Alternative procedures to discriminate non nested multivariate linear regression models

Araujo, MI; FERNANDES, M; Pereira, BB

Informa Group plc

DISCRIMINATING NONNESTED LINEAR REGRESSIONS

Maria Ivanilde Araújo

Department of Statistics, Universidade Federal do Amazonas

Marcelo Fernandes

Economics Department, Queen Mary, University of London

Basilio de Bragança Pereira

Faculdade de Medicina and COPPE, Universidade Federal do Rio de Janeiro

Abstract: This paper builds classical and Bayesian testing procedures for choosing between nonnested multivariate regression models. Although there are several classical tests for discriminating univariate regressions, only the Cox test is able to consistently handle the multivariate case. We then derive the limiting distribution of the Cox statistic in such a context, correcting an earlier derivation in the literature. Further, we show how to build alternative Bayes factors for the testing of nonnested multivariate linear regression models. In particular, we compute expressions for the posterior Bayes factor, the fractional Bayes factor, and the intrinsic Bayes factor.

Acknowledgments: The authors are grateful to Aziz Lazraq and Robert Cléroux for valuable comments. Fernandes and Pereira respectively thank research fellowships from CNPq and CAPES, Brazil. The usual disclaimer applies.

1 Introduction

Applied researchers constantly face the fundamental problem of choosing among alternative statistical models. The aim is to evince whether the models give significantly different fits to the data, while keeping each model on equal footing. The Neyman-Pearson theory of hypothesis testing only applies if the models belong to the same family of distributions. Special procedures are thus called for if models belong to nonnested families.

Furthermore, the use of Bayes factors are not without difficulties. First, the priors for one model must be coherent with the priors for the other models. In fact, if the parameter spaces have different dimensions and there is no simple relation linking the parameters of each model, performing a Bayesian analysis is a daunting task. Second, if prior information is weak so that one employs an improper prior, the usual Bayes factor is not properly defined. Alternative Bayes factors are thereof necessary.

The literature tackling the issue of nonnested families of hypothesis has a long pedigree, dating back to the seminal works of Cox (1961, 1962). For reviews and further references on nonnested hypothesis testing, see Gourieroux and Monfort (1994), McAleer (1995), Pereira (1977b, 1981, 1998), and Pesaran and Weeks (2001). Literature reviews focusing on the particular case of regression models appear in Pesaran (1974), MacKinnon (1983), McAleer and Pesaran (1986), and McAleer (1987). Further references in time series are Godfrey and Tremayne (1988), Burke, Godfrey and Tremayne (1990), Silvapulle and King (1993), and Chambers (1993), whereas Zellner (1984), Rossi (1985), De Jong (1993), Wiginton (1974), Allenby (1990), and Rust and Schmittlein (1985) develop Bayesian approaches. Other econometric applications are Harvey (1977), Aurikko (1985), and Pesaran and Pesaran (1995).

The contributions of this paper are twofold. First, we note that, even though the resulting expression is correct, there is a flaw in Pesaran and Deaton's (1978) derivation of the asymptotic variance of the Cox statistic for nonnested multivariate regressions. Second, we show how to build alternative Bayes factors for the testing of nonnested multivariate linear regression models. In particular, we compute expressions for Aitkin's (1991) posterior Bayes factor, O'Hagan's (1995) fractional Bayes factor, and Berger and Pericchi's (1996) intrinsic Bayes factor.

The plan of the paper is as follows. Section 2 reviews the classical and Bayesian solutions for the general problem of nonnested hypothesis testing. Section 3 introduces the multivariate regression problem as well as the classical test statistics. It also corrects, for linear regression models, the derivation of the asymptotic variance of the generalized Cox test advanced by Pesaran and Deaton (1978). Section 4 describes alternative Bayes factors for the multivariate regression problem, while Section 5 offers some concluding remarks.

2 Classical and Bayesian analyses

2.1 Cox procedure and alternatives

Let $y = (y_1, \dots, y_n)$ denote a random sample from some unknown distribution. The null hypothesis H_0 and alternative hypothesis H_1 respectively specify the parametric densities $f_0(y|\alpha_0)$ and $f_1(y|\alpha_1)$ for the random vector y, where α_0 and α_1 are unknown parameter vectors. Assume further that the density families are nonnested in the sense that arbitrary members of one family cannot be obtained as a limit of members of the other.

The asymptotic tests developed by Cox (1961, 1962) relies on a modification of the Neyman-Pearson likelihood ratio principle. The test statistic for H_0 against H_1 is

$$T_{01} = \Delta\ell_{01}(\hat{\alpha}_0, \hat{\alpha}_1) - n \left[\text{plim}_{n\to\infty} \frac{\Delta\ell_{01}(\hat{\alpha}_0, \hat{\alpha}_1)}{n} \right]_{\alpha_0 = \hat{\alpha}_0}, \tag{1}$$

where the probability limit is taken under H_0, $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are respectively the maximum likelihood estimators of α_0 and α_1, and $\Delta\ell_{01}(\alpha_0, \alpha_1)$ is the log-likelihood ratio $\ell_0(\alpha_0) - \ell_1(\alpha_1)$.

Cox shows that, asymptotically, T_{01} has a negative mean under the alternative hypothesis, whereas, under the null hypothesis, it is normally distributed with mean zero and variance

$$V_0(T_{01}) = V_0[\Delta\ell_{01}(\alpha_0, \alpha_{1|0})] - C_0' I_0^{-1} C_0, \tag{2}$$

where $\alpha_{1|0}$ is the plim of $\hat{\alpha}_1$ under H_0, $C_0 \equiv n \frac{\partial}{\partial \alpha_0} \left[\text{plim}_{n\to\infty} \frac{1}{n} \Delta\ell_{01}(\hat{\alpha}_0, \hat{\alpha}_1) \right]$, and I_0 is the information matrix of α_0 (see also White, 1982). If one wishes to test H_1 against H_0, analogous results hold for the statistic T_{10}. It then follows that $T^*_{01} = T_{01} V_0^{-1/2}(T_{01})$ and $T^*_{10} = T_{10} V_1^{-1/2}(T_{10})$ are asymptotically standard normal under H_0 and H_1, respectively.

Note that the possible outcomes of the Cox test include inconclusive results such as the rejection, as well as the nonrejection, of both hypotheses.

As an alternative, Cox (1961) suggests combining the two models in a general nesting specification of which they would both be special cases. Assuming that the density is proportional to the exponential mixture

$$[f_0(y|\alpha_0)]^\lambda \, [f_1(y|\alpha_1)]^{1-\lambda}, \tag{3}$$

it suffices to conduct inference about λ to test the relative merits of the two models.[1] Unfortunately, this sort of testing procedure, which is widely used in econometrics (see, for example, Fisher, 1983), cannot be extended to simultaneous equations as shown by Pesaran (1982).

Sawyer (1984) uses the distributional results of Cox (1961) to propose a statistic for the multiple testing of k alternative models. Let $f_i(y|\alpha_i)$, $i = 1, \ldots, k$ be the densities under scrutiny and denote by T_{ij} the $k-1$ Cox statistics for testing the null hypothesis H_i against each alternative hypothesis H_j $(j \neq i)$. Letting $T_i' = (T_{i,1}, \ldots, T_{i,i-1}, T_{i,i+1}, \ldots, T_{i,k})$, one may test H_i against all others H_j $(j \neq i)$ using $T_i' \Sigma^{-1} T_j$, which is asymptotically χ^2_{k-1} under H_i. Here Σ is the covariance matrix $C_i(T_{ij}, T_{i1})$ that is readily derived from Cox's results. Regularity conditions and properties are given in Pereira (1977a), White (1982), Loh (1985), Pesaran (1987), Pace and Salvan (1990), Rukhin (1993), and Zabel (1993).

2.2 Bayesian analysis

Another general approach suggested by Cox (1961) rests on Bayesian inference. The posterior odds for H_0 versus H_1 is

$$\frac{\pi_0}{\pi_1}\frac{q_0(y)}{q_1(y)} \equiv \frac{\pi_0}{\pi_1} B_{01}(y) \tag{4}$$

where π_j is the prior probability of H_j and

$$q_j(y) \equiv \int f_j(y|\alpha_j)\pi_j(\alpha_j)\,\mathrm{d}\alpha_j \tag{5}$$

[1] Atkinson (1970) develops a general theory for testing $H_0 : \lambda = 0$ against $H_1 : \lambda = 1$. To estimate α_1 under the null hypothesis H_0, Atkinson uses $\alpha_{1|0}$ rather than $\hat{\alpha}_1$. However, Atkinson's procedure does not always entail consistent tests (see Pereira, 1977a).

denotes the predictive distribution with prior probability $\pi_j(\alpha_j)$ for the parameters under H_j $(j = 0, 1)$. The Bayes factor $B_{01}(y)$ represents the weight of evidence in the data favoring H_0 over H_1. Cox also provides a general expression considering a loss function in the posterior odds, and describes the large-sample approximation for the distribution of the Bayes factor.

This approach has two main limitations. First, the prior knowledge expressed by π_0 and $\pi_0(\alpha_0)$ must be coherent with that of π_1 and $\pi_1(\alpha_1)$. If the parameter spaces have different dimensions, for instance, there is no simple relation between the parameters. Second, if the prior information is weak and one applies an improper prior, the usual Bayes factor is not well defined (see Aitkin, 1991; O'Hagan, 1995). To overcome these difficulties, some alternatives have been recently proposed (see Kass and Raftery, 1995).

Aitkin (1991) proposes the posterior Bayes factor that compares the posterior means of the likelihood function under H_0 and H_1. More formally, the posterior density under H_j $(j = 0, 1)$ is

$$\pi_j(\alpha_j|y) \equiv \frac{f_j(y|\alpha_j)\pi_j(\alpha_j)}{\int f_j(y|\alpha_j)\pi_j(\alpha_j)\,\mathrm{d}\alpha_j}, \tag{6}$$

which yields

$$q_j^P(y) = \int f_j(y|\alpha_j)\pi_j(\alpha_j|y)\,\mathrm{d}\alpha_j \tag{7}$$

as the posterior mean of the likelihood function under H_j $(j = 0, 1)$. The posterior Bayes factor then corresponds to the ratio of the posterior means, namely $B_{01}^P(y) = q_0^P(y)/q_1^P(y)$.

O'Hagan (1991) derives a modification to the posterior Bayes factor of Aitkin (1991) so as to avoid problems with improper priors. Consider the partition $y = (x, z)$ of the sample. From the subsample x, one obtains proper posterior densities $\pi_0(\alpha_0|x)$ and $\pi_1(\alpha_1|x)$ to use as the priors for the subsample z. The partial Bayes factor $B_{01}(z|x)$ then equals

$$B_{01}(z|x) \equiv \frac{\int f_0(z|x,\alpha_0)\pi_0(\alpha_0|x)\,\mathrm{d}\alpha_0}{\int f_1(z|x,\alpha_1)\pi_1(\alpha_1|x)\,\mathrm{d}\alpha_1} = \frac{q_0(y)/q_0(x)}{q_1(y)/q_1(x)} = \frac{B_{01}(y)}{B_{01}(x)}. \tag{8}$$

The idea is that improper priors affect $B_{01}(y)$ and $B_{01}(x)$ in the same fashion, and hence the effect on $B_{01}(z|x)$ is null.

Berger and Pericchi (1996) first define that the training sample x in the sample partition $y = (x, z)$ is minimal if the posteriors for α_0 and α_1 are proper and there is no subset of

x that entails a proper posterior. There are usually many, say R, partitions featuring a minimal training sample. Berger and Pericchi then derive the intrinsic Bayes factor $B_{01}^{I}(y)$ as the average of the partial Bayes factors $\{B_{01}(z_r|x_r);\ r = 1, \ldots, R\}$ obtained from the R minimal training samples.[2]

O'Hagan (1995) develop an alternative Bayes factor, which also relies on the use of a training sample. Let $b = n_x/n$ denote the training fraction, where n_x is the size of the training sample x. The fractional Bayes factor then is $B_{01}^{[b]}(y) = q_0^{[b]}(y)/q_1^{[b]}(y)$, with

$$q_j^{[b]}(y) = \frac{\int f_j(y|\alpha_j)\pi_j(\alpha_j)\,\mathrm{d}\alpha_j}{\int [f_j(y|\alpha_j)]^b \pi_j(\alpha_j)\,\mathrm{d}\alpha_j} \tag{9}$$

for $j = 0, 1$. The key to understand the fractional Bayes factor resides in the fact that, for n_x large enough, the likelihood for the full sample is approximately equal to the power b of the likelihood for the training sample. O'Hagan (1995) shows that the fractional Bayes factor is consistent provided that b shrinks to zero as n grows.

3 Test statistics for multivariate linear regressions

Most results and applications in the literature deals with the testing of univariate nonnested regressions with homoskedastic error terms. Exceptions are due to Pesaran and Deaton (1978) and Davidson and MacKinnon (1983), who deal with nonlinear systems of equations.

It is surprising that only Pesaran (1982) and Davidson and MacKinnon (1983) mention the identification problem that impedes the generalization of nonnested hypotheses testing procedures other than the Cox test to systems of equations without imposing unrealistic assumptions. It is even more surprising that there is no mention in the literature to the fact that, although Pesaran and Deaton's (1978) expression for the asymptotic variance of the Cox test statistic is correct, their derivation is defective. In this section, we describe the multivariate linear regression setting that we are interested in and then provide a correct derivation for the variance of the Cox test statistic.

[2] One could also use other descriptive statistic, e.g. the geometric mean and the median, to summarize the information given by the R partial Bayes factors. Moreover, if R is too large, one may randomly select a sample from the collection of possible training samples.

3.1 Notation and definition of the setup

We consider two nonnested multivariate linear regression models $H_0: \; Y = XB_0 + U_0$ and $H_1: \; Y = ZB_1 + U_1$, where Y is a $n \times m$ matrix of regressands, X and Z are respectively $n \times p$ and $n \times q$ matrices of regressors, and B_0 and B_1 are respectively $p \times m$ and $q \times m$ matrices of parameters. The error terms U_0 and U_1 have rows that are independent and identically distributed (iid) as normal random vectors with means zero and covariance matrices Σ_0 and Σ_1, respectively. We also assume that X and Z are of full rank (p and q, respectively), with $n \geq m + p$ and $n \geq m + q$. It thus follows that: $U_0 \sim \mathcal{N}(0, I_n \otimes \Sigma_0)$ and $U_1 \sim \mathcal{N}(0, I_n \otimes \Sigma_1)$, whereas $Y \sim \mathcal{N}(XB_0, I_n \otimes \Sigma_0)$ under H_0 and $Y \sim \mathcal{N}(ZB_1, I_n \otimes \Sigma_1)$ under H_1.

The matrices of regressors X and Z are fixed and nonnested in the sense that it is not possible to obtain the columns of X from the columns of Z, and vice versa. We further assume that the matrices $\Sigma_{X'X} \equiv \lim_{n\to\infty} \frac{1}{n} X'X$ and $\Sigma_{Z'Z} \equiv \lim_{n\to\infty} \frac{1}{n} Z'Z$ are nonsingular, and that $\Sigma_{X'Z} \equiv \lim_{n\to\infty} \frac{1}{n} X'Z$ is a nonzero matrix.

We respectively denote by $\ell_0(\alpha_0)$ and $\ell_1(\alpha_1)$ the log-likelihood functions under H_0 and H_1, where $\alpha_0 = (\mathrm{vec}\, B_0, \mathrm{vec}\, \Sigma_0)'$ and $\alpha_1 = (\mathrm{vec}\, B_1, \mathrm{vec}\, \Sigma_1)'$. Recall that the **vec** operator stacks the columns, whereas the **vech** operator stacks only the elements on and under the diagonal. Depending on the context, we may consider $\alpha_0^* = (\mathrm{vec}\, B_0, \mathrm{vech}\, \Sigma_0)'$ and $\alpha_1^* = (\mathrm{vec}\, B_1, \mathrm{vech}\, \Sigma_1)'$ so as to avoid singularities in the information matrix.

We also take benefit from the fact that $\mathrm{vec}\, \Sigma_0 = D_m \mathrm{vech}\, \Sigma_0$, where D_m is the full ranked $m^2 \times m(m+1)/2$ duplication matrix (see Magnus and Neudecker, 1999, pages 49-50). It then follows that

$$\alpha_0 = \begin{pmatrix} \mathrm{vec}\, B_0 \\ \mathrm{vec}\, \Sigma_0 \end{pmatrix} = \begin{pmatrix} I_{pm} & 0 \\ 0 & D_m \end{pmatrix} \begin{pmatrix} \mathrm{vec}\, B_0 \\ \mathrm{vech}\, \Sigma_0 \end{pmatrix} = G_m \alpha_0^*, \tag{10}$$

where G_m is a $(pm + m^2) \times \left(pm + \frac{m(m+1)}{2}\right)$ matrix. Because $D_m' D_m$ is invertible, it then holds that $(D_m' D_m)^{-1} D_m' \mathrm{vec}\, \Sigma_0 = \mathrm{vech}\, \Sigma_0$. This yields

$$\alpha_0^* = \begin{pmatrix} \mathrm{vec}\, B_0 \\ \mathrm{vech}\, \Sigma_0 \end{pmatrix} = \begin{pmatrix} I_{pm} & 0 \\ 0 & (D_m' D_m)^{-1} D_m' \end{pmatrix} \begin{pmatrix} \mathrm{vec}\, B_0 \\ \mathrm{vec}\, \Sigma_0 \end{pmatrix} = H_m \alpha_0,$$

where H_m is a $\left(pm + \frac{m(m+1)}{2}\right) \times (pm + m^2)$ matrix. The matrices G_m and H_m are important because they reveal the close link between the information matrices I_0 and I_0^* of α_0 and α_0^*, respectively, namely $I_0^* = G_m' I_0 G_m$ and $I_0 = H_m' I_0^* H_m$.

We next consider the test statistic T_{01} given in (1), whose limiting distribution is normal with mean zero and variance $V_0(T_{01})$ as in (2). In the context of the multivariate linear regression model, the log-likelihoods of the two nonnested models are

$$\begin{aligned}\ell_0(\alpha_0) &= -\frac{n}{2}\log|\Sigma_0| - \frac{mn}{2}\log(2\pi) - \frac{1}{2}\operatorname{tr}(Y - XB_0)\Sigma_0^{-1}(Y - XB_0)' \\ \ell_1(\alpha_1) &= -\frac{n}{2}\log|\Sigma_1| - \frac{mn}{2}\log(2\pi) - \frac{1}{2}\operatorname{tr}(Y - ZB_1)\Sigma_1^{-1}(Y - ZB_1)',\end{aligned}$$

and hence $\Delta\ell_{01}(\hat\alpha_0, \hat\alpha_1) = \frac{n}{2}\left(\log|\hat\Sigma_1| - \log|\hat\Sigma_0|\right)$, where $\hat\Sigma_0 = \frac{1}{n}\hat U_0'\hat U_0$ and $\hat\Sigma_1 = \frac{1}{n}\hat U_1'\hat U_1$. It is well known that, under the null H_0, $\hat U_0 = M_X U_0$ and $\hat U_1 = M_Z U_0 + M_Z X B_0$, where $M_W = I - W(W'W)^{-1}W'$, and that

$$\begin{aligned}\hat\Sigma_1 &= \frac{1}{n}\hat U_1'\hat U_1 = \frac{1}{n}(M_Z U_0 + M_Z X B_0)'(M_Z U_0 + M_Z X B_0) \\ &= \frac{1}{n}(U_0' M_Z U_0 + B_0' X' M_Z U_0 + U_0' M_Z X B_0 + B_0' X' M_Z X B_0).\end{aligned} \tag{11}$$

The latter has asymptotic expectation under H_0 equal to $\Sigma_{1|0} \equiv \Sigma_0 + B_0'\bar\Sigma B_0$, where $\bar\Sigma \equiv \Sigma_{X'X} - \Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'X}$.

As $\hat\Sigma_0$ converges to Σ_0 in probability under H_0, it follows that

$$n\,\text{plim}_{n\to\infty}\frac{\Delta\ell_{01}(\hat\alpha_0, \hat\alpha_1)}{n} = \frac{n}{2}\left(\log|\Sigma_0 + B_0'\bar\Sigma B_0| - \log|\Sigma_0|\right), \tag{12}$$

and hence

$$\begin{aligned}T_{01} &= \Delta\ell_{01}(\hat\alpha_0, \hat\alpha_1) - n\left[\text{plim}_{n\to\infty}\frac{\Delta\ell_{01}(\hat\alpha_0, \hat\alpha_1)}{n}\right]_{\alpha_0 = \hat\alpha_0} \\ &= \frac{n}{2}\left(\log|\hat\Sigma_1| - \log\left|\hat\Sigma_0 + \frac{\hat B_0' X' M_Z X \hat B_0}{n}\right|\right),\end{aligned}$$

where $\frac{1}{n}X'M_Z X$ consistently estimates $\bar\Sigma$.

Remark: In finite samples, the distribution of T_0 under H_0 depends on the unknown parameters. In fact, it is well known that $\hat B_0 = B_0 + (X'X)^{-1}X'U_0$ and hence

$$\frac{|\hat\Sigma_1|}{|\hat\Sigma_{1|0}|} = \frac{|U_0'M_Z U_0 + U_0'M_Z X B_0 + B_0'X'M_Z U_0 + B_0'X'M_Z X B_0|}{\left|U_0'M_X U_0 + [B_0 + (X'X)^{-1}X'U_0]'\, X'M_Z X\, [B_0 + (X'X)^{-1}X'U_0]\right|}, \tag{13}$$

which boils down to a quite complex function of B_0. If the models are nested, and thus $M_Z X = 0$, all terms depending on B_0 disappear in (13).

3.2 The asymptotic variance of T_{01}

We start with the derivation of the first term of (2). It follows from

$$\begin{aligned}\Delta\ell_{01}(\alpha_0,\alpha_{1|0}) &= \ell_0(\alpha_0)-\ell_1(\alpha_{1|0})\\ &= \frac{1}{2}\operatorname{tr}\left[\left(XB_0-ZB_{1|0}+U_0\right)\Sigma_{1|0}^{-1}\left(XB_0-ZB_{1|0}+U_0\right)'\right]-\frac{1}{2}\operatorname{tr}\left[U_0\Sigma_0^{-1}U_0'\right],\end{aligned}$$

where $B_{1|0}=\Sigma_{Z'Z}^{-1}\Sigma_{Z'X}B_0$, that

$$\begin{aligned}V_0\Big(\Delta\ell_{01}(\alpha_0,\alpha_{1|0})\Big) &= V_0\left\{\frac{1}{2}\operatorname{tr}\Big[(XB_0-ZB_{1|0})\Sigma_{1|0}^{-1}(XB_0-ZB_{1|0})'\Big]\right.\\ &\qquad\left.+\frac{1}{2}\operatorname{tr}\Big[U_0\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)U_0'\Big]+\operatorname{tr}\Big[(XB_0-ZB_{1|0})\Sigma_{1|0}^{-1}U_0'\Big]\right\}\\ &= V_0\left\{\frac{1}{2}\operatorname{tr}\Big[U_0\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)U_0'\Big]+\operatorname{tr}\Big[(XB_0-ZB_{1|0})\Sigma_{1|0}^{-1}U_0'\Big]\right\},\end{aligned}$$

given the nonstochastic nature of the first term. To solve this variance, we use the following set of lemmata.

Lemma 1: $\operatorname{tr}\left[U_0\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)U_0'\right]=\sum_{i=1}^n U_{0(i)}\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)U_{0(i)}'$, where $U_{0(i)}$ denotes the i-th row of U_0.

Proof: From Lemma 2.2.3 of Muirhead (1982, p. 76), it follows that

$$\operatorname{tr}U_0\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)U_0'=\left(\operatorname{vec}U_0'\right)'\left[I\otimes\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)\right]\operatorname{vec}U_0'. \tag{14}$$

From $U_0'=\left(U_{0(1)}',\ldots,U_{0(n)}'\right)$ and $(\operatorname{vec}U_0')'=\left(U_{0(1)},\ldots,U_{0(n)}\right)$, it then ensures that

$$\operatorname{tr}\Big[U_0\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)U_0'\Big]=\left(U_{0(1)}\cdots U_{0(n)}\right)\left[I_n\otimes\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)\right]\begin{pmatrix}U_{0(1)}'\\ \vdots\\ U_{0(n)}'\end{pmatrix},$$

which equals $\sum_{i=1}^n U_{0(i)}\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)U_{0(i)}'$. ■

Lemma 2: $\operatorname{tr}(XB_0-ZB_{1|0})\Sigma_{1|0}^{-1}U_0'=\sum_{i=1}^n K_{(i)}U_{0(i)}'$, where $K_{(i)}$ is the i-th row of the $p\times m$ matrix $K\equiv(XB_0-ZB_{1|0})\Sigma_{1|0}^{-1}$.

The latter result is trivial, hence we omit the proof. Lemmata 1 and 2 then yield

$$V_0\Big(\Delta\ell_{01}(\alpha_0,\alpha_{1|0})\Big)=V_0\left\{\frac{1}{2}\sum_{i=1}^n U_{0(i)}\left(\Sigma_{1|0}^{-1}-\Sigma_0^{-1}\right)U_{0(i)}'+\sum_{i=1}^n K_{(i)}U_{0(i)}'\right\}.$$

It is also interesting to observe that, by the normality assumption, the row vectors $U_{0(i)}$ are iid $\mathcal{N}(0, \Sigma_0)$. To complete the derivation, we use two well-known results that we collect in the next lemma.

Lemma 3: If $X \sim \mathcal{N}(\mu, \Sigma)$ and A is symmetric, then $X'AX$ has covariance matrix $2\,\mathrm{tr}(A\Sigma)^2 + 4\,\mu' A\Sigma A\mu$. Further, if a is a vector, then the linear combination $a'X$ has variance $a'\Sigma a$.

Applying Lemma 3 then gives way to

$$V_0\Big(\Delta\ell_{01}(\alpha_0, \alpha_{1|0})\Big) = \mathrm{tr}\Big[(XB_0 - ZB_{1|0})\,\Sigma_{1|0}^{-1}\,\Sigma_0\,\Sigma_{1|0}^{-1}\,(XB_0 - ZB_{1|0})'\Big] + \frac{n}{2}\,\mathrm{tr}\Big[\,\Big(\Sigma_{1|0}^{-1} - \Sigma_0^{-1}\Big)\,\Sigma_0\Big]^2. \tag{15}$$

We now turn our attention to the next quantity in (2), namely

$$\begin{aligned} C_0 &\equiv n\,\frac{\partial}{\partial\alpha_0}\Big[\mathrm{plim}_{n\to\infty}\frac{1}{n}\,\Delta\ell_{01}(\hat\alpha_0, \hat\alpha_1)\Big] \\ &= \frac{n}{2}\left[\frac{\partial}{\partial\alpha_0}\log|\Sigma_0 + B_0'\bar\Sigma B_0| - \frac{\partial}{\partial\alpha_0}\log|\Sigma_0|\right]. \end{aligned}$$

To avoid singularities in the information matrix, we will work with α_0^* rather than α_0. We apply another well-known result from matrix calculus, which we state in the next lemma.

Lemma 4: If A is symmetric, then $\frac{\partial}{\partial A}\log|A| = 2A^{-1} - \mathrm{diag}\,A^{-1}$.

It then follows from Lemma 4 that $\frac{\partial}{\partial\Sigma_0}\log|\Sigma_0| = 2\Sigma_0^{-1} - \mathrm{diag}\,\Sigma_0^{-1}$ and $\frac{\partial}{\partial\mathrm{vech}\,\Sigma_0}\log|\Sigma_0| = \mathrm{vech}\,\big(2\Sigma_0^{-1} - \mathrm{diag}\Sigma_0^{-1}\big)$. Lemma 4 also implies that $\frac{\partial}{\partial B_0}\log|\Sigma_0| = \frac{\partial}{\partial\mathrm{vec}\,B_0}\log|\Sigma_0| = 0$, whereas

$$\frac{\partial\log|\Sigma_0 + B_0'\bar\Sigma B_0|}{\partial\Sigma_0} = 2\left(\Sigma_0 + B_0'\bar\Sigma B_0\right)^{-1} - \mathrm{diag}\left(\Sigma_0 + B_0'\bar\Sigma B_0\right)^{-1} = 2\Sigma_{1|0}^{-1} - \mathrm{diag}\,\Sigma_{1|0}^{-1} \tag{16}$$

and

$$\begin{aligned} \frac{\partial\log|\Sigma_0 + B_0'\bar\Sigma B_0|}{\partial\mathrm{vech}\,\Sigma_0} &= \mathrm{vech}\left[2\left(\Sigma_0 + B_0'\bar\Sigma B_0\right)^{-1} - \mathrm{diag}\left(\Sigma_0 + B_0'\bar\Sigma B_0\right)^{-1}\right] \\ &= \mathrm{vech}\left[2\Sigma_{1|0}^{-1} - \mathrm{diag}\,\Sigma_{1|0}^{-1}\right]. \end{aligned} \tag{17}$$

Similarly,

$$\frac{\partial\log|\Sigma_0 + B_0'\bar\Sigma B_0|}{\partial B_0} = 2\bar\Sigma B_0\left(\Sigma_0 + B_0'\bar\Sigma B_0\right)^{-1} = 2\bar\Sigma B_0\Sigma_{1|0}^{-1} \tag{18}$$

and

$$\frac{\partial \log|\Sigma_0 + B_0'\bar{\Sigma}B_0|}{\partial \mathrm{vec}\, B_0} = \mathrm{vec}\, 2\bar{\Sigma}B_0\Sigma_{1|0}^{-1}. \tag{19}$$

We are now ready to compute C_0 using α_0^*:

$$\begin{aligned} C_0 &= \frac{n}{2}\left[\frac{\partial}{\partial\alpha_0^*}\log|\Sigma_0 + B_0'\bar{\Sigma}B_0| - \frac{\partial}{\partial\alpha_0^*}\log|\Sigma_0|\right] \\ &= \frac{n}{2}\begin{pmatrix} 2\,\mathrm{vec}\,\bar{\Sigma}B_0\Sigma_{1|0}^{-1} \\ 2\,\mathrm{vech}\left(\Sigma_{1|0}^{-1} - \Sigma_0^{-1}\right) - \mathrm{vech}\,\mathrm{diag}\left(\Sigma_{1|0}^{-1} - \Sigma_0^{-1}\right) \end{pmatrix}, \end{aligned}$$

which reduces to Pesaran's (1974) result if $m = 1$.

The last step is to compute the information matrix of α_0^*, though we start deriving the information matrix I_0 of $\alpha_0 = (B_0, \Sigma_0)$. Letting $\Lambda_0 \equiv \Sigma_0^{-1}$ yields

$$\ell_0(\alpha_0) = -\frac{mn}{2}\log(2\pi) + \frac{n}{2}\log|\Lambda_0| - \frac{1}{2}\mathrm{tr}\,(Y - XB_0)'(Y - XB_0)\,\Lambda_0, \tag{20}$$

whereas $\frac{\partial}{\partial B_0}\ell_0(\alpha_0) = X'(Y - XB_0)\Sigma_0^{-1}$ and $\frac{\partial^2}{\partial B_0 \partial B_0'}\ell_0(\alpha_0) = -X'X \otimes \Sigma_0^{-1}$. For n large enough, the latter also equals $-n\Sigma_{X'X} \otimes \Sigma_0^{-1}$. Moreover, the Hessian is block diagonal since $\frac{\partial^2}{\partial B_0 \partial \Lambda_0}\ell_0(\alpha_0) \to 0$. It is also possible to demonstrate that

$$\frac{\partial}{\partial\Lambda_0}\ell_0(\alpha_0) = \frac{n}{2}(2\Sigma_0 - \mathrm{diag}\,\Sigma_0) - \frac{1}{2}(2U_0'U_0 - \mathrm{diag}\,U_0'U_0), \tag{21}$$

and that

$$\frac{\partial^2}{\partial\Lambda_0\partial\Lambda_0}\ell_0(\alpha_0) = \frac{n}{2}\left[2\tilde{\Sigma}_0 - |\Sigma_0|\,\mathrm{diag}\,(A_1M_{01}, \ldots, A_mM_{0m})\right], \tag{22}$$

where $\tilde{\Sigma}_0(i,i) = -\Sigma_0 J_{ii}\Sigma_0$, $\tilde{\Sigma}_0(i,j) = -\Sigma_0(J_{ij} + J_{ji})\Sigma_0$ for $i \neq j$, and J_{ij} is the $m \times m$ matrix with elements that take value one in the position (i,j), zero otherwise. Further, $A_i = |\Lambda_{0ii}|$ is the cofactor (i,i) of Λ_0 with corresponding matrix Λ_{0ii} and

$$M_{0i} = \left[2\Lambda_{0ii}^{-1} - \mathrm{diag}\Lambda_{0ii}^{-1}\right]^+ - \left(2\Lambda_0^{-1} - \mathrm{diag}\Lambda_0^{-1}\right).$$

Finally, the operator $[\ \cdot\ ]^+$ forms an $m \times m$ matrix with 0's on the i-th row and i-th column and fills the remainder terms with the $(m-1)\times(m-1)$ matrix in the brackets. It is not surprising to observe that, for $m = 1$, the information matrices of B_0 and $\Lambda_0 = \Sigma_0^{-1}$ reduce to those derived by Pesaran (1974).

It rests to compute the information matrix of Σ_0. Because $\frac{\partial\Lambda_0}{\partial\Sigma_0} = \breve{\Sigma}$, where

$$\breve{\Sigma}(i,j) = \begin{cases} \Sigma_0^{-1}J_{ii}\Sigma_0^{-1} & \text{if } i = j \\ -\Sigma_0^{-1}(J_{ij} + J_{ji})\,\Sigma_0^{-1} & \text{if } i \neq j, \end{cases}$$

it follows that the information matrix of Σ_0 is

$$\frac{\partial^2}{\partial\Sigma_0\partial\Sigma_0}\ell_0(\alpha_0) = \frac{n}{2}\,\breve{\Sigma}\Big[2\tilde{\Sigma}_0 - |\Sigma_0|\,\mathrm{diag}\,(A_1 M_{011},\ldots,A_m M_{0mm})\Big]\breve{\Sigma}.$$

This matrix is singular, however. We therefore remove the singularities by dealing with the information matrix of $\mathrm{vech}\,\Sigma_0$, which is given by

$$\frac{n}{2}\,D_m'\breve{\Sigma}\Big[2\tilde{\Sigma}_0 - |\Sigma_0|\,\mathrm{diag}(A_1 M_{011},\ldots,A_m M_{0mm})\Big]\breve{\Sigma}D_m.$$

The information matrix I_0^* of α_0^* then reads

$$I_0^* = \begin{pmatrix} n\Sigma_{X'X}\otimes\Sigma_0^{-1} & 0 \\ 0 & \frac{n}{2}\,D_m'\breve{\Sigma}\Big[|\Sigma_0|\,\mathrm{diag}(A_i M_{0ii})_{i=1}^m - 2\tilde{\Sigma}_0\Big]\breve{\Sigma}D_m \end{pmatrix}, \tag{23}$$

and hence $n(I_0^*)^{-1} = \mathrm{diag}(\Delta_1,\Delta_2)$ with $\Delta_1 \equiv \Sigma_{X'X}^{-1}\otimes\Sigma_0$ and

$$\Delta_2 \equiv 2\left\{D_m'\breve{\Sigma}\Big[|\Sigma_0|\,\mathrm{diag}(A_1\,M_{011},\ldots,A_m\,M_{0mm}) - 2\tilde{\Sigma}_0\Big]\breve{\Sigma}D_m\right\}^{-1}.$$

We are now ready to derive an expression for the second term in (2):

$$\frac{1}{n}\,C_0'\,\mathrm{plim}_{n\to\infty}(nI_0^{-1})\,C_0 = \frac{n}{4}\,\Big[\mu_1'\Delta_1\mu_1 + \mu_2\Delta_2\mu_2\Big], \tag{24}$$

where $\mu_1 \equiv 2\,\mathrm{vec}\,\bar{\Sigma}B_0\Sigma_{1|0}^{-1}$ and $\mu_2 \equiv 2\,\mathrm{vech}\left(\Sigma_{1|0}^{-1} - \Sigma_0^{-1}\right) - \mathrm{vech}\,\mathrm{diag}\left(\Sigma_{1|0}^{-1} - \Sigma_0^{-1}\right)$. It then suffices to combine Equations (15) and (24) to obtain the asymptotic variance $V_0(T_{01})$ of the test statistic.

Remark: Instead of applying Lemma 4, Pesaran and Deaton (1978) use a simpler matrix calculus result that holds only for nonsymmetric matrices. The resulting expression for the asymptotic variance of the Cox statistic is nonetheless correct for the errors in the derivation of C_0 and I_0 cancel out.[3]

In practice, one must replace the parameters by consistent estimates to estimate the asymptotic variance of the Cox statistic. It is however straightforward to find consistent estimators for the key parameters by plugging their least-squares counterparts.

[3] We thank M. H. Pesaran for supplying a proof that rests on the properties of the duplication matrix.

4 Bayes factors for multivariate linear regressions

In this section, we extend the results of Aitkin (1991), O'Hagan (1995), and Berger and Pericchi (1996) to the context of multivariate linear regressions. From Section 2.2, the posterior odds for H_0 against H_1 is $(\pi_0/\pi_1)\, B_{01}$. Suppose that one uses improper priors for the parameters such that $\pi_0(\alpha_0)$ and $\pi_1(\alpha_1)$ are respectively proportional to constants K_0 and K_1. The Bayes factor B_{01} then is proportional to K_0/K_1 and is not well defined. For the multivariate regression models, Jeffreys diffuse prior is given by

$$\pi_0(\alpha_0) = \pi_0(B_0)\, \pi_0(\Sigma_0) = K_0\, |\Sigma_0|^{-\frac{m+1}{2}}, \tag{25}$$

giving way to the following predictive distribution under the null hypothesis

$$q_0(Y) = K_0 \int |\Sigma_0|^{-\frac{n+m+1}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr} S_0 \Sigma_0^{-1}\right] \times \left\{ \int \exp\left[-\frac{1}{2}\,\mathrm{tr}(B_0 - \hat{B}_0)' X'X (B_0 - \hat{B}_0)\Sigma_0^{-1}\right] \mathrm{d}B_0 \right\} \mathrm{d}\Sigma_0,$$

where $S_0 \equiv \hat{U}_0'\hat{U}_0 = \left(Y - X\hat{B}_0\right)' \left(Y - X\hat{B}_0\right)$ and $\hat{B}_0 = (X'X)^{-1}X'Y$. If one rewrites the integrand of the second integral as

$$\exp\left[-\frac{1}{2}\,\mathrm{vec}(B_0 - \hat{B}_0)'\, \Sigma_0^{-1} \otimes (X'X)\, \mathrm{vec}(B_0 - \hat{B}_0)\right],$$

it becomes apparent that it must integrate to $(2\pi)^{mp/2}|\Sigma_0|^{p/2}\left|(X'X)^{-1}\right|^{m/2}$ given that it resembles a normal density with mean $\mathrm{vec}\,\hat{B}_0$ and covariance matrix $\Sigma_0 \otimes (X'X)^{-1}$ and that $\left|\Sigma_0 \otimes (X'X)^{-1}\right|^{1/2} = |\Sigma_0|^{p/2}\left|(X'X)^{-1}\right|^{m/2}$.

The predictive distribution then becomes

$$q_0(Y) = (2\pi)^{mp/2}\, K_0\, |X'X|^{-m/2} \int |\Sigma_0|^{-\frac{n+m+1-p}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr} S_0 \Sigma_0^{-1}\right] \mathrm{d}\Sigma_0,$$

which has an integrand that is proportional to the density function of the inverted Wishart with parameters S_0 and $n - p$. This means that integrating with respect to Σ_0 results in

$$q_0(Y) = \pi^{\frac{m(2n-2p-m+1)}{4}}\, K_0\, |X'X|^{-m/2}\, |S_0|^{-\frac{n-p}{2}} \prod_{s=1}^{m} \Gamma\left(\frac{n-p-s+1}{2}\right). \tag{26}$$

A similar expression holds for the alternative model $Y = ZB_1 + U_1$. The resulting Bayes factor then is

$$B_{01}(Y) = \pi^{m(p-q)/2}\, \frac{K_0}{K_1} \left(\frac{|Z'Z|}{|X'X|}\right)^{m/2} \frac{|S_1|^{(n-q)/2}}{|S_0|^{(n-p)/2}} \prod_{s=1}^{m} \frac{\Gamma\left(\frac{n-p-s-1}{2}\right)}{\Gamma\left(\frac{n-q-s-1}{2}\right)}, \tag{27}$$

where $S_1 \equiv (Y - Z\hat{B}_1)'(Y - Z\hat{B}_1)$ and $\hat{B}_1 = (Z'Z)^{-1}Z'Y$. It is clear from (27) that the Bayes factor is not well defined for it depends on the unknown ratio K_0/K_1.

From (26) and (27), it is now possible to derive the alternative Bayes factor that we discuss in Section 2.2. For instance, the posterior Bayes factor $B_{01}^P(Y)$ of Aitkin (1991) results from the ratio between

$$q_0^P(Y) = (2\sqrt{\pi})^{-mn} |S_0|^{-n/2} \prod_{s=1}^m \frac{\Gamma\left(\frac{2n-p-s+1}{2}\right)}{\Gamma\left(\frac{n-p-s+1}{2}\right)}$$

and

$$q_1^P(Y) = (2\sqrt{\pi})^{-mn} |S_1|^{-n/2} \prod_{s=1}^m \frac{\Gamma\left(\frac{2n-q-s+1}{2}\right)}{\Gamma\left(\frac{n-q-s+1}{2}\right)}.$$

It therefore ensues that

$$B_{01}^P(Y) = \left(\frac{|S_1|}{|S_0|}\right)^{n/2} \prod_{s=1}^m \frac{\Gamma\left(\frac{2n-p-s+1}{2}\right)\Gamma\left(\frac{n-q-s+1}{2}\right)}{\Gamma\left(\frac{2n-q-s+1}{2}\right)\Gamma\left(\frac{n-p-s+1}{2}\right)}. \tag{28}$$

The arithmetic version of the intrinsic Bayes factor of Berger and Pericchi (1996) becomes

$$B_{01}^I(Y) = \frac{1}{R}\sum_{r=1}^R \frac{B_{01}(Y)}{B_{01}(Y_{(r)})} = B_{01}(Y)\,\frac{1}{R}\sum_{r=1}^R B_{10}(Y_{(r)})$$

where $Y_{(r)}$ is a minimal training sample with design matrices $X_{(r)}$ and $Z_{(r)}$ under H_0 and H_1, respectively. By definition, $Y_{(r)}$ is matrix such that both $X_{(r)}'X_{(r)}$ and $Z_{(r)}'Z_{(r)}$ are nonsingular. It has dimension $\bar{n} \times m$, where $\bar{n} = \lceil (m+1)/2 \rceil + \max(p,q)$ and $\lceil \cdot \rceil$ returns the smallest integer greater than its argument. From (27), it follows that

$$\begin{aligned} B_{01}^I(Y) = {} & \left(\frac{|Z'Z|}{|X'X|}\right)^{m/2} \frac{|S_1|^{\frac{n-q}{2}}}{|S_0|^{\frac{n-p}{2}}} \prod_{s=1}^m \frac{\Gamma\left(\frac{n-p-s+1}{2}\right)\Gamma\left(\frac{\bar{n}-q-s+1}{2}\right)}{\Gamma\left(\frac{n-q-s+1}{2}\right)\Gamma\left(\frac{\bar{n}-p-s+1}{2}\right)} \\ & \times \frac{1}{R}\sum_{r=1}^R \left(\frac{|X_{(r)}'X_{(r)}|}{|Z_{(r)}'Z_{(r)}|}\right)^{m/2} \frac{|S_{0(r)}|^{(\bar{n}-p)/2}}{|S_{1(r)}|^{(\bar{n}-q)/2}}, \end{aligned} \tag{29}$$

where $S_{j(r)}$ is analogous to S_j for the r-th minimal training set ($j = 0, 1$).

Finally, the fractional Bayes factor of O'Hagan (1995) results from the ratio between

$$q_0^{[b]}(Y) = \pi^{mn(1-b)/2} b^{mnb/2} |S_0|^{-n(1-b)/2} \prod_{s=1}^m \frac{\Gamma\left(\frac{n-p-s+1}{2}\right)}{\Gamma\left(\frac{nb-p-s+1}{2}\right)}$$

and

$$q_1^{[b]}(Y) = \pi^{mn(1-b)/2} b^{mnb/2} \left|S_1\right|^{-n(1-b)/2} \prod_{s=1}^{m} \frac{\Gamma\left(\frac{n-q-s+1}{2}\right)}{\Gamma\left(\frac{nb-q-s+1}{2}\right)}. \quad (30)$$

It then holds that

$$B_{01}^{[b])}(Y) = \left(\frac{|S_1|}{|S_0|}\right)^{n(1-b)/2} \prod_{s=1}^{m} \frac{\Gamma\left(\frac{n-p-s+1}{2}\right)\Gamma\left(\frac{nb-q-s+1}{2}\right)}{\Gamma\left(\frac{n-q-s+1}{2}\right)\Gamma\left(\frac{nb-p-s+1}{2}\right)}. \quad (31)$$

5 Conclusion

This paper aims at contributing to the literature of testing nonnested multivariate linear regression models. We first correct Pesaran and Deaton's (1978) derivation of the asymptotic variance of the Cox statistic and then show how to apply Aitkin's (1991) posterior Bayes factor, O'Hagan's (1995) fractional Bayes factor, and Berger and Pericchi's (1996) intrinsic Bayes factor in such a context. Topics for future research include the extension of these results to the context of nonnested multivariate nonlinear regression models.

Appendix

Proof of (16) and (18): Let $W = [w_{ij}] = \Sigma_0 + B_0'\bar{\Sigma}B_0$ with $w_{ii} = \sigma_{ii} + B_{0(i)}'\bar{\Sigma}B_{0(i)}$ and $w_{ij} = \sigma_{ij} + B_{0(i)}'\bar{\Sigma}B_{0(j)}$, where $\Sigma_0 = [\sigma_{(ij)}]$ and $B_{0(i)}$ is the i-th column of B_0. For any parameter θ, it holds that

$$\frac{\partial|W|}{\partial\theta} = \sum_i\sum_j \frac{\partial|W|}{\partial w_{ij}}\frac{\partial w_{ij}}{\partial\theta} = \sum_i \frac{\partial|W|}{\partial w_{ii}}\frac{\partial w_{ii}}{\partial\theta} + \sum_{i\neq j}\frac{\partial|W|}{\partial w_{ij}}\frac{\partial w_{ij}}{\partial\theta}.$$

However, $\frac{\partial w_{ii}}{\partial\sigma_{ii}} = 1$ for every i, $\frac{\partial w_{ii}}{\partial\sigma_{ij}} = 0$ for $i \neq j$, $\frac{\partial w_{ij}}{\partial\sigma_{ij}} = 1$ for all (i,j), and $\frac{\partial w_{ij}}{\partial\sigma_{i'j'}} = 0$ for all $(i,j) \neq (i',j')$. Thus, for every i, $\frac{\partial|W|}{\partial\sigma_{ii}} = \frac{\partial|W|}{\partial w_{ii}} = W_{ii}$, where W_{ii} is the cofactor associated to w_{ii}. Similarly, for every (i,j) such that $i \neq j$, it follows that $\frac{\partial|W|}{\partial\sigma_{ij}} = 2W_{ij}$, where W_{ij} is the cofactor associated to w_{ij}. This completes the proof of (16). Analogous manipulations show that $\frac{\partial w_{ii}}{\partial B_{0(i)}} = 2\bar{\Sigma}B_{0(i)}$ for all i, $\frac{\partial w_{ii}}{\partial B_{0(j)}} = 0$ and $\frac{\partial w_{ij}}{\partial B_{0(i)}} = \bar{\Sigma}B_{0(j)}$ for $i \neq j$, and $\frac{\partial w_{ij}}{\partial B_{0(k)}} = 0$ for all $i \neq k$ and $j \neq k$. It then ensues that, for every i, $\frac{\partial|W|}{\partial B_{0(i)}} = 2\sum_j W_{ij}\bar{\Sigma}B_{0(j)}$. Thus,

$$\frac{\partial|W|}{\partial B_0} = 2\bar{\Sigma}B_0W^{-1}|W| = 2\bar{\Sigma}B_0\left(\Sigma_0 + B_0'\bar{\Sigma}B_0\right)^{-1}|\Sigma_0 + B_0'\bar{\Sigma}B_0|.$$

Taking the log yields (18). ■

Proof of (21) and (22): The first result readily follows from the fact that $\frac{\partial \mathrm{tr} XY}{\partial X} = Y + Y' - \mathrm{diag}\, Y$ if X is symmetric. To derive the second result, observe that $\frac{\partial^2}{\partial\Lambda_0\partial\Lambda_0}\ell_0(\alpha_0) = \frac{n}{2}\left[2\frac{\partial}{\partial\Lambda_0}\Lambda_0^{-1} - \frac{\partial}{\partial\Lambda_0}\mathrm{diag}\,\Lambda_0^{-1}\right]$. It is a standard result in matrix calculus that $\frac{\partial}{\partial\Lambda_0}\Lambda_0^{-1} = \tilde{\Sigma}_0$. Next, denote by λ_{ij} the typical element of Λ_0. To obtain $\frac{\partial}{\partial\Lambda_0}\mathrm{diag}\,\Lambda_0^{-1}$, observe that

$$\mathrm{diag}\Lambda_0^{-1} = \frac{1}{|\Lambda_0|}\begin{pmatrix} A_1 & 0 & 0 & \cdots & 0 \\ 0 & A_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_m \end{pmatrix}$$

where A_i is the cofactor of λ_{ii}. However, A_i is the determinant of the symmetric matrix Λ_{0ii}. For instance, $A_1 = |\Lambda_{011}|$ and

$$\Lambda_0 = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & & & \\ \vdots & & \begin{pmatrix} & \Lambda_{011} & \end{pmatrix} & \\ \lambda_{m1} & & & \end{pmatrix},$$

which implies that

$$\frac{\partial A_1}{\partial \Lambda_0} = \begin{pmatrix} 0 & 0 \quad 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \left(|\Lambda_{011}| \left[2\Lambda_{011}^{-1} - \operatorname{diag} \Lambda_{011}^{-1} \right] \right) & \\ 0 & & & \end{pmatrix}.$$

In general $\frac{\partial A_i}{\partial \Lambda_0} = A_i \left[2\Lambda_{0ii}^{-1} - \operatorname{diag} \Lambda_{0ii}^{-1} \right]^+$ for all $i = 1, \dots, m$. Now,

$$\begin{aligned} \frac{\partial}{\partial \Lambda_0} \frac{A_i}{|\Lambda_0|} &= \frac{1}{|\Lambda_0|^2} \left(\frac{\partial A_i}{\partial \Lambda_0} |\Lambda_0| - A_i \frac{\partial |\Lambda_0|}{\partial \Lambda_0} \right) \\ &= \frac{A_i}{|\Lambda_0|} \left\{ \left[2\Lambda_{0ii}^{-1} - \operatorname{diag} \Lambda_{0ii}^{-1} \right]^+ - \left(2\Lambda_0^{-1} - \operatorname{diag} \Lambda_0^{-1} \right) \right\} \\ &= \frac{A_i \, M_{0ii}}{|\Lambda_0|}, \end{aligned}$$

so that

$$\frac{\partial^2}{\partial \Lambda_0 \partial \Lambda} \ell_0(\alpha_0) = \frac{n}{2} \left[2\tilde{\Sigma}_0 - \frac{1}{|\Lambda_0|} \operatorname{diag}(A_1 M_{011}, \dots, A_m M_{0mm}) \right],$$

completing the proof. ■

References

Aitkin, M. (1991), Posterior Bayes factors (with discussion), *Journal of the Royal Statistical Society B* **53**, 111–142.

Allenby, G. M. (1990), Hypothesis testing with scanner data: The advantage of Bayesian methods, *Journal of Marketing Research* **27**, 379–389.

Atkinson, A. C. (1970), A method for discriminating between models (with discussion), *Journal of the Royal Statistical Society B* **32**, 323–353.

Aurikko, E. (1985), Testing the functional form of Finnish aggregate imports, *Economic Letters* **18**, 223–228.

Berger, J. O. and Pericchi, L. R. (1996), The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association* **91**, 109–122.

Burke, S. P., Godfrey, L. G. and Tremayne, A. R. (1990), Testing AR(1) against MA(1) disturbances in the linear regression models: An alternative procedure, *Review of Economic Studies* **57**, 135–145.

Chambers, M. J. (1993), A nonnested approach to testing continous time models against discrete alternatives, *Journal of Econometrics* **57**, 319–343.

Cox, D. R. (1961), Tests of separate families of hypotheses, *Proceedings of the 4th Berkeley Symposium* **1**, 105–123.

Cox, D. R. (1962), Further results on tests of separate families of hypotheses, *Journal of the Royal Statistical Society B* **24**, 406–423.

Davidson, R. and MacKinnon, J. D. (1983), Testing the specification of multivariate models in the presence of alternative hypotheses, *Journal of Econometrics* **23**, 301–313.

De Jong, D. N. (1993), Bayesian inference in limited dependent variable models: An application to measuring strike duration, *Journal of Applied Econometrics* **30**, 345–361.

Fisher, G. R. (1983), Tests for two separate regressions, *Journal of Econometrics* **21**, 117–132.

Godfrey, L. G. and Tremayne, A. R. (1988), Checks for model adequacy for univariate time series models and their application to econometric relationship, *Econometric Reviews* **7**, 1–64.

Gourieroux, C. and Monfort, A. (1994), Testing non–nested hypotheses, *Handbook of Econometrics*, Vol. 4, Elsevier, pp. 2585–2637.

Harvey, A. C. (1977), Discrimination between CES and VES production functions, *Annals of Economic and Social Measurement* **4**, 463–471.

Kass, R. E. and Raftery, A. (1995), Bayes factors, *Journal of the American Statistical Association* **90**, 773–795.

Loh, W. Y. (1985), A new method for testing separate families of hypotheses, *Journal of the American Statistical Association* **80**, 361–368.

MacKinnon, J. G. (1983), Model specification tests against non–nested alternatives, *Econometric Reviews* **2**, 85–110.

Magnus, J. and Neudecker, H. (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, Chichester.

McAleer, M. (1987), Specification tests for separate models: A survey, *Specification Analysis in the Linear Model*, Routledge and Kehan Paul, London, pp. 146–196.

McAleer, M. (1995), The significance of testing empirical non–nested models, *Journal of Econometrics* **65**, 149–171.

McAleer, M. and Pesaran, M. H. (1986), Statistical inference in non–nested econometric models, *Applied Mathematics and Computation* **20**, 271–311.

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley, New York.

O'Hagan, A. (1991), Contribution to the discussion of posterior Bayes factor by Aitkin (1991), *Journal of the Royal Statistical Society B* **53**, 137.

O'Hagan, A. (1995), Fractional Bayes factors for model comparison, *Journal of the Royal Statistical Society B* **57**, 99–138.

Pace, L. and Salvan, A. (1990), Best conditional tests for separate families of hypotheses, *Journal of the Royal Statistical Society B* **52**, 125–134.

Pereira, B. d. B. (1977a), A note on the consistency and on finite sample comparisons of some Tests of separate families of hypotheses, *Biometrika* **64**, 109–113 (correction in volume 64, page 655).

Pereira, B. d. B. (1977b), Discriminating among separate models: A bibliography, *International Statistical Review* **45**, 163–172.

Pereira, B. d. B. (1981), Discriminating among separate models: An additional bibliography, *International Statistical Information* **62**, 3.

Pereira, B. d. B. (1998), Separate families of hypotheses, *in* P. Armitage and T. Colton (eds), *Encyclopedia of Biostatistics*, Vol. 5, John Wiley, pp. 4069–4074.

Pesaran, B. and Pesaran, M. H. (1995), A non–nested test of level–differenced versus log–differenced stationary models, *Econometric Reviews* **14**, 213–227.

Pesaran, M. H. (1974), On the general problem of model selection, *Review of Economic Studies* **41**, 153–171.

Pesaran, M. H. (1982), On the comprehensive method for testing non–nested regression models, *Journal of Econometrics* **18**, 263–274.

Pesaran, M. H. (1987), Global and partial non–nested hypotheses and asymptotic local power, *Econometric Theory* **3**, 69–97.

Pesaran, M. H. and Deaton, A. S. (1978), Testing non–nested regression models, *Econometrica* **46**, 677–694.

Pesaran, M. H. and Weeks, M. (2001), Non–nested hypothesis testing: An overview, *in* B. H. Baltagi (ed.), *Companion to Theoretical Econometrics*, Basil Blackwell, Oxford.

Rossi, P. (1985), Comparison of alternative functional forms in production, *Journal of Econometrics* **30**, 345–361.

Rukhin, A. L. (1993), Bahadur efficiency of tests of separate hypotheses and adaptive test statistics, *Journal of the American Statistical Association* **88**, 161–165.

Rust, R. T. and Schmittlein, D. S. (1985), A Bayesian cross–validated likelihood method for comparing alternative specifications of quantitative models, *Marketing Science* **4**, 20–40.

Sawyer, K. R. (1984), Multiple hypotheses testing, *Journal of the Royal Statistical Society B* **46**, 419–424.

Silvapulle, P. and King, M. L. (1993), Nonnested testing for autocorrelation in the linear regression model, *Journal of Econometrics* **58**, 295–314.

White, H. (1982), Regularity conditions for Cox's test of non–nested hypotheses, *Journal of Econometrics* **19**, 301–318.

Wiginton, J. C. (1974), A Bayesian approach to discrimination among economic models, *Decisions Sciences* **5**, 182–193.

Zabel, J. E. (1993), A comparison of nonnested tests for misspecified models using the method of approximate slopes, *Journal of Econometrics* **57**, 205–232.

Zellner, A. (1984), Posterior odds ratios for regression hypotheses: General considerations and some specific results, *in* A. Zellner (ed.), *Basic Issues in Econometrics*, Chicago University Press, pp. 275–305.