EUI WORKING PAPERS IN ECONOMICS

EUI Working Paper ECO No. 95/37

Simulation-Based Encompassing for Non-Nested Models: A Monte Carlo Study of Alternative Simulated Cox Test Statistics

CHIARA MONFARDINI

European University Institute, Florence

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
ECONOMICS DEPARTMENT

EUI Working Paper ECO No. 95/37

Simulation-Based Encompassing for Non-Nested Models: A Monte Carlo Study of Alternative Simulated Cox Test Statistics

CHIARA MONFARDINI

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in November 1995
European University Institute
Badia Fiesolana
I – 50016 San Domenico (FI)
Italy

Simulation-Based Encompassing for Non-Nested Models: a Monte Carlo Study of Alternative Simulated Cox Test Statistics

Chiara Monfardini*
European University Institute
E-mail: monfardi@datacomm.iue.it

September 1995

Abstract

We consider the possibility of using simulated methods in order to compute test statistics for non-nested models, in the general framework represented by the encompassing principle. Simulation techniques are useful whenever the evaluation of a test statistic involves a difficult integration problem, and represent a (simpler) alternative to numerical procedures. Focusing on the Cox test statistic, we extend a procedure proposed in the literature and calculate both its numerator and its denominator through simulation, getting a refinement of previous proposals which considered only simulation of the numerator. Some Monte Carlo experiments are then conducted taking as an example the test of linear *versus* loglinear models. Our results suggest that simulating both the quantities in the numerator and the denominator leads to a finite sample distribution of the test statistic closer to the asymptotic one.

Keywords: Non-nested hypotheses, Cox test statistic, simulation.

*I am indebted to my supervisor Grayham Mizon. I would also like to thank Stefan Dillmann, Massimiliano Marcellino and Stefano Nardelli for their suggestions on a previous version of this paper. Any remaining errors are mine.

1 Introduction

Simulation methods have found important applications in econometrics since the availability of fast and powerful computers has allowed the replication of drawings from a given experimental design. Traditionally, Monte Carlo experiments have been used to study the bias and the mean square error of an estimator or the size and power of a test statistic, investigating their finite sample properties by repeatedly drawing samples of simulated data[1]. The increasing power of computers has also been the basis for Bootstrap methods, which, starting from a model estimated on an observed sample of data, use the technique of resampling through simulations to derive the empirical distribution of the statistic of interest[2]. A different use of Monte Carlo techniques has been widespread in Bayesian econometrics to calculate the posterior moments of the parameters of an econometric model[3].

Recently, simulation methods have been introduced in the context of the classical inference process, to solve computational problems arising when the derivation of an estimator or a test statistic involves the calculation of integrals which are difficult to evaluate or which do not have a closed form solution. Such computational difficulties characterise, for example, models that exhibit non-linear functional form or contain unobservable variables. The major advantage offered by simulation techniques is the possibility of replacing numerical integration with approximation methods whose properties can be easily derived in terms of statistical theory. This is possible because the source of approximation is the generation of pseudo-random values from specified statistical distributions and the integrals to be computed are interpreted as expectations of certain random variables.

While in the last few years the study and application of simulation

[1] See Hendry (1984) for a discussion on Monte Carlo experiments in econometrics.

[2] An up-to-date survey on the literature on the Bootstrap can be found in Bertail (1992).

[3] Van Dijk (1987) describes Monte Carlo procedures in Bayesian estimation.

methods for estimation have received increasing attention[4], there still seems to be a lot to be investigated with regard to the role of simulation methods in the context of hypothesis testing. In this paper we consider some simulation-based procedures to compute the Cox test statistic for non-nested models, taking as an example the choice of the functional form between a linear and a loglinear specification. This example, which has been widely investigated in the literature, involves, despite its apparent simplicity, difficult computational problems which can be overcome by resorting to simulation techniques. Such problems stem from the presence of an expected value in the numerator of the statistic, which is likely not to have analytical solution and has to be solved numerically.

Either calculated analytically (if possible) or numerically, the Cox statistic has a limiting distribution of reference which is an asymptotic one. The simulation approach introduces a further level of approximation in finite samples, which is justified on the ground of asymptotic properties. Understanding the validity of the approximations represented by the finite sample distribution of the statistics obtained through simulation is a fundamental step toward the legitimation of their application. Thus, we use some Monte Carlo experiments to investigate the finite sample behaviour of the simulated versions of the Cox test and to address the issue of evaluating the loss (if any) in terms of finite sample properties we incur when using simulation instead of numerical integration. Starting from some of the simulated versions of the Cox statistic proposed by Pesaran and Pesaran (1993 and 1995), which rely on simulations only for the computation of the numerator of the statistic, we assess through a Monte Carlo experiment their performance in finite samples for the models under scrutiny and find it to be poor.

Then, generalizing a procedure suggested by Lu and Mizon (1990), we extend the use of simulations to the computation of both the numerator and the denominator of the Cox statistic, getting a simulated version of it which has a satisfactory behaviour in small samples. Although conditional on the specific Monte Carlo experiment considered, our findings

[4]See, for example Mc Fadden (1989), Pakes and Pollard (1989), Gourieroux and Monfort (1991), Laroque and Salanie (1993).

support the simulation of the covariance matrix in addition to simulating the numerator, since simulating both quantities leads to a finite sample distribution of the statistic closer to the asymptotic one.

The structure of the paper is as follows. Section 2 introduces non-nested hypothesis testing and the importance of finding some computational procedure to extend the practical implementability of the Cox statistic. In section 3 we illustrate the motivations for a simulation approach to non-nested hypothesis testing in the general framework of the encompassing principle and describe different simulation techniques for the Cox test statistic. Section 4 is devoted to the Monte Carlo experiment and to the interpretation of the results. Section 5 concludes.

2 Testing non-nested hypotheses

To define the statistical framework in which econometric models are evaluated through hypothesis testing, assume that T observations on the random variables y_t and z_t are available, and call this information set $W_T = (\underline{y}_T, \underline{z}_T)$. Let the unknown conditional distribution characterising the data be $h(y_t \mid x_t; \theta)$, where $\theta \in \Theta$ represents a parameter vector and $x_t = (z_t, \underline{y}_{t-1}, \underline{z}_{t-1})$ contains the conditioning variables (contemporaneous exogenous variables, z_t, and lagged endogenous and exogenous variables from time 1 up to time $t-1$, $\underline{y}_{t-1}$ and $\underline{z}_{t-1}$). For the conditional density above to be the basis of valid inference about θ the hypothesis of weak exogeneity of z_t with respect to θ must hold, meaning that the marginal process $(z_t \mid y_{t-1}, z_{t-1})$ can be ignored (see Engle *et alias* (1983)). Moreover, for the analysis of the following sections it is convenient to assume the condition of Granger non-causality of y_t to hold (stating that there is no feedback from lagged values of y to z_t) for the analysis of encompassing to be performed conditionally on x_t (see Mizon and Richard (1986)). Let us define the class of probability distributions put forward by the econometrician as an adequate statistical representation of the data generating process, $h(y_t \mid x_t; \theta)$, be $m(y_t \mid x_t; \gamma), \gamma \in \Gamma$ (this corresponds to the statistical model in the terminology of Spanos (1986)).

Given this setting, in the process of hypothesis testing a model is taken as null hypothesis, H_0, and is subject to statistical test against a rival model, taken as the alternative hypothesis, H_1, on the basis of a rule which indicates whether to accept or not H_0 as valid according to the observed data.

In this general context the two concepts of nested and non-nested hypothesis testing are casted. In the first case, the two models under scrutiny are such that one can be obtained by imposing some parametric restrictions on the other, i.e. the hypothesis to be tested can be expressed as $\gamma = \gamma^*$, with $\gamma^* \in \Gamma^* \subset \Gamma$. In the usual practice, the above hypothesis is taken as the null, and the procedure aims at accepting as valid a simpler model, say $m^*(y_t \mid x_t; \gamma^*)$, than the alternative one, $m(y_t \mid x_t; \gamma)$. Following Spanos (1986), nested hypothesis testing performed in this direction amounts to specification testing, as it is based on the assumption of correct specification of the more general statistical model under the alternative. It is well-known that there are three main asymptotically equivalent testing principles for nested hypotheses, namely the likelihood ratio test (LR), the Wald test (WALD) and the Lagrange multiplier test (LM). Such tests are applicable in the maximum likelihood framework, i.e. under the classical set of regularity conditions on the likelihood function ensuring the applicability of some central limit theorems and the use of Taylor series expansions [5].

On the other hand, in non-nested hypothesis testing the two rival models are such that neither can be obtained by imposing some parametric restrictions on the other. This means that the modeler is interested in comparing two separate families of probability distributions, which are both statements about the data generating process $h(y_t \mid x_t; \theta)$[6]. Following the usual notation for non-nested hypotheses, let two statistical

[5] See Godfrey (1991) for an exhaustive presentation of the three tests.

[6] It can be noticed that it is always possible to consider two non-nested models as obtained by imposing different parametric restrictions on a more general model, cfr. Lu and Mizon (1992) who describe how to build a model embedding two non-linear non-nested models.

models be defined by the following classes of probability distributions:

$$M_1: \qquad \{f(y_t \mid x_t; \alpha), \alpha \in A\} \tag{1}$$

$$M_2: \qquad \{g(y_t \mid x_t; \delta), \delta \in \Delta\}. \tag{2}$$

Non-nested hypothesis testing is to be seen as a way of checking the validity of M_1 through the evidence provided by a rival model, M_2, for the explanation of the same phenomenon. MacKinnon (1983) emphasizes that non-nested tests applied to two rival models are tests of misspecification of the model, in which the two models are taken in turn as the null hypothesis, with three possible outcomes: one can be rejected while the other cannot (providing evidence in favour of one model), both models can be rejected (indicating misspecification of both), or neither can be rejected (signaling that the information contained in the data is not enough for the models to be distinguished or that the models are observationally equivalent reparametrizations). In the econometric literature various theoretical approaches have been developed for testing a null model against one (or more) non-nested alternatives. A recent overview is given in McAleer (1995), which aims to evaluate the relevance of non-nested hypothesis testing in empirical modelling.

Most of the literature on non-nested hypothesis testing is based on a procedure proposed by Cox in two seminal papers (1961,1962), which generalizes the LR test for the case of non-nested hypotheses. Extension of the LM and WALD principles is also possible, as shown by Gourieroux and Monfort (1983). The role of the WALD test for non-nested hypotheses has been emphasized in the framework of the encompassing theory (Mizon and Richard (1986)), which is mainly based on a Wald Encompassing Test (WET). We will see in the next section that the Cox test statistic, which is the focus of our attention throughout the paper, can be interpreted as a particular case of the WET.

The proposal of Cox starts from the consideration of the usual LR test. Since, in case of non-nested hypotheses, the LR test divided by the sample size tends to a non zero limit, this limit is consistently estimated and used to to "center" the LR statistic. The Cox test presents the great advantage of being very broadly applicable, allowing, for example, the comparison of linear and non linear regression models characterised by different joint distributions of the relevant variables and therefore by different distributions of the error terms. The assumptions required for the validity of the Cox test statistic are set out in White (1982): the framework in which the models to be compared are cast is still the maximum likelihood one[7]. Unfortunately, however, simplicity of implementation does not correspond to the theoretical generality of the Cox procedure. Actually, the computational difficulties arising in the calculation of the Cox test statistic limited considerably its application. The main developments of the Cox test statistic are made by Pesaran (1974), who derives its analytical expression for the case of univariate regression models with normal- distributed errors, and Pesaran and Deaton (1978), who extend the calculation of the statistic to nonlinear and possibly multivariate normal regression models. Another interesting extension of the Cox test statistic is proposed by Aneuryn-Evans and Deaton (1980) for the case of the test of linear *versus* loglinear regression models, which we will consider in section 4.

3 Simulation approaches to encompassing for non-nested models

3.1 Motivations for a simulation approach

The encompassing principle is concerned with the capability of an econometric model to explain the characteristics of rival models, and has been formalized by Mizon (1984) and Mizon and Richard (1986) to develop a test-generating procedure unifying the literature on nested and non-nested hypothesis testing. The role of the encompassing principle as a provider of a comprehensive framework for the comparison of non-nested

[7]The assumptions include the regularity conditions needed for consistency and asymptotic normality of quasi-maximum likelihood estimators, i.e. estimators obtained maximizing a likelihood function which may not correspond to the joint density of observations.

models is of interest for the following analysis because it covers situations in which computational problems are likely to arise. Consider the two rival models introduced in (1) and (2) of the previous section. Generally speaking, M_1 is said to encompass M_2 if it can explain the results obtained by M_2. It follows that the evaluation of a given encompassing test for M_1, against M_2 involves the estimation of the expected value of a certain statistic of interest within the context of M_2 under M_1. Therefore, an integration problem has to be solved, the degree of complexity of which will depend on the form of the densities characterising the two models.

Given that model M_1, as well as model M_2, is only an approximation of the unknown process which has generated the sample y_T, its usefulness can be assessed by asking whether it can mimic the data generating process in being able to predict the behavior of statistics which are of interest in the analysis of M_2. Indicating by $\hat{\alpha}$ and $\hat{\delta}$ the pseudo maximum likelihood estimators (PML)[8] obtained by maximizing the loglikelihood functions corresponding respectively to the densities in (1) and (2), the statistic of interest within M_2 is denoted by $\hat{b} = b(\underline{W}_T, \hat{\delta})$, its expectation under M_1 by $b_\alpha = E_\alpha(\hat{b})$ and a consistent estimator of the latter is given by $b_{\hat{\alpha}} = E_{\hat{\alpha}}(\hat{b})$, obtained by evaluating the expected value at $\alpha = \hat{\alpha}$. b_α is called the pseudo-true value of $\hat{b}$. Notice that b_α depends on the sample size T, its and its evaluation requires that all the variables involved in the expression of $\hat{b}$ are given a status in M_1. Letting $b = E_\delta(\hat{b})$, the encompassing hypothesis is given by $b - b_\alpha = 0$. Such hypothesis can be tested by the difference:

$$\hat{\phi} = \hat{b} - b_{\hat{\alpha}} \tag{3}$$

which, if M_1 encompasses M_2 with respect to $\hat{b}$, should not differ significantly from zero. That is, the estimate of the prediction of $\hat{b}$ made by M_1, $b_{\hat{\alpha}}$, should coincide with the observed statistic in M_2, $\hat{b}$. Mizon and Richard (1986) indicate the conditions under which $\hat{\phi}$ has a limiting normal distribution with zero mean and variance $V_\alpha(\hat{\phi})$ under M_1, and define

[8]See Gourieroux, Monfort and Trognon (1984). Briefly, PML are MLE estimators allowing for the possibility that the likelihood function which is maximized does not correspond to the joint density of observations.

the Wald encompassing test statistic (WET), as the quadratic form:

$$\eta_W(\hat{b}) = T\hat{\phi}' V_\alpha(\hat{\phi})^+ \hat{\phi} \tag{4}$$

(where $V_\alpha(\hat{\phi})^+$ indicates a generalized inverse) has a limiting $\chi^2(\nu, 0)$ distribution under M_1, with $\nu = \text{rank}(V_\alpha(\hat{\phi}))$.

The above definition of WET is the most general one. It includes the case of complete parametric encompassing (CPE), when $\hat{b} = \hat{\delta}$ and, more relevantly for our purposes, the Cox test statistic, when the choice of $\hat{b}$ is:

$$\hat{b}_{Cox} = b_{Cox}(\underline{W}_T, \hat{\delta}) = \frac{1}{T}[L_f(\hat{\alpha}) - L_g(\hat{\delta})], \tag{5}$$

where $L_f(.)$ and $L_g(.)$ denote the loglikelihood functions associated with the whole sample in the respective models. Such a choice of $b(.)$ gives:

$$\begin{aligned} \hat{\phi}_{Cox} &= \hat{b}_{Cox} - E_{\hat{\alpha}}(\hat{b}_{Cox}) \\ &= \frac{1}{T}[L_f(\hat{\alpha}) - L_g(\hat{\delta})] - E_{\hat{\alpha}}\frac{1}{T}[L_f(\hat{\alpha}) - L_g(\hat{\delta})]. \end{aligned} \tag{6}$$

It is well known that the Cox test statistic, say S_{Cox}, it is asymptotically distributed under M_1 as a standard normal, that is:

$$S_{Cox} = \frac{\sqrt{T}\hat{\phi}_{Cox}}{V_\alpha(\hat{\phi}_{Cox})^{\frac{1}{2}}} \xrightarrow[M_1]{d} N(0,1). \tag{7}$$

It is now clear that what motivates the use of simulation techniques for the evaluation of the WET in (4) is the computation of the pseudo-true value $b_{\hat{\alpha}} = E_{\hat{\alpha}}(\hat{b})$ when the integration of the statistic $\hat{b}$ under M_1 is difficult, or even impossible to perform analytically. Analogously, a problem can arise for the computation of $E_{\hat{\alpha}}(\hat{b}_{Cox})$ when the test S_{Cox} in (7) has to be evaluated. Such computational difficulty is likely to put a binding constraint on the complexity of the formulation of the models to be compared, forcing the application of non-nested testing and encompassing for non-nested hypotheses to linear or simple non-linear models.

The first proposal of a simulation approach to encompassing was made by Hendry and Richard (1987), and focused on the objective of obtaining the empirical distribution of $\hat{\phi}$ with reference to the case of CPE (i.e. taking as statistic of interest in M_2 the estimate of the parameter $\hat{\delta}$, and $\hat{\phi} = \hat{\delta} - \delta_{\hat{\alpha}}$). Following their suggestion, it is possible to get a simulated version of the pseudo-true value $\delta_{\hat{\alpha}}$. The idea is to replace $\delta_{\hat{\alpha}} = E_{\hat{\alpha}}(\hat{\delta})$, by an approximation obtained by averaging over a certain number of replications of the statistic $\hat{\delta}$, calculated on observations artificially generated under M_1 and in correspondence to the PML estimator $\hat{\alpha}$. This suggestion was implemented for the Cox test statistic by Pesaran and Pesaran (1989, 1993, 1995). Lu and Mizon (1990) extended the method to the simulation of the covariance matrix of $\sqrt{T}\hat{\phi}$, always referring to the case of CPE. Gourieroux and Monfort (1995) analyse the WET for CPE for non nested dynamic models assuming explicitly that the true distribution of the data does not belong to one of the competing models and propose some simulated versions of it relying on analytical expressions of the asymptotic covariance matrices.

Generally speaking, the simulation approach to encompassing is valid under fairly general conditions, and can be useful for a vast range of "complex" models exhibiting non-linear functional form. Beside the generality over the kind of models considered, we want to stress the generality that the simulation procedure can reach across the number of test statistics that can be used in the comparison of two models. The simulation approach can in fact be extended from the CPE case to the more general case of the WET, considering different choices for the statistic of interest $\hat{b}$. The remainder of the paper will deal with the particular case represented by the Cox test statistic.

3.2 Simulation techniques for the Cox test statistic

Before turning to the description of the simulation procedures it is convenient to introduce some notation. For simplicity restate models (1) and (2) of section 2 as:

$$H_0: \qquad f_t(\alpha)$$

$$H_1: \qquad g_t(\delta)$$

$t = 1, \ldots, T$ and focus on the case in which $x_t = z_t$, i.e. the two models are static, and the stochastic process $\{y_t, x_t\}$ is *i.i.d.*. We assume in the following analysis that the regularity conditions set out in White (1982) on the densities $f(.)$ and $g(.)$ hold, in order to ensure that the PML $\hat{\alpha}$ and $\hat{\delta}$ converge to appropriate limits and have asymptotic normal distributions when appropriately normalized. We also define the average loglikelihood functions corresponding to the two models as:

$$\overline{L}_f = \frac{1}{T}\sum_{t=1}^{T} \log f_t(\alpha), \qquad \overline{L}_g = \frac{1}{T}\sum_{t=1}^{T} \log g_t(\delta).$$

The probability limits under H_0 of $\hat{\alpha}$ and $\hat{\delta}$ are indicated respectively by: $\alpha_0 = \text{plim}_{H_0}\hat{\alpha}$ and $\delta_{\alpha_0} = \text{plim}_{H_0}\hat{\delta}$; the latter being the asymptotic pseudo-true value of $\hat{\delta}$ under H_0[9]. The test statistic we consider is S_{Cox} in (7), but we drop henceforth the subscript "Cox" from both the numerator and the denominator to simplify the notation, i.e. we refer to:

$$S = \frac{\sqrt{T}\hat{\phi}}{V_{\alpha_0}(\hat{\phi})^{\frac{1}{2}}} \xrightarrow[H_0]{d} N(0, 1). \tag{8}$$

Gourieroux and Monfort (1992) show that the test based on the critical region:

$$C_T = S < u_a,$$

where u_a is the quantile of order a of the standard normal, is asymptotically of level a. The one-sided nature of the critical region stems from the fact that, under the alternative model H_1, $\hat{\phi}$ tends to $-\infty$.

3.2.1 Simulation of $\hat{\phi}$

Let us write the quantity in the numerator of the Cox test statistic as:

$$\hat{\phi}_1 = \left[\overline{L}_f(\hat{\alpha}) - \overline{L}_g(\hat{\delta})\right] - E_{\hat{\alpha}}\left[\overline{L}_f(\hat{\alpha}) - \overline{L}_g(\hat{\delta})\right]. \tag{9}$$

Notice that the first term in (9), is the loglikelihood ratio, and that its probability limit under H_0 is non zero in the non-nested case:

[9] Cfr. Gourieroux and Monfort (1983)

$$\overline{L}_f(\hat{\alpha}) - \overline{L}_g(\hat{\delta}) \xrightarrow[H_0]{p} E_{\alpha_0} \log f(\alpha_0) - \log g(\delta_{\alpha_0}).$$

The second term in (9) is a consistent estimate of the above probability limit and is needed to center the asymptotic distribution of the statistic.

Simulations can be used to compute the estimated pseudo-true value $\delta_{\hat{\alpha}} = E_{\hat{\alpha}}\hat{\delta}$ and $E_{\hat{\alpha}}[\overline{L}_f(\hat{\alpha}) - \overline{L}_g(\hat{\delta})]$. Both quantities involve the evaluation of an estimate of an expected value under H_0, i.e. with respect to the conditional density function $f_t(\hat{\alpha})$. Such an expected value can be replaced by a sample mean of "objects" generated by drawing from the distribution $f_t(\hat{\alpha})$. Under H_0, and in correspondence of the PML $\hat{\alpha}$, a sample of T simulated observations for y, $\underline{y}_T^h$, can be independently generated H times, keeping $\underline{x}_T$ fixed[10]. For each $h = 1, \ldots, H$, it is then possible to use $\underline{y}_T^h$ to compute the PML of δ, say $\hat{\delta}_{\hat{\alpha}}^h$, where the dependence on $\hat{\alpha}$ is induced by the simulated observations, $\underline{y}_T^h$. Averaging $\hat{\delta}_{\hat{\alpha}}^h$ over the H replications leads to a simulated estimate of the pseudo-true value $\delta_{\hat{\alpha}}$:

$$\hat{\delta}_{\hat{\alpha}}^H = \frac{1}{H}\sum_{h=1}^{H} \hat{\delta}_{\hat{\alpha}}^h. \tag{10}$$

For the following developments about the statistical properties of the simulated pseudo-true value it is convenient to introduce some further notation and definitions. Let ϵ be the error term characterising the model under H_0 and express the PML $\hat{\delta}$ under H_0 as $\hat{\delta} = \hat{\delta}(x, \epsilon, \alpha)$, evidencing the error term as the source of randomness [11]. Moreover, we henceforth treat x as fixed regressor as far as we study the statistical properties linked with the "simulation dimension" H[12]. Defining the error term of the model in such a way that it is independent of the parameter vector

[10] Taking as an example the nonlinear regression model $H_0 : y_t = h(x_t; \vartheta) + \epsilon_t$, where $\epsilon_t \sim N.I.D.(0, \sigma^2)$, y_t^h is computed as $h(x_t; \hat{\vartheta}) + \epsilon_t^h$, $t = 1, \ldots, T$ by drawing ϵ_t^h from $N(0, \hat{\sigma}^2)$. This is then independently repeated H times.

[11] When H_0 is true y is a function of (x, ϵ, α).

[12] This assumption is in line with the actual procedure followed to obtain the simulated observations, since for the generation of $y_t^h, h = 1, \ldots, H$, x_t is kept fixed throughout the H repeated drawings of the error term. Equivalently, one can consider the following analysis as performed conditionally on x.

α, call $\varphi(\epsilon)$ the (known) distribution of ϵ for given x. The estimated pseudo-true value of δ is then redefined as:

$$\delta_{\hat{\alpha}} = E_{\hat{\alpha}}\hat{\delta} = \int_{\epsilon} \hat{\delta}(x, \epsilon, \hat{\alpha})\varphi(\epsilon)d\epsilon.$$

With the above notation, it is possible to define the concept of *unbiased simulator*, adapting to our case the definition of Gourieroux and Monfort (1994).

Definition 1

Conditionally on $\hat{\alpha}$ and under H_0, let $\tilde{\delta} = \tilde{\delta}(x, \epsilon; \hat{\alpha})$ be a simulator of $\delta_{\hat{\alpha}}$. $\tilde{\delta}(x, \epsilon; \hat{\alpha})$ is said to be an *unbiased simulator* of $\delta_{\hat{\alpha}}$ if:

$$E_{\epsilon}\tilde{\delta} = \int_{\epsilon} \tilde{\delta}(x, \epsilon, \alpha)\varphi(\epsilon)d\epsilon = \delta_{\hat{\alpha}}.$$

The following propositions, proved in the Appendix, provide some finite sample and asymptotic properties of the simulated pseudo-true value under H_0.

Proposition 1

$\hat{\delta}^h_{\hat{\alpha}}$, the PML estimator of δ evaluated in $\underline{y}^h_T$, is an unbiased simulator of $\delta_{\hat{\alpha}}$ $\forall h = 1, \ldots, H$.

This finite sample property is an immediate consequence of Definition 1 and it allows to state the following asymptotic property of the simulated pseudo-true value.

Proposition 2

As H tends to infinity $\hat{\delta}^H_{\hat{\alpha}}$ converges in probability to $\delta_{\hat{\alpha}}$.

Proposition 2 states that for H going to infinity the effect of simulation on $\hat{\delta}^H_{\hat{\alpha}}$ disappears, so that the usual asymptotic theory for T going to infinity on the estimated pseudo-true value can be safely applied for H sufficiently high.

As far as the properties for T going to infinity when H is kept fixed are concerned, the main result is consistency of the simulated pseudo-true value for δ_{α_0}, while a finite value of H involve a loss of efficiency, as stated by the two following propositions. To obtain these two results concerning the asymptotics for T when H is kept fixed it is appropriate to (temporarily) abandon the assumption of x being fixed regressor and to consider the available observations as generated by the stochastic process $\{y_t, x_t\}$.

Proposition 3

As T tends to infinity, $\hat{\delta}_{\hat{\alpha}}^H$ converges in probability to δ_{α_0} for each value of H.

Proposition 4

$\hat{\delta}_{\hat{\alpha}}^H$ is T-asymptotically normal around δ_{α_0} for each value of H:

$$\sqrt{T}(\hat{\delta}_{\hat{\alpha}}^H - \delta_{\alpha_0}) \underset{T\to\infty}{\overset{d}{\to}} N(0, V^H(\delta_{\alpha_0})),$$

where the asymptotic covariance matrix is given by:

$$V^H(\delta_{\alpha_0}) = J_{gg}^{-1}[\frac{I_{gg}}{H} + (1+\frac{2}{H})I_{gf}J_{ff}^{-1}I_{gf}]J_{gg}^{-1},$$

with the matrices $I_{ij}, J_{ij}, i, j = f, g$ defined with usual notation for the asymptotic covariance matrix of PML estimators $(\hat{\alpha}, \hat{\delta})$ under H_0[13], given in the Appendix.

From the result in the above proposition it is apparent that the loss of efficiency tends to zero as H is increased without bound, and that $\hat{\delta}_{\hat{\alpha}}^H$ and $\delta_{\hat{\alpha}}$ are asymptotically equivalent as both H and T go to infinity, having asymptotic covariance matrix equal to $J_{gg}^{-1}I_{gf}J_{ff}^{-1}I_{fg}J_{gg}^{-1}$[14].

Pesaran and Pesaran (1993) proposed to use the simulated observations under H_0, $\underline{y}_T^h$, $h = 1, \ldots, H$ and the simulated pseudo-true value $\hat{\delta}_{\hat{\alpha}}^H$ to get the following simulated version of $\hat{\phi}_1$, in which an empirical mean replace an expected value under H_0 again, cfr. (9):

$$\hat{\phi}_1^H = \left[\overline{L}_f(\hat{\alpha}) - \overline{L}_g(\hat{\delta})\right] - \frac{1}{H}\sum_{h=1}^{H}\left[\overline{L}_f^h(\hat{\alpha}) - \overline{L}_g^h(\hat{\delta}_{\hat{\alpha}}^H)\right], \qquad (11)$$

[13] See, for example, Mizon and Richard (1986).

[14] cfr. the results on the asymptotics of the analytical estimated pseudo-true value $\delta_{\hat{\alpha}}$ of Gourieroux, Monfort and Trognon (1983).

where $\overline{L}_f^h(\hat{\alpha}) = \frac{1}{T}\sum_{t=1}^T \log f(y_t^h; \hat{\alpha})$ and $\overline{L}_g^h(\hat{\delta}_{\hat{\alpha}}^H) = \frac{1}{T}\sum_{t=1}^T \log g(y_t^h; \hat{\delta}_{\hat{\alpha}}^H)$.

To investigate the properties of the simulated numerator, it is useful to analyse the dependence of $\hat{\phi}_1$ on the error term ϵ under H_0. Express first the estimated loglikelihood ratio associated with the single observation as:

$$\log f_t(\hat{\alpha}) - \log g_t(\hat{\delta}) = \hat{b}(x_t, \epsilon_t, \alpha; \hat{\alpha}, \hat{\delta}). \tag{12}$$

Its simulated version under H_0 is:

$$\log f(y_t^h; \hat{\alpha}) - \log g(y_t^h; \hat{\delta}_{\hat{\alpha}}^H) = \hat{b}^h(x_t, \epsilon_t^h, \hat{\alpha}; \hat{\alpha}, \hat{\delta}_{\hat{\alpha}}^H). \tag{13}$$

Substituting (12) in (9), $\hat{\phi}_1$ is expressed under H_0 as:

$$\hat{\phi}_1 = \frac{1}{T}\sum_{t=1}^T \left[\hat{b}(x_t, \epsilon_t, \alpha; \hat{\alpha}, \hat{\delta}) - E_{\hat{\alpha}}\hat{b}(x_t, \epsilon_t, \alpha; \hat{\alpha}, \delta_{\hat{\alpha}})\right], \tag{14}$$

where:

$$E_{\hat{\alpha}}\hat{b}(x_t, \epsilon_t, \alpha; \hat{\alpha}, \delta_{\hat{\alpha}}) = \int_\epsilon \hat{b}(x_t, \epsilon, \hat{\alpha}; \hat{\alpha}, \delta_{\hat{\alpha}})\varphi(\epsilon)d\epsilon.$$

From (11) and (13) the simulated analogue of (14) can be written as:

$$\begin{aligned}\hat{\phi}_1^H &= \frac{1}{T}\sum_{t=1}^T \left[\hat{b}(x_t, \epsilon_t, \alpha; \hat{\alpha}, \hat{\delta}) - \frac{1}{H}\sum_{h=1}^H \hat{b}^h(x_t, \epsilon_t^h, \hat{\alpha}; \hat{\alpha}, \hat{\delta}_{\hat{\alpha}}^H)\right] \\ &= \frac{1}{T}\sum_{t=1}^T \left[\hat{b}_t(\hat{\alpha}, \hat{\delta}) - \hat{b}_t^H(\hat{\alpha}, \hat{\delta}_{\hat{\alpha}}^H)\right]. \end{aligned} \tag{15}$$

The following propositions, proved in the Appendix, guarantee that the above simulated numerator converges to its analytical counterpart as H is increased without bound, keeping T fixed, providing the theoretical justification for the simulation procedure.

Proposition 5

$\hat{b}^h(x, \epsilon^h, \hat{\alpha}; \hat{\alpha}, \delta_{\hat{\alpha}})$ is an unbiased simulator of $E_{\hat{\alpha}}\hat{b}(x, \epsilon, \alpha; \hat{\alpha}, \delta_{\hat{\alpha}})$.

Proposition 6

As H tends to infinity, $\hat{b}^H(\hat{\alpha}, \hat{\delta}_{\hat{\alpha}}^H)$ converges in probability to $E_{\hat{\alpha}}\hat{b}(x, \epsilon, \alpha; \hat{\alpha}, \delta_{\hat{\alpha}})$.

Pesaran and Pesaran (1989) indicated an asymptotically equivalent expression for (9), obtained by simplifying the terms involving $L_f(\hat{\alpha})$ in the expression of $\hat{\phi}_1$, as the probability limit of their difference is zero under H_0, i.e.:

$$\hat{\phi}_2 = E_{\hat{\alpha}} \overline{L}_g(\hat{\delta}) - \overline{L}_g(\hat{\delta}). \tag{16}$$

Accordingly, the simulated version of it they proposed is:

$$\hat{\phi}_2^H = \frac{1}{H} \sum_{h=1}^{H} \overline{L}_g^h(\hat{\delta}_{\hat{\alpha}}^H) - \overline{L}_g(\hat{\delta}). \tag{17}$$

Moreover, they noticed that it is possible to avoid the two stages described above by getting firstly $\hat{\delta}_{\hat{\alpha}}^H$, secondly $\overline{L}_g^h(\hat{\delta}_{\hat{\alpha}}^H)$. Two further possibilities are then given by the following "1 step" versions of the simulated numerator:

$$\hat{\phi}_{1,1step}^H = [\overline{L}_f(\hat{\alpha}) - \overline{L}_g(\hat{\delta})] - \frac{1}{H} \sum_{h=1}^{H} [\overline{L}_f^h(\hat{\alpha}) - \overline{L}_g^h(\hat{\delta}_{\hat{\alpha}}^h)], \tag{18}$$

$$\hat{\phi}_{2,1step}^H = \frac{1}{H} \sum_{h=1}^{H} \overline{L}_g^h(\hat{\delta}_{\hat{\alpha}}^h) - \overline{L}_g(\hat{\delta}). \tag{19}$$

However, from the propositions above it is clear that these one-step choices are not convenient in finite samples. In fact, they result in a loss of efficiency, since the asymptotic variance of $\hat{\delta}_{\hat{\alpha}}^H$ is minimal when H tends to infinity. Notice that expressions (11), (17), (18), (19), although asymptotically equivalent when both H and T go to infinity, might be in practice the source of different behaviours of the simulated Cox test statistic as a consequence of the finite size of the sample, T, and of a finite number, H, of simulations of the endogenous variable.

3.2.2 The analytical estimators of $V_{\alpha_0}(\hat{\phi})$

The derivation of the the asymptotic variance of $\sqrt{T}\hat{\phi}$ under H_0, say $V_{\alpha_0}(\phi)$, can be found in Gourieroux and Monfort (1987). Pesaran and Pesaran (1989,1993) suggest a consistent estimator of it which does not require any simulation. Letting $\hat{b}_t = \log f_t(\hat{\alpha}) - \log g_t(\hat{\delta})$ and $\bar{b} = \frac{1}{T}\sum_{t=1}^{T}\hat{b}_t$, the estimator has the form:

$$\hat{V}(\hat{\phi}) = \frac{1}{T}\sum_{t=1}^{T}(\hat{b}_t - \bar{b})^2 - \Psi_T'(\hat{\alpha},\hat{\delta})F_T(\hat{\alpha})^{-1}\Psi_T(\hat{\alpha},\hat{\delta}), \qquad (20)$$

where:

$$\Psi_T(\hat{\alpha},\hat{\delta}) = \frac{1}{T}\sum_{t=1}^{T}\hat{b}_t \frac{\partial \log f_t(\alpha)}{\partial \alpha}_{|\alpha=\hat{\alpha}},$$

$$F_T(\hat{\alpha}) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log f_t(\alpha)}{\partial \alpha}\frac{\partial \log f_t(\alpha)}{\partial \alpha'}_{|\alpha=\hat{\alpha}}.$$

In a more recent paper, Pesaran and Pesaran (1995) put forward two further possibilities, asymptotically equivalent to (20) for estimatiug the denominator of the test statistic[15].

One estimator, which we will denote $\hat{V}^a(\hat{\phi})$, corresponds to (20) when evaluated in the simulated pseudo-true value $\hat{\delta}^H_{\hat{\alpha}}$ instead of the PML $\hat{\delta}$ and considers a small sample correction by substituting T for $T - k_f - 1$, k_f being the number of explanatory variables of the model H_0.

The third estimator of the variance ignores the second part of (20), corresponding to the sampling uncertainty associated with the parameter estimates and it is given by:

$$\hat{V}^c(\hat{\phi}) = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{b}_t - \bar{b})^2.$$

with $\hat{b}$ and $\bar{b}$ evaluated in $\hat{\delta}^H_{\hat{\alpha}}$.

3.2.3 Simulation of $V_{\alpha_0}(\hat{\phi})$

In this section we generalize the method proposed by Lu and Mizon (1990) for the simulation of the covariance matrix of the CPE test statistic. The proposal we make to simulate the asymptotic variance of $\sqrt{T}\hat{\phi}$

[15] Pesaran and Pesaran indicate a third "outer product" version of the variance matrix, which we neglet given that it need not be non-negative.

consists in replicating H times the evaluation of the simulated $\hat{\phi}^H$ under H_0, according to whatever expression chosen among the ones above presented. The asymptotic variance can then be approximated by the empirical variance of the H replications of $\hat{\phi}^H$, appropriately scaled by T to approximate the asymptotic value. We now describe the simulation procedure for the variance when the numerator is simulated by $\hat{\phi}_1^H$, in (15), which is the one involving the greater number of steps. The procedure for the other versions of the simulated numerator is similar.

STEP 1 Following the procedure indicated in section 3.2.1, the simulated observations $\underline{y}_T^h$, $h = 1, \ldots, H$, obtained conditionally on $\hat{\alpha}$, are used to evaluate: firstly, $\hat{\delta}_{\hat{\alpha}}^H$, secondly, the quantities:

$$\hat{b}_t^h = \hat{b}^h(x_t, \epsilon_t^h, \hat{\alpha}; \hat{\alpha}, \hat{\delta}_{\hat{\alpha}}^H),$$

finally, the estimates $\hat{\alpha}^h$, $h = 1, \ldots, H$, using the observations $\underline{y}_T^h$ in the PML formula for α.

STEP 2 Conditionally on each $\hat{\alpha}^h$ obtained at step 1, M values $\underline{y}_T^{hm}$, $m = 1, \ldots, M$ are generated by drawing independently ϵ_t^m, $m = 1, \ldots, M$. Similarly to what indicated to evaluate expression (10), these artificial samples are used to compute the PML $\hat{\delta}_{\hat{\alpha}^h}^m$. Averaging over the M replications obtained for each h, we have the analogue of (10), that is:

$$\hat{\delta}_{\hat{\alpha}^h}^M = \frac{1}{M} \sum_{m=1}^{M} \hat{\delta}_{\hat{\alpha}^h}^m.$$

The quantities obtained above are then used to evaluate:

$$\hat{b}_t^{hM} = \frac{1}{M} \sum_{m=1}^{M} \hat{b}^{hm}(x_t, \epsilon_t^m, \hat{\alpha}^h; \hat{\alpha}^h, \hat{\delta}_{\hat{\alpha}^h}^M)$$

STEP 3 Combining the outcomes of step 1 and step 2, the h-th replication of the simulated numerator is (compare with (15)):

$$\hat{\phi}_1^{hM} = \frac{1}{T} \sum_{t=1}^{T} (\hat{b}_t^h - \hat{b}_t^{hM}).$$

STEP 4 Repeating step 2 H times, we get H replications of the quantity in step 3, which are used to estimate $V_{\alpha_0}(\hat{\phi})$ through the sample variance:

$$\hat{V}^H(\hat{\phi}_1) = T[\frac{1}{H}\sum_{h=1}^{H}(\hat{\phi}_1^{hM})^2 - (\frac{1}{H}\sum_{h=1}^{H}\hat{\phi}_1^{hM})^2]. \tag{21}$$

The following proposition, proved in the Appendix, states consistency of the simulator of the variance of the Cox test statistic.

Proposition 7

For M, H, T going to infinity, $\hat{V}^H(\hat{\phi}_1)$ is a consistent estimate of $V_{\alpha_0}(\hat{\phi})$.

4 Some Monte Carlo experiments on the simulated Cox statistic

4.1 Testing linear versus loglinear models

As a specific example of application of the methods described above, we consider a case which has been widely investigated in the literature, i.e. the testing of linear *versus* loglinear models. In applied econometrics, when specifying a regression model in which the dependent variable is always positive, the logarithmic transformation is very often applied in order to achieve stabilization of the variance. Therefore, some methodology allowing a rigorous (post-estimate) evaluation of both models performed on statistical grounds is advocated.

The hypotheses to be tested are given by:

$$\begin{array}{lll} H_0: & y_t = x_t'\vartheta + \epsilon_t & \epsilon_t \sim N.I.I.D.(0,\sigma^2) \\ H_1: & \log y_t = z_t'\beta + v_t & v_t \sim N.I.I.D.(0,\tau^2) \end{array} \tag{22}$$

$t = 1, \ldots, T$, where x_t and z_t are vectors $k \times 1$ containing an intercept and k-1 explanatory variables at time t, and the variables in z_t are the logarithms of the variables in x_t. Notice that the hypothesis of normality in H_0 is, in principle, untenable as, for the comparison with H_1 to be sensible, the linear model cannot generate negative values of y_t. The distribution of the error term ϵ_t in (18) has to be chosen so that it ensures that the model cannot generate negative values, for example truncated normal. In this case it is often assumed that the effect of truncation can be neglected and that the distribution of ϵ_t becomes well approximated by the normal.

Choosing one of the above options implies acceptance of quite strong assumptions, relevant from the point of view of economic theory, on the relationship between the y and x variables. Namely in the first case we assume constancy of the response of y to changes of x (constant slope), while in the second one we assume constant elasticity of y to x. This is the reason why in the econometric literature a number of different methods for testing the linear and the loglinear specifications against each other have been developed. Given the existence of a number of possibilities on the way of performing the test and the difficulty encountered in understanding the finite sample properties of the different test statistics, an interesting experiment has been conducted by Godfrey, McAleer and McKenzie (1988) (GMM), with the aim of examining the properties of various tests for linear and loglinear models. They include among the examined tests the Jarque and Bera (1980)(JB) normality test on the residuals. A remarkable result is that this test seems not to be able to reject the hypothesis of normality of the error term whenever this hypothesis should be rejected[16]. This indicates that the JB normality test has no power as a test of this form of functional form misspecification, contrary to other possible procedures. Alternative procedures include artificial nesting approaches based on the Box-Cox (1964) regression model (embedding the linear and the log-linear specifications as special cases), tests of functional form misspecification like the Reset[17] test beside the Cox test statistic. An analytical comparison of the asymptotic powers of several tests of linear and loglinear regression models is provided by Kobayashi and McAleer (1995), who show that the Lagrange Multiplier test in the Box-Cox model framework (Godfrey and Wickens, 1981) has

[16]It is evident that if u_t is normal, v_t can't be normal and vice versa.

[17]Ramsey (1969).

the highest power, while the Cox test can be ranked before the PE test of MacKinnon, the test of Bera and McAleer (1989) and the Andrews test (1971). The Cox test statistic for the models in (18) has been particularly investigated by Aneuryn-Evans and Deaton (1980) (AD), who derive its form assuming symmetric truncation of the normal distribution of ϵ_t. The calculations required in order to implement the test are quite complicated and very specific for the case under scrutiny. For the test of H_1 *versus* H_0 an analytical expression (albeit complicated) is obtainable, while numerical integration is required in order to perform the test in the opposite direction. AD find through a Monte Carlo experiment that the small sample behaviour of the Cox test statistic is close to the asymptotic one and that the test can detect misspecification of the model even when the null hypothesis is not taken as the true mechanism generating the data. In order to compare the performance of this test with that of alternative methods, GMM consider in their simulation the same data generating processes as AD[18], and reach the conclusion that the computational effort required by the Cox statistic is a sufficient justification to prefer other approaches to the problem, although they present different drawbacks in terms of statistical properties. Simulating the Cox test statistic, although it requires some programming, represents a simplification of the calculation and provides a method whose applicability is extendible to a broad class of models. What it is to be investigated is the dimension of the loss in terms of finite sample properties that the simulated procedure implies. This is the object of the following Monte Carlo experiments.

4.2 Monte Carlo experiments and results

4.2.1 The data generating processes

In our experiment we generate the data as in two of the experiments (DGP 1, DGP 2) performed by AD (1980), for comparison purposes. The models considered are:

[18] The results concerning the Cox test statistic are directly taken from the AD results.

$$
\begin{array}{lll}
H_0: & y_t = \vartheta_0 + \vartheta_1 x_t + \epsilon_t & \epsilon_t \sim N.I.I.D.(0, \sigma^2) \\
H_1: & \log y_t = \beta_0 + \beta_1 \log x_t + v_t & v_t \sim N.I.I.D.(0, \tau^2)
\end{array} \tag{23}
$$

$t = 1, \ldots, T$.

When the linear model is true, the series $\{y_t, x_t, t = 1, \ldots, T\}$ are generated by letting:

DGP 1: $\vartheta_0 = 500; \vartheta_1 = 5; \sigma = 31; x_t = 10 + 0.9x_{t-1} + u_t; u_t \sim N(0, 8^2)$;

DGP 2: $\vartheta_0 = 500; \vartheta_1 = 5; \sigma = 61; x_t = 10 + 0.9x_{t-1} + u_t; u_t \sim N(0, 16^2)$.

The above choice of the parameters justifies the assumption of normality on the residuals of H_0[19]. The following analysis is limited to the case of testing the linear model, H_0, against the loglinear one, H_1, as it is for this case that numerical integration is needed even for this simplest specification. When evaluating the power of the test, the data are generated under H_1, setting:

DGP 1a: $\beta_0 = 4.6; \beta_1 = 0.5; \tau = 0.031; \log x_t = 0.46 + 0.9 \log x_{t-1} + w_t$;
$w_t \sim N(0, 0.08^2)$;

DGP 2a: $\beta_0 = 4.6; \beta_1 = 0.5; \tau = 0.061; \log x_t = 0.46 + 0.9 \log x_{t-1} + w_t$;
$w_t \sim N(0, 0.16^2)$.

The same experiments are used in Pesaran and Pesaran (1995) to evaluate the finite sample performances of some simulated versions of the Cox test statistic based on the analytical expressions of the asymptotic variance in section 3.2.2 which we will compare with the Cox test obtained by simulation of both the numerator and its variance following the procedure in section 3.2.3. We would like to emphasize that the autoregressive nature of x_t in the above DGPs violates the assumption of i.i.d. processes formulated in section 3.2. However, stationarity and ergodicity of the series x_t should ensure the validity of the properties of the simulation procedure above stated.

[19] i.e. the distribution of ϵ_t can in princple assumed to be truncated normal, but the truncation needed to ensure the positiveness of y_t is negligible.

4.2.2 Summary of previous results

The interest in examining the possibility of improving the performance in finite samples of the simulated Cox test statistic derives from a previous simulation experiment (Monfardini, 1994) conducted with the same models[20]. Three simulated versions of the Cox statistic were considered, using respectively expressions $\hat{\phi}_1^H, \hat{\phi}_2^H, \hat{\phi}_{2,1step}^H$ for the numerator (see section 3.2.1). The denominator in the three cases was not simulated, but based on the consistent estimate of $V_{\alpha_0}(\hat{\phi})$ given in section 3.2.2. The experiment evidenced a very different behaviour of the three test statistics in finite sample, enabling us to draw the following conclusions:

1. the performance of the statistic, in terms of closeness to the standard normal distribution, improves when the two stages characterising the calculation of $\hat{\phi}_2^H$ are used, i.e. the statistic using $\hat{\phi}_2^H$, performs better than the statistic using $\hat{\phi}_{2,1step}^H$;

2. avoiding the simplification in the numerator characterising $\hat{\phi}_2^H$ gives rise to a better behaviour of the statistic, i.e. the statistic using $\hat{\phi}_1^H$ performs better than the statistic using $\hat{\phi}_2^H$.

Such considerations are in favour of approximating the numerator of the test statistic using $\hat{\phi}_1^H$, a choice supported also by the theoretical argument concerning the efficiency presented in Proposition 4.

4.2.3 Monte Carlo results

In this section we present various results corresponding to different Monte Carlo experiments whose detailed description is given below case by case. The routines to generate the series $\{y_t, x_t, t = 1, \ldots, T\}$ and to calculate the different versions of the simulated Cox test statistic are written in Gauss 3.0 and use its pseudo-random (normal) number generator on a

[20]The data generating process was the same as above except from the variances of the error terms ϵ_t and v_t, i.e. we had $\sigma = 61$ and $\tau = 0.061$. We have verified that this change has a minor effect on the results.

Hewlett Packard 700/RX machine operating under Unix. For each sample, the first 80 observations of the autoregressive process x_t were discarded to eliminate the effect of the choice of the initial values. Under DGP 2, as the probability of drawing negative values of x_t is slightly different from zero, repeated drawing from the distribution of x_t conditional on x_{t-1} has been performed until a positive value was obtained [21].

Indicating by S^n, $n = 1, \dots, 5000$, the n^{th} replication of the simulated test, the closeness of the finite sample distribution of the simulated test to the standard normal (i.e. the asymptotic distribution) is judged on the basis of the following indicators:

- $P_{0.05}$: percentage of rejections at a 5% nominal level, based on the rule $S^n < -1.64$;
- $P_{0.01}$: percentage of rejections at a 1% nominal level, based on the rule $S^n < -2.33$;
- $P^l_{0.025}$: percentage of values falling to the left of $u_{0.025}$, the 2.5% quantile of the normal distribution, based on the rule $S^n < -1.96$;
- $P^r_{0.025}$: percentage of values falling to the right of $-u_{0.025}$, based on the rule $S^n > 1.96$;
- mean: mean of the values of the test statistic over the 5000 replications;
- s.d.: standard deviation of the values of the test statistic over the 5000 replications;
- t-value(mean): test t for the hypothesis that the mean is equal to zero;
- P(sk): observed p-value of the Doornik-Hansen (1994) (DH) test of normality based on the skewness, performed on the N replications of the test statistic;

[21] The number of the cases in which such redrawing was necessary was negligible.

- P(ku): observed p-value of the DH test of normality based on the kurtosis;

- P(j): observed p-value of the Doornik-Hansen (1994) (DH) test of normality based on the joint consideration of skewness and kurtosis.

Tables 1 and 2 display some evidence on the performance in finite samples for DGPs 1 and 2 of the simulated tests statistics which do not resort to simulation of the denominator, according to the analytical expressions described in 3.2.2, namely:

$$S_1 = \frac{\sqrt{T}\hat{\phi}_1^H}{\hat{V}(\hat{\phi})^{\frac{1}{2}}}, \qquad S_1^a = \frac{\sqrt{T}\hat{\phi}_1^H}{\hat{V}^a(\hat{\phi})^{\frac{1}{2}}}, \qquad S_1^c = \frac{\sqrt{T}\hat{\phi}_1^H}{\hat{V}^c(\hat{\phi})^{\frac{1}{2}}}.$$

Different experiments have been conducted for different values of the sample size, T, and of the number of drawings of pseudo-random numbers used for its computation, H. The results concern the distribution of the test under H_0, so that the empirical rejection frequencies are estimates of the size of the test. For both the DGPs under study, in correspondence of $T = 80$ the empirical rejection frequency is considerably (almost two times) higher than the size, at both the 5% and the 1% level, for each value of H considered. An unsatisfactory pattern can also be found in the asymmetry of the empirical distribution of the statistic revealed by the values of $P^l_{0.025}$ and $P^r_{0.025}$, as the first one is systematically greater than the latter. An explanation for that seems to be the negative, and significantly different from zero, mean of the empirical distribution. When the sample size is increased to 200, the frequencies of rejection get closer to the theoretical levels, but the discrepancy between the empirical distribution of the statistic and the asymptotic distribution still appears to be large. The DH normality tests confirm that the empirical distribution is significantly different from the theoretical one in all cases. Beside the bad performance of this simulated test for finite T, results in Tables 1 and 2 evidence that it is the sample size, T, and not the level of accuracy of the approximation of the test statistic, H, the fundamental dimension to be augmented in order to approach the asymptotic distribution. This characteristic is taken into account in the following experiments, which investigate the impact of increasing T on the performance of the test.

Tables 3 to 8 are devoted to the comparison of the test as calculated in Tables 1 and 2, and the version of the test which uses both the simulated numerator and the simulated variance, as described in section 3.2.3. Moreover, we consider a further possibility, given by replacing the empirical variance in (21) with an empirical second moment of the replicated numerator, given its null expected value under H_0. Calling this alternative simulator of the variance $\hat{V}^{H,a}(\hat{\phi}_1)$, the two "new" versions of the simulated Cox test are:

$$S_2 = \frac{\sqrt{T}\hat{\phi}_1^H}{\hat{V}^H(\hat{\phi}_1)^{\frac{1}{2}}}, \qquad S_2^a = \frac{\sqrt{T}\hat{\phi}_1^H}{\hat{V}^{H,a}(\hat{\phi}_1)^{\frac{1}{2}}}.$$

In order to perform an efficient comparison, the five test statistics are computed at each replication on the same sample $\{y_t, x_t, t = 1, \ldots, T\}^n, n = 1, \ldots, 5000$. To reduce the variance across the experiments corresponding to different sample sizes, we generated samples of the maximum dimension considered, i.e. $T = 280$, and then selected appropriate subsamples of 200, 80, 40[22]. Moreover, we apply this variance reduction technique not only to the generation of samples of different sizes, but also to the generation of the set of drawings used to simulate the test statistic. Because of the high computational time and the indication of the experiments in Tables 1 and 2, we keep the number of random drawings needed to approximate both the numerator and the denominator quite low, by fixing:

- $H = 50$ for S_1, S_1^a, S_1^c,
- $H = 50, M = 50$ for S_2, S_2^a.

More interestingly, we consider the finite sample distribution in correspondence of a greater range of values of T.

Tables 3 and 4 display the results concerning the distribution of the test under H_0 for DGP 1 and DGP 2 respectively. The results we get are quite interesting: the test statistics $S_2's$ in fact systematically outperform the test statistic $S_1's$ as far as the empirical distribution under

[22]See Hendry (1983) for this variance reduction technique.

H_0 is concerned. The empirical rejection frequencies and the estimated quantiles are much closer to the theoretical values than the correspondig ones obtained through S_1, especially for small values of T ($T = 40, 80$). Moreover, the empirical distribution of $S_2's$ have a mean which is more often accepted to be significantly closer to zero, and the normality tests on $S_2's$ do evidence less problems than the ones on $S_1's$ in both tables. A remarkable feature concerning S_1^c is that limiting the analysis to the estimated 5% quantile or to the sum of the areas in columns labelled $P_{0.025}^l$ and $P_{0.025}^r$ (the latter corresponding to the choice of giving a two-sided nature to the test, contrary to what stated in section 3.2) leads to judge satisfactory its behaviour, while the pattern emerging from Tables 3 and 4 clearly indicates that the distribution of this statistic is seriously skewed to the left. This argument also applies to the Monte Carlo results in Pesaran and Pesaran (1995), from which S_1^c appears to perform relatively well.

As far as the estimated standard deviation of the statistics is concerned, notice that for S_1 and S_1^a it remains greater than one and considerably far from it compared with the other test statistics for $T = 40, T = 80$. This suggests that the particularly bad performance of these two statistics is linked to the expression of the estimate of the variance they involve. In order to understand better the ability of the different estimates of the variance in the denominator of the test to capture the actual variation of its numerator and other features of the statistics, some measures have been computed which are contained in **Tables 5 and 6**. The figures in the columns are, in order, the mean (1), the standard deviation (2) and the variance (3) of $\hat{\phi}_1^H$ over the 5000 replication, the t-test (4) for the hypothesis that (1) is equal to zero, the average over the 5000 replications of the estimated standard errors (5)[23], the average of the estimated variance (6) [24], the ratio (7) between (2) and (5), the ratio (8) between (3) and (6) (both ratios are wanted as close as possible to one)' and, finally, the value of the test statistic -approximately distributed as

[23] calibrated by division for $\sqrt{T}$ to get comparability with the quantity in column (2).

[24] devided by T for comparability with the value in column (3).

a a normal- for the hypothesis that the estimated probability is equal to its nominal value of 0.05. The results appears to be quite different between the two DGPs considered. For DGP 1 (Table 5) the mean of the numerator is reasonably close to zero, the ratios in columns (7) and (8) signal that the formulas for the variance used by S_1 and S_1^a lead serious underevaluation of the actual standard deviation of $\hat{\phi}_1^H$ for small T. This tendence disappears as T is increased. The other three statistics exhibit similar outcomes as far as the estimated variance is concerned. The last column confirms the results already emerged by Tables 3 and 4, indicating in favour of the new simulated versions of the test. In case of DGP 2 (Table 6) the most striking feature is that all analytical expressions tend to overevaluate the variance of the numerator for high T, while the simulated ones keep closest for all values of T. From such considerations it can be concluded that, although with different effects across the DGPs considered simulating the variance of the statistics leads to better results.

In **Tables 7 and 8** we let the data generating processes to follow the loglinear model, getting the empirical distribution of the statistic under H_1, and therefore some "power" results. The rejection frequencies of the false model H_0, indicating the power of the test, have been calculated using the 5% quantile of the standard normal distribution. This seems a reasonable choice for S_2 and S_2^a, whose distribution approximates the standard normal quite well, while in the case of the statistics S_1's, the quantile should be corrected in order to ensure the control on the first-type error. However, it must be stressed that after such a correction the rejection frequencies of the two tests may become closer. Therefore, the result of S_2's having apparent less power than S_1's (particularly evident for DGP1) could be even reversed. Notice the failure of acceptance of normality of the distribution in all cases but one.

A brief reference to the results of AD (1980) is of interest for our analysis: the case corresponding to our DGP 1[25], they get an empirical rejection frequency equal to 0.053 for $T = 40$ and to 0.066 for $T = 80$ (to be compared, for example with 0.045 and 0.060 obtained for S_2 in Table

[25]The results concerning DGP 2 are not presented in their paper.

3[26]). It seems therefore, at least in the particular case analysed, that the simulation method we propose for the Cox test statistic does not present disadvantages with respect to the numerical procedure used by AD[27].

Tables 9 to 12 contain some results on alternative tests which could be used to perform the comparison between H_0 and H_1. The first possibility (Tables 9 and 11) is simply to run a normality test on the residuals of the estimated models. We would like to be able to reject the normality of the residuals of the loglinear model when the data are generated by the linear one. The DH test of normality (distributed asymptotically as a χ_2^2) appears to have very little discriminatory power for DGP 1,with the rejection frequency of the hypothesis of normality increasing only sligtly with T. The power is higher for DGP 2, probably due to the fact that it is associated with a non null probability of negative values of the dependent variable in the linear model. On the other hand, when the data are generated according to the loglinear model, the rejection frequencies of the true hypothesis are close to the nominal size of the test for each T for both DGPs. The second possibility we consider (Tables 10 and 12) is the Reset test of functional misspecification form. In particular we calculate the Reset (1) (distributed as an $F_{1,T-3}$ in our case), which is a test for adding the square of the estimated dependent variable to the model in H_0. When the linear model is true, the test rejects the hypothesis of correct specification of H_0 with a frequency close to the size. On the other hand, the rejection frequencies of the same hypothesis when the true model is the loglinear one indicate the lower power of the Reset test with respect to the Cox one.

5 Conclusions

Simulation-based methods represent a useful tool for the extension of the encompassing to a vast range of non-nested models, allowing the model-

[26]These values are obtained by summing $P_{0.025}^l$ and $P_{0.025}^r$, in order to be comparable with AD findings, referred to a two-sided critical region.

[27]Their experiment differs from ours in the number of replications, they put equal to 500.

ler to escape the limit of considering only simple specifications and to avoid numerical integration. Although such methods require intensive computation, the increasing power of computers make their implementation feasible. However, before they can be used for applied econometrics, a better understanding of their finite sample properties is needed. In this paper we focus on a particular encompassing test statistic, the Cox test statistic, and examine different possibilities of computing it by simulation suggested in the literature. We then indicate how to simulate both the numerator of the test and its denominator, i.e. its standard error. The method obtained can be easily extended to encompassing test statistics for non-nested models other than Cox. A Monte Carlo experiment on a particular case of non-nested models, that is linear *versus* loglinear models, shows that simulating the denominator beside the numerator leads to a finite sample distribution of the test statistic considerably closer to the asymptotic one. Although very specific to the Monte Carlo experiment considered, our results suggest that simulating the variance could be important even in cases when an analytical expression and a consistent estimate for it is available. Moreover, comparison with previous Monte Carlo results on the same models supports the conclusion that our simulation-based procedure seems not involve any loss with respect to numerical methods.

Table 1: **DGP 1. Finite sample behaviour of S_1, S_1^a, S_1^c. Size results.**

H_0 **is true**										
	$P_{0.05}$	$P_{0.01}$	$P^l_{0.025}$	$P^r_{0.025}$	mean	s.d.	t-value (mean)	P(sk)	P(ku)	P(j)
T=80										
H=50										
S_1	0.107	0.044	0.071	0.046	-0.077	1.246	-4.355	0.000	0.185	0.000
S_1^a	0.117	0.051	0.079	0.030	-0.134	1.225	-7.726	0.000	0.140	0.000
S_1^c	0.083	0.025	0.054	0.015	-0.103	1.061	-6.836	0.000	0.000	0.000
H=250										
S_1	0.106	0.043	0.071	0.044	-0.094	1.233	-5.383	0.000	0.198	0.000
S_1^a	0.115	0.050	0.080	0.027	-0.150	1.214	-8.723	0.000	0.155	0.000
S_1^c	0.088	0.027	0.053	0.014	-0.118	1.063	-7.872	0.000	0.000	0.000
H=500										
S_1	0.107	0.043	0.072	0.043	-0.098	1.232	-5.626	0.000	0.256	0.000
S_1^a	0.115	0.051	0.080	0.026	-0.154	1.212	-8.958	0.000	0.087	0.000
S_1^c	0.090	0.027	0.053	0.014	-0.122	1.064	-8.107	0.000	0.000	0.000
T=200										
H=50										
S_1	0.084	0.028	0.053	0.025	-0.079	1.110	-5.024	0.000	0.840	0.000
S_1^a	0.091	0.035	0.063	0.013	-0.123	1.110	-7.825	0.000	0.002	0.000
S_1^c	0.075	0.020	0.045	0.011	-0.096	1.024	-6.608	0.000	0.000	0.000
H=250										
S_1	0.084	0.026	0.052	0.023	-0.088	1.098	-5.676	0.000	0.907	0.000
S_1^a	0.093	0.034	0.060	0.013	-0.132	1.099	-8.474	0.000	0.001	0.000
S_1^c	0.076	0.019	0.045	0.009	-0.105	1.016	-7.320	0.000	0.000	0.000
H=500										
S_1	0.084	0.026	0.053	0.023	-0.090	1.096	-5.803	0.000	0.916	0.000
S_1^a	0.094	0.034	0.060	0.013	-0.133	1.097	-8.598	0.000	0.002	0.000
S_1^c	0.075	0.020	0.044	0.009	-0.107	1.015	-7.459	0.000	0.000	0.000

Table 2: **DGP 2. Finite sample behaviour of S_1, S_1^a, S_1^c. Size results.**

H_0 **is true**										
	$P_{0.05}$	$P_{0.01}$	$P^l_{0.025}$	$P^r_{0.025}$	mean	s.d.	t-value (mean)	P(sk)	P(ku)	P(j)
T=80										
H=50										
S_1	0.124	0.054	0.086	0.028	-0.171	1.252	-9.645	0.000	0.953	0.000
S_1^a	0.144	0.073	0.107	0.011	-0.298	1.276	-16.535	0.000	0.000	0.000
S_1^c	0.100	0.348	0.062	0.005	-0.196	1.032	-13.455	0.000	0.000	0.000
H=250										
S_1	0.120	0.053	0.083	0.026	-0.177	1.241	-10.091	0.000	0.876	0.000
S_1^a	0.140	0.072	0.103	0.009	-0.302	1.268	-16.894	0.000	0.000	0.000
S_1^c	0.101	0.036	0.062	0.005	-0.201	1.028	-13.866	0.000	0.000	0.000
H=500										
S_1	0.120	0.054	0.085	0.026	-0.180	1.241	-10.232	0.000	0.802	0.000
S_1^a	0.140	0.072	0.104	0.009	-0.304	1.266	-17.014	0.000	0.000	0.000
S_1^c	0.100	0.036	0.063	0.005	-0.203	1.027	-13.974	0.000	0.000	0.000
T=200										
H=50										
S_1	0.080	0.031	0.050	0.008	-0.143	1.011	-10.004	0.000	0.250	0.000
S_1^a	0.105	0.047	0.075	0.002	-0.251	1.095	-16.203	0.000	0.000	0.000
S_1^c	0.076	0.026	0.047	0.001	-0.178	0.920	-13.705	0.000	0.000	0.000
H=250										
S_1	0.079	0.029	0.051	0.007	-0.145	1.000	-10.256	0.000	0.728	0.000
S_1^a	0.103	0.048	0.071	0.002	-0.251	1.083	-16.415	0.000	0.000	0.000
S_1^c	0.076	0.024	0.045	0.002	-0.180	0.911	-13.950	0.000	0.000	0.000
H=500										
S_1	0.079	0.029	0.050	0.007	-0.146	1.000	-10.329	0.000	0.706	0.000
S_1^a	0.103	0.048	0.073	0.002	-0.252	1.083	-16.479	0.000	0.000	0.000
S_1^c	0.076	0.024	0.045	0.002	-0.181	0.911	-14.020	0.000	0.000	0.000

Table 3: **DGP 1. Finite sample behaviour of $S_1, S_1^a, S_1^c, S_2, S_2^a$. Size results.**

H_0 **is true**										
	$P_{0.05}$	$P_{0.01}$	$P^l_{0.025}$	$P^r_{0.025}$	mean	s.d.	t-value (mean)	P(sk)	P(ku)	P(j)
T=40										
S_1	0.131	0.060	0.094	0.078	-0.044	1.425	-2.198	0.000	0.541	0.002
S_1^a	0.133	0.060	0.095	0.052	-0.108	1.361	-5.607	0.000	0.293	0.000
S_1^c	0.084	0.025	0.049	0.015	-0.084	1.072	-5.576	0.000	0.000	0.000
S_2	0.048	0.010	0.023	0.022	-0.009	0.981	-0.621	0.227	0.809	0.468
S_2^a	0.046	0.009	0.021	0.019	-0.009	0.962	-0.629	0.229	0.970	0.483
T=80										
S_1	0.107	0.044	0.071	0.046	-0.077	1.246	-4.355	0.000	0.185	0.000
S_1^a	0.117	0.051	0.079	0.030	-0.134	1.248	-7.726	0.000	0.140	0.000
S_1^c	0.083	0.025	0.054	0.015	-0.103	1.061	-6.836	0.000	0.000	0.000
S_2	0.059	0.013	0.030	0.030	-0.025	1.040	-1.716	0.801	0.157	0.357
S_2^a	0.056	0.012	0.028	0.027	-0.025	1.020	-1.701	0.790	0.135	0.316
T=200										
S_1	0.084	0.028	0.053	0.025	-0.079	1.110	-5.024	0.000	0.840	0.000
S_1^a	0.091	0.035	0.063	0.013	-0.123	1.111	-7.825	0.000	0.002	0.000
S_1^c	0.075	0.020	0.045	0.011	-0.096	1.024	-6.608	0.000	0.000	0.000
S_2	0.058	0.012	0.030	0.030	-0.017	1.050	-1.122	0.828	0.002	0.010
S_2^a	0.056	0.011	0.028	0.028	-0.016	1.028	-1.106	0.676	0.004	0.016
T=280										
S_1	0.085	0.028	0.051	0.020	-0.097	1.087	-6.323	0.000	0.726	0.000
S_1^a	0.093	0.035	0.062	0.012	-0.137	1.091	-8.934	0.000	0.000	0.000
S_1^c	0.079	0.023	0.046	0.011	-0.113	1.021	-7.803	0.000	0.000	0.000
S_2	0.063	0.015	0.035	0.031	-0.033	1.052	-2.216	0.966	0.004	0.014
S_2^a	0.059	0.013	0.033	0.029	-0.032	1.031	-2.203	0.876	0.005	0.020

Table 4: **DGP 2. Finite sample behaviour of $S_1, S_1^a, S_1^c, S_2, S_2^a$. Size results.**

H_0 is true										
	$P_{0.05}$	$P_{0.01}$	$P^l_{0.025}$	$P^r_{0.025}$	mean	s.d.	t-value (mean)	P(sk)	P(ku)	P(j)
T=40										
S_1	0.155	0.075	0.116	0.063	-0.162	1.481	-7.744	0.000	0.000	0.000
S_1^a	0.166	0.085	0.126	0.052	-0.296	1.425	-14.698	0.000	0.000	0.000
S_1^c	0.109	0.041	0.072	0.009	-0.198	1.108	-12.670	0.000	0.000	0.000
S_2	0.062	0.016	0.033	0.023	-0.065	1.035	-4.477	0.026	0.175	0.034
S_2^a	0.058	0.014	0.031	0.021	-0.065	1.012	-4.532	0.015	0.174	0.020
T=80										
S_1	0.124	0.054	0.086	0.028	-0.171	1.252	-9.644	0.000	0.953	0.000
S_1^a	0.144	0.073	0.107	0.011	-0.298	1.275	-16.535	0.000	0.000	0.000
S_1^c	0.100	0.035	0.062	0.005	-0.196	1.032	-13.455	0.000	0.000	0.000
S_2	0.062	0.015	0.038	0.028	-0.047	1.053	-3.157	0.118	0.001	0.002
S_2^a	0.059	0.014	0.034	0.025	-0.047	1.030	-3.201	0.082	0.002	0.002
T=200										
S_1	0.080	0.031	0.050	0.008	-0.143	1.011	-10.004	0.000	0.250	0.000
S_1^a	0.105	0.047	0.075	0.002	-0.251	1.095	-16.203	0.000	0.000	0.000
S_1^c	0.076	0.026	0.047	0.001	-0.178	0.920	-13.705	0.000	0.000	0.000
S_2	0.063	0.015	0.034	0.027	-0.033	1.058	-2.195	0.466	0.054	0.120
S_2^a	0.057	0.014	0.029	0.025	-0.032	1.034	-2.219	0.334	0.176	0.251
T=280										
S_1	0.063	0.025	0.040	0.006	-0.136	0.936	-10.277	0.000	0.468	0.000
S_1^a	0.087	0.041	0.060	0.002	-0.229	1.014	-15.978	0.000	0.000	0.000
S_1^c	0.064	0.025	0.041	0.001	-0.173	0.875	-13.963	0.000	0.000	0.000
S_2	0.064	0.016	0.036	0.027	-0.035	1.059	-2.398	0.042	0.555	0.106
S_2^a	0.060	0.015	0.034	0.034	-0.036	1.036	-2.430	0.020	0.622	0.058

Table 5: **DGP 1. Analysis of the test staitistics under H_0.**

H_0 **is true**	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Mean(\hat{\phi}_1^H)$	$SD(\hat{\phi}_1^H)$	$Var(\hat{\phi}_1^H)$	$t-val(\hat{\phi}_1^H)$	$\overline{SD}/\sqrt{T}$	$\overline{Var}/T$	(2)/(5)	(3)/(6)	$T(P_{0.05})$
T=40									
S_1	-0.00046	0.02626	0.00069	-1.22692	0.01765	0.00039	1.48799	1.77390	17.0065
S_1^a	*	*	*	*	0.01879	0.00045	1.39772	1.53019	17.2527
S_1^c	*	*	*	*	0.02247	0.00063	1.16877	1.09303	8.45023
S_2	*	*	*	*	0.02458	0.00071	1.06875	0.98148	-0.79704
S_2^a	*	*	*	*	0.02506	0.00073	1.04823	0.94337	-1.35018
T=80									
S_1	-0.00056	0.02294	0.00053	-1.71313	0.01775	0.00038	1.29243	1.39479	13.97394
S_1^a	*	*	*	*	0.01839	0.00041	1.24805	1.27295	14.67340
S_1^c	*	*	*	*	0.02027	0.00050	1.13209	1.05012	8.54198
S_2	*	*	*	*	0.02079	0.00049	1.10337	1.06700	2.81201
S_2^a	*	*	*	*	0.02121	0.00051	1.08197	1.02534	1.78675
T=200									
S_1	-0.00028	0.01707	0.00029	-1.17811	0.01513	0.00026	1.12840	1.10396	8.62550
S_1^a	*	*	*	*	0.01540	0.00027	1.10818	1.04586	10.1584
S_1^c	*	*	*	*	0.01617	0.00030	1.05571	0.94793	6.80231
S_2	*	*	*	*	0.01570	0.00027	1.08740	1.07393	2.47661
S_2^a	*	*	*	*	0.01602	0.00028	1.06577	1.03174	1.78675
T=280									
S_1	-0.00046	0.01524	0.00023	-2.14607	0.01386	0.00022	1.10001	1.06060	8.87423
S_1^a	*	*	*	*	0.01405	0.00023	1.08520	1.01574	10.3918
S_1^c	*	*	*	*	0.01460	0.00025	1.04402	0.93892	7.60221
S_2	*	*	*	*	0.01397	0.00021	1.09121	1.09608	3.67795
S_2^a	*	*	*	*	0.01425	0.00022	1.06980	1.05385	2.81201

Table 6: **DGP 2. Analysis if the test statistics under H_0.**

H_0 **is true**	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Mean(\hat{\phi}_1^H)$	$SD(\hat{\phi}_1^H)$	$Var(\hat{\phi}_1^H)$	$t-val(\hat{\phi}_1^H)$	$\overline{SD}/\sqrt{T}$	$\overline{Var}/T$	(2)/(5)	(3)/(6)	$T(P_{0.05})$
T=40									
S_1	-0.00039	0.05851	0.00342	-4.72522	0.03905	0.00199	1.49853	1.72050	20.54370
S_1^a	*	*	*	*	0.04313	0.00263	1.35660	1.30033	22.04480
S_1^c	*	*	*	*	0.05134	0.00364	1.13966	0.93957	13.45620
S_2	*	*	*	*	0.05216	0.00319	1.12165	1.07277	3.57186
S_2^a	*	*	*	*	0.05332	0.00334	1.09734	1.02488	2.42011
T=80									
S_1	-0.00222	0.05082	0.00258	-3.08537	0.04384	0.00258	1.15922	0.99936	15.81260
S_1^a	*	*	*	*	0.04707	0.00325	1.07972	0.79356	18.93190
S_1^c	*	*	*	*	0.05279	0.00397	0.96260	0.65016	11.74840
S_2	*	*	*	*	0.04607	0.00240	1.10306	1.07601	3.57186
S_2^a	*	*	*	*	0.04708	0.00250	1.07951	1.02997	2.70089
T=200									
S_1	-0.00132	0.03958	0.00157	-2.35728	0.04936	0.00341	0.80191	0.45823	7.90553
S_1^a	*	*	*	*	0.05103	0.00386	0.77566	0.40599	12.65100
S_1^c	*	*	*	*	0.05441	0.00427	0.72754	0.36622	7.02749
S_2	*	*	*	*	0.03671	0.00145	1.07817	1.07945	3.67795
S_2^a	*	*	*	*	0.03754	0.00151	1.05441	1.03187	2.19235
T=280									
S_1	-0.00107	0.03592	0.00129	-2.10103	0.05146	0.00373	0.69791	0.34584	3.78345
S_1^a	*	*	*	*	0.05280	0.00410	0.68023	0.21463	9.36398
S_1^c	*	*	*	*	0.05545	0.00443	0.64769	0.29120	3.99273
S_2	*	*	*	*	0.03336	0.00118	1.07655	1.09240	3.99273
S_2^a	*	*	*	*	0.03411	0.00123	1.05294	1.04445	3.03228

Table 7: **DGP 1a. Finite sample behaviour of $S_1, S_1^a, S_1^c, S_2^a, S_2^c$. Power results.**

H_1 **is true**						
	$P_{0.05}$	mean	s.d.	P(sk)	P(ku)	P(j)
T=80						
S_1	0.544	-1.817	1.370	0.002	0.000	0.000
S_1^a	0.560	-1.975	1.557	0.000	0.000	0.000
S_1^c	0.480	-1.574	1.141	0.431	0.138	0.244
S_2	0.462	-1.563	1.204	0.000	0.003	0.000
S_2^a	0.452	-1.533	1.181	0.000	0.002	0.000
T=200						
S_1	0.864	-3.066	1.324	0.000	0.879	0.000
S_1^a	0.873	-3.417	1.632	0.000	0.151	0.000
S_1^c	0.846	-2.796	1.146	0.343	0.252	0.331
S_2	0.846	-2.977	1.373	0.000	0.827	0.000
S_2^a	0.838	-2.918	1.344	0.000	0.664	0.000
T=280						
S_1	0.949	-3.721	1.321	0.000	0.482	0.000
S_1^a	0.953	-4.170	1.670	0.000	0.017	0.000
S_1^c	0.942	-3.416	1.150	0.000	0.045	0.000
S_2	0.942	-3.702	1.436	0.000	0.171	0.000
S_2^a	0.940	-3.630	1.409	0.000	0.344	0.000

Table 8: **DGP 2a. Finite sample behaviour of $S_1, S_1^a, S_1^c, S_2^a, S_2^c$. Power results.**

H_1 **is true**						
	$P_{0.05}$	mean	s.d.	P(sk)	P(ku)	P(j)
T=80						
S_1	0.891	-3.590	1.658	0.000	0.000	0.000
S_1^a	0.901	-4.539	2.537	0.000	0.000	0.000
S_1^c	0.864	-3.044	1.346	0.000	0.000	0.000
S_2	0.850	-3.280	1.686	0.000	0.000	0.000
S_2^a	0.843	-3.215	1.651	0.000	0.000	0.000
T=200						
S_1	0.998	-6.002	1.739	0.000	0.146	0.000
S_1^a	0.998	-8.158	3.140	0.000	0.000	0.000
S_1^c	0.998	-5.169	1.446	0.000	0.422	0.000
S_2	0.997	-6.265	2.242	0.000	0.000	0.000
S_2^a	0.997	-6.140	2.193	0.000	0.000	0.000
T=280						
S_1	1.000	-7.224	1.765	0.000	0.956	0.000
S_1^a	1.000	-10.028	3.354	0.000	0.000	0.000
S_1^c	1.000	-6.178	1.499	0.000	0.001	0.000
S_2	1.000	-7.778	2.471	0.000	0.000	0.000
S_2^a	1.000	-7.635	2.425	0.000	0.000	0.000

Table 9: **DGP 1, DGP 1a. Rejection frequencies of the DH normality test on the residuals of the loglinear model.**

H_0 **is true** *(power results)*	
	$P_{0.05}^{DH}$
T=40	0.055
T=80	0.067
T=200	0.117
T=280	0.137
H_1 **is true** *(size results)*	
	$P_{0.05}^{DH}$
T=80	0.044
T=200	0.053
T=280	0.053

$P_{0.05}^{DH}$ is the rejection frequency of the DH test at a 5% nominal level, based on the χ_2^2 distribution.

Table 10: **DGP 1, DGP 1a. Rejection frequencies of the $Reset(1)$ test.**

H_0 **is true** *(size results)*	
	$P_{0.05}^{R}$
T=40	0.053
T=80	0.054
T=200	0.049
T=280	0.052
H_1 **is true** *(power results)*	
	$P_{0.05}^{R}$
T=80	0.218
T=200	0.565
T=280	0.729

$P_{0.05}^{R}$ is the rejection frequency of the Reset (1) test at a 5% nominal level, based on the $F_{1,T-3}$ distribution.

Table 11: **DGP 2, DGP 2a. Rejection frequencies of the DH normality test on the residuals of the loglinear model.**

H_0 **is true** *(power results)*	
	$P^{DH}_{0.05}$
T=40	0.101
T=80	0.233
T=200	0.598
T=280	0.731
H_1 **is true** *(size results)*	
	$P^{DH}_{0.05}$
T=80	0.044
T=200	0.054
T=280	0.053

$P^{DH}_{0.05}$ is the rejection frequency of the DH test at a 5% nominal level, based on the χ^2_2 distribution.

Table 12: **DGP 2, DGP 2a. Rejection frequencies of the $Reset(1)$ test.**

H_0 **is true** *(size results)*	
	$P^{R}_{0.05}$
T=40	0.056
T=80	0.048
T=200	0.060
T=280	0.049
H_1 **is true** *(power results)*	
	$P^{R}_{0.05}$
T=80	0.538
T=200	0.924
T=280	0.981

$P^{R}_{0.05}$ is the rejection frequency of the Reset (1) test at a 5% nominal level, based on the $F_{1,T-3}$ distribution.

Appendix

Proof of Proposition 1

$$\hat{\delta}_{\hat{\alpha}}^h = \hat{\delta}(x, \epsilon^h, \hat{\alpha}) \Longrightarrow E_{\epsilon^h}\hat{\delta} = \int_{\epsilon^h} \hat{\delta}(x, \epsilon^h, \hat{\alpha})\varphi(\epsilon^h)d\epsilon^h = \delta_{\hat{\alpha}},$$

as the error terms ϵ^h are identically distributed for $\forall h = 1, \ldots, H$.

Proof of Proposition 2

From the independence and identical distribution of the drawings $\epsilon^h, h = 1, \ldots, H$, it follows that the terms $\hat{\delta}_{\hat{\alpha}}^h$ are i.i.d. conditionally on $\hat{\alpha}$. Hence the Weak Law of the Large Numbers and Proposition 1 imply:

$$\hat{\delta}_{\hat{\alpha}}^H = \frac{1}{H}\sum_{h=1}^{H} \hat{\delta}_{\hat{\alpha}}^h \underset{H\to\infty}{\xrightarrow{p}} \int_{\epsilon^h} \hat{\delta}(x, \epsilon^h, \hat{\alpha})\varphi(\epsilon^h)d\epsilon^h = \delta_{\hat{\alpha}}.$$

Proof of Proposition3

Consider the definitions of the following matrices:

$$\begin{aligned}
I_{ff} &= E_x E_{\alpha_0}\left[\frac{\partial \log f_t(\alpha)}{\partial \alpha}\frac{\partial \log f_t(\alpha)}{\partial \alpha'}\right] = J_{ff}, \\
I_{fg} &= E_x E_{\alpha_0}\left[\frac{\partial \log f_t(\alpha)}{\partial \alpha}\frac{\partial \log g_t(\delta)}{\partial \delta'}\right], \\
I_{gg} &= E_x E_{\alpha_0}\left[\frac{\partial \log g_t(\delta)}{\partial \delta}\frac{\partial \log g_t(\delta)}{\partial \delta'}\right], \\
J_{gg} &= -E_x E_{\alpha_0}\left[\frac{\partial^2 \log g_t(\delta)}{\partial \delta \partial \delta'}\right],
\end{aligned}$$

evaluated respectively in α_0 and δ_{α_0} and where the expectations are taken with respect to the true density generating the x and the true conditional density of y given x, i.e. $f_t(\alpha_0)$. Notice that when $f_t(\alpha)$ and $g_t(\delta)$ are replaced by their simulated counterparts, i.e. $f_t^h = f(y_t^h(\alpha);\alpha)$ and $g_t^h = g(y_t^h(\alpha);\delta)$, the expected value of the above quantities is unchanged. The first order conditions (f.o.c.) to get the PMLE $\hat{\delta}_{\hat{\alpha}}^h$ are:

$$\frac{\partial \overline{L}_g^h(\delta(\alpha))}{\partial \delta}_{|\delta=\hat{\delta}_{\hat{\alpha}}^h} = 0$$

valid $\forall\alpha$. Expanding these f.o.c. (multiplied by $\sqrt{T}$) around the pseudo-true value δ_{α_0} we have:

$$\hat{\delta}_{\alpha_0}^h = \delta_{\alpha_0} + \frac{1}{\sqrt{T}} J_{gg}^{-1} \frac{1}{\sqrt{T}} \frac{\sum_{t=1}^T \partial \log g(y_t^h(\alpha_0);\delta_{\alpha_0})}{\partial \delta} + o_p(1),$$

where:

$$J_{gg}^{-1} \frac{1}{\sqrt{T}} \frac{\sum_{t=1}^T \partial \log g(y_t^h(\alpha_0);\delta_{\alpha_0})}{\partial \delta} \xrightarrow[T\to\infty]{d} N(0, J_{gg}^{-1} I_{gg} J_{gg}^{-1}).$$

This allows us to write:

$$\hat{\delta}_{\alpha_0}^h = \delta_{\alpha_0} + O_p(\frac{1}{\sqrt{T}}).$$

On the other hand, expansion of $\hat{\delta}_{\hat{\alpha}}^h$ around $\hat{\delta}_{\alpha_0}^h$ gives:

$$\hat{\delta}_{\hat{\alpha}}^h = \hat{\delta}_{\alpha_0}^h + \frac{\partial \delta_{\alpha_0}}{\partial \alpha'}(\hat{\alpha} - \alpha_0) + o_p(1),$$

where the second term in the right handside is $O_p(\frac{1}{\sqrt{T}})$ [28] , so that substituting for $\hat{\delta}_{\alpha_0}^h$ the expression above found we can write:

$$\hat{\delta}_{\hat{\alpha}}^h = \delta_{\alpha_0} + O_p(\frac{1}{\sqrt{T}}),$$

showing that $\hat{\delta}_{\hat{\alpha}}^h$ is consistent for δ_{α_0} as T goes to infinity for one single drawing of the the simulated observation ($H = 1$). It is now evident that the same consistency result applies to $\hat{\delta}_{\hat{\alpha}}^H = \frac{1}{H}\sum_{h=1}^H \hat{\delta}_{\hat{\alpha}}^h$ for every value of H.

[28] cfr. Gourieroux, Monfort and Trognon (1983).

Proof of Proposition 4

Starting from the case of $H = 1$, combining the two expansions of the previous proof we have the expression:

$$\sqrt{T}(\hat{\delta}^h_{\hat{\alpha}} - \delta_{\alpha_0}) = J_{gg}^{-1} \frac{1}{\sqrt{T}} \frac{\sum_{t=1}^T \partial \log g(y_t^h(\alpha_0); \delta_{\alpha_0})}{\partial \delta} + \frac{\partial \delta_{\alpha_0}}{\partial \alpha'} \sqrt{T}(\hat{\alpha} - \alpha_0) + o_p(1),$$

where:

$$\sqrt{T}(\hat{\alpha} - \alpha_0) = J_{ff}^{-1} \frac{1}{\sqrt{T}} \frac{\sum_{t=1}^T \partial \log f(y_t; \alpha_0)}{\partial \alpha} + o_p(1),$$

and:

$$J_{ff}^{-1} \frac{1}{\sqrt{T}} \frac{\sum_{t=1}^T \partial \log f(y_t; \alpha_0)}{\partial \alpha} \underset{T \to \infty}{\overset{d}{\to}} N(0, J_{ff}^{-1}).$$

Recalling that the asymptotic variance of the first term of the expansion of $\sqrt{T}(\hat{\delta}^h_{\hat{\alpha}} - \delta_{\alpha_0})$ is $J_{gg}^{-1} I_{gg} J_{gg}^{-1}$, the equality $\frac{\partial \delta_{\alpha_0}}{\partial \alpha'} = J_{gg}^{-1} I_{gf}$ [29] implies that the second term of the same expansion has asymptotic variance $J_{gg}^{-1} I_{gf} J_{ff}^{-1} I_{fg} J_{gg}^{-1}$, and that the latter expression is also equal to the asymptotic covariance matrix between the two above mentioned terms. This leads to the result:

$$\sqrt{T}(\hat{\delta}^h_{\hat{\alpha}} - \delta_{\alpha_0}) \underset{T \to \infty}{\overset{d}{\to}} N(0, J_{gg}^{-1}(I_{gg} + 3 I_{gf} J_{ff} I_{fg}) J_{gg}^{-1}).$$

Consider now the case of H greater than one, in which, combining again the two expansions of the previous proof, the simulated pseudo-true value, $\hat{\delta}^H_{\hat{\alpha}} = \frac{1}{H} \sum_{h=1}^H \hat{\delta}^h_{\hat{\alpha}}$ can be written as:

$$\sqrt{T}(\hat{\delta}^H_{\hat{\alpha}} - \delta_{\alpha_0}) = \frac{1}{H} \sum_{h=1}^H \sqrt{T}(\hat{\delta}^h_{\alpha_0} - \delta_{\alpha_0}) + \frac{\partial \delta_{\alpha_0}}{\partial \alpha'} \sqrt{T}(\hat{\alpha} - \alpha_0) + o_p(1),$$

where the first term has a $T-$asymptotic variance equal to $\frac{1}{H} J_{gg}^{-1} I_{gg} J_{gg}^{-1}$, the asymptotic variance of the second one is given above, and the asymptotic covariance matrix between the two is given by $\frac{1}{H} J_{gg}^{-1} I_{gf} J_{ff}^{-1} I_{fg} J_{gg}^{-1}$. It follows that:

$$\sqrt{T}(\hat{\delta}^H_{\hat{\alpha}} - \delta_{\alpha_0}) \underset{T \to \infty}{\overset{d}{\to}} N(0, J_{gg}^{-1}[\frac{I_{gg}}{H} + (1 + \frac{2}{H}) I_{gf} J_{ff}^{-1} I_{fg}] J_{gg}^{-1}).$$

[29] cfr. Gourieroux, Monfort, Trognon (1983).

Proof of Proposition 5

$$E_{\epsilon^h}\hat{b}^h(x,\epsilon^h,\hat{\alpha};\hat{\alpha},\delta_{\hat{\alpha}}) = \int_{\epsilon^h}\hat{b}^h(x,\epsilon^h,\hat{\alpha};\hat{\alpha},\delta_{\hat{\alpha}})\varphi(\epsilon^h)d\epsilon^h = E_{\hat{\alpha}}\hat{b}(x,\epsilon,\alpha,\hat{\alpha};\delta_{\hat{\alpha}}).$$

as the error terms ϵ^h are identically distributed for $\forall h = 1,\ldots,H$.

Proof of Proposition 6

From the independence and identical distribution of the drawings $\epsilon^h, h = 1,\ldots,H$, it follows that the terms $\hat{b}^h(\hat{\alpha},\hat{\delta}^H_{\hat{\alpha}})$ are *i.i.d.* conditionally on $\hat{\alpha}$ and $\hat{\delta}^H_{\hat{\alpha}}$. Hence the Weak Law of the Large Numbers, Proposition 5 and Slutzky theorem imply:

$$\hat{b}^H(\hat{\alpha},\hat{\delta}^H_{\hat{\alpha}}) \underset{H\to\infty}{\overset{p}{\to}} E_{\hat{\alpha}}\hat{b}(x,\epsilon,\alpha;\hat{\alpha},\delta_{\hat{\alpha}}).$$

Proof of proposition 7

Conditionally on $\hat{\alpha}^h$, given the result:

$$\hat{\delta}^M_{\hat{\alpha}^h} = \frac{1}{M}\sum_{m=1}^{M}\hat{\delta}^m_{\hat{\alpha}^h} \underset{M\to\infty}{\overset{p}{\to}} \delta_{\hat{\alpha}^h},$$

valid with arguments similar to the ones of Proposition 2, for the Weak Law of Large Numbers and the Slutzky theorem, for M tending to infinity the following convergence in probability holds (cfr. Proposition 5):

$$\hat{b}^{hM} \underset{M\to\infty}{\overset{p}{\to}} \int_{\epsilon^m}\hat{b}^{hm}(x,\epsilon^m,\hat{\alpha}^h;\hat{\alpha}^h,\delta_{\hat{\alpha}^h})\varphi(\epsilon^m)d\epsilon^m = E_{\hat{\alpha}^h}\hat{b}(x,\epsilon,\alpha;\hat{\alpha}^h,\delta_{\hat{\alpha}^h}).$$

Accordingly, call $\hat{\phi}^{h*}_1$ the probability limit of $\hat{\phi}^{hM}_1$ as $M\to\infty$:

$$\hat{\phi}^{hM}_1 \underset{M\to\infty}{\overset{p}{\to}} \hat{\phi}^{h*}_1 = \frac{1}{T}\sum_{t=1}^{T}(\hat{b}^h_t - E_{\hat{\alpha}^h}\hat{b}(x,\epsilon,\alpha;\hat{\alpha}^h,\delta_{\hat{\alpha}^h})).$$

Write (21) as $\hat{V}^H(\hat{\phi}_1) = T\hat{V}^{HM}$, and $\hat{V}^{HM} = \frac{1}{H}\sum_{h=1}^{H}(\hat{\phi}_1^{hM} - \bar{\phi}_1^M)^2$, where $\bar{\phi}_1^M = \frac{1}{H}\sum_{h=1}^{H}\hat{\phi}_1^{hM}$. Then Slutzky theorem implies:

$$\hat{V}^{HM} \underset{M\to\infty}{\overset{p}{\rightarrow}} \hat{V}^{H*} = \frac{1}{H}\sum_{h=1}^{H}(\hat{\phi}_1^{h*} - \bar{\phi}_1^*)^2,$$

where $\bar{\phi}_1^* = \frac{1}{H}\sum_{h=1}^{H}\hat{\phi}_1^{h*}$. Notice that $\hat{V}^{H*}$ is a function of H PMLE $\hat{\alpha}^h, h = 1,\ldots,H$. This results in a problem if the limit in probability of $\hat{V}^{H*}$ for $H \to \infty$ is taken. To overcome this, recalling that $\hat{\phi}_1^{h*}$, $h = 1,\ldots,H$ are function of ϵ^h and $\hat{\alpha}$, it is possible to express them as $\tilde{\phi}_1^h(\hat{\alpha})$, $h = 1,\ldots,H$, *i.i.d.* terms for given $\hat{\alpha}$. Call their variance for finite T $V^T(\tilde{\phi}_1(\hat{\alpha})$. Letting H tend to infinity we have then:

$$\hat{V}^{H*} \underset{H\to\infty}{\overset{p}{\rightarrow}} V^T(\tilde{\phi}_1(\hat{\alpha})).$$

Finally, we have to cope with the sample size T, the dependence from which is implicit in the notation of the above quantities. Note that as T goes to infinity $\hat{\phi}_1^{h*}$ and $\tilde{\phi}_1^h(\hat{\alpha})$ have the same limit in probability as the numerator of the Cox test statistic $\hat{\phi}_1$ in formula (9), say $\hat{\phi}_1(\alpha_0)$, whose asymptotic variance is $V_{\alpha_0}(\hat{\phi}_1)$. Therefore, summarizing the above results we have:

$$\hat{V}^H(\hat{\phi}_1) \underset{M,H\to\infty}{\overset{p}{\rightarrow}} TV^T(\tilde{\phi}_1(\hat{\alpha}) \underset{T\to\infty}{\overset{p}{\rightarrow}} V_{\alpha_0}(\hat{\phi}_1).$$

References

[1] **Aneuryn-Evans,G. and Deaton,A. (1980)**, "Testing Linear versus Logarithmic Regression Models", *Review of Economic Studies*, 47, 257-291.

[2] **Atkinson,A.C. (1970)**, "A Method for Discriminating between Models", *Journal of the Royal Statistical Society*, Series B, 32, 211-243.

[3] **Bera,A.K. and Jarque,C.M. (1980)**, "Efficient Tests for Normality, Heteroskedasticity and Serial Independence of Regressor Residuals", *Economic Letters*, 6, 255-259.

[4] **Bera, A.K. and McAleer,M. (1989)**, "Nested and Non-Nested Procedures for Testing Linear and Loglinear Regression Models", Sankhya B, 51, 212-224.

[5] **Bertail,P.(1992)**, "Le Bootstrap: une Revue de la Literature", *CREST-Departement de la Recherche, INSEE*, Document de Travail No. 9201.

[6] **Box,G.E.P. and Cox,D.R. (1964)**, "An Analysis of Transformations", *Journal of the Royal Statistical Society*, Series B, 26, 211-252.

[7] **Cox,D.R. (1961)**, "Tests of Separate Families of Hypotheses", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistic and Probability*, 1, 105-123, Berkeley, University of California Press.

[8] **Cox,D.R. (1962)**, "Further Results on Tests of Separate Families of Hypotheses, *Journal of the Royal Statistical Society*, Series B, 24, 406-424.

[9] **Davidson,R. and MacKinnon,J.G. (1993)**, "Estimation and Inference in Econometrics", Oxford University Press.

[10] **Davidson,R. and MacKinnon,J.G. (1981)**, "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica*, 49, 781-793.

[11] **Doornik,J.A. and Hansen,H. (1994)**, "A Practical Test for Univariate and Multivariate Normality", mimeo, Nuffield College, Oxford.

[12] **Godfrey,L.G. (1988)**, "Misspecification Tests in Econometrics. The Lagrange Multiplier Principle and Other Approaches", *Econometric Society Monographs*, No 16, Cambridge University Press.

[13] **Godfrey,L.G., McAleer,M. and McKenzie,C.R. (1988)**, "Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models", *Review of Economics and Statistics*, 70, 492-503.

[14] **Godfrey,L.G. and Wickens, (1981)**, "Testing Linear and Loglinear regressions for functional form", Review of Economic Studies, 48, 487-496.

[15] **Gourieroux,C. and Monfort,A. (1987)**, "Statistique et Models Econometriques", Paris, Economica.

[16] **Gourieroux,C. and Monfort,A. (1991)**, "Simulation Based Econometrics in Models with Heterogeneity", *Annales d'Economie et de Statistique*, No. 20/21.

[17] **Gourieroux,C. and Monfort,A. (1992)**, "Testing Non-Nested Hypotheses", *CREST-Departement de la Recherche, INSEE*, Document de Travail No. 9207.

[18] **Gourieroux,C. and Monfort,A. (1994)**, "Simulation Based Econometric Methods", *CORE Lecture Series*, Oxford University Press.

[19] **Gourieroux,C. and Monfort,A. (1995)**, "Testing, Encompassing and Simulating Dynamic Econometric Models", *Econometric Theory*, 11, 195-228.

[20] **Gourieroux,C., Monfort,A. and Trognon,A. (1983)**, "Testing Non-Nested Hypotheses",*Journal of Econometrics*, 21, 83-115.

[21] **Gourieroux,C., Monfort,A. and Trognon,A. (1984)**, "Pseudo Maximum Likelihood Methods: Theory",*Econometrica*, 52, 681-700.

[22] **Hendry,D.F. (1984)**, "Monte Carlo Experimentation in Econometrics", in *Handbook of Econometrics, vol 2*, Griliches,Z. and Intriligator,M.D. (eds), North Holland, Amsterdam, 937-976.

[23] **Hendry,D.F. (1987)**, "Econometric Methodology: a Personal Perspective", Ch. 10 in *Advances in Econometrics*, Bewley,T.F. (ed), Cambridge University Press, 29-48.

[24] **Hendry,D.F. and Richard,J.F.(1989)**, "Recent Developments in the Theory of Encompassing",in *Contributions to Operation Research and Econometrics. The Twentieth Anniversary of CORE*, Cornet,B. and Tulkens,H. (eds.), Cambridge, MIT Press, 393-440.

[25] **Kobayashi,M. and McAleer,M. (1995)**, "Power Comparison of Tests of Linear and Loglinear Regression Models", CentER W.P., Tilburg University.

[26] **Laroque,G. and Salanie,B. (1993)**,"Simulation-Based Estimation of Models with Lagged Latent Variables", *CREST-Departement de la recherche, INSEE*, Document de travail No. 9315.

[27] **Lu,M. and Mizon,G.E. (1990)**,"Monte Carlo Encompassing: Simulation of Pseudo True Values and Covariance Matrices", mimeo, Southampton University.

[28] **Lu,M. and Mizon,G.E. (1993)**,"The Encompassing Principle and Specification Tests", WP 93/11 European University Institute.

[29] **MacKinnon,J.G. (1983)**, "Model Specification against Non-Nested Alternatives", *Econometric Reviews*, 2(1), 85-110.

[30] **McAleer,M. (1995)**, "The Significance of Testing Empirical Non-Nested Models", *Journal of Econometrics*, 67, 149-171.

[31] **McFadden,D. (1989)**, "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, 57, 995-1026.

[32] **Mizon,G.E. (1984)**, "The Encompassing Approach in Econometrics", in D.F. Hendry and K.F. Wallis (eds.) *Econometrics and Quantitative Economics*, Oxford, Basil Blackwell, 135-172.

[33] **Mizon,G.E. and Richard,J.F. (1986)**, "The Encompassing Principle and its Applications to Testing Non-Nested Hypotheses", *Econometrica*, 54, 657-678.

[34] **Monfardini,C. (1994)**, "Simulation-Based Econometric Methods and the Analysis of Encompassing", mimeo, European University Institute.

[35] **Pakes,A. and Pollard,D. (1989)**, "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57, 1027-1057.

[36] **Pesaran,M.H. (1974)**, "On the General Problem of Model Selection", *Review of Economic Studies*, 41, 153-171.

[37] **Pesaran,M.H. and Deaton,A.S. (1978)**, "Testing Nonlinear Regression Models", *Econometrica*, 46, 677-694.

[38] **Pesaran,M.H. and Pesaran,B. (1989)**, "A Simulation Approach to the Problem of Computing Cox's Statistic for Testing Non-Nested Models", *DAE* W.P. No. 8912, University of Cambridge.

[39] **Pesaran,M.H. and Pesaran,B. (1993)**, "A Simulation Approach to the Problem of Computing Cox's Statistic for Testing Non-Nested Models", *Journal of Econometrics*, 57, 377-392.

[40] **Pesaran,M. H. and Pesaran,B. (1995)**, " A Non-Nested Test of Level-Differenced versus Log-Differenced Stationary Models", *Econometric Reviews*, 14(2), 213-227.

[41] **Quandt,R.E. (1974)**, "A Comparison of Methods for Testing Non-nested Hypotheses", *Review of Economics and Statistics*, 56(1).

[42] **Ramsey,J.B. (1969)**, "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis", *Journal of the Royal Statistical Society*, Series B, 31, 350-371.

[43] **Spanos,A. (1986)**, "Statistical Foundations of Econometric Modelling", Cambridge University Press.

[44] **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, 215-261.

[45] **White,H. (1982)**, "Regularity Conditions for Cox's Test of Non-Nested Hypotheses, *Journal of Econometrics*,1, 301-318.

EUI WORKING PAPERS

EUI Working Papers are published and distributed by the European University Institute, Florence

Copies can be obtained free of charge – depending on the availability of stocks – from:

The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

Please use order form overleaf

Publications of the European University Institute

Department of Economics Working Paper Series

To Department of Economics **WP**
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
E-mail: publish@datacomm.iue.it
Italy

From Name . . . .

Address. . . . .

. . . .

. . . .

. . . .

. . . .

(Please print)

❒ Please enter/confirm my name on EUI Economics Dept. Mailing List
❒ Please send me a complete list of EUI Working Papers
❒ Please send me a complete list of EUI book publications
❒ Please send me the EUI brochure Academic Year 1996/97

Please send me the following EUI **ECO** Working Paper(s):

No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .

Date . . . . Signature . . . .

Working Papers of the Department of Economics Published since 1993

ECO No. 93/1
Carlo GRILLENZONI
Forecasting Unstable and Non-Stationary Time Series

ECO No. 93/2
Carlo GRILLENZONI
Multilinear Models for Nonlinear Time Series

ECO No. 93/3
Ronald M. HARSTAD/Louis PHLIPS
Futures Market Contracting When You Don't Know Who the Optimists Are

ECO No. 93/4
Alan KIRMAN/Louis PHLIPS
Empirical Studies of Product Markets

ECO No. 93/5
Grayham E. MIZON
Empirical Analysis of Time Series: Illustrations with Simulated Data

ECO No. 93/6
Tilman EHRBECK
Optimally Combining Individual Forecasts From Panel Data

ECO NO. 93/7
Víctor GÓMEZ/Agustín MARAVALL
Initializing the Kalman Filter with Incompletely Specified Initial Conditions

ECO No. 93/8
Frederic PALOMINO
Informed Speculation: Small Markets Against Large Markets

ECO NO. 93/9
Stephen MARTIN
Beyond Prices Versus Quantities

ECO No. 93/10
José María LABEAGA/Angel LÓPEZ
A Flexible Demand System and VAT Simulations from Spanish Microdata

ECO No. 93/11
Maozu LU/Grayham E. MIZON
The Encompassing Principle and Specification Tests

ECO No. 93/12
Louis PHLIPS/Peter MØLLGAARD
Oil Stocks as a Squeeze Preventing Mechanism: Is Self-Regulation Possible?

ECO No. 93/13
Pieter HASEKAMP
Disinflation Policy and Credibility: The Role of Conventions

ECO No. 93/14
Louis PHLIPS
Price Leadership and Conscious Parallelism: A Survey

ECO No. 93/15
Agustín MARAVALL
Short-Term Analysis of Macroeconomic Time Series *

ECO No. 93/16
Philip Hans FRANSES/Niels HALDRUP
The Effects of Additive Outliers on Tests for Unit Roots and Cointegration

ECO No. 93/17
Fabio CANOVA/Jane MARRINAN
Predicting Excess Returns in Financial Markets

ECO No. 93/18
Iñigo HERGUERA
Exchange Rate Fluctuations, Market Structure and the Pass-through Relationship

ECO No. 93/19
Agustín MARAVALL
Use and Misuse of Unobserved Components in Economic Forecasting

ECO No. 93/20
Torben HOLVAD/Jens Leth HOUGAARD
Measuring Technical Input Efficiency for Similar Production Units: A Survey of the Non-Parametric Approach

*out of print

ECO No. 93/21
Stephen MARTIN/Louis PHLIPS
Product Differentiation, Market Structure and Exchange Rate Passthrough

ECO No 93/22
F. CANOVA/M. FINN/A. R. PAGAN
Evaluating a Real Business Cycle Model

ECO No 93/23
Fabio CANOVA
Statistical Inference in Calibrated Models

ECO No 93/24
Gilles TEYSSIÈRE
Matching Processes in the Labour Market in Marseilles. An Econometric Study

ECO No 93/25
Fabio CANOVA
Sources and Propagation of International Business Cycles: Common Shocks or Transmission?

ECO No. 93/26
Marco BECHT/Carlos RAMÍREZ
Financial Capitalism in Pre-World War I Germany: The Role of the Universal Banks in the Financing of German Mining Companies 1906-1912

ECO No. 93/27
Isabelle MARET
Two Parametric Models of Demand, Structure of Market Demand from Heterogeneity

ECO No. 93/28
Stephen MARTIN
Vertical Product Differentiation, Intra-industry Trade, and Infant Industry Protection

ECO No. 93/29
J. Humberto LOPEZ
Testing for Unit Roots with the k-th Autocorrelation Coefficient

ECO No. 93/30
Paola VALBONESI
Modelling Interactions Between State and Private Sector in a "Previously" Centrally Planned Economy

ECO No. 93/31
Enrique ALBEROLA ILA/J. Humberto LOPEZ/Vicente ORTS RIOS
An Application of the Kalman Filter to the Spanish Experience in a Target Zone (1989-92)

ECO No. 93/32
Fabio CANOVA/Morten O. RAVN
International Consumption Risk Sharing

ECO No. 93/33
Morten Overgaard RAVN
International Business Cycles: How much can Standard Theory Account for?

ECO No. 93/34
Agustín MARAVALL
Unobserved Components in Economic Time Series *

ECO No. 93/35
Sheila MARNIE/John MICKLEWRIGHT
Poverty in Pre-Reform Uzbekistan: What do Official Data Really Reveal? *

ECO No. 93/36
Torben HOLVAD/Jens Leth HOUGAARD
Measuring Technical Input Efficiency for Similar Production Units: 80 Danish Hospitals

ECO No. 93/37
Grayham E. MIZON
A Simple Message for Autocorrelation Correctors: DON'T

ECO No. 93/38
Barbara BOEHNLEIN
The Impact of Product Differentiation on Collusive Equilibria and Multimarket Contact

ECO No. 93/39
H. Peter MØLLGAARD
Bargaining and Efficiency in a Speculative Forward Market

***

*out of print

ECO No. 94/1
Robert WALDMANN
Cooperatives With Privately Optimal Price Indexed Debt Increase Membership When Demand Increases

ECO No. 94/2
Tilman EHRBECK/Robert WALDMANN
Can Forecasters' Motives Explain Rejection of the Rational Expectations Hypothesis?

ECO No. 94/3
Alessandra PELLONI
Public Policy in a Two Sector Model of Endogenous Growth *

ECO No. 94/4
David F. HENDRY
On the Interactions of Unit Roots and Exogeneity

ECO No. 94/5
Bernadette GOVAERTS/David F. HENDRY/Jean-François RICHARD
Encompassing in Stationary Linear Dynamic Models

ECO No. 94/6
Luigi ERMINI/Dongkoo CHANG
Testing the Joint Hypothesis of Rationality and Neutrality under Seasonal Cointegration: The Case of Korea

ECO No. 94/7
Gabriele FIORENTINI/Agustín MARAVALL
Unobserved Components in ARCH Models: An Application to Seasonal Adjustment *

ECO No. 94/8
Niels HALDRUP/Mark SALMON
Polynomially Cointegrated Systems and their Representations: A Synthesis

ECO No. 94/9
Mariusz TAMBORSKI
Currency Option Pricing with Stochastic Interest Rates and Transaction Costs: A Theoretical Model

ECO No. 94/10
Mariusz TAMBORSKI
Are Standard Deviations Implied in Currency Option Prices Good Predictors of Future Exchange Rate Volatility?

ECO No. 94/11
John MICKLEWRIGHT/Gyula NAGY
How Does the Hungarian Unemployment Insurance System Really Work? *

ECO No. 94/12
Frank CRITCHLEY/Paul MARRIOTT/Mark SALMON
An Elementary Account of Amari's Expected Geometry

ECO No. 94/13
Domenico Junior MARCHETTI
Procyclical Productivity, Externalities and Labor Hoarding: A Reexamination of Evidence from U.S. Manufacturing

ECO No. 94/14
Giovanni NERO
A Structural Model of Intra-European Airline Competition

ECO No. 94/15
Stephen MARTIN
Oligopoly Limit Pricing: Strategic Substitutes, Strategic Complements

ECO No. 94/16
Ed HOPKINS
Learning and Evolution in a Heterogeneous Population

ECO No. 94/17
Berthold HERRENDORF
Seigniorage, Optimal Taxation, and Time Consistency: A Review

ECO No. 94/18
Frederic PALOMINO
Noise Trading in Small Markets *

ECO No. 94/19
Alexander SCHRADER
Vertical Foreclosure, Tax Spinning and Oil Taxation in Oligopoly

ECO No. 94/20
Andrzej BANIAK/Louis PHLIPS
La Pléiade and Exchange Rate Pass-Through

ECO No. 94/21
Mark SALMON
Bounded Rationality and Learning; Procedural Learning

*out of print

ECO No. 94/22
Isabelle MARET
Heterogeneity and Dynamics of Temporary Equilibria: Short-Run Versus Long-Run Stability

ECO No. 94/23
Nikolaos GEORGANTZIS
Short-Run and Long-Run Cournot Equilibria in Multiproduct Industries

ECO No. 94/24
Alexander SCHRADER
Vertical Mergers and Market Foreclosure: Comment

ECO No. 94/25
Jeroen HINLOOPEN
Subsidising Cooperative and Non-Cooperative R&D in Duopoly with Spillovers

ECO No. 94/26
Debora DI GIOACCHINO
The Evolution of Cooperation: Robustness to Mistakes and Mutation

ECO No. 94/27
Kristina KOSTIAL
The Role of the Signal-Noise Ratio in Cointegrated Systems

ECO No. 94/28
Agustín MARAVALL/Víctor GÓMEZ
Program SEATS "Signal Extraction in ARIMA Time Series" - Instructions for the User

ECO No. 94/29
Luigi ERMINI
A Discrete-Time Consumption-CAP Model under Durability of Goods, Habit Formation and Temporal Aggregation

ECO No. 94/30
Debora DI GIOACCHINO
Learning to Drink Beer by Mistake

ECO No. 94/31
Víctor GÓMEZ/Agustín MARAVALL
Program TRAMO "Time Series Regression with ARIMA Noise, Missing Observations, and Outliers" - Instructions for the User

ECO No. 94/32
Ákos VALENTINYI
How Financial Development and Inflation may Affect Growth

ECO No. 94/33
Stephen MARTIN
European Community Food Processing Industries

ECO No. 94/34
Agustín MARAVALL/Christophe PLANAS
Estimation Error and the Specification of Unobserved Component Models

ECO No. 94/35
Robbin HERRING
The "Divergent Beliefs" Hypothesis and the "Contract Zone" in Final Offer Arbitration

ECO No. 94/36
Robbin HERRING
Hiring Quality Labour

ECO No. 94/37
Angel J. UBIDE
Is there Consumption Risk Sharing in the EEC?

ECO No. 94/38
Berthold HERRENDORF
Credible Purchases of Credibility Through Exchange Rate Pegging: An Optimal Taxation Framework

ECO No. 94/39
Enrique ALBEROLA ILA
How Long Can a Honeymoon Last? Institutional and Fundamental Beliefs in the Collapse of a Target Zone

ECO No. 94/40
Robert WALDMANN
Inequality, Economic Growth and the Debt Crisis

ECO No. 94/41
John MICKLEWRIGHT/
Gyula NAGY
Flows to and from Insured Unemployment in Hungary

*out of print

ECO No. 94/42
Barbara BOEHNLEIN
The Soda-ash Market in Europe: Collusive and Competitive Equilibria With and Without Foreign Entry

ECO No. 94/43
Hans-Theo NORMANN
Stackelberg Warfare as an Equilibrium Choice in a Game with Reputation Effects

ECO No. 94/44
Giorgio CALZOLARI/Gabriele FIORENTINI
Conditional Heteroskedasticity in Nonlinear Simultaneous Equations

ECO No. 94/45
Frank CRITCHLEY/Paul MARRIOTT/ Mark SALMON
On the Differential Geometry of the Wald Test with Nonlinear Restrictions

ECO No. 94/46
Renzo G. AVESANI/Giampiero M. GALLO/Mark SALMON
On the Evolution of Credibility and Flexible Exchange Rate Target Zones *

***

ECO No. 95/1
Paul PEZANIS-CHRISTOU
Experimental Results in Asymmetric Auctions - The 'Low-Ball' Effect

ECO No. 95/2
Jeroen HINLOOPEN/Rien WAGENVOORT
Robust Estimation: An Example

ECO No. 95/3
Giampiero M. GALLO/Barbara PACINI
Risk-related Asymmetries in Foreign Exchange Markets

ECO No. 95/4
Santanu ROY/Rien WAGENVOORT
Risk Preference and Indirect Utility in Portfolio Choice Problems

ECO No. 95/5
Giovanni NERO
Third Package and Noncooperative Collusion in the European Airline Industry

ECO No. 95/6
Renzo G. AVESANI/Giampiero M. GALLO/Mark SALMON
On the Nature of Commitment in Flexible Target Zones and the Measurement of Credibility: The 1993 ERM Crisis *

ECO No. 95/7
John MICKLEWRIGHT/Gyula NAGY
Unemployment Insurance and Incentives in Hungary

ECO No. 95/8
Kristina KOSTIAL
The Fully Modified OLS Estimator as a System Estimator: A Monte-Carlo Analysis

ECO No. 95/9
Günther REHME
Redistribution, Wealth Tax Competition and Capital Flight in Growing Economies

ECO No. 95/10
Grayham E. MIZON
Progressive Modelling of Macroeconomic Time Series: The LSE Methodology *

ECO No. 95/11
Pierre CAHUC/Hubert KEMPF
Alternative Time Patterns of Decisions and Dynamic Strategic Interactions

ECO No. 95/12
Tito BOERI
Is Job Turnover Countercyclical?

ECO No. 95/13
Luisa ZANFORLIN
Growth Effects from Trade and Technology

ECO No. 95/14
Miguel JIMÉNEZ/Domenico MARCHETTI, jr.
Thick-Market Externalities in U.S. Manufacturing: A Dynamic Study with Panel Data

ECO No. 95/15
Berthold HERRENDORF
Exchange Rate Pegging, Transparency, and Imports of Credibility

*out of print

ECO No. 95/16
Günther REHME
Redistribution, Income cum Investment Subsidy Tax Competition and Capital Flight in Growing Economies

ECO No. 95/17
Tito BOERI/Stefano SCARPETTA
Regional Dimensions of Unemployment in Central and Eastern Europe and Social Barriers to Restructuring

ECO No. 95/18
Bernhard WINKLER
Reputation for EMU - An Economic Defence of the Maastricht Criteria

ECO No. 95/19
Ed HOPKINS
Learning, Matching and Aggregation

ECO No. 95/20
Dorte VERNER
Can the Variables in an Extended Solow Model be Treated as Exogenous? Learning from International Comparisons Across Decades

ECO No. 95/21
Enrique ALBEROLA-ILA
Optimal Exchange Rate Targets and Macroeconomic Stabilization

ECO No. 95/22
Robert WALDMANN
Predicting the Signs of Forecast Errors

ECO No. 95/23
Robert WALDMANN
The Infant Mortality Rate is Higher where the Rich are Richer

ECO No. 95/24
Michael J. ARTIS/Zenon G. KONTOLEMIS/Denise R. OSBORN
Classical Business Cycles for G7 and European Countries

ECO No. 95/25
Jeroen HINLOOPEN/Charles VAN MARREWIJK
On the Limits and Possibilities of the Principle of Minimum Differentiation

ECO No. 95/26
Jeroen HINLOOPEN
Cooperative R&D Versus R&D-Subsidies: Cournot and Bertrand Duopolies

ECO No. 95/27
Giampiero M. GALLO/Hubert KEMPF
Cointegration, Codependence and Economic Fluctuations

ECO No. 95/28
Anna PETTINI/Stefano NARDELLI
Progressive Taxation, Quality, and Redistribution in Kind

ECO No. 95/29
Ákos VALENTINYI
Rules of Thumb and Local Interaction

ECO No. 95/30
Robert WALDMANN
Democracy, Demography and Growth

ECO No. 95/31
Alessandra PELLONI
Nominal Rigidities and Increasing Returns

ECO No. 95/32
Alessandra PELLONI/Robert WALDMANN
Indeterminacy and Welfare Increasing Taxes in a Growth Model with Elastic Labour Supply

ECO No. 95/33
Jeroen HINLOOPEN/Stephen MARTIN
Comment on Estimation and Interpretation of Empirical Studies in Industrial Economics

ECO No. 95/34
M.J. ARTIS/W. ZHANG
International Business Cycles and the ERM: Is there a European Business Cycle?

ECO No. 95/35
Louis PHLIPS
On the Detection of Collusion and Predation

ECO No. 95/36
Paolo GUARDA/Mark SALMON
On the Detection of Nonlinearity in Foreign Exchange Data

*out of print

ECO No. 95/37
Chiara MONFARDINI
Simulation-Based Encompassing for Non-Nested Models: A Monte Carlo Study of Alternative Simulated Cox Test Statistics

***out of print**