with respect to the J matrix (or the disturbance covariance matrix) is remarkable. Note also that standard (finite-sample) results on hypothesis testing on the MLR model impose the assumption that the errors follow a Gaussian [see, e.g. Anderson (1984) and Rao (1973)] or an elliptically symmetric distribution (see Kariya, 1981, 1985).[1]

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

$$H_0\colon RBC = D, \tag{3.1}$$

where R is a known $r \times K$ matrix of rank $r \leqslant K$, C is a known $p \times c$ matrix of rank $c \leqslant p$, and D is a known $r \times c$ matrix. An important special case of this problem consists in testing

$$H_{01}\colon Rb_j = \delta_j, \quad j = 1, \ldots, p, \tag{3.2}$$

which corresponds to H_0 with $C = I_p$. In this context, the most commonly-used criteria are: the LR criterion, the Lawley–Hotelling (LH) trace criterion, the Bartlett–Nanda–Pillai (BNP) trace criterion and the maximum root (MR) criterion.[2] All these test criteria are functions of the roots $m_1, m_2, \ldots, m_p$ of the equation

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

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

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

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

$$LH = \sum_{i=1}^{p}(1 - m_i)/m_i, \quad BNP = \sum_{i=1}^{p}(1 - m_i), \quad MR = \max_{1 \leqslant i \leqslant p}(1 - m_i)/m_i. \tag{3.5}$$

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

$$\begin{aligned}\hat{U}'\hat{U} &= U'MU = J[U(J^{-1})']'M[U(J^{-1})']J' \\ &= JW'MWJ',\end{aligned} \tag{3.6}$$

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

$$\hat{U}_0'\hat{U}_0 = JW'M_0WJ', \tag{3.7}$$

[1] As shown by Kariya (1981), tests derived under the Gaussian distributional assumption usually remain valid under the more general assumption of elliptical symmetry on U. Since elliptical symmetric precludes independence between $U_1, \ldots, U_p$, except in the Gaussian case, assumption (2.2) considered here is not covered by elliptical symmetry.

[2] For references, see Rao (1973, Chapter 8) or Anderson (1984, Chapters 8 and 13). 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) and Stewart (1997).

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*

$$|W'MW - mW'M_0W| = 0, \tag{3.8}$$

where M *is defined as in* (2.5), M_0 *as in* (3.7), $W = U(J^{-1})'$ *and the roots are put in descending order in both cases.*

Proof. From (3.6) and (3.7), we have

$$\hat{U}'\hat{U} = JW'MWJ', \quad \hat{U}_0'\hat{U}_0 = JW'M_0WJ'.$$

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

$$|JW'MWJ' - mJW'M_0WJ'| = 0,$$

hence

$$|J||W'MW - mW'M_0W||J'| = 0$$

and

$$|W'MW - mW'M_0W| = 0.$$

Since the vector $w = vec(W_1, \ldots, W_N)$ has a completely specified distribution, the roots of Eq. (3.8) have a joint distribution which does not involve *any* unknown parameter. □

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 hypothesis 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' **L** criterion can be readily established.

Corollary 3.2 (Distribution of Wilks' statistic). *Under the assumptions of Theorem* 3.1, *Wilks'* **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'* **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:

$$Y_c = XB_c + U_c, \tag{3.9}$$

where $Y_c = YC, B_c = BC$ and $U_c = UC$. The corresponding null hypothesis takes the form $RB_c = D$. The proof then proceeds as for Theorem 3.1. Thus the null distribution of Wilks' $\mathbf{L}$ criterion may depend on X, R and C (although not D), but does not involve any unknown nuisance parameter. 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. To the best of our knowledge, Theorem 3.1 has not been stated in the earlier literature on inference in the MLR model. [3]

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.

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

[3] The distributions of various test criteria proposed in this area are almost invariably derived for Gaussian (or elliptically symmetric) MLR models; see, for example, Rao (1973), Arnold (1981, Chapter 19), Anderson (1984, Chapter 8), and Kariya (1981,1985). This holds, in particular, for the pivotal character of the roots of (3.3). Although one may argue that some of these invariance results do not require the Gaussian (or elliptical symmetry) assumption, the fact remains that normality is explicitly imposed and the way the methods can be extended to general non-Gaussian parametric error distributions has not apparently been discussed in the statistical literature. In the context of univariate linear regressions, Breusch (1980) has also provided some interesting invariance results with respect to regression coefficients and *one* scale parameter (σ). However, these do not apply to multivariate regressions where one needs to show invariance with respect to the J matrix which involves both scale parameters and coefficients representing dependence (e.g. correlations) between disturbance terms in different equations.

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

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

where $\hat{\Sigma}_0^*$ and $\hat{\Sigma}$ 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 (2.6). Further, consider a set of UL restrictions $H_0^{**}: \tilde{R}BC = D$ such that $H_0^{**} \subseteq H_0^*$.[4] Now define $L(H_0^*)$, $L(H_0^{**})$ to be the MLF under H_0^* and H_0^{**}, respectively. Under assumption (2.3),

$$L(H_0^*) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln(|\hat{\Sigma}_0^*|) - \frac{np}{2}, \tag{4.3}$$

$$L(H_0^{**}) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln(|\hat{\Sigma}_0^{**}|) - \frac{np}{2}, \tag{4.4}$$

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

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

Using (4.3)–(4.5), we see that

$$\Lambda^* \leqslant \Lambda^{**}, \tag{4.6}$$

where

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

It follows that $P[\Lambda^* \geqslant x] \leqslant P[\Lambda^{**} \geqslant x]$ $\forall x$, where $P[\Lambda^{**} \geqslant x]$, 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).

[4] Such a set of UL restrictions does always exist. For example, in all cases, we can take $\tilde{R} = I_K$, $C = I_p$, and D equal to the true value of B, so H_0^{**} takes the form $B = D$. However, we can get tighter bounds if we choose the number of rows in $\tilde{R}$ and the number of columns in C as small as possible.

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 any restrictions of the form* $\tilde{R}BC = D$ *that satisfy* $R^*b \in \Delta_0$, *where* $\tilde{R}$ *and* C *are* $r \times K$ *and* $p \times c$ *matrices such that* $r = rank(\tilde{R})$ *and* $c = rank(C)$. *Let* Λ^{**} *be the inverse of Wilks criterion for testing the latter restrictions. Then under the null hypothesis,* $P[\Lambda^* \geqslant \lambda^{**}(\alpha)] \leqslant a$, *for all* $0 \leqslant a \leqslant 1$, *where* $\lambda^{**}(\alpha)$ *is determined such that* $P[\Lambda^{**} \geqslant \lambda^{**}(\alpha)] = \alpha$.

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

$$P[\Psi(v_1, v_2, v_3) \geqslant \Psi_\alpha(v_1, v_2, v_3)] = \alpha, \quad 0 \leqslant a \leqslant 1, \tag{4.8}$$

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 have 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* $\tilde{R}BC = D$ *that satisfy* $R^*b \in \Delta_0$, *where* $\tilde{R}$ *and* C *are* $r \times K$ *and* $p \times c$ *such that* $r = rank(\tilde{R})$ *and* $c = rank(C)$. *Then, under the null hypothesis, for all* $0 \leqslant \alpha \leqslant 1$, $P[\Lambda^* \geqslant \Psi_\alpha(n - K, p, \tilde{q})] \leqslant \alpha$, *where* $\tilde{q} = \min(r, c)$ *and* $\Psi_\alpha(\cdot)$ *is defined by* (4.8).

The latter theorem follows on observing that the bounding statistic Λ^{**} is distributed like $\Psi(n - K, p, \tilde{q})$; see Theorem A.1 in Appendix A. Then, using (4.6) and (4.8), we have

$$P[\Lambda^* \geqslant \Psi_\alpha(n - K, p, \tilde{q})] \geqslant \alpha, \quad 0 \leqslant \alpha \leqslant 1. \tag{4.9}$$

Consequently, the critical value $Q_a = \Psi_\alpha(n - K, p, \tilde{q})$ is conservative at level α. Of course, one should seek the smallest critical bound possible. This would mean expressing $\tilde{R}$ so that $\tilde{q}$ is as small as possible.

To conclude, we 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 to α, then it follows from Theorem 4.1 that the exact p-value certainly rejects the null hypothesis 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 and 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 $K = p + 1; \tilde{\mathrm{H}}_0^{D1}$: $(0, 1, \ldots, 1)B = 0$; $p = 5, 7, 8; n = 20, 25, 40, 50, 100$.

D2. MLR system, cross-equation UL constraints. Model: (2.1); H_0^{D2}: (3.1) with the coefficients of R, B and C selected according to a N(0, 1) distribution; $p = 11, 12, 13; K = 12, 13; r = 12, 13; c = 11, 12, 13; n = 25$.

D3. MLR system, cross-equation constraints. Model: (2.1); $\mathrm{H}_0^{D3} : b_{jj} = b_{11}, j = 2, \ldots, p$ and $b_{kj} = 0, j \neq k, j, k = 1, \ldots, p$; $p = 3, 5; n = 25$.

D4. MLR system, nonlinear constraints. Model: (2.1) with $K = 2; \mathrm{H}_0^{D4} : b_{0j} = \gamma (1 - b_{1j})$, $j = 1, \ldots, p$, γ unknown; $p = 40; n = 60$.

Experiments Dl and D2 illustrate the UL case. Dl 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) in Appendix A is valid only asymptotically; in the subsequent tables, the latter test is denoted $F_{\text{asy}}^{\text{RAO}}$. 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

Table 1
Coefficients for the simulation experiments

D1	$b_{kj} = \begin{cases} 0.1, & j = 1,\ldots,I[p/2], \\ 0.2, & j = I[p/2]+1,\ldots,p, \end{cases}$	$k = 1,\ldots,p-1$
	$b_{pj} = \sum_{k=1}^{p-1} b_{kj},\ j = 1,\ldots,p,$	
	$b_{0j} = \begin{cases} 1.2, & j = 1,\ldots,I[p/2] \\ 1.8, & j = I[p/2]+1,\ldots,p \end{cases}$	
D2	The elements of the matrices R, B, C were selected (once) independently from the $\mathrm{N}(0,1)$ distribution	
D3	$p = 3$ $B = \begin{bmatrix} 1.2 & 0.8 & -1.1 \\ 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$	$p = 5$ $B = \begin{bmatrix} 1.2 & 0.8 & -1.1 & 1.9 & -0.2 \\ 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.1 \end{bmatrix}$
D4	$\gamma = 0.009$ and $b_{1j},\ j = 1,\ldots,p$, drawn (once) as NID $(0, 0.16)$	

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 $\mathrm{N}(0, \Sigma)$ with $\Sigma = JJ'$ and the elements of J drawn (once) from a normal distribution.[5] The regression coefficients are reported in Table 1. The power of the tests in $(\mathrm{D1}, n = 25, p = 8)$, 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). 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, MC bounds tests, LMC tests (bootstrap), and MC tests based pivotal statistics (in the case of uniform linear hypotheses). The BMC test performed in D3 is based on the LR statistic which corresponds to the UL constraints setting all

[5] The BMC and PMC invariant tests are do not depend on the choice for J. The results on standard tests are conditional on the chosen value for J; the values of the regression coefficients also intervene in the nonlinear example. We have performed several experiments with various parameter choices (not reported for space considerations) with qualitatively similar results.

Table 2
Empirical levels of various tests: experiment D1

n	$p=5$			$p=7$			$p=8$		
	LR_{asy}	LR_c	LR_{PMC}	LR_{asy}	LR_c	LR_{PMC}	LR_{asy}	LR_c	LR_{PMC}
20	0.295	0.100	0.050	0.599	0.250	0.042	0.760	0.404	0.051
25	0.174	0.075	0.045	0.384	0.145	0.036	0.492	0.190	0.045
40	0.130	0.066	0.052	0.191	0.068	0.045	0.230	0.87	0.049
50	0.097	0.058	0.049	0.138	0.066	0.041	0.191	0.073	0.054
100	0.070	0.052	0.050	0.078	0.051	0.049	0.096	0.052	0.053

Table 3
Test powers: experiment $D1$ $n=25$, $p=8; \mathrm{H}_0: b_{11}=0.1$

b_{11}	0.2	0.4	0.8	1.0	1.4
LR_{asy}	0.055	0.176	0.822	0.965	1.0
LR_{PMC} $(N=19)$	0.054	0.165	0.688	0.881	0.991
LR_{PMC} $(N=99)$	0.056	0.173	0.799	0.950	0.999

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	0.198	0.047
11,12,12,11	1.00	0.096	0.054
12,12,12,12	1.00	0.114	0.048
12,13,13,12	1.00	0.225	0.038

coefficients except the intercepts to specific values. In the case of D4, the BMC test corresponds to the following UL restrictions: $b_{0j}=\gamma(1-b_{1j})$, $j=1,\ldots,p,\gamma$ 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 2–6. 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 (0.047 and 0.038) compatible with their nominal 5% level. Our results show the following.

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}
0.122	0.055	0.036	0.310	0.044	0.029

Table 6
Test powers: experiment D3 $H_0: b_{11} = 0.1$

	$N=19$					$N=99$				
$p=3$										
b_{11}	0.3	0.5	0.7	0.9	1.0	0.3	0.5	0.7	0.9	1.0
LR_{asy}	0.140	0.522	0.918	0.995	1.0	0.140	0.522	0.918	0.955	1.0
LR_{LMC}	0.137	0.468	0.849	0.987	0.991	0.135	0.539	0.912	0.995	1.0
LR_{BMC}	0.095	0.404	0.799	0.963	0.987	0.099	0.441	0.861	0.986	0.999
$p=5$										
b_{11}	0.3	0.5	0.7	0.9	1.1	0.3	0.5	0.7	0.9	1.1
LR_{asy}	0.128	0.515	0.904	0.995	1.0	0.128	0.515	0.904	0.995	1.0
LR_{LMC}	0.138	0.467	0.937	0.967	1.0	0.137	0.537	0.904	0.994	1.0
LR_{BMC}	0.120	0.427	0.792	0.958	0.995	0.110	0.484	0.877	0.990	1.0

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 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 LMC test works better then standard asymptotic approximations. This observation is consistent with results showing (under appropriate regularity conditions) that the bootstrap can deliver asymptotic refinements similar to Edgeworth expansions (see Hall, 1992).[6] As predicted by theory, the levels of the BMC test are adequate in all experiments.

5.2.2. Test powers

Experiment Dl 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

[6] Of course, there is no general theoretical guarantee that such as an approximation cannot lead to over-rejections in finite samples for the MLR model considered here. For some illustrative evidence, involving nonlinear restrictions on MLR models, see Dufour and Khalaf (1998).

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 exhibits power very close to that of the other procedures. 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 could be expected given that the bound is conservative by construction, this experiment shows that the bound has relatively good power. It is important to recall however that the LMC test may not always satisfy the level constraint in finite samples; for illustrative evidence, see Dufour and Khalaf (1998). 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 linear hypotheses, which include many types of restrictions important in practice, we studied both 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 imposed in order to perform 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).

Acknowledgements

This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), 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 Torrés, 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).

Appendix A. 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 under normality.

Theorem A.1 (Distribution of Wilks' statistic under Gaussian models). *Under* (2.1)–(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 $\min(r, c) \leqslant 2$, 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

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

and

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

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.

Appendix B. 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 Λ_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 and 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 Λ_j, $j=0,\ldots,N$ in non-decreasing order and obtain the MC p-value $\hat{p}_N(\Lambda_0)$ where

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

with

$$\hat{G}_N(x) = \frac{1}{N}\sum_{i=1}^{N} I_{[0,\infty]}(\Lambda_i - x), \quad I_A(z) = \begin{cases} 1 & \text{if } z \in A, \\ 0 & \text{if } z \notin A. \end{cases} \tag{B.2}$$

Then the test's critical region corresponds to

$$\hat{p}_N(\Lambda_0) \leqslant \alpha, \quad 0 < \alpha < 1. \tag{B.3}$$

In the pivotal statistic case, the latter critical region is provably exact, i.e. $P[\hat{p}_N(\Lambda_0) \leqslant \alpha] \leqslant \alpha$ with $P[\hat{p}_N(\Lambda_0) \leqslant \alpha] = \alpha$ when there is an integer k such that $\alpha = k/(N+1)$. Thus $\hat{p}_N(\Lambda_0)$ 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.[7] 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_0^* 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 $\hat{\theta}_n^0$ of θ [i.e. an estimator $\hat{\theta}_n^0$ of θ estimator such that the data generating process associated with $\theta=\hat{\theta}_n^0$ satisfies H_0, and $\hat{\theta}_n^0 \xrightarrow{p} \theta$ as $n \to \infty$ under H_0]. Using $\hat{\theta}_n^0$, generate N simulated samples and, from them, N simulated values of the test statistic: Λ_j^*, $j = 0, \ldots, N$. Then compute $\hat{p}_N(\Lambda_0^* | \hat{\theta}_n^0)$, where $\hat{p}_N(x|\bar{\theta})$ refers to $\hat{p}_N(x)$ based on realizations of Λ^* generated given $\theta = \bar{\theta}$ and $\hat{p}_N(x)$ is defined in (B.1), replacing Λ_j, $j = 0, \ldots, N$ by Λ_j^* in (B.2). A MC test may be based on the critical region

$$\hat{p}_N(T_0 | \hat{\theta}_n^0) \leqslant \alpha, \quad \alpha \leqslant 0 \leqslant 1.$$

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,

$$\lim_{n\to\infty} \{P[\hat{p}_N(\Lambda_0^* | \hat{\theta}_n^0) \leqslant \alpha] - P[\hat{p}_N(\Lambda_0^* | \theta) \leqslant \alpha]\} = 0, \tag{B.4}$$

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 $N \to \infty$. Furthermore, as shown by Hall (1992), or Davidson and MacKinnon (1999) (among others), the bootstrap p-value so defined provides a sizable refinement on the precision of asymptotic tests. 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 and 3.3, where M_0 is chosen conformably with H_0^{**} which Should be constructed as outlined in Section 4.

References

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

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis, 2nd Edition. Wiley, New York.

Arnold, S.F., 1981. The Theory of Linear Models and Multivariate Analysis. Wiley, New York.

Atkinson, S.E., 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.

[7] 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).

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.

Berndt, E.R., 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: Proschan, F., Serfling, R.J. (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., 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., MacKinnon, J.G., 1999. 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., Khalaf, L., 1998. Simulation-based finite and large sample inference methods in simultaneous equations. Technical Report, C.R.D.E., Université de Montréal.

Dufour, J.-M., Khalaf, L., 2002. Finite sample tests in seemingly unrelated regressions. In: Giles, D.E.A., Ed., Computer-Aided Econometrics. Marcel Dekker, New York, forthcoming.

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

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

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

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

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

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

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer, New York.

Hashimoto, N., 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.

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

Kariya, T., 1981. Robustness of multivariate tests. The Annals of Statistics 6, 1267–1275.

Kariya, T., 1985. Testing in the Multivariate General Linear Model, Economic Research Series, Vol. 22. The Institute of Economic Research, Hitotsubashi University, Japan. Kinokuniya Company Ltd., Tokyo.

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

Ohtani, K., 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.

Rao, C.R., 1973. Linear Statistical Inference and its Applications, 2nd Edition. Wiley, New York.

Rilstone, P., 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.

Shao, S., Tu, D., 1995. The Jackknife and Bootstrap. Springer, New York.

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

Taylor, T.G., Shonkwiler, J.S., Theil, H., 1986. Monte Carlo and bootstrap testing of demand homogeneity. Economics Letters 20, 55–57.

Theil, H., Shonkwiler, J.S., Taylor, T.G., 1985. A Monte Carlo test of Slutsky symmetry. Economics Letters 19, 331–332.

Theil, H., Taylor, T., Shonkwiler, J.S., 1986. Monte Carlo testing in systems of equations, in: Slottje, D.J., Rhodes Jr., G.F. (Eds)., Advances in Econometrics, Innovation in Quantitative economics: Essays in Honor of Robert L. Basmann, Vol. 5, JAI Press Greenwich, CT, pp. 227–239.

Williams, M.A., 1986. An economic application of bootstrap statistical methods: addyston pipe revisited. American Economic Review 30, 52–58.