Journal of Econometrics 57 (1993) 377–392. North-Holland

A simulation approach to the problem of computing Cox's statistic for testing nonnested models*

M. Hashem Pesaran

Cambridge University, Cambridge, UK
University of California, Los Angeles, CA 90024, USA

Bahram Pesaran

Bank of England, London, UK

Received August 1990, final version received April 1992

This paper proposes a new procedure for computing the Cox statistic for tests of nonnested hypotheses using the method of stochastic simulation. The procedure is applicable to a wide class of probability distributions, is relatively simple to implement, and does not require an analytic derivation of the pseudo-true estimators that enter the Cox test statistic. The paper also contains an application of the proposed method to the test of probit versus logistic formulations of the univariate binary choice models. The empirical results show that the simulation method works reasonably well even for a moderate number of replications.

1. Introduction

The application of Cox's (1960, 1961) nonnested test in econometrics has been mainly confined to linear or simple nonlinear regression models [see, for example, Pesaran (1974), Pesaran and Deaton (1978), and Davidson and MacKinnon (1981)]. This has been mainly due to the complex and often intractable derivations that are usually involved in the computation of the numerator of the Cox statistic in nonregression situations.[1] The present paper

Correspondence to: M. Hashem Pesaran, Trinity College, Cambridge CB2 1TQ, United Kingdom.

* We would like to thank David Cox, Mike McAleer, Simon Potter, Keunkwan Ryu, Pravin Trivedi, and a referee for constructive comments and suggestions. The first author gratefully acknowledges partial financial support from the ESRC and the Isaac Newton Trust of Trinity College, Cambridge.

[1] The computational difficulties associated with application of the Cox test has led some authors, notably MacKinnon et al. (1983) and Davidson and MacKinnon (1984), to develop an alternative framework which is particularly convenient in the context of generalized regression models.

offers a solution to this problem by the method of stochastic simulation which is applicable generally, and does not require an analytic derivation of the pseudo-true estimators that enter the Cox statistic.

The stochastic simulation method, also known as the Monte Carlo method, has found important applications in econometrics and statistics. In the past the simulation method has been used primarily to obtain information on the small sample properties of asymptotically valid estimators and test statistics, or to calibrate the distribution of test statistics as in the 'bootstrap' method.[2] More recently, simulation methods have also been employed to solve difficult problems involving numerical integration which often arise in the computation of estimators or test statistics in econometrics. Examples of this include the use of Monte Carlo integration in Bayesian econometrics [cf. Geweke (1989), van Dijk (1987) and Zellner et al. (1988)] and the method of simulated moments proposed recently by Pakes and Pollard (1989) and applied to the multinomial probit model by McFadden (1989). Our use of stochastic simulation in this paper is similar to the latter set of applications and allows us to compute the numerator of the Cox statistic in the case of most probability distributions without having to carry out difficult numerical integrations.

The Monte Carlo method can also be used to obtain a reference distribution for the observed values of a given test statistic as in Williams (1970) and Atkinson (1986). Williams is the first to use such a Monte Carlo method for testing nonnested hypotheses. He considers the test of a smooth exponential model against a segmented or piecewise linear model. Atkinson (1986) applies the Monte Carlo method to the test of the exponential against the lognormal distribution. In this approach a reference distribution is simulated for a given test statistic, invariably the log-likelihood ratio statistic, using artificially generated observations under the null hypothesis. The observed value of the test statistic is then compared with the simulated reference distribution. This approach relies solely on simulations to calibrate the distribution of the log-likelihood ratio and may not have satisfactory asymptotic properties in the case of nonnested models where the log-likelihood ratio statistic is not centered even in large samples. The method proposed in this paper employs stochastic simulation only as a method of computing the Cox statistic and does not aim at making small sample adjustments to it. Although, in principle it should be possible to use the simulation method both for the computation of the Cox statistic and for a derivation of a better small sample approximation to its distribution under the null.

The plan of the paper is as follows. Section 2 sets out the general nonnested hypotheses testing problem and pinpoints the main difficulty involved in the computation of the Cox statistic. It shows that the problem lies with the

[2] See the excellent surveys by Hall (1988) and Hinkley (1988).

computation of the numerator of test statistic. Section 3 considers the computation of the pseudo-true estimators that enter the Cox statistic, and section 4 examines alternative procedures for the computation of the numerator of the Cox statistic. The proposed simulation method is applied in section 5 to the problem of testing probit versus logistic formulations of the univariate binary choice model. An empirical application to the problem of choice of health insurance in Australia is presented in section 6.

2. Cox's test statistic

Consider the two nonnested conditional probability distributions

$$\mathrm{H}_f:\ f(y, \boldsymbol{\theta} \mid \boldsymbol{x}), \qquad \mathrm{H}_g:\ g(y, \gamma \mid z),$$

and assume that there are N observations on y, namely Y_t, $t = 1, 2, \ldots, N$, and that the conditional distributions of Y_t's are i.i.d. (The case of dependent and heterogeneous observations can be dealt with in a similar fashion.) The vectors $\boldsymbol{x}$ and z represent the conditioning variables, and $\boldsymbol{\theta}$ and γ are the $p \times 1$ and $q \times 1$ vectors of unknown parameters under H_f and H_g, respectively. A version of the Cox statistics for the test of H_f against H_g, which is relatively simple to compute, is given by Cox (1961, 1962) as

$$T_f(\hat{\boldsymbol{\theta}}, \hat{\gamma}) = \bar{d} - \hat{\mathrm{E}}_f(\bar{d}), \tag{1}$$

where $\mathrm{E}_f(\cdot)$ stands for the conditional expectations operator under H_f,

$$\bar{d} = N^{-1} \sum_{t=1}^{N} d_t, \tag{2}$$

$$d_t = \log f(Y_t, \hat{\boldsymbol{\theta}} \mid \boldsymbol{x}_t) - \log g(Y_t, \hat{\gamma} \mid z_t), \tag{3}$$

$\hat{\boldsymbol{\theta}}$ and $\hat{\gamma}$ are the maximum likelihood (ML) estimators of $\boldsymbol{\theta}$ and γ under H_f and H_g, respectively, and $\hat{\mathrm{E}}_f(\bar{d})$ stands for a consistent estimator of $\bar{d}$ under H_f. In the case of nested hypothesis, we have $\mathrm{E}_f(\bar{d}) = 0$, and Cox's statistic reduces to the familiar log-likelihood ratio statistic. But in the general case where $\mathrm{E}_f(\bar{d})$ is nonzero, Cox's statistic centers the log-likelihood ratio statistic so that $T_f(\hat{\boldsymbol{\theta}}, \hat{\gamma})$ has mean zero in large samples [cf. McAleer and Pesaran (1986)]. The asymptotic variance of $\sqrt{n}\, T_f(\hat{\boldsymbol{\theta}}, \hat{\gamma})$ under H_f is given by[3]

$$v_f^2(\boldsymbol{\theta}_0, \gamma_*) = \mathrm{V}_f\{\log f(y, \boldsymbol{\theta}_0) - \log g(y, \gamma_*)\}$$
$$- \psi'(\boldsymbol{\theta}_0, \gamma_*) F(\boldsymbol{\theta}_0)^{-1} \psi(\boldsymbol{\theta}_0, \gamma_*), \tag{4}$$

[3] For notational simplicity we have dropped explicit reference to the conditional nature of the densities under H_f and H_g.

where $V_f(\cdot)$ stands for the conditional variance operator under $f(y, \boldsymbol{\theta}_0)$,

$$\psi(\boldsymbol{\theta}_0, \gamma_*) = \mathrm{E}_f[\{\partial \log f(y, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}\}\{\log f(y, \boldsymbol{\theta}_0) - \log g(y, \gamma_*)\}], \tag{5}$$

$$\begin{aligned} F(\boldsymbol{\theta}_0) &= -\mathrm{E}_f\{\partial^2 \log f(y, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'\} \\ &= \mathrm{E}_f[\{\partial \log f(y, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}_0\}\{\partial \log f(y, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}_0'\}], \end{aligned} \tag{6}$$

$\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ defined as the probability limit of $\hat{\boldsymbol{\theta}}$ under H_f, γ_* is the probability limit of $\hat{\gamma}$ under H_f, and is a function of $\boldsymbol{\theta}_0$.

In what follows we assume that the regularity conditions set out, for example, in White (1982) hold for the densities $f(y, \boldsymbol{\theta})$ and $g(y, \gamma)$ and that we have:

Proposition 1. Under H_f, $(\hat{\boldsymbol{\theta}}, \hat{\gamma})$ converges in probability to $(\boldsymbol{\theta}_0, \gamma_*)$, and under H_g, $(\hat{\boldsymbol{\theta}}, \hat{\gamma})$ converges in probability to $(\boldsymbol{\theta}_*, \gamma_0)$.

Proposition 2. Under H_f, $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ and $\sqrt{N}(\hat{\gamma} - \gamma_*)$ are distributed as asymptotic normal. Similarly, under H_g, $\sqrt{N}(\hat{\gamma} - \gamma_0)$ and $\sqrt{N}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_*)$ are distributed as asymptotic normal.

Proposition 3. Under H_f, the standardized Cox statistic,

$$N_f(\hat{\boldsymbol{\theta}}, \hat{\gamma}) = \sqrt{N}\, T_f(\hat{\boldsymbol{\theta}}, \hat{\gamma})/\hat{v}_f(\hat{\boldsymbol{\theta}}, \hat{\gamma}), \tag{7}$$

is asymptotically distributed as N(0, 1), *where $\hat{v}_f(\hat{\boldsymbol{\theta}}, \hat{\gamma})$ stands for a consistent estimator of $v_f(\boldsymbol{\theta}_0, \gamma_*)$ defined in (4).*

Proposition 4. $v_f^2(\boldsymbol{\theta}_0, \gamma_*)$ can be consistently estimated either by

$$\tilde{v}_f^2 = N^{-1} \sum_{t=1}^{N} (d_t - \bar{d})^2 - \psi'_N(\hat{\boldsymbol{\theta}}, \hat{\gamma}) A_N(\hat{\boldsymbol{\theta}})^{-1} \psi_N(\hat{\boldsymbol{\theta}}, \hat{\gamma}), \tag{8}$$

where

$$\psi_N(\boldsymbol{\theta}, \gamma) = N^{-1} \sum_{t=1}^{N} d_t\{\partial \log f(Y_t, \boldsymbol{\theta})/\partial \boldsymbol{\theta}\},$$

$$A_N(\boldsymbol{\theta}) = \left\{ -N^{-1} \sum_{t=1}^{N} \{\partial^2 \log f(Y_t, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'\} \right\},$$

or by

$$\hat{v}_f^2 = N^{-1} \sum_{t=1}^{N} (d_t - \bar{d})^2 - \psi'_N(\hat{\theta}, \hat{\gamma}) B_N(\hat{\theta})^{-1} \psi_N(\hat{\theta}, \hat{\gamma}), \tag{9}$$

where

$$B_N(\theta) = \left\{ N^{-1} \sum_{t=1}^{N} \{\partial \log f(Y_t, \theta)/\partial \theta\} \{\partial \log f(Y_t, \theta)/\partial \theta'\} \right\}.$$

The difference between (8) and (9) lies in the different consistent estimators of $F(\theta_0)$ used in (4). Formula (8) utilizes the 'outer-product' expression for the information matrix, while (9) uses the 'inner-product' expression. Under H_f both formulae yield consistent estimators of $F(\theta_0)$, but for computational purposes (9) is preferable. To see this, note that since $N^{-1}\sum_{t=1}^{N} \partial \log f(Y_t, \hat{\theta})/\partial \theta_i = 0$, $i = 1, 2, \ldots, p$, then (9) can also be written as

$$\hat{v}_f^2 = N^{-1} \boldsymbol{d}' \{\boldsymbol{I}_N - \boldsymbol{R}(\hat{\theta})[\boldsymbol{R}'(\hat{\theta})\boldsymbol{R}(\hat{\theta})]^{-1} \boldsymbol{R}'(\hat{\theta})\} \boldsymbol{d}, \tag{10}$$

where $\boldsymbol{d}' = (d_1, d_2, \ldots, d_N)$, and $\boldsymbol{R}(\theta)$ is the $N \times (p+1)$ matrix:

$$\boldsymbol{R}(\theta) = \begin{bmatrix} 1 & \partial \log f(Y_1, \theta)/\partial \theta_1 & \ldots & \partial \log f(Y_1, \theta)/\partial \theta_p \\ 1 & \partial \log f(Y_2, \theta)/\partial \theta_1 & \ldots & \partial \log f(Y_2, \theta)/\partial \theta_p \\ \vdots & \vdots & & \vdots \\ 1 & \partial \log f(Y_N, \theta)/\partial \theta_1 & \ldots & \partial \log f(Y_N, \theta)/\partial \theta_p \end{bmatrix}. \tag{11}$$

Therefore, $\hat{v}_f^2$ can be computed by running the regression of d_i (the estimated log-likelihood ratio for the ith observation) on a constant term and the derivatives $\partial \log f(Y_i, \theta)/\partial \theta_1$, $\partial \log f(Y_i, \theta)/\partial \theta_2, \ldots,$ $\partial \log f(Y_i, \theta)/\partial \theta_p$. The estimated standard error of this regression (not corrected for the loss of degrees of freedom) is equal to $\hat{v}_f$. This seems to be computationally the most simple method of obtaining a consistent estimate of $v_f(\theta_0, \gamma_*)$ and will be used in the applications to follow.

Turning now to the numerator of the standardized Cox statistic in (7), we again note that there are many ways of consistently estimating $E_f(\bar{d})$ in (2). The procedure commonly employed in the literature is to first approximate $E_f(\bar{d})$ by the Kullback–Leibler measure of 'closeness' of H_f with respect to H_g [see Pesaran (1987)]:

$$C(\theta_0, \gamma_*) = \int \log\{f(y, \theta_0)/g(y, \gamma_*)\} f(y, \theta_0)\, \mathrm{d}y, \tag{12}$$

and then estimate $C(\boldsymbol{\theta}_0, \gamma_*)$ under H_f by $C(\hat{\boldsymbol{\theta}}, \hat{\gamma}_*)$, where $\hat{\gamma}_*$ stands for a consistent estimate of γ_* under H_f. There are two major difficulties with this procedure which have prevented the application of the Cox test to a large number of interesting problems in econometrics. First, an analytical derivation of the integral in (12) is often not possible. Second, even if (12) can be derived analytically, the derivation of the pseudo-ML estimator of γ_* is not a straightforward matter.[4] The usual method of computing $\hat{\gamma}_*$ involves two stages. In the first stage $\gamma_*(\boldsymbol{\theta})$ is derived analytically, and in the second stage $\hat{\gamma}_*$ is computed by evaluating $\gamma_*(\boldsymbol{\theta}_0)$ at $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}$. The difficulty with this approach is that the derivation of γ_* is usually complicated and in many cases a closed analytic expression for γ_* may not be available. In view of these problems, a computational procedure that does not require the analytic derivations of the integral (12), and the estimators of the pseudo-true parameters γ_* is desirable. Here we propose such a procedure.

In the next section we discuss the computation of $\hat{\gamma}_*$, and in section 4 we consider the computation of the numerator of Cox's statistics.

3. Computation of $\gamma_*(\hat{\boldsymbol{\theta}})$ by simulation

In view of the fact that $\hat{\gamma}_*$ is a consistent estimator of γ_* under H_f, a simple but capital-intensive method of computing $\hat{\gamma}_*$ would be to compute the ML estimates of γ using observations on y that are 'artificially' simulated assuming $f(y, \hat{\boldsymbol{\theta}})$ is the data generation process. Let $\boldsymbol{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{Nj})'$ be the $N \times 1$ vector of independent observations generated artificially according to H_f with $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Using $\boldsymbol{Y}_j$, compute the ML estimate of γ under H_g and denote it by $\hat{\gamma}_j$. Then estimate γ_* by

$$\hat{\gamma}_*(R) = \frac{1}{R} \sum_{j=1}^{R} \hat{\gamma}_j. \tag{13}$$

Note that $\hat{\gamma}_j$ is a function of $\boldsymbol{Y}_j$ and hence $\hat{\boldsymbol{\theta}}$. It is now easily seen that since the $N \times 1$ observation vectors $\boldsymbol{Y}_j$ are independently simulated, then $\hat{\gamma}_j$, $j = 1, 2, \ldots, R$, are also independently and identically distributed, and by the Law of Large Numbers we have $\hat{\gamma}_*(R) \xrightarrow{p} \gamma_*(\hat{\boldsymbol{\theta}})$ as R, the number of replications, increases without bound.

It is useful to demonstrate the above procedure in the context of nonnested linear regression models where the exact analytic expression for γ_* is known.

[4] Notice also that the computation of the complete parametric encompassing tests, proposed, for example, by Dastoor (1983) and Mizon and Richard (1986), are also subject to similar computational difficulties.

Consider the nonnested linear regression models

$$\mathrm{H}_f\colon\ \boldsymbol{y} = X\boldsymbol{\alpha} + \boldsymbol{u}_f, \quad \boldsymbol{u}_f \sim \mathrm{N}(0, \sigma^2 \boldsymbol{I}_N), \quad 0 < \sigma^2 < \infty,$$

$$\mathrm{H}_g\colon\ \boldsymbol{y} = Z\boldsymbol{\beta} + \boldsymbol{u}_g, \quad \boldsymbol{u}_g \sim \mathrm{N}(0, \omega^2 \boldsymbol{I}_N), \quad 0 < \omega^2 < \infty,$$

analyzed extensively in the literature,[5] where $\boldsymbol{y}$ is an $N \times 1$ vector of observations on the dependent variable, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $N \times K_f$ and $N \times K_g$ matrices of observations on the regressors of the models H_f and H_g, respectively, and $\boldsymbol{u}_f$ and $\boldsymbol{u}_g$ are the $N \times 1$ vectors of the disturbances. In terms of our general notations the parameter vectors under H_f and H_g are, respectively, $\boldsymbol{\theta}' = (\boldsymbol{\alpha}', \sigma^2)$ and $\gamma' = (\boldsymbol{\beta}', \omega^2)$. Denote the OLS estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\alpha}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$ and those of σ^2 and ω^2 by $\hat{\sigma}^2 = N^{-1}\boldsymbol{y}'\boldsymbol{M}_x\boldsymbol{y}$ and $\hat{\omega}^2 = N^{-1}\boldsymbol{y}'\boldsymbol{M}_z\boldsymbol{y}$, where $\boldsymbol{M}_x = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{M}_z = \boldsymbol{I}_N - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. Then it is well-known that

$$\hat{\boldsymbol{\beta}}_* = \hat{\mathrm{E}}_f(\hat{\boldsymbol{\beta}}) = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\hat{\boldsymbol{\alpha}}, \tag{14}$$

$$\hat{\omega}_*^2 = \hat{\mathrm{E}}_f(\hat{\omega}^2) = N^{-1}(N - K_g)\hat{\sigma}^2 + N^{-1}\hat{\boldsymbol{\alpha}}'(\boldsymbol{X}'\boldsymbol{M}_z\boldsymbol{X})^{-1}\hat{\boldsymbol{\alpha}}. \tag{15}$$

The simulation approach to computing $\hat{\boldsymbol{\beta}}_*$ and $\hat{\omega}_*^2$ involves first generating independent sets of $N \times 1$ samples of y with mean vector $\boldsymbol{X}\hat{\boldsymbol{\alpha}}$ and the variance $\hat{\sigma}^2\boldsymbol{I}_N$. Let the jth such sample be $\boldsymbol{y}_j$. Then

$$\hat{\boldsymbol{\beta}}_j = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}_j, \qquad \hat{\omega}_j^2 = N^{-1}(\boldsymbol{y}_j'\boldsymbol{M}_z\boldsymbol{y}_j), \tag{16}$$

and the simulation estimators of $\boldsymbol{\beta}_*$ and ω_*^2 will be given by

$$\hat{\boldsymbol{\beta}}_*(R) = \frac{1}{R}\sum_{j=1}^{R}\hat{\boldsymbol{\beta}}_j = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\left\{\frac{1}{R}\sum_{j=1}^{R}\boldsymbol{y}_j\right\}, \tag{17}$$

$$\hat{\omega}_*^2(R) = \frac{1}{R}\sum_{j=1}^{R}\hat{\omega}_j^2 = \frac{1}{NR}\left(\sum_{j=1}^{R}\boldsymbol{y}_j'\boldsymbol{M}_z\boldsymbol{y}_j\right). \tag{18}$$

Since the observations $\boldsymbol{y}_j$ are generated as random samples with the same mean $(\boldsymbol{X}\hat{\boldsymbol{\alpha}})$ and the same variance $(\hat{\sigma}^2\boldsymbol{I}_N)$, it then follows from the Law of Large Numbers for the i.i.d. observations that, as $R \to \infty$,

$$\operatorname*{plim}_{R\to\infty}\left\{\frac{1}{R}\sum_{j=1}^{R}\boldsymbol{y}_j\right\} = \mathrm{E}_R(\boldsymbol{y}_j) = \boldsymbol{X}\hat{\boldsymbol{\alpha}}, \tag{19}$$

[5] See, for example, Pesaran (1974), Davidson and MacKinnon (1981), and Mizon and Richard (1986).

and, therefore, from (14) and (17) it follows that $\operatorname{plim}_{R\to\infty} [\hat{\boldsymbol{\beta}}_*(R)] = \hat{\boldsymbol{\beta}}_*$. Similarly, since $\boldsymbol{y}_j = \boldsymbol{X}\hat{\boldsymbol{a}} + \boldsymbol{u}_j$ and $\boldsymbol{u}_j \sim \mathrm{N}(\boldsymbol{0}, \hat{\sigma}^2 \boldsymbol{I}_N)$, then

$$\frac{1}{R}\sum_{j=1}^{R} \boldsymbol{y}_j'\boldsymbol{M}_z\boldsymbol{y}_j = \hat{\boldsymbol{a}}'\boldsymbol{X}'\boldsymbol{M}_z\boldsymbol{X}\hat{\boldsymbol{a}} + 2\hat{\boldsymbol{a}}'\boldsymbol{X}'\boldsymbol{M}_z\left(\frac{1}{R}\sum_{j=1}^{R} \boldsymbol{u}_j\right) + \frac{1}{R}\sum_{j=1}^{R} (\boldsymbol{u}_j'\boldsymbol{M}_z\boldsymbol{u}_j)$$

and

$$\operatorname*{plim}_{R\to\infty}\left\{\frac{1}{R}\sum_{j=1}^{R} \boldsymbol{y}_j'\boldsymbol{M}_z\boldsymbol{y}_j\right\} = \hat{\boldsymbol{a}}'\boldsymbol{X}'\boldsymbol{M}_z\boldsymbol{X}\hat{\boldsymbol{a}} + \hat{\sigma}^2(N - K_g).$$

Hence, using (18) it follows that as $R \to \infty$, $\hat{\omega}_*^2(R)$ approaches $\hat{\omega}_*^2$ given by (15). It is clear that in this simple example the simulation estimators (17) and (18) do in fact tend to the estimators (14) and (15) obtained analytically, as the number of replications is allowed to increase without bounds.[6]

4. Computation of the numerator of Cox's statistic by simulation

Let $L_f(Y, \hat{\theta}) = N^{-1}\sum_{t=1}^{N} \log f(Y_t, \hat{\theta})$ and $L_g(Y, \hat{\gamma}) = N^{-1}\sum_{t=1}^{N} \log g(Y_t, \hat{\gamma})$ be the average maximized log-likelihood functions under H_f and H_g, respectively. Then the numerator of the Cox's statistic can be consistently estimated by

$$T_f(\hat{\theta}, \hat{\gamma}) = L_f(Y, \hat{\theta}) - L_g(Y, \hat{\gamma}) - C(\hat{\theta}, \hat{\gamma}_*), \tag{20}$$

when $Y' = (Y_1, Y_2, \ldots, Y_N)$, and $C(\hat{\theta}, \hat{\gamma}_*)$ stands for a consistent estimator of the Kullback–Leibler measure of 'closeness' of H_f and H_g defined in (12). One possible simulation estimator of $C(\hat{\theta}, \hat{\gamma}_*)$ is given by

$$C_R(\hat{\theta}, \hat{\gamma}_*(R_*)) = \frac{1}{R}\sum_{j=1}^{R} \{L_f(Y_j, \hat{\theta}) - L_g(Y_j, \hat{\gamma}_*(R))\}, \tag{21}$$

where, as in section 2, Y_j is the $N \times 1$ artificially generated sample from the distribution of Y under H_f with $\theta_0 = \hat{\theta}$, and $\hat{\gamma}_*(R)$ stands for the simulation

[6] An alternative simulation estimator of ω^2 is given by

$$\tilde{\omega}_*^2(R) = \frac{1}{NR}\sum_{j=1}^{R} (\boldsymbol{y}_j - \boldsymbol{Z}\hat{\boldsymbol{\beta}}_*(R))'(\boldsymbol{y}_j - \boldsymbol{Z}\hat{\boldsymbol{\beta}}_*(R)).$$

It is easy to show that both simulation estimators, namely $\hat{\omega}_*^2(R)$ and $\tilde{\omega}_*^2(R)$, tend to the same limit as $R \to \infty$.

estimator of γ_* described in the previous section and is given by (13). Again the terms $L_f(Y_j, \hat{\theta}) - L_g(Y_j, \hat{\gamma}_*(R))$, $j = 1, 2, \ldots, R$, in (21) are independently and identically distributed, and hence by the Law of Large Numbers the simulation estimator $C_R(\hat{\theta}, \hat{\gamma}_*(R))$ converges to $C(\hat{\theta}, \hat{\gamma}_*)$, as the number of replications is allowed to increase without bounds.

The simulation procedure advocated here is the same as the method of Monte Carlo integration applied to evaluate the integral

$$C_R(\hat{\theta}, \hat{\gamma}_*(R))$$
$$= \int \{L_f(Y, \hat{\theta}) - L_g(Y, \hat{\gamma}_*(R))\} f(Y_1, \hat{\theta}) f(Y_2, \hat{\theta}) \cdots f(Y_N, \hat{\theta}) \mathrm{d}Y,$$

where $\hat{\theta}$ and $\hat{\gamma}_*(R)$ are treated as fixed parameters.[7] The Monte Carlo integration method has been applied in econometrics in the Bayesian analysis when the posterior distribution is not analytically tractable [see, for example, van Dijk (1987), Geweke (1989), and Zellner et al. (1988)]. The simulation estimator (21) follows closely the analytic procedure adopted in the literature for the estimation of $\mathrm{E}_f\{L_g(Y, \hat{\theta}) - L_g(Y, \hat{\gamma})\}$ and its computation involves two stages. In the first stage γ_* is estimated by $\hat{\gamma}_*(R)$, and in the second stage $\mathrm{E}_f[L_f(Y, \hat{\theta}) - L_g(Y, \hat{\gamma}_*(R))\}$ is estimated by the Monte Carlo integration method treating $\hat{\gamma}_*(R)$ as given.

The closeness measure $C(\theta, \gamma_*)$ can sometimes be derived analytically, even though an analytic solution for γ_* may not be available. This happens, for example, in the case of testing the log-normal against the exponential distribution or the logistic against the probit model. See example 2 in Pesaran (1987) and section 5 below, respectively. In such cases the closeness measure can be estimated consistently by $C(\hat{\theta}, \hat{\gamma}_*(R))$ and the second round of simulations required to compute $C_R(\hat{\theta}, \hat{\gamma}_*(R))$ will not be needed.

In the case of the nonnested regression models set out in section 3, using (21), we have

$$C_R(\hat{\theta}, \hat{\gamma}_*(R)) = -\tfrac{1}{2}\log\left[\frac{\hat{\sigma}^2}{\tilde{\omega}_*^2(R)}\right] - \tfrac{1}{2}\left[\frac{\tilde{\sigma}^2(R)}{\hat{\sigma}^2} - \frac{\tilde{\omega}_*^2(R)}{\tilde{\omega}_*^2(R)}\right],$$

where[8]

$$\tilde{\omega}_*^2(R) = \frac{1}{NR}\sum_{j=1}^{R}(y_j - Z\hat{\beta}_*(R))'(y_j - Z\hat{\beta}_*(R)),$$

[7] For a useful review of the Monte Carlo integration method, see Ripley (1987).

[8] Notice that $\tilde{\omega}_*^2(R)$ is the alternative simulation estimator of ω_*^2 discussed in footnote 7.

$$\tilde{\sigma}^2(R) = \frac{1}{NR} \sum_{j=1}^{R} (y_j - X\hat{a})'(y_j - X\hat{a}),$$

Now, noting that $y_j \sim \mathrm{N}(X\hat{a}, \hat{\sigma}^2 I_N)$, it is easily seen that

$$\operatorname*{plim}_{R \to \infty} C_R(\hat{\theta}, \hat{\gamma}_*(R)) = C(\hat{\theta}, \hat{\gamma}_*) = \tfrac{1}{2}\log(\hat{\omega}_*^2/\hat{\sigma}^2),$$

as required.

5. Testing probit versus logit models

In the analysis of binary choice data, the functional form chosen for the underlying probability model is often based on computational convenience and the alternative specifications are rarely made subject to formal statistical tests. In this section we show how the simulation procedure can be applied to compute Cox's statistic for the test of the probit against the logit model, and *vice versa*. This example is of interest in itself as it is generally believed that it is impossible to distinguish between these models empirically [see, for example, Cramer (1991, p. 17)].

Consider the following univariate binary choice models:

$$\mathrm{H}_f: \quad \mathrm{P}(y_t = 1 \mid x_t) = \Lambda(\theta' x_t) = \frac{\exp(\theta' x_t)}{1 + \exp(\theta' x_t)}, \tag{22}$$

$$\mathrm{H}_g: \quad \mathrm{P}(y_t = 1 \mid z_t) = \Phi(\gamma' z_t) = \int_{-\infty}^{\gamma' z_t} \phi(v)\,\mathrm{d}v, \tag{23}$$

where $\Lambda_t = \Lambda(\theta' x_t)$ and $\Phi_t = \Phi(\gamma' z_t)$ are the probability distribution functions of the standard logistic and normal distribution functions respectively, and $\phi(\cdot)$ is the density function of a standard normal variate. The relevant expressions for the average log-likelihood function under H_f and H_g are given by

$$\mathrm{H}_f: \quad L_f(Y, \theta \mid x) = N^{-1} \sum_{t=1}^{N} \{ Y_t \log \Lambda_t + (1 - Y_t)\log(1 - \Lambda_t) \}, \tag{24}$$

$$\mathrm{H}_g: \quad L_g(Y, \gamma \mid z) = N^{-1} \sum_{t=1}^{N} \{ Y_t \log \Phi_t + (1 - Y_t)\log(1 - \Phi_t) \}, \tag{25}$$

where $Y = (Y_1, Y_2, \ldots, Y_N)'$ is the $N \times 1$ vector of observations on y.

The variance of the Cox statistic for the test of H_f against H_g can be computed using (10), noting that

$$d_t = Y_t \log\left(\frac{\hat{\Lambda}_t}{\hat{\Phi}_t}\right) + (1 - Y_t)\log\left[\frac{1 - \hat{\Lambda}_t}{1 - \hat{\Phi}_t}\right], \qquad t = 1, 2, \ldots, N, \tag{26}$$

and

$$\boldsymbol{R}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} 1 & x_{11}e_{1f} & \cdots & x_{K_f 1}e_{1f} \\ 1 & x_{12}e_{2f} & \cdots & x_{K_f 2}e_{2f} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1N}e_{Nf} & \cdots & x_{K_f N}e_{Nf} \end{bmatrix}, \tag{27}$$

where $e_{tf} = Y_t - \hat{\Lambda}_t$.

In the present application where the 'closeness' measure $C(\boldsymbol{\theta}, \gamma_*)$ can be derived analytically, the numerator of the Cox's statistic can be computed in two ways:

$$T_f(R) = L_f(Y, \hat{\boldsymbol{\theta}}) - L_g(Y, \hat{\gamma}) - C_R(\hat{\boldsymbol{\theta}}, \hat{\gamma}_*(R)), \tag{28}$$

$$\tilde{T}_f(R) = L_f(Y, \hat{\boldsymbol{\theta}}) - L_g(Y, \hat{\gamma}) - C(\hat{\boldsymbol{\theta}}, \hat{\gamma}_*(R)), \tag{29}$$

where

$$L_f(Y, \hat{\boldsymbol{\theta}}) - L_g(Y, \hat{\gamma}) = N^{-1} \sum_{t=1}^{N} \left\{ Y_t \log\left(\frac{\hat{\Lambda}_t}{\hat{\Phi}_t}\right) + (1 - Y_t)\log\left(\frac{1 - \hat{\Lambda}_t}{1 - \hat{\Phi}_t}\right)\right\},$$

$$C_R(\hat{\boldsymbol{\theta}}, \hat{\gamma}_*(R)) = (NR)^{-1} \sum_{j=1}^{R} \sum_{t=1}^{N} \left\{ Y_{tj} \log\left(\frac{\hat{\Lambda}_t}{\Phi_t^*(R)}\right) + (1 - Y_{tj})\log\left(\frac{1 - \hat{\Lambda}_t}{1 - \Phi_t^*(R)}\right)\right\},$$

$$C(\hat{\boldsymbol{\theta}}, \hat{\gamma}_*(R)) = \frac{1}{N} \sum_{t=1}^{N} \left\{ \hat{\Lambda}_t \log\left(\frac{\hat{\Lambda}_t}{\Phi_t^*(R)}\right) + (1 - \hat{\Lambda}_t)\log\left(\frac{1 - \hat{\Lambda}_t}{1 - \Phi_t^*(R)}\right)\right\},$$

$\Phi_t^*(R) = \Phi(z_t' \hat{\gamma}_*(R))$, $\hat{\Lambda}_t = \Lambda(x_t' \hat{\boldsymbol{\theta}})$, and Y_{tj}, $j = 1, 2, \ldots, R$, $t = 1, 2, \ldots, N$, are independently simulated binary responses ($Y_{tj} = 1$ or $Y_{tj} = 0$) under H_f, with $\mathrm{P}(Y_{tj} = 1) = \hat{\Lambda}_t$.[9]

It is clear that for large enough R, the two statistics $T_f(R)$ and $\tilde{T}_f(R)$ will yield identical results. However, $\tilde{T}_f(R)$ which requires only the estimation of the pseudo-ML estimators, $\gamma_*(R)$, is preferable to $T_f(R)$. The standardized simulated Cox statistics, namely $S_f(R) = \sqrt{n}\,T_f(R)/\hat{v}_f$ and $\tilde{S}_f(R) = \sqrt{n}\,\tilde{T}_f(R)/\hat{v}_f$, can now be computed using the standard error of the regression of d_t on $\boldsymbol{R}(\boldsymbol{\theta})$ as an estimator of $\hat{v}_f$. In the case of the present application d_t and $\boldsymbol{R}(\boldsymbol{\theta})$ are defined in (26) and (27), respectively. Similar results can also be obtained for the test of H_g against H_f.

Using the above procedure, the Cox test can also be computed for the more interesting problem of testing the multinomial probit model against the multinomial logit model, and *vice versa*. Advances in Monte Carlo integration methods may make the maximum likelihood estimation of the multinomial probit model more feasible in the future, and such ML simulation techniques can be combined with the simulation method proposed in this paper to overcome the difficulties associated with computation of ML estimators of the parameter of the multinomial probit model and the computation of the numerator of the Cox statistic.[10] Notice that due to the linear form of the log-likelihood function of the multinomial probit or logit model in the discrete choice variable, the functional form of the closeness measure $C(\boldsymbol{\theta}, \gamma_*)$ can be derived analytically, and one only needs to use the simulation technique for the estimation of γ_*.

6. An empirical application: Insurance choice in Australia

In this section we report the results of estimating the probit and the logit models using data on the choice of health insurance in Australia. We then apply the simulation procedure developed in the previous section to test the probit and the logic models against one another. The data set used in the estimations cover 1331 single-person households at or below the age of 65 in the New South

[9] A simple method of simulating Y_{tj}'s is to set

$$Y_{tj} = 1 \quad \text{if} \quad \hat{\Lambda}_t > U,$$
$$= 0 \quad \text{otherwise},$$

where U is drawn from a uniform distribution on (0, 1). In the empirical application we generated U's by the Wichman and Hill (1982) algorithm which is particularly suitable for use on personal computers. As seed for initialization of the algorithm we used $IX = 7$, $IY = 23$ and $IZ = 73$.

[10] For an earlier attempt at estimation of the multinomial probit model by simulation, see Lerman and Manski (1981).

Wales area and contains the following variables:[11]

$$y_t = \begin{cases} 1 = \text{purchase of health insurance}, \\ 0 = \text{nonpurchase of health insurance}, \end{cases}$$

$$x_{t1} = 1,$$

$$x_{t2} = \text{sex dummy} \quad (1 \text{ if male}),$$

$$x_{t3} = \text{age} \quad (\text{years divided by } 100),$$

$$x_{t4} = \text{income} \quad (\text{in } \$1000),$$

$$x_{t5} = \text{income squared}.$$

The ML estimators of the parameters of the logit (H_f) and the probit (H_g) models and their asymptotic standard errors are given in the first columns of tables 1 and 2, respectively. Not surprisingly the two models give very similar results. The estimates are statistically significant and in particular reaffirm Cameron and Trivedi's (1991) conclusion that controlling for the size of the insurance premium, income is an important determinant of health insurance decision.[12]

As far as the choice between the probit and the logit is concerned, a crude comparison of the maximized values of the log-likelihood function favours the logit model. However, the difference between the two log-likelihood values is only 0.716, and it is not clear whether this difference is statistically significant. To provide a formal test of the logit and the probit models against one another, in tables 1 and 2 we also report the simulated standardized Cox statistics and the pseudo-ML estimators, $\hat{\theta}_*(R)$ and $\hat{\gamma}_*(R)$ computed using the following number of replications: $R = 50$, 100, 200, 500, 600, 800. The results are quite encouraging. The simulated estimates show only a small degree of variations as the number of replications is increased from 50 to 100 and settle down relatively quickly. In the present application, $R = 200$ seems to be adequate for a reasonably precise estimate of the Cox statistic. As R is allowed to increase beyond the 200 mark, only the integers in the second decimal place in the Cox statistic register a change. Furthermore, as pointed our earlier, the two versions of the simulated Cox statistics, $\{S_f(R), S_g(R)\}$ and $\{\tilde{S}_f(R), \tilde{S}_g(R)\}$, tend to the same

[11] The original source of this data is *Australian Health Insurance Survey*, 1983. The data used here is a subset of the data analyzed in Cameron and Trivedi (1991) and interested readers should consult this paper for more details. We are grateful to Pravin Trivedi for making this data set available to us.

[12] Since 1983 the insurance premium has been the same across all individuals resident in the same state. This fact explains the absence of a price variable in the logit or in the probit model.

Table 1
Maximum likelihood and pseudo-ML estimators of the parameters of the logit model (H_f).[a]

Independent variables	ML estimators	Pseudo-ML estimators $\hat{\theta}_*(R)$					
		Number of replications (R)					
	$\hat{\theta}$	50	100	200	500	600	800
Intercept	− 4.6499 (0.4279)	− 4.6066	− 4.5819	− 4.5865	− 4.6241	− 4.6183	− 4.6252
Sex	0.4397 (0.0971)	0.4679	0.4591	0.4584	0.4631	0.4604	0.4599
Age	− 0.7165 (0.3196)	− 0.7191	− 0.7271	− 0.7074	− 0.7118	− 0.7154	− 0.7203
Income	8.6234 (0.8045)	8.4538	8.4288	8.4397	8.5219	8.5145	8.5354
$(\text{Income})^2$	− 3.0416 (0.3628)	− 2.9504	− 2.9456	− 2.9517	− 2.9889	− 2.9843	− 2.9956
LL	− 617.382	− 619.051	− 620.852	− 620.874	− 619.600	− 619.509	− 619.427
		Simulated nonnested test statistics					
$S_f(R)$	—	1.5898	1.6618	1.6463	1.6842	1.6615	1.6371
$\tilde{S}_f(R)$	—	1.5279	1.5671	1.6022	1.5988	1.6025	1.6045

[a]The dependent variable is a dummy variable representing the purchase or nonpurchase of health insurance in the New South Wales, Australia. Figures in parentheses are standard erors. The logit coefficients have been multiplied by 0.625 to make them comparable to the probit coefficient [see Amemiya (1981)]. LL is the maximized value of the log-likelihood function under H_f. The LL values under the pseudo-ML estimators refer to $R^{-1}\sum_{j=1}^{R} NL_f(Y_j, \hat{\theta}_*(R))$, where Y_j are generated under H_g. $S_f(R)$ and $\tilde{S}_f(R)$ are the simulated standardized Cox statistics for test of the logit model against the probit model.

values as the number of replications is allowed to increase. It is also interesting to note that the statistics $\tilde{S}_f(R)$, $\tilde{S}_g(R)$ that are based on the analytic expressions for the closeness measure settle down more quickly and give an accurate estimate of the Cox statistic even after 100 replications. On the basis of these results, the Cox test reject (at the 5 percent level) the probit against the logit model, but not *vice versa*. However, the rejection probability of the probit model is not very high, and given the known tendency for the Cox test to overreject in small samples,[13] it seems reasonable to conclude that in the present application the logit model is preferable to the probit mode, but only marginally.

7. Concluding remarks

In this paper we have proposed a simple method of computing the Cox statistic which is applicable to a wide class of probability distributions. The

[13] See, for example, Godfrey and Pesaran (1983).

Table 2
Maximum likelihood and pseudo-ML estimators of the parameters of the probit model (H_g).[a]

Independent variables	ML estimators $\hat{\gamma}$	Pseudo-ML estimators $\hat{\gamma}_*(R)$ — Number of replications (R): 50	100	200	500	600	800
Intercept	− 4.4169 (0.3849)	− 4.4554	− 4.4342	− 4.4411	− 4.4737	− 4.4694	− 4.4752
Sex	0.4123 (0.0868)	0.4056	0.4010	0.3998	0.4052	0.4031	0.4021
Age	− 0.6558 (0.2944)	− 0.6445	− 0.6603	− 0.6472	− 0.6554	− 0.6594	− 0.6639
Income	8.1389 (0.7135)	8.1973	8.1785	8.1999	8.2733	8.2687	8.2874
$(\text{Income})^2$	− 2.8606 (0.3212)	− 2.8794	− 2.8754	− 2.8866	− 2.9190	− 2.9168	− 2.9261
LL	− 618.098	− 618.085	− 619.404	− 619.215	− 617.531	− 617.442	− 617.337
		Simulated nonnested test statistics					
$S_g(R)$	—	− 1.8923	− 1.9691	− 2.0591	− 2.1190	− 2.1053	− 2.0783
$\tilde{S}_g(R)$	—	− 2.1334	− 2.0782	− 2.0524	− 2.0542	− 2.0519	− 2.0514

[a] The dependent variable is a dummy variable representing the purchase or nonpurchase of health insurance in the New South Wales, Australia. Figures in parentheses are standard errors. LL is the maximized value of the log-likelihood function under H_f. The LL values under the pseudo-ML estimators refer to $R^{-1}\sum_{j=1}^{R} NL_g(Y_j, \hat{\gamma}_*(R))$, where Y_j are generated under H_f. $S_g(R)$ and $\tilde{S}_g(R)$ are the simulated standardized Cox statistics for test of the probit model against the logit model.

empirical example discussed in the previous section shows that quite reasonable results can be obtained only after a relatively small number of replications. The method has important potential for testing multinomial logit and probit models, and can be readily applied to nonnested limited dependent variable models, choice of distribution functions, and nonlinear and non-Gaussian time series models.[14]

References

Amemiya, T., 1981, Qualitative response models: A survey, Journal of Economic Literature 19, 1483–1536.

Atkinson, A.C., 1986, Monte Carlo tests of separate families of hypotheses, Unpublished manuscript (Imperial College, London).

[14] Bera and Higgins (1991) apply the simulation technique proposed here to test bilinear and ARCH models against one another. Pesaran and Pesaran (1992) apply the technique to testing nonnested generalized regression models with different transformations of the dependent variable such as linear versus log-linear models, or level-differenced models against the log-differenced models.

Bera, A. and L.M. Higgins, 1991, ARCH and bilinearity as competing models for non-linear dependence, Unpublished manuscript (University of Illinois, Urbana-Champaign, IL).
Cameron, A.C. and P.K. Trivedi, 1991, The role of income and health risk in the choice of health insurance: Evidence from Australia, Journal of Public Economics, forthcoming.
Cox, D.R., 1962, Further results on tests of separate families of hypotheses, Journal of the Royal Statistical Society B 24, 406–424.
Cox, D.R., 1961, Tests of separate families of hypotheses, Proceedings of the 4th Berkeley symposium, Vol. 1 (University of California Press, Berkeley, CA) 105–123.
Cramer, J.S., 1991, The logit model (Edward Arnold, London).
Dastoor, N.K., 1983, Some aspects of testing non-nested hypotheses, Journal of Econometrics 21, 213–228.
Davidson, R. and J.G. MacKinnon, 1981, Several tests for model specification in the presence of alternative hypotheses, Econometrica 49, 781–793.
Davidson, R. and J.G. MacKinnon, 1984, Model specification tests based on artificial linear regressions, International Economic Review 25, 485–502.
Geweke, J., 1989, Bayesian inference in econometric models using Monte Carlo integration, Econometrica 57, 1317–1339.
Godfrey, L.G. and M.H. Pesaran, 1983, Tests of non-nested regression models: Small sample adjustments and Monte Carlo evidence, Journal of Econometrics 21, 133–154.
Hall, P., 1988, Theoretical comparison of bootstrap confidence intervals, Annals of Statistics 16, 927–953.
Hinkley, D.V., 1988, Bootstrap methods, Journal of the Royal Statistical Society B 50, 321–337.
Lerman, S.R. and C.F. Manski, 1981, On the use of simulated frequencies to approximate choice probabilities, in: C.F. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT Press, Cambridge, MA) 305–319.
MacKinnon, J.G., H. White, and R. Davidson, 1983, Test for model specification in the presence of alternative hypotheses: Some further results, Journal of Econometrics 21, 53–70.
McAleer, M. and M.H. Pesaran, 1986, Statistical inference in non-nested econometric models, Applied Mathematics and Computation 20, 271–311.
McFadden, D., 1989, A method of simulated moments for estimation of discrete response models without numerical integration, Econometrica 57, 995–1026.
Mizon, G.E. and J.F. Richard, 1986, The encompassing principle and its application to non-nested hypotheses, Econometrica 54, 657–678.
Pakes, A. and D. Pollard, 1989, Simulation and the asymptotics of optimization estimators, Econometrica 57, 1027–1057.
Pesaran, M.H., 1974, On the general problem of model selection, Review of Economic Studies 41, 153–171.
Pesaran, M.H., 1987, Global and partial non-nested hypotheses and asymptotic local power, Econometric Theory 3, 69–97.
Pesaran, M.H. and A.S. Deaton 1978, Testing non-nested nonlinear regression models, Econometrica 46, 677–694.
Pesaran, B. and M.H. Pesaran, 1992, A non-nested test of first-difference versus log-difference stationary models, Unpublished manuscript (Cambridge University, Cambridge).
Ripley, B.D., 1987, Stochastic simulation (Wiley, New York, NY).
van Dijk, H.K., 1987, Some advances in Bayesian estimation methods using Monte Carlo integration, in: T.B. Fomby and G.F. Rhodes, eds., Advances in econometrics (JAI Press, Greenwich, CT).
White, H., 1982, Regularity conditions for Cox's test of non-nested hypotheses, Journal of Econometrics 19, 301–318.
Williams, D.A., 1970, Discrimination between regression models to determine the pattern of enzyme synthesis in synchronous cultures, Biometrics, March, 23–32.
Witchman, B.A. and I.D. Hill, 1982, An efficient and portable pseudo-random number generator (algorithm AS183), Applied Statistics 31, 188–190.
Zellner, A., L. Bauwens, and H.K. van Dijk, 1988, Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods, Journal of Econometrics 38, 39–72.