Simulation Based Finite and Large Sample Inference Methods in Multivariate Regressions and Seemingly Unrelated Regressions [1]

Jean-Marie Dufour [2] and Lynda Khalaf [3]

November 1996
This version: February 1998

[1] This work was supported by 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 and Michael Veall for 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, and the Canadian Economics Association.

[2] C.R.D.E, Université de Montréal, C.P. 6128 Succursale Centre Ville, Montréal, Québec, Canada H3C 3J7. TEL: (514) 343 6557. FAX: (514) 343 5831.
e-mail: dufour@plgcn.umontreal.ca;
Web page: http://www.crde.umontreal.ca/personnel/dufour.html

[3] 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 regression (MLR) and seemingly unrelated regressions (SURE) models it is well known that commonly employed asymptotic test criteria are seriously biased towards overrejection. In this paper, we propose finite and large sample likelihood based test procedures for possibly non-linear hypotheses on the coefficients of MLR and SURE systems. Two complementary approaches are described. First, we derive general nuisance-parameter free bounds on the distribution of standard likelihood ratio criteria. Even though it may be difficult to compute these bounds analytically, they can easily be obtained by simulation, in such a way that the resulting bounds test has the desired level in finite samples. Second, we propose Monte Carlo tests which can be run whenever the bounds are not conclusive. These include, in particular, quasi-likelihood ratio criteria based on non-maximum-likelihood estimators (which may be simpler to compute). Illustrative Monte Carlo experiments show that: (i) the bounds are sufficiently tight to yield conclusive results in a large proportion of cases, and (ii) the randomized procedures correct all the usual size distortions in such contexts. We also present an extension of standard tests of uniform linear hypotheses in MLR contexts to non-Gaussian error distributions; in fact, the normality assumption is not necessary for most of the results we obtain. The procedures proposed are finally applied to test restrictions on a factor demand model.

Key words: multivariate linear regression; seemingly unrelated regressions; Monte Carlo test; bounds test; nonlinear hypothesis; finite sample test; exact test; bootstrap; capital asset pricing model (CAPM); factor demand; cost function.

RÉSUMÉ

Il est bien connu que les critères de test asymptotiques usuels dans le contexte des modèles de régression multivariés (MLR) et des régressions empilées (SURE) tendent à rejeter trop souvent par rapport à leur niveau nominal. Dans ce texte, nous proposons des procédures de type quotient de vraisemblance pour des hypothèses possiblement non linéaires sur les coefficients de modèles MLR et SURE. Nous décrivons deux approches complémentaires. Premièrement, nous obtenons des bornes libres de paramètres de nuisance sur la distribution des statistiques de quotient de vraisemblance. Même si ces bornes se prêtent difficilement à une évaluation par des méthodes analytiques, on peut les approximer aisément par simulation et utiliser l'approximation simulée d'une façon telle que le niveau du test fondé sur la borne soit contrôlé peu importe la taille d'échantillon. Deuxièmement, nous proposons des tests de Monte Carlo qui peuvent s'appliquer lorsque le test à borne n'est pas concluant. Ces derniers incluent notamment des procédures de type quasi-quotient de vraisemblance fondées sur des estimateurs plus faciles à calculer que les estimateurs du maximum de vraisemblance. Nous présentons des résultats d'expériences de Monte Carlo montrant que: (i) les bornes sont suffisamment serrées pour produire des résultats concluants dans une grande majorité des cas, et (ii) les procédures de tests de Monte Carlo corrigent les problèmes de taille de test dans de tels contextes. Nous présentons aussi une extension des tests standards d'hypothèses uniformes linéaires dans le contexte des modèles MLR au cas d'erreurs non-gaussiennes. De fait, l'hypothèse de normalité des erreurs n'est pas nécessaire pour la plupart des résultats que nous obtenons. Nous présentons un application des techniques proposées afin de tester des restrictions sur un modèle de demande de facteurs de production.

Mots-clefs: modèle de régression multivarié; régressions empilées (SURE); test de Monte Carlo; test à bornes; hypothèse non linéaire; test à distance finie; test exact; bootstrap; modèle de prix d'actifs financiers (CAPM); demande de facteurs; fonction de coût.

Contents

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. Extensive discussion of this problem can be found in the statistics literature on multivariate analysis of variance (MANOVA) and the econometric literature on seemingly unrelated regressions (SURE). The MLR model can be viewed as a special case of the SURE model where the regressor matrices for the different equations are the same. Conversely, the SURE specification may be viewed as a special case of the MLR model constrained by various exclusion restrictions on the different equations.

In the MLR framework, several finite sample procedures have been proposed for testing linear restrictions. These include, in particular, tests based on 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). Yet the use of these methods is limited to very specific problems: tests of *uniform mixed linear* hypotheses [Berndt and Savin (1977)]. Examples of *uniform mixed linear* constraints include: (i) the case where the same transformations of the regression coefficients are set to given values, within or across equations, and (ii) the hypothesis that a single parameter equals zero. Further, in most instances, exact distributional results are difficult to exploit and approximate distributions are suggested. Thus far less restrictive testing problems have not apparently been considered from a finite sample perspective, with perhaps the notable exception of the 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 decomposition of the OLS residuals. These observations suggest that this test has limited practical interest.

In connection with the SURE model introduced by Zellner (1962), the standard literature on hypothesis tests is asymptotic; see, for example, Srivastava and Giles (1987). Very few analytical finite sample results are available. A rare exception is provided by Harvey and Phillips (1982, Section 3) who derived independence tests between the disturbances of an equation and those of the other equations of a SURE model. The tests involve conventional F-statistics and are based on the residuals obtained from regressing each dependent variable on all independent variables of the system. Of course this problem is a very specific one. In a different vein, Phillips (1985) derived the exact distribution of a two-stage SURE estimator using a fractional matrix calculus. However, the analytical expressions obtained are very complex and, more importantly, involve unknown nuisance parameters, namely the elements of the error covariance matrix. The latter fact makes the application of Phillips' distributional results to practical hypothesis testing problematic.

Asymptotic Wald, Lagrange multiplier and likelihood ratio tests are available and commonly employed in econometric applications of the MLR model; see for example, 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 moderate to large. Well known examples include Laitinen (1978), Meisner (1979), Bera, Byron and Jarque (1981) and Theil and Fiebig (1985) in the context of homogeneity and symmetry testing in demand systems. Further evidence is reported in relation to multivariate tests of the CAPM; see for example Stambaugh (1982), Jobson and Korkie (1982), Amsler and Schmidt (1985) and MacKinlay (1987). These and other references are discussed in Stewart (1997).

It is clear that standard asymptotic approximations are quite unsatisfactory in this context. Attempts to improve those 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 and SURE 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 proper alternative to standard asymptotic theory, the bootstrap only has an asymptotic justification when the null distribution of the test statistic involves nuisance parameters, hence the finite sample validity of resulting inference remains to be established. This point should be born in mind while interpreting results on the usefulness of the bootstrap. For general discussion of bootstrap methods, the reader may consult Hall (1992), Efron and Tibshirani (1993) and

Shao and Tu (1995); about some econometric applications see Jeong and Maddala (1993), Vinod (1993) and Davidson and MacKinnon (1996*a*, 1996*b*). 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.

In this paper, we propose finite sample likelihood based tests for possibly non-linear hypotheses on the coefficients of seemingly unrelated regressions. We discuss two approaches that can be applied on their own or sequentially, namely: (i) a conservative bounds test, and (ii) Monte Carlo tests. Practical implementation of both procedures is simple. The methods we propose are best 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, it is shown that likelihood ratio (LR) statistics have null distributions which are boundedly pivotal, *i.e.* which admit nuisance-parameter-free bounds. Even though it may be difficult to compute analytically these bounds, they can easily be obtained by simulation. Here, we will apply this result in the context of MLR and SURE systems. The implications for hypothesis testing are two-fold. First, the finite sample bounds on the LR criterion easily yield conservative tests. Second, bootstrap techniques can lead to tests with correct levels.

To be more specific, we give at this point a preliminary discussion of the proposed conservative bound with regards to SURE systems. First, we reconsider the testing problem within the framework of an appropriate MLR model, namely the MLR setup of which the model on hand is a restricted form. Secondly, we introduce, in the relevant MLR framework, a uniform linear hypothesis that is a special case of the general restrictions in the null. The intuition behind this suggestion follows from the fact that exact nuisance-parameter free critical values for the LR criterion are available when the null is uniform linear within a MLR. Indeed, it turns out that the LR criterion for testing the suggested uniform linear hypothesis conveniently bounds the LR statistic for testing the general constraints.

In addition, we propose alternative Monte Carlo tests [see Dwass (1957), Barnard (1963), Jöckel (1986) or Dufour (1995)] that can be run whenever the bounds tests are not conclusive. We consider: (i) an asymptotically valid procedure that may be interpreted as a parametric bootstrap, and (ii) a method which is exact for any sample size, following Dufour (1995). While the normality assumption underlies the motivation for the statistics we consider, this is not necessary for most of the results obtained. In fact, we discuss an extension of standard tests of uniform linear hypotheses in MLR contexts to non-Gaussian distributions. Further, in situations where maximum likelihood (ML) methods may be computationally expensive, we introduce LR-type test criteria based on non-ML estimators. In particular, we consider two-stage statistics or estimators at any step of the process by which the likelihood is maximized iteratively. We emphasize that Monte Carlo and bounds tests should be viewed as complementary rather than alternative procedures.

The paper is organized as follows. Section 2 develops the notation and definitions. Section 3 discusses the known distributional results pertaining to the test criteria in the context of the MLR model and provides an extension of standard tests of uniform linear hypothesis to non-Gaussian distributions. Section 4 presents test statistics for general linear hypotheses in the MLR model and establishes bounds on the significance points for these statistics. We also discuss how to apply the results to non-linear and inequality restrictions. The generalization to the SURE model is discussed in Section 5. Simulation results are reported in Section 6. Section 7 illustrates the procedures proposed by applying them to test restrictions on a factor demand model, and Section 8 concludes.

2 Framework

In this section, we introduce the models and notations to be used in the paper. The first model we consider is the MLR model. Then, we focus our attention on the SURE model, which can be viewed as a special case of the MLR model obtained by imposing different exclusion restrictions on the different equations of a MLR model.

2.1 The multivariate linear regression model

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, $B = [\beta_1, \ldots, \beta_p]$ 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 with covariance matrix Σ where $\det(\Sigma) \neq 0$. 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; for further reference, let $W = [W_1, \ldots, W_n]' = UG'$, where $G = J^{-1}$. In particular, this condition will be satisfied when

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

in which case the covariance matrix of $\widetilde{U}_i$ is $JJ' = (G'G)^{-1}$. 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 likelihood function (MLF) over the unrestricted parameter space is

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

To derive the distribution of the relevant test statistics, we shall exploit the following decomposition of the sum of squared error (SSE) matrix $\widehat{U}'\widehat{U}$:

$$\begin{aligned} \widehat{U}'\widehat{U} &= U'MU = G^{-1}\left(UG'\right)'M\left(UG'\right)\left(G^{-1}\right)' \\ &= G^{-1}W'MW\left(G^{-1}\right)' \end{aligned} \tag{2.8}$$

where the matrix $W = UG'$ has a distribution that does not involve nuisance parameters.

2.2 The seemingly unrelated regressions model

Let us now consider the following p equation regression model:

$$Y_i = X_i\beta_i + U_i\ ,\ i = 1,\ \ldots\ ,\ p\ , \tag{2.9}$$

where X_i is a $n \times k_i$ full-column rank matrix of fixed regressors and $U_1, U_2\ ,\ \ldots\ , U_p$ satisfy the same distributional assumptions as in (2.2) - (2.3). This system is known as the SURE model. Let

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ Y_p \end{bmatrix}\ ,\ X^* = \begin{bmatrix} X_1 & 0 & \cdot & 0 \\ 0 & X_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & X_p \end{bmatrix}\ ,\ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_p \end{bmatrix}\ ,\ u = \begin{bmatrix} U_1 \\ U_2 \\ \cdot \\ U_p \end{bmatrix}\ .$$

Then an alternative compact representation of the model is

$$y = X^*\beta + u\ . \tag{2.10}$$

In this case, β is $k^* \times 1$ with $k^* = \sum_{i=1}^{p} k_i$. The likelihood function associated with (2.10) is:

$$L_1 = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln(|\Sigma|) - \frac{1}{2}(y - X^*\beta)'(\Sigma^{-1} \otimes I_n)(y - X^*\beta) \quad (2.11)$$

which is usually maximized using iterative numerical procedures.

To develop finite sample tests for the SURE model, we will find useful to explicate the relation between SURE and MLR models. Let $\overline{X}_i,\ i = 1, \ldots, p$, denote the explanatory variables excluded from the i-th equation. Further, define Z as any full column rank $n \times k$ matrix which spans the same space as $[X_1,\ X_2,\ \ldots,\ X_p]$, and $J_i,\ \overline{J}_i$ selection matrices such that $ZJ_i = X_i,\ Z\overline{J}_i = \overline{X}_i,\ i = 1, \ldots, p$. Then (2.9) may be written as

$$Y_i = ZJ_i\beta_i + Z\overline{J}_i\overline{\beta}_i + U_i\ ,\ i = 1,\ \ldots\ ,\ p\ , \quad (2.12)$$

with the restriction

$$\overline{\beta} = \left[\overline{\beta}_1',\ \overline{\beta}_2',\ \ldots\ ,\ \overline{\beta}_p'\right]' = 0\ . \quad (2.13)$$

To simplify notation, we will rewrite the latter model as

$$y = (I_p \otimes Z)\widetilde{\beta} + u \quad (2.14)$$

where $\widetilde{\beta} = \left[\widetilde{\beta}_1',\ \widetilde{\beta}_2',\ \ldots\ ,\ \widetilde{\beta}_p'\right]'$, and $\widetilde{\beta}_i$ is a $k \times 1$ vector which includes all the elements of β_i and $\overline{\beta}_i$ (though possibly not in the same order).

3 Uniform linear hypotheses in the multivariate linear model

In this section, we shall review known finite sample distributional results pertaining to various criteria for testing a general linear hypothesis in the context of the MLR model (2.1), and provide some extensions that will allow analogous tests to be performed in a large set of models with Gaussian or non-Gaussian errors, provided the latter have a distribution which is specified up to the unknown matrix J. Finite sample procedures are available only for the case where the constraints take the special uniform linear (UL) form

$$H_0 : RBC = D \quad (3.1)$$

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

$$H_{01} : R\beta_i = \delta_i\ ,\ i = 1,\ \ldots\ ,\ p\ ,$$

which corresponds to $C = I_p$. In this context, the most commonly used criteria are: the likelihood ratio (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)]. All these test criteria are functions of the roots $m_1, m_2, \ldots, m_p$ of the equation

$$\left|\widehat{U}'\widehat{U} - m\widehat{U}_0'\widehat{U}_0\right| = 0 \tag{3.2}$$

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

$$LR = -n\ln(\mathbf{L})\ ,\ \mathbf{L} = |\widehat{U}'\widehat{U}|/|\widehat{U}_0'\widehat{U}_0| = \prod_{i=1}^{p} m_i\ ; \tag{3.3}$$

$$LH = \sum_{i=1}^{p}(1 - m_i)/m_i\ ; \tag{3.4}$$

$$BNP = \sum_{i=1}^{p}(1 - m_i)\ ; \tag{3.5}$$

$$MR = \max_{1\leq i\leq p}\ (1 - m_i)/m_i\ . \tag{3.6}$$

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).

In Section 2, we saw that $\widehat{U}'\widehat{U}$ can be expressed as $\widehat{U}'\widehat{U} = G^{-1}W'MW(G^{-1})'$ which depends on Σ only through G. Similarly, $\widehat{U}_0'\widehat{U}_0$ can be expressed as

$$\widehat{U}_0'\widehat{U}_0 = G^{-1}W'M_0W(G^{-1})' \tag{3.7}$$

where $M_0 = I - X(X'X)^{-1}(X'X - R'(R(X'X)^{-1}R')^{-1}\ R)\ (X'X)^{-1}X'$. These observations yield the following basic result which allows one to derive finite sample tests based on the above criteria.

Theorem 1 *Under* (2.1), (2.2) *and* H_{01}, *the vector* $(m_1, m_2, \ldots, m_p)'$ *of the roots of* (3.2) *is distributed like the vector of the corresponding roots of*

$$\left|W'M\,W - mW'M_0W\right| = 0 \tag{3.8}$$

where M *is defined as in* (2.6), M_0 *as in* (3.7), $W = UG'$ *with* $G = J^{-1}$, *and the roots are put in descending order in both cases.*

PROOF: From (2.8) and (3.7), we have:

$$\widehat{U}'\widehat{U} = G^{-1}W'MW(G^{-1})' ,$$
$$\widehat{U}_0'\widehat{U}_0 = G^{-1}W'M_0W(G^{-1})' .$$

Consequently, the determinantal equation (3.2) can be expressed as

$$\left|G^{-1}W'MW(G^{-1})' - mG^{-1}W'M_0W(G^{-1})'\right| = 0 ,$$

hence

$$\left|G^{-1}\right| \left|W'MW - mW'M_0W\right| \left|(G^{-1})'\right| = 0$$

and

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

Since the vector $w = vec(W_1 , \ldots , W_N)$ has a completely specified distribution, the roots of equation (3.8) have distributions which does not involve Σ. 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 the assumption (2.2). Although Theorem 1 is not explicitly stated by Anderson (1984) or Rao (1973), it can be obtained easily by looking at their demonstrations. Since an explicit proof of Theorem 1 is not apparently available, we supply one above. On the basis of this theorem, the distribution of the Wilks' **L** criterion can be readily established.

Corollary 2 *Under the assumptions of Theorem 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 2 as follows.

Corollary 3 *Under the assumptions of Theorem 1, Wilks'* **L** *statistic for testing* H_{01} *is distributed like* $|W'MW| / |W'M_0W|$.

Note that the above characterization of the exact distribution does not require the normality assumption. Eventually, when the normality hypothesis (2.3) holds, the distribution of the Wilks criterion, as stated in Theorem 4 below, is well known [see Anderson (1984)].

Theorem 4 *Under* (2.1), (2.2), (2.3) *and* H_{01}, *Wilks'* **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.

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 cannot be assessed analytically. However, the above results can be used to obtain randomized or Monte Carlo tests that are applicable given the distributional assumption

(2.2). Such 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), Dufour and Kiviet (1996, 1998), Kiviet and Dufour (1997), and Dufour, Farhat, Gardiol and Khalaf (1998).

Let T_0 denote the observed test statistic T, where T is the adopted test criterion, for instance LR, as defined in (3.3). By Monte Carlo methods and for a given number N of replications, generate T_j, $j = 1, \ldots, N$ independent realizations of the statistic in question, under the null hypothesis. 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 T_j, $j = 0, \ldots, N$ in non-decreasing order and obtain the MC p-value $\widehat{p}_N(T_0)$ where

$$\widehat{p}_N(x) = \frac{N\widehat{G}_N(x) + 1}{N + 1}\,, \tag{3.9}$$

with

$$\widehat{G}_N(x) = \tfrac{1}{N}\sum_{i=1}^{N} I_{[0,\infty]}\,(T_i - x), \quad I_A(z) = \begin{cases} 1, & if\ z \in A \\ 0, & if\ z \notin A \end{cases}\,. \tag{3.10}$$

Then the test's critical region corresponds to

$$\widehat{p}_N(T_0) \le \alpha\,,\ 0 < \alpha < 1. \tag{3.11}$$

Under the assumptions of Theorems 1 or 4, the latter critical region is provably exact.

We now turn to the general UL hypothesis (3.1). In this case, the constrained MLE may simply be obtained by maximizing the likelihood associated with

$$Y_c = XB_c + U_c \tag{3.12}$$

where $Y_c = YC$, $B_c = BC$ and $U_c = UC$ with covariance $C'\Sigma C$, subject to $RB_c = D$. The resulting Wilks test statistic will satisfy the assumptions of Theorems 1 or 4. This extends the results established above to the general UL case.

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

$$\left(\frac{\rho\tau - 2\lambda}{rc}\right)\frac{1 - \mathbf{L}^{1/\tau}}{\mathbf{L}^{1/\tau}} \sim F(rc\,,\ \rho\tau - 2\lambda) \tag{3.13}$$

where

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

and

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

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 (3.13) holds asymptotically [Rao (1973, Chapter 8)]. Stewart (1997) provides an extensive discussion of these special F tests; see also Shukur and Edgerton (1994).

In Section 6, we report simulations on Monte Carlo tests based on the above finite sample theory. For a proof of Theorem 4 and a review of asymptotic results pertaining to the criteria (3.3) - (3.6), the reader may consult Anderson (1984, Chapter 8) or Rao (1973, Chapter 8). Finally, recall that not all linear hypotheses can be expressed as in H_0; we discuss other linear hypotheses in the following section.

4 General hypotheses in the multivariate linear model

In the context of (2.4) consider the general hypothesis

$$H_{01} : R^* b \in \Delta_0 \tag{4.1}$$

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 non-linear and inequality constraints. The relevant LR statistic is:

$$LR = n \ln(\Lambda^*) \ , \quad \Lambda^* = |\widehat{\Sigma}_{01}| / |\widehat{\Sigma}| \tag{4.2}$$

where $\widehat{\Sigma}_{01}$ and $\widehat{\Sigma}$ are the MLE of Σ imposing and ignoring H_{01}. The null distribution of LR is nuisance parameter dependent [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.

Consider the MLR model (2.4) and let $L(H_{12})$ denote the unrestricted MLF. In the Gaussian model, $L(H_{12})$ is expressed by (2.7). Further, suppose we can find another set of UL restrictions $H_{02} : \widetilde{R}BC = D$ such that $H_{02} \subseteq H_{01}$. Now define $L(H_{0i}),\ i = 1,\ 2$, to be the MLF under H_{0i}. Given assumption (2.3),

$$L(H_{0i}) = -\ \frac{np}{2} \ln(2\pi) \ - \ \frac{n}{2} \ \ln(\left|\widehat{\Sigma}_{0i}\right|) \ - \frac{np}{2},\ i = 1,\ 2\,, \tag{4.3}$$

where $\widehat{\Sigma}_{02}$ is the MLE under H_{02}. Then it is straightforward to see that

$$L(H_{02}) \leq L(H_{01}) \leq L(H_{12}) \ . \tag{4.4}$$

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

$$\Lambda^* \leq \Lambda^*_c \tag{4.5}$$

where

$$\Lambda_c^* = |\widehat{\Sigma}_{02}|/|\widehat{\Sigma}| \, . \tag{4.6}$$

It follows that $P[\Lambda^* \geq x] \leq P[\Lambda_c^* \geq x]$, $\forall x$, where $P[\Lambda_c^* \geq x]$ is nuisance-parameter free and may be obtained in finite samples as shown in Section 3. Under (2.3) the null distribution of Λ_c^* 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 null distribution of Λ_c^* may thus be easily obtained by simulation. Let $\Psi_\alpha(\cdot)$ be such that

$$P[\Psi(v_1, v_2, v_3) \geq \Psi_\alpha(v_1, v_2, v_3)] = \alpha \, , \; 0 \leq \alpha \leq 1 \, , \tag{4.7}$$

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 (4.7) may be rewritten as

$$P[\Lambda^* \geq \Psi_\alpha(n-k, \; p, \; \widetilde{q})] \leq \alpha \, , \; 0 \leq \alpha \leq 1 \, , \tag{4.8}$$

where $\widetilde{q} = \min(r, c), \; r = rank(\widetilde{R}), \; c = rank(C)$. Consequently, the critical value Q_α defined by

$$Q_\alpha = \Psi_\alpha(n-k, \; p, \; \widetilde{q}) \tag{4.9}$$

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. We proceed next to state our main conclusion for the Gaussian model.

Theorem 5 *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* $q^* \times k$ *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* $R^*b \in \Delta_0$. *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* $\widetilde{q} = \min(r, c), \; r = rank(\widetilde{R}), \; c = rank(C)$ *and* $\Psi_\alpha(\cdot)$ *is defined by* (4.7).

At this point, it is worth noting that normality [hypothesis (2.3)] by no way constitutes a necessary assumption in this case. Indeed, the critical values of the bounding statistic may still be determined by simulation under the general assumption (2.2). For further reference, we call the MC test based on the conservative bound a BMC test. Inequality (4.5) follows from the properties of least squares estimation irrespective of the density function. For the sake of generality, we now restate our main result for model (2.4) given the distributional assumption (2.2).

Theorem 6 *Consider the MLR model* (2.4) *with* (2.2). *Let* Λ^* *be the Wilks statistic defined by* (4.2) *for testing* $R^*b \in \Delta_0$ *where* R^* *is a* $q^* \times k$ *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* $R^*b \in \Delta_0$ *with* $\widetilde{q} = \min(r, c), \; r = rank(\widetilde{R}), \; c = rank(C)$. *Let* Λ_c^* *be the Wilks criterion for testing the latter restrictions. Then under the null hypothesis,* $P[\Lambda^* \geq \lambda_c^*(\alpha)] \leq \alpha$, *for all* $0 \leq \alpha \leq 1$, *where* $\lambda_c^*(\alpha)$ *is determined such that* $P[\Lambda_c^* \geq \lambda_c^*(\alpha)] = \alpha$.

Clearly, the above results hold when the hypothesis is linear of the form $R^*b = \delta_0$. In addition, the fact that the null distribution of the LR statistic can be bounded (in a non trivial way) as in Theorems 5 and 6 implies that MC test techniques may be used to obtain valid inference based on the statistic in (4.2) when the bounds test is not conclusive. In Section 3, we described the MC test procedure in the context of pivotal statistics. Dufour (1995) discusses extensions of MC tests in the presence of nuisance parameters. We now briefly outline the underlying methodology.

Consider a test statistic T for testing an hypothesis H_0, and suppose the null distribution of T depends on an unknown parameter vector θ. From the observed data, compute: (i) the test statistic T_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. Then compute $\widehat{p}_N(T_0|\widehat{\theta}_n^0)$, where $\widehat{p}_N(x|\overline{\theta})$ refers to $\widehat{p}_N(x)$ based on realizations of T generated given $\theta = \overline{\theta}$ and $\widehat{p}_N(x)$ is defined in (3.9). A MC test may be based on the critical region

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

For further reference, we denote the latter procedure a *local* Monte Carlo (LMC) test. Dufour (1995) gives general conditions under which LMC tests have the correct level asymptotically (as $n \to \infty$), *i.e.,* under H_0,

$$\lim_{n\to\infty} \left\{ P[\widehat{p}_N(T_0|\widehat{\theta}_n^0) \leq \alpha] - P[\widehat{p}_N(T_0|\theta) \leq \alpha] \right\} = 0 \ . \tag{4.10}$$

In particular, these conditions are usually met whenever the test criterion involved is asymptotically pivotal. To obtain an exact critical region, the MC p-value ought to be maximized with respect to the intervening parameters. Specifically, Dufour (1995) demonstrates that the test (henceforth denoted *maximized* Monte Carlo (MMC) test) based on the critical region

$$\sup_{\theta \ \in \ M_0} [\widehat{p}_N(T_0|\theta)] \leq \alpha$$

where M_0 is the nuisance parameter space under the null is exact at level α.

The LMC test procedure is closely related to a parametric bootstrap, with however a fundamental difference. Whereas bootstrap tests are valid as $N \to \infty$, the number of simulated samples used in MC tests is explicitly taken into account. Further the LMC p-value may be viewed as exact in a *liberal* sense, *i.e.* if the LMC fails to reject, we can be sure that the exact test involving the maximum p-value is not significant at level α. We emphasize the fact that the MMC test can be implemented in complementarity with the above defined bounds tests. Indeed, if the BMC test rejects the null then the MMC test is most certainly significant. These issues will be further studied in Section 6, in the context of a Monte Carlo experiment. We next employ a well known example from the finance literature to illustrate how the above results may be used.

A fundamental problem in financial economics involves testing the efficiency of a candidate benchmark portfolio. Let R_{jt}, R_{kt}, $j = 1, \ldots, p$, $k = 1, \ldots, K$ be security returns for period t, $t = 1, \ldots, T$. The hypothesis of interest is that some portfolio of the K security subset is efficient with respect to the total set of $p + K$ securities. If it is assumed that a riskless asset R_F exists, then efficiency can be tested using the following multivariate regression:

$$r_{jt} = \alpha_j + \sum_{k=1}^{K} \beta_{jk} r_{kt} + \epsilon_{jt}\ ,\ j = 1, \ldots, p\ ,\ t = 1, \ldots, T\ , \tag{4.11}$$

where $r_{jt} = R_{jt} - R_{Ft}$, $r_{kt} = R_{kt} - R_{Ft}$. The hypothesis of efficiency implies that the intercepts α_j are jointly equal to zero, *i.e.*

$$\alpha_j = 0\ ,\ j = 1\ , \ldots, p\ . \tag{4.12}$$

A well known example of (4.11) is the capital asset pricing model (CAPM)

$$R_{jt} - R_{Ft} = \alpha_j + \beta_j(R_{Mt} - R_{Ft}) + \epsilon_{jt}\ ,\ j = 1\ , \ldots, p\ ,\ t = 1\ , \ldots, T\ ,$$

where R_{Mt} are the returns on the market benchmark. Gibbons, Ross and Shanken (1989, GRS) show that a transformation of the LR criterion to test (4.12) has an exact F distribution given normality of asset returns. MacKinlay (1987) proposes a similar statistic in the context of a single beta CAPM. Specifically, GRS suggest the following test statistic:

$$Q = \frac{T\widehat{\alpha}'\widehat{S}^{-1}\widehat{\alpha}}{1 + \overline{r}'\widehat{\Delta}^{-1}\overline{r}} \tag{4.13}$$

where $\widehat{\alpha}$ is the vector of intercept OLS estimates, $\widehat{S} = \frac{T}{T-K-1}\widehat{\Sigma}$ is the unbiased estimator of Σ, $\overline{r} = (\overline{r}_1, \ldots, \overline{r}_K)'$ is the vector of time series means for $r_t = (r_{1t}, \ldots, r_{Kt})'$, and $\widehat{\Delta}$ is the sample covariance matrix for r_t. Under (4.12), Q has the *Hotelling* $T^2(p,\ T - K - 1)$ distribution or alternatively,

$$\frac{Q(T - K - p)}{p(T - K - 1)} \sim F(p, T - K - p)\ . \tag{4.14}$$

Let Λ_c denote the statistic defined by (4.2) in this context. It can be shown [see, for example Stewart (1997)] that Λ_c is related to the GRS criterion as follows:

$$\Lambda_c - 1 = \frac{Q}{T - K - 1}\ . \tag{4.15}$$

The econometric analysis is more complicated when the zero beta intercept is unknown and must be inferred using the return data [see, for example Gibbons (1982)]. In this case, the excess-return MLR becomes

$$R_{jt}-\gamma = \alpha_j+\sum_{k=1}^{K}\beta_{jk}(R_{kt}-\gamma) + \epsilon_{jt}\ ,\ j = 1\ ,\ \ldots\ ,\ p\ ,\ t = 1\ ,\ldots,\ T\ , \tag{4.16}$$

where γ is the unknown zero-beta intercept. Under the hypothesis of mean-variance efficiency, there exists a scalar γ such that the vector of α_j is equal to zero. This implies the following non-linearly restricted form of the basic MLR for the returns:

$$R_{jt} = \alpha_j + \sum_{k=1}^{K}\beta_{jk}R_{kt} + \epsilon_{jt}\ ,\ \ j = 1\ ,\ \ldots\ ,\ p\ ,\ t = 1\ ,\ \ldots\ ,\ T\ , \tag{4.17}$$

with

$$\alpha_j = \gamma\left(1-\sum_{k=1}^{K}\beta_{jk}\right)\ ,\ j = 1\ ,\ \ldots\ ,\ p\ . \tag{4.18}$$

Suppose that we wish to test (4.18) and let Λ^* denote the associated statistic from (4.2). Exact tests for this specific problem have been studied by Shanken (1986) and more recently by Stewart (1997). In what follows, we show that the exact procedures in question may be obtained as an application of our general methodology. Shanken (1986) employs the statistic $Q(\widehat{\gamma})$, where, in the context of (4.16),

$$Q(\gamma) = \frac{T\widehat{\alpha}'(\gamma)\widehat{S}^{-1}\widehat{\alpha}(\gamma)}{1+(\overline{R}-\gamma\iota_K)'\widehat{\Delta}^{-1}(\overline{R}-\gamma\iota_K)}\ ,$$

$$\widehat{\gamma} = \underset{\gamma}{\mathrm{ARGMIN}}\ Q(\gamma)\ ,\ \widehat{\alpha}(\gamma) = \widehat{a}-\gamma(\iota_p-\widehat{\beta}\iota_K)\ ,$$

$\widehat{a}$ is the vector of intercept estimates, $\widehat{\beta}$ is the (p, K) matrix of OLS estimates, $\widehat{\Sigma}$ is the unbiased estimate of Σ, $\overline{R} = (\overline{R}_1,\ \ldots\ ,\ \overline{R}_K)'$ is the vector of time series means for $R_t = (R_{1t},\ \ldots\ ,\ R_{Kt})'$, $\widehat{\Delta}$ is the sample covariance matrix for R_t and ι_J denotes a vector of J 1's. Shanken shows that: (i) the LR statistic for testing (4.18) is a transformation of $Q(\widehat{\gamma})$, (ii) $\widehat{\gamma}$ is the constrained MLE of γ, and (iii) the null distribution of $Q(\widehat{\gamma})$ may be bounded by the Hotelling $T^2(N, T-K-1)$ distribution. Turning to our proposed bound on the statistic Λ^*, we suggest to consider the statistic Λ_c associated with the special case of (4.18) where γ is any known constant. By (4.13), (4.15) and using (4.5), this naturally leads to the use of conservative critical points involving the $F(p, T-K-N)$ distribution. This is the same result as the one obtained by Shanken (1986) and Stewart (1997).

5 Hypothesis testing in seemingly unrelated regressions

This section considers testing hypotheses about the parameters of the SURE model. Indeed, the results for the MLR model provide interesting applications for systems inference in SURE model. We consider the problem of testing in the context of model (2.10) a general hypothesis of the form

$$H_0 : C^* \beta \in \Delta_0^* \tag{5.1}$$

where C^* is a full row-rank $v_0^* \times k^*$ matrix, Δ_0^* is a non-empty subset of $\mathbb{R}^{v_0^*}$. In terms of the MLR model (2.14) which includes (2.10) as a special case, H_0 may be stated as

$$H_{01} : C\widetilde{\beta} \in \Delta_0 \tag{5.2}$$

where C is a full row-rank $v_0 \times (pk)$ matrix, Δ_0 is a non-empty subset of $\mathbb{R}^{v_0}$ and C is expressed so that it incorporates the SURE restrictions (2.13). The associated LR statistic is

$$LR = n \ln(\Lambda) \ , \quad \Lambda = |\widehat{\Sigma}_{01}| / |\widehat{\Sigma}_{11}| \tag{5.3}$$

where $\widehat{\Sigma}_{01}$ and $\widehat{\Sigma}_{11}$ are the MLE imposing (5.2) and (2.13). For the purpose of deriving the conservative bound, we also consider

$$LR = n \ln(\Lambda^*), \quad \Lambda^* = |\widehat{\Sigma}_{01}| / |\widehat{\Sigma}| \tag{5.4}$$

where $\widehat{\Sigma}$ is the unrestricted MLE. As it stands, testing H_{01} based on (5.4) is exactly the type of problem discussed in Section 4. Indeed, consider a UL hypothesis H_{02} such that $H_{02} \subseteq H_{01}$ and the associated statistic

$$\Lambda_c^* = |\widehat{\Sigma}_{02}| / |\widehat{\Sigma}| \tag{5.5}$$

where $\widehat{\Sigma}_{02}$ is the MLE under H_{02}. As established in Section 3, the exact null distribution of Λ_c^* is nuisance-parameter-free and may easily be simulated. Conformably with the notation in Section 4, let $L(H_{0i})$, $i = 1,\ 2$ be the MLF under H_{0i}. Further, let $L(H_{1i})$, $i = 1,\ 2$ refer to the MLF under (2.13) and the unrestricted maximum, respectively. Hence, the following inequality holds under the distributional assumption (2.3):

$$L(H_{02}) \leq L(H_{01}) \leq L(H_{11}) \leq L(H_{12}). \tag{5.6}$$

Consequently, it is straightforward to see that

$$\Lambda \leq \Lambda^* \leq \Lambda_c^*. \tag{5.7}$$

The critical bound may accordingly be obtained from the null distribution of Λ_c^* as described in Section 4. To facilitate the analysis, we shall, in the following, provide an illustrative example.

Example 1 ***Three equations SURE model***

In the following SURE model with Gaussian errors,

$$\begin{aligned} Y_1 &= \beta_{10} + \beta_{11} X_1 + U \ , \\ Y_2 &= \beta_{20} + \beta_{22} X_2 + U_2 \ , \\ Y_3 &= \beta_{30} + \beta_{33} X_3 + U_3 \ , \end{aligned} \tag{5.8}$$

consider testing

$$H_0 : \beta_{11} = \beta_{22} = \beta_{33} \ . \tag{5.9}$$

In the framework of the corresponding MLR model

$$\begin{aligned} Y_1 &= \beta_{10} + \beta_{11} X_1 + \beta_{12} X_2 + \beta_{13} X_3 + U_1 \ , \\ Y_2 &= \beta_{20} + \beta_{21} X_1 + \beta_{22} X_2 + \beta_{23} X_3 + U_2 \ , \\ Y_3 &= \beta_{30} + \beta_{31} X_1 + \beta_{32} X_2 + \beta_{33} X_3 + U_3 \ , \end{aligned} \tag{5.10}$$

H_0 is equivalent to the joint hypothesis $H_0^* : \beta_{11} = \beta_{22} = \beta_{33}$ and $\beta_{12} = \beta_{13} = \beta_{21} = \beta_{23} = \beta_{31} = \beta_{32} = 0$. In order to use the above results on the conservative bound, we need to construct a set of UL restrictions in the sense of Section 3 that satisfy the hypothesis in question. It is easy to see that the constraints setting the coefficients $\beta_{ij}, \ i, \ j = 1, \ldots, 3$ to specific values meet this purpose. All that remains is to calculate Wilks' statistic conforming with (5.3) and use the critical value defined by Theorems 5 - 6 as a conservative cut-off point.

In the next section, we examine the performance of LMC and BMC tests in SURE contexts given linear and non-linear restrictions. In the linear case, we also consider LMC tests based on standard Wald-type criteria and several alternative statistics justified on the basis of computational cost as opposed to those relying on full maximum likelihood estimation.

6 Simulation study

This section reports an investigation, by simulation, of the performance of the various proposed statistics in the context of MLR and SURE models. We considered the following designs.

D1. **MLR system, within-equation UL constraints**
Model: $Y_{ij} = \alpha_j + \sum_{k=1}^{p} \beta_{jk} X_{ik} \ , \ i = 1 \ , \ \ldots \ , \ n, \quad j = 1 \ , \ \ldots \ , \ p \ ;$
$H_0 : \sum_{k=1}^{p} \beta_{jk} = 0 \ , \ j = 1 \ , \ \ldots \ , \ p \ ;$
$p = 5, 7, 8 \ ; \quad n = 20, 25, 40, 50, 100 \, .$

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

D3. **MLR system, cross-equation constraints**
Model: $Y_{ij} = \alpha_j + \sum_{k=1}^{p} \beta_{jk} X_{ik}$, $i = 1$, ... , n , $j = 1$, ... , p ;
$H_0 : \beta_{jj} = \beta_{11}$, $j = 2$, ... , p and $\beta_{jk} = 0$, $j \neq k$, $j,\ k = 1$, ... , p ;
$p = 3, 5$; $n = 25$.

D4. **SURE system, cross-equation constraints**
Model: $Y_{ij} = \alpha_j + \beta_{jj} X_{ij}$, $i = 1$, ... , n , $j = 1$, ... , p ;
$H_0 : \beta_{jj} = \beta_{11}$, $j = 2$, ... , p ;
$p = 3, 5$; $n = 25$.

D5. **MLR system, non-linear constraints**
Model: $Y_{ij} = \alpha_j + \beta_j X_i$, $i = 1$, ... , n , $j = 1$, ... , p;
$H_0 : \alpha_j = \gamma(1 - \beta_j)$, $j = 1$, ... , p , γ unknown;
$p = 40$; $n = 60$.

D6. **SURE system, non-linear constraints**
Model: $Y_{ij} = \alpha_j + \beta_{j1} X_{1i} + \beta_{j2} X_{2ij}$, $i = 1$, ... , n , $j = 1$, ... , p ;
$H_0 : \beta_{j1} = \gamma \beta_{j2}$, $j = 1$, ... , p , γ unknown;
$p = 7$; $n = 25$.

For each model, the 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 coefficients for all experiments are reported in Table 1. The statistics examined are the relevant LR criteria defined by (3.3), (4.2), (5.3) and (5.4). The subscripts asy, BMC, LMC and MMC refer respectively to the standard asymptotic tests, the MC bounds tests, the *local* MC test and the *maximized* MC; because of the computational cost involved, the latter was only applied in the context of D3 and D4 with $p = 3$ and $N = 19$. For each test statistic, the LMC randomized procedure is based on simulations that use a restricted estimator similar to the estimator(s) involved in the corresponding test statistic: a restricted ML (or quasi-ML) estimator for LR or Wald-type tests based on ML (quasi-ML) estimators, restricted feasible GLS estimators for tests on GLS estimators. In D1 we have also considered the Bartlett-corrected LR test [Attfield (1995, Section 3.3)] which we denote LR_c. In D2 the asymptotic F test (3.13) was also assessed. In D4 we have studied, in addition to the standard LR criteria, the χ^2 and F Wald tests and Wald-type criteria based on feasible generalized least squares (FGLS) estimates. Specifically, we considered the statistic suggested in Srivastava and Giles (1987, Chapter 10):

$$\tau = \left(\frac{v_1}{v_0^*}\right) \frac{\left(C^*\widehat{\beta}\right)' \left[C^* \left(X^{*'} \left(S^{-1} \otimes I_n\right) X^*\right)^{-1} C^{*'}\right]^{-1} \left(C^*\widehat{\beta}\right)}{(y - X^*\widehat{\beta})' \left(S^{-1} \otimes I_n\right) (y - X^*\widehat{\beta})} \tag{6.1}$$

where C^* is the selection matrix implied by the null, $v_1 = np - k^*$, and S and $\widehat{\beta}$ are the feasible generalized least squares parameter estimates. Under the null hypothesis, $v_0^*\tau$, has a $\chi^2(v_0^*)$ asymptotic distribution. Theil (1971, Chapter 6) suggests that the $F(v_0^*, v_1)$ distribution better captures the statistics finite sample distribution. Yet this claim is not supported by either analytical or simulation evidence. Maximum likelihood estimators may be substituted for $\widehat{\beta}$ and S in the formulae for the Wald criterion. For further reference, the latter asymptotic tests and their MC counterparts are denoted W-MLE$_j$, W-GLS$_j$, $j \in \{\chi^2, F\}$ and W-MLE$_{LMC}$, W-GLS$_{LMC}$ respectively. We also introduced the following quasi-LR (QLR) statistics:

$$QLR_{(j)} = n \ln(\Lambda_{(j)}) \ , \ \Lambda_{(j)} = \left|\widetilde{\Sigma}_{\mathbf{0}}^{(j)}\right| / \left|\widetilde{\Sigma}^{(j)}\right| \tag{6.2}$$

where $\widetilde{\Sigma}_{\mathbf{0}}^{(j)}$ and $\widetilde{\Sigma}^{(j)}$ denote the constrained and unconstrained iterative estimators of Σ and the superscript j refers to the number of iterations involved. Though we did not analytically establish the asymptotic distribution of the latter criteria, we assessed their asymptotic significance using the χ^2 reference distribution for the usual LR statistic. We append the subscript LMC to the notation for the QLR test to refer to the corresponding LMC test. Further, criteria inspired by those suggested in Theil et al. (1985) were also studied. For the model with three equations, we considered:

$$\begin{aligned}
\mu_{31} &= |\widehat{\beta}_{11} - \widehat{\beta}_{22}| + |\widehat{\beta}_{22} - \widehat{\beta}_{33}| \quad , \\
\mu_{32} &= |\widehat{\beta}_{11} - \widehat{\beta}_{33}| + |\widehat{\beta}_{22} - \widehat{\beta}_{33}| \quad , \\
\mu_{33} &= |\widehat{\beta}_{11} - \widehat{\beta}_{22}| + |\widehat{\beta}_{11} - \widehat{\beta}_{33}| \quad .
\end{aligned}$$

In the five-equations case, the following were selected among the many possible choices:

$$\begin{aligned}
\mu_{51} &= |\widehat{\beta}_{11} - \widehat{\beta}_{22}| + |\widehat{\beta}_{22} - \widehat{\beta}_{33}| + |\widehat{\beta}_{33} - \widehat{\beta}_{44}| + |\widehat{\beta}_{44} - \widehat{\beta}_{55}| \quad , \\
\mu_{52} &= |\widehat{\beta}_{22} - \widehat{\beta}_{33}| + |\widehat{\beta}_{33} - \widehat{\beta}_{44}| + |\widehat{\beta}_{44} - \widehat{\beta}_{55}| + |\widehat{\beta}_{55} - \widehat{\beta}_{11}| \quad , \\
\mu_{53} &= |\widehat{\beta}_{33} - \widehat{\beta}_{44}| + |\widehat{\beta}_{44} - \widehat{\beta}_{55}| + |\widehat{\beta}_{55} - \widehat{\beta}_{11}| + |\widehat{\beta}_{11} - \widehat{\beta}_{22}| \quad , \\
\mu_{54} &= |\widehat{\beta}_{44} - \widehat{\beta}_{55}| + |\widehat{\beta}_{55} - \widehat{\beta}_{11}| + |\widehat{\beta}_{11} - \widehat{\beta}_{22}| + |\widehat{\beta}_{22} - \widehat{\beta}_{33}| \quad , \\
\mu_{55} &= |\widehat{\beta}_{55} - \widehat{\beta}_{11}| + |\widehat{\beta}_{11} - \widehat{\beta}_{22}| + |\widehat{\beta}_{22} - \widehat{\beta}_{33}| + |\widehat{\beta}_{33} - \widehat{\beta}_{44}| \quad .
\end{aligned}$$

The results are summarized in Tables 2 to 10. We report the empirical frequencies of type I errors, based on a nominal size of 5% and 1000 replications. In addition, the power of the tests in (D1, $n = 25$, $p = 8$), D3 and D4 were investigated by simulating the model with the same parameter values except for β_{11}. For the purpose of power comparisons, the sizes of the asymptotic tests were *locally corrected*, *i.e.* an independent simulation was conducted for the same parameter choices as the initial experiment to determine empirical 5% cut-off points. Of course, such corrected tests only constitute theoretical benchmarks and are not feasible in practice. The MC tests were applied with 19 and 99 replications and restricted estimates were used to generate the simulated samples. All the experiments were conducted using Gauss-386i VM version 3.1. Our results show the following.

Table 1. Coefficients for the simulation experiments

1.	$\beta_{kj} = \begin{cases} .1\,, & j=1\,,\ \dots\,,\ I[p/2] \\ .2, & j=I[p/2]+1\,,\ \dots\,,\ p \end{cases}\,,\ k=1\,,\ \dots\,,\ p-1$ $\beta_{pj} = \sum_{k=1}^{p-1}\beta_{kj},\ j=1\,,\ \dots\,,\ p\,,\quad \beta_{0j} = \begin{cases} 1.2\,, & j=1\,,\ \dots\,,\ I[p/2] \\ 1.8, & j=I[p/2]+1\,,\ \dots\,,\ p \end{cases}$
2.	R, B, C drawn (once) as $NID(0,I)$
3.	$B_{(3EQ)} = \begin{bmatrix} 1.2 & .8 & -1.1 \\ .1 & 0 & 0 \\ 0 & .1 & 0 \\ 0 & 0 & .1 \end{bmatrix}\,,\quad B_{(5EQ)} = \begin{bmatrix} 1.2 & .8 & -1.1 & 1.9 & -.2 \\ .1 & 0 & 0 & 0 & 0 \\ 0 & .1 & 0 & 0 & 0 \\ 0 & 0 & .1 & 0 & 0 \\ 0 & 0 & 0 & .1 & 0 \\ 0 & 0 & 0 & 0 & .1 \end{bmatrix}$
4.	$\beta_{(3EQ)} = \begin{pmatrix} 1.2 & .1 & .8 & .1 & -1.1 & .1 \end{pmatrix}'$ $\beta_{(5EQ)} = \begin{pmatrix} 1.2 & .1 & .8 & .1 & -1.1 & .1 & 1.9 & .1 & -.2 & .1 \end{pmatrix}'$
5.	$\gamma = .009$ and β_j, $j=1\,,\ \dots\,,\ p$, drawn (once) as $NID(0,.16)$
6.	$\gamma = .009$ and α_j, β_j, $j=1\,,\ \dots\,,\ p$, drawn (once) as $NID(0,.16)$

Table 2. Empirical levels of various tests: experiment D1

Sample size	5 equations			7 equations			8 equations		
	LR_{asy}	LR_c	LR_{MC}	LR_{asy}	LR_c	LR_{MC}	LR_{asy}	LR_c	LR_{MC}
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

$H_0 : \beta_{11} = .1$

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

Table 4. Empirical levels of various tests: experiment D2

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

Table 5. Empirical levels of various tests: experiment D3

3 equations				5 equations		
LR_{asy}	LR_{LMC}	LR_{BMC}	LR_{MMC}	LR_{asy}	LR_{LMC}	LR_{BMC}
.122	.055	.036	.036	.310	.044	.029

Table 6. Power of various tests: experiment D3. $H_0 : \beta_{11} = .1$

	19 replications					99 replications				
	3 equations									
β_{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_{MMC}	.095	.404	.799	.963	.987	.099	.441	.861	.986	.999
LR_{BMC}	.095	.404	.799	.963	.987	.099	.441	.861	.986	.999
	5 equations									
β_{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

Table 7. Empirical levels of various tests: experiment D4

Asymptotic tests			MC tests					
Test	3EQ	5EQ	Test	3EQ	5EQ	Test	3EQ	5EQ
$W-GLS_{\chi^2}$	.061	.130	$W-GLS_{LMC}$	.049	.047	μ_{31}	.058	-
$W-MLE_{\chi^2}$	.124	.254	$W-MLE_{LMC}$	.047	.049	μ_{32}	.051	-
$W-GLS_F$	.052	.121	LR_{LMC}	.047	.043	μ_{33}	.055	-
$W-MLE_F$	.111	.242	LR_{MMC}	.038		μ_{51}	-	.027
LR_{asy}	.094	.143	LR^*_{MMC}	.036		μ_{52}	-	.026
$QLR_{(0)}$	.068	.077	LR_{BMC}	.036	.029	μ_{53}	-	.025
$QLR_{(1)}$	.088	.131	$QLR_{(0)-LMC}$	.045	.052	μ_{54}	-	.011
$QLR_{(2)}$	.094	.143	$QLR_{(1)-LMC}$	.048	.052	μ_{55}	-	.025
			$QLR_{(2)-LMC}$	.047	.044			

TABLE 8. POWER OF THE BOUNDS TESTS. EXPERIMENT D4

	3 equations					5 equations				
β_{11}	.3	.5	.7	.9	1	.3	.5	.7	.9	1.0
p_1	.065	.383	.791	.963	.987	.082	.416	.792	.958	.995
p_2	.171	.324	.171	.034	.013	.249	.497	.207	.042	.005
p_3	.030	.021	.008	0.00	0.00	.038	.011	0.00	0.00	0.00
p_4	.734	.272	.030	.003	0.00	.631	.076	.001	0.00	0.00
p_1	.077	.434	.858	.986	.999	.075	.474	.877	.990	1.0
p_2	.204	.372	.127	.014	.001	.256	.439	.122	.010	0.00
p_3	.022	.007	.003	0.00	0.00	.035	.010	0.00	0.00	0.00
p_4	.697	.187	.012	0.00	0.00	.634	.077	.001	0.00	0.00

Note: p_1 is the empirical probability that LR_{LMC} and LR_{BMC} reject, p_2 measures the probability that LR_{BMC} fails to reject and LR_{LMC} rejects, p_3 measures the probability that LR_{BMC} rejects and LR_{LMC} fails to reject and p_4 is the empirical probability that both tests fail to reject.

TABLE 9. POWER OF VARIOUS TESTS: EXPERIMENT D4, 3 EQUATIONS

	19 replications					99 replications				
β_{11}	.3	.5	.7	.9	1.0	.3	.5	.7	.9	1.0
$W-GLS_{asy}$	.192	.647	.939	.993	.999	.192	.647	.939	.993	.999
$W-MLE_{asy}$	.264	.787	.984	1.0	1.0	.264	.787	.984	1.0	1.0
LR_{asy}	.281	.806	.985	1.0	1.0	.281	.806	.985	1.0	1.0
$W-GLS_{LMC}$	.185	.579	.884	.974	.986	.202	.640	.934	.990	.998
$W-MLE_{LMC}$	.225	.704	.958	.997	1.00	.260	.774	.985	1.00	1.00
LR_{LMC}	.236	.707	.962	.997	1.00	.262	.779	.985	1.00	1.00
$QLR_{(0)}$	.227	.689	.950	.993	.988	.256	.762	.977	.997	.999
$QLR_{(1)}$	.238	.709	.961	.997	1.00	.259	.776	.986	1.00	1.00
$QLR_{(2)}$	.236	.707	.962	.997	1.00	.262	.776	.985	1.00	1.00
LR^*_{MMC}	.095	.404	.799	.963	.987	.099	.441	.861	.986	.999
LR_{MMC}	.054	.388	.804	.978	.993	-	-	-	-	-
LR_{BMC}	.095	.404	.799	.963	.987	.099	.441	.861	.986	.999
μ_{31}	.076	.108	.148	.216	.259	.064	.108	.165	.219	.268
μ_{32}	.197	.552	.869	.974	.992	.210	.641	.935	.995	.998
μ_{33}	.093	.183	.307	.432	.489	.088	.184	.328	.503	.601

Table 10. Power of various tests: experiment D4, 5 equations

	19 replications					99 replications				
β_{11}	.3	.5	.7	.9	1.1	.3	.5	.7	.9	1.1
$W-GLS_{asy}$	.200	.703	.961	.994	.999	.200	.703	.961	.994	.999
$W-GLS_{asy}$	.317	.918	1.0	1.0	1.0	.317	.918	1.0	1.0	1.0
LR_{asy}	.331	.913	.999	1.0	1.0	.331	.913	.999	1.0	1.0
$W-GLS_{LMC}$	.162	.619	.918	.982	.998	.186	.684	.946	.990	.999
$W-GLS_{LMC}$	.265	.832	.991	.999	1.00	.297	.903	1.0	1.00	1.00
LR_{LMC}	.286	.841	.999	.999	1.00	.328	.908	.998	1.00	1.00
$QLR_{(0)}$	.265	.806	.971	.998	1.00	.316	.864	.983	.999	1.00
$QLR_{(1)}$	.290	.849	.988	.998	1.00	.334	.900	.997	1.00	1.00
$QLR_{(2)}$	.287	.842	.991	.999	1.00	.331	.908	.997	1.00	1.00
LR_{BMC}	.120	.427	.792	.958	.995	.110	.484	.877	.990	1.00
μ_{51}	.029	.034	.038	.041	.048	.032	.036	.039	.041	.044
μ_{52}	.031	.036	.039	.042	.045	.031	.034	.038	.040	.041
μ_{53}	.042	.085	.154	.258	.359	.035	.077	.152	.241	.397
μ_{54}	.023	.071	.159	.289	.456	.025	.067	.175	.302	.512
μ_{55}	.031	.050	.071	.118	.170	.033	.056	.092	.128	.180

6.1 MLR model with uniform linear hypothesis

Experiment D1 yields three main conclusions. First, it is evident from Table 2 that the LR_{asy} tests overreject substantially. As emphasized in Section 2, this problem is indeed well documented. Second, the Bartlett correction, though providing some improvement, fails in larger systems. In this regard, it is worth noting that Attfield (1995) had conducted a similar Monte Carlo study to demonstrate the effectiveness of Bartlett adjustments in this framework; however, the example analyzed was restricted to a two-equations model. Finally, the MC tests achieve perfect size control and 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. This experiment may be viewed as an illustration of homogeneity tests in demand systems.

Experiment D2 was designed to investigate the accuracy of the asymptotic F test (3.13) where $r > 2$ and $c > 2$. From the results in Table 4, we can see that LR test based on the asymptotic χ^2 critical value is severely oversized; the asymptotic F test performs better but size correction is still needed. The size of the MC test corresponds closely to 5%. As shown in Section 3, MC tests in the context of D1 and D2 are provably exact.

6.2 MLR model with cross-equation restrictions

Tables 5 - 6 summarize our findings in this case. A striking observation is that the conservative bound provides conclusive results in a large proportion of cases. Further, the LR_{BMC} and LR_{MMC} test yield equivalent decisions for all cases examined. LMC tests provide substantial improvement over conventional asymptotics: the randomized procedure corrects test sizes with no substantial power loss. Increasing the number of equations does not have a great effect on the relative performance of the 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.

6.3 Monte Carlo evidence: SURE with linear restrictions

Two Gaussian SURE models, modelled after Example 1, were used for experiment D4. Systems involving three and five equations were considered to which we will refer as the 3EQ and the 5EQ models. Our results show the following (see Tables 7 - 10):

1. The asymptotic criteria have an upward bias in size; as can be seen in Table 7, rejection of the null is repeatedly many times more than what it should be. The bias clearly worsens in the 5EQ example. Across the cases examined, the Wald-type statistics have larger sizes when based on their asymptotic χ^2 critical values. Although the F approximation seems to correct the problem in the 3EQ model, it clearly fails to do so in the 5EQ case.

2. The BMC test was found to be well behaved. Power gains are possible in other test problems where a tighter critical bound is available. Indeed, we have observed reasonable power even if we have experimented with the worst scenario, in the sense that bounding test statistics correspond to a null hypothesis which fixes the values of all regression coefficients (except the intercept). Furthermore, as in experiment D3, we found that the BMC and the MMC tests based on LR^* yield equivalent decisions for all cases examined; the MMC test based on LR performs marginally better. This illustrates the value of the conservative bounds test as a tool to be used in conjunction with LMC test methods and not necessarily as an alternative to those methods. As emphasized earlier, the bounds procedure is computationally inexpensive and exact. In addition, whenever the bounds test rejects, inference may be made without further appeal to randomized tests.

3. There is no indication of overrejection for all LMC tests considered. While the critical values used, conditional on the particular choice of consistent estimator for the error covariance matrix, are only asymptotically justified, the procedure was remarkably effective in correcting the bias. Whether this conclusion would carry to quite larger systems is indeed an open question. In this regard, note that

available simulation evidence on the SURE model, specifically the experiment in Rocke (1989) on *large* systems is limited to three-equations at best.

4. The LMC tests performed noticeably well in terms of power in all instances, even when the number of replications was as low as 19. We emphasize that the size-corrected asymptotic tests are *unavailable* in practical testing situations since the local correction it entails requires Σ to be known. The statistic with the best power properties across the alternatives examined was the randomized LR.

5. While they did exhibit adequate sizes, the statistics inspired by Theil et al. (1985) did not fare well in terms of power. For the 3EQ model, the performance was dramatically poor for μ_{32} and μ_{33} but less so in the case of μ_{31}. Even then, as compared to the randomized LR, the performance is less than satisfactory.

6. Simulation evidence does not favor the randomized usual LR tests over those based on $\Lambda_{(j)}$ typically involving fewer iterations, although we are uncertain as to the asymptotic equivalence of both procedures. This observation has an important bearing on empirical practice. The simplicity of the method based $\Lambda_{(j)}$ on has much to recommend it for larger models in which statistics requiring full MLE may be quite expensive to randomize.

6.4 Nonlinear hypotheses

Experiment D5 was modeled after the MLR system (4.11) under the null hypothesis (4.12). We considered 40 equations with 60 observations following the empirical example analyzed in Stewart (1997). To derive the LR statistic, the constrained MLE was numerically computed according to Shanken (1986). As was the case with linear hypotheses, the asymptotic tests severely overreject. Indeed, 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. We observed similar results in the context of the non-linear SURE model D6. In this case, we considered a seven equations system with 25 observations. The observed empirical sizes of the LR statistics were 12.5% whereas the levels of the bounds test (2.6%) is adequate.

7 Empirical illustration

In this section, we present an empirical application that illustrates the results presented in this paper. We consider testing restrictions on the parameters of a generalized Leontief cost function. We use the data from Berndt and Wood (1975) and the factor demand system from Berndt (1991, pp. 460-462). The model imposes constant returns to scale and linear homogeneity in prices, and includes four inputs: capital (K), labor (L), energy (E) and non-energy intermediate materials (M). If we denote the output by Y and the input prices P_j, $j = K, L, E, M$, the stochastic cost minimizing input-output KLEM equations are:

$$\frac{K}{Y} = d_{KK} + d_{KL}\,(P_L/P_K)^{1/2} + d_{KE}\,(P_E/P_K)^{1/2} + d_{KM}\,(P_M/P_K)^{1/2} + e_K$$

$$\frac{L}{Y} = d_{LL} + d_{LK}\,(P_K/P_L)^{1/2} + d_{LE}\,(P_E/P_L)^{1/2} + d_{LM}\,(P_M/P_L)^{1/2} + e_L$$

$$\frac{E}{Y} = d_{EE} + d_{EK}\,(P_K/P_E)^{1/2} + d_{EL}\,(P_L/P_E)^{1/2} + d_{EM}\,(P_M/P_E)^{1/2} + e_E$$

$$\frac{M}{Y} = d_{MM} + d_{MK}\,(P_K/P_M)^{1/2} + d_{ML}\,(P_L/P_M)^{1/2} + d_{ME}\,(P_E/P_M)^{1/2} + e_M$$

where the error terms $e_K,\ e_L,\ e_E,\ e_M$ satisfy the distributional assumptions (2.3). We focus on testing the symmetry restrictions entailed by microeconomic theory, *i.e.*

$$H_{01} : \begin{cases} d_{KL} = d_{LK}, & d_{KM} = d_{MK} \\ d_{KE} = d_{EK}, & d_{LM} = d_{ML} \\ d_{LE} = d_{EL}, & d_{EM} = d_{ME} \end{cases}$$

as well as a subset of these constraints

$$H_{02} : \quad d_{EM} = d_{ME}, \quad d_{KM} = d_{MK} \quad .$$

Conforming with the procedures described above, we reconsider the testing problem in the context of the MLR model of which the KLEM system is a restricted form. The latter model's individual equations include the 32 price ratios $(P_i/P_j)^{1/2}$, $i, j = K, L, E, M$, as regressors. The unrestricted MLE SURE estimates using the data provided in Berndt (1991) on the manufacturing US. sector 1947-71 are given below (with asymptotic standard errors in parentheses):

$$\widehat{\frac{K}{Y}} = \underset{(.0143)}{.0263} + \underset{(.0088)}{.0036}\,(P_L/P_K)^{1/2} + \underset{(.0301)}{.0649}\,(P_E/P_K)^{1/2} - \underset{(.0426)}{.0443}\,(P_M/P_K)^{1/2}$$

$$\widehat{\frac{L}{Y}} = -\underset{(.0157)}{.0719} + \underset{(.0245)}{.0517}\,(P_K/P_L)^{1/2} + \underset{(.0476)}{.2200}\,(P_E/P_L)^{1/2} + \underset{(.0676)}{.0264}\,(P_M/P_L)^{1/2}$$

$$\widehat{\frac{E}{Y}} = \underset{(.0183)}{.0403} - \underset{(.0088)}{.0111}\,(P_K/P_E)^{1/2} - \underset{(.0053)}{.0048}\,(P_L/P_E)^{1/2} + \underset{(.0259)}{.0150}\,(P_M/P_E)^{1/2}$$

$$\widehat{\frac{M}{Y}} = \underset{(.1214)}{.7401} - \underset{(.0420)}{.0542}\,(P_K/P_M)^{1/2} - \underset{(.0258)}{.1374}\,(P_L/P_M)^{1/2} + \underset{(.0855)}{.0399}\,(P_E/P_M)^{1/2}$$

For both hypotheses, we obtain the FGLS and MLE Wald statistics (6.1), the LR and LR^* criteria as defined in (5.3) and (5.4) and the QLR statistics (6.2). In the case of the Wald and QLR test, we obtain the asymptotic χ^2 and LMC p-values using 19 and 99 simulated samples. The exact BMC and MMC p-values are also obtained for the LR criteria. The bounding statistic $LR_c = n\ln(\Lambda_c^*)$, corresponds the UL hypothesis that sets all the coefficients of the MLR model (except the intercepts) to specific values. As emphasized in Section 5, the BMC procedure based on LR^* yields tighter bounds [see inequality (5.7)]. The results are summarized below.

It is evident that the symmetry hypothesis H_{01} is rejected using all asymptotic and exact tests. In the case of H_{02}, all tests against the unconstrained SURE specification are not significant. However, the asymptotic χ^2 and LMC tests LR^* are significant at the 5% level. Although the bounds p-value is $> .05$, the MMC test is significant at the 5% level, even with 19 simulated samples. It is worth noting that the QLR and the LR LMC tests yield equivalent decisions for both testing problems. Moreover, all MC tests based on 19 and 99 replications also yield similar decisions.

8 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. In view of this, we have combined the bounds and Monte Carlo test approaches to provide p-values for test statistics that are more accurate than those based on asymptotic approximations. The basic results were stated in terms of arbitrary hypotheses in MLR contexts. We have also focused on special cases, namely uniform and general linear hypotheses, and have extended the methodology to the SURE framework. We have reported the results of an extensive Monte Carlo experiment that covered uniform linear, cross-equation and non-linear restrictions in MLR and SURE models. The feasibility of the test strategy was also illustrated with an empirical application. 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 fail. In contrast, the various tests we have proposed displayed excellent size and power properties.

Table 11A: Cross-equation symmetry tests

$H_{01}: \begin{cases} d_{KL}=d_{LK},\ d_{KM}=d_{MK},\ d_{KE}=d_{EK}, \\ d_{LM}=d_{ML},\ d_{LE}=d_{EL},\ d_{EM}=d_{ME} \end{cases}$								
		LR^*	LR	$QLR_{(0)}$	$QLR_{(1)}$	$QLR_{(2)}$	$Wald_{GLS}$	$Wald_{MLE}$
Statistic		176.582	74.159	75.545	75.140	74.911	239.597	238.777
Asymptotic p-value		.000	.000	.000	.000	.000	.000	.000
Reps	MC p-value							
19	BMC	.05	.70	-	-	-	-	-
99		.01	.67	-	-	-	-	-
19	MMC	.05	05	-	-	-	-	-
99		.01	.01	-	-	-	-	-
19	LMC	.05	.05	.05	.05	.05	.05	.05
99		.01	.01	.01	.01	.01	.01	.01

Note: Under H_{01}, $LR^* \overset{asy}{\sim} \chi^2(42)$ while the other statistics have asymptotic $\chi^2(6)$ distributions. The LR^* statistic tests the symmetry restrictions (6 constraints) jointly with the SURE exclusion restrictions (36 constraints) _ a total of 42 restrictions _ against the unrestricted MLR model. *Reps* stands for *replications*.

Table 11B: Cross-equation symmetry tests (partial)

$H_{02}: \quad d_{EM}=d_{ME}, \quad d_{KM}=d_{MK}$								
		LR^*	LR	$QLR_{(0)}$	$QLR_{(1)}$	$QLR_{(2)}$	$Wald_{GLS}$	$Wald_{MLE}$
Statistic		102.574	.15179	.15180	.15179	.15179	.1283	.1279
Asymptotic p-value		.000	.927	.927	.927	.927	.937	.938
Reps	MC p-value							
19	BMC	.15	1.0	-	-	-	-	-
99		.17	1.0	-	-	-	-	-
19	MMC	.05	1.0	-	-	-	-	-
99		.05	1.0	-	-	-	-	-
19	LMC	.05	.90	90	90	90	90	90
99		.04	.94	.94	.94	.94	.94	.94

Note: Under H_{02}, $LR^* \overset{asy}{\sim} \chi^2(38)$ while the other statistics have asymptotic $\chi^2(2)$ distributions. LR^* tests a subset of symmetry restrictions (2 constraints) jointly with the SURE exclusion restrictions (36 constraints) _ 38 restrictions in all _ against the unrestricted MLR model.

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. (1991), *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley, Reading (MA).

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

Berndt, E. R. and Wood, D. O. (1975), 'Technology, prices and the derived demand for energy', *Review of Economics and Statistics* **53**, 259–268.

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. (1996*a*), The power of bootstrap tests, Technical report, Department of Economics, Queen's University, Kingston (Ontario).

Davidson, R. and MacKinnon, J. G. (1996*b*), The size distortion of bootstrap tests, Technical report, Department of Economics, Queen's University, Kingston (Ontario).

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 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.

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

Gibbons, M. R., Ross, S. A. and Shanken, J. (1989), 'A test of the efficiency of a given portfolio', *Econometrica* **57**, 1121–1152.

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.

Harvey, A. and Phillips, G. D. A. (1982), 'Testing for contemporaneous correlation of disturbances in systems of regression equations', *Bulletin of Economic Research* **34(2)**, 79–81.

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* 'Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability', 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. 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, Volume 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.

Phillips, P. C. B. (1985), 'The exact distribution of LIML: II', *International Economic Review* **26**, 21–36.

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*), 'Bootstrap tests for the generalized least squares regression models', *Economics Letters* **34**, 261–265.

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.

Srivastava, V. K. and Giles, D. E. (1987), *Seemingly Unrelated Regression Equations Models. Estimation and inference*, Marcel Dekker, New York.

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. (1971), *Principles of Econometrics*, John Wiley & Sons, New York.

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, Volume 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.

Zellner, A. (1962), 'An efficient method for estimating seemingly unrelated regressions and tests for aggregate bias', *Journal of the American Statistical Association* **57**, 348–368.