

WILEY

Indirect Inference

Author(s): C. Gourieroux, A. Monfort and E. Renault

Source: *Journal of Applied Econometrics*, Vol. 8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques (Dec., 1993), pp. S85-S118

Published by: [Wiley](#)

Stable URL: <http://www.jstor.org/stable/2285076>

Accessed: 19/01/2015 14:27

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Wiley is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Applied Econometrics*.

<http://www.jstor.org>

INDIRECT INFERENCE

C. GOURIEROUX

CREST and CEPREMAP, 15 bvd Gabriel-Péri, 92245 Malakoff, Cedex France

A. MONFORT

CREST-INSEE, 15 bvd Gabriel-Péri, 92245 Malakoff, Cedex France

AND

E. RENAULT

GREMAQ-IDEI, 75 bvd de la Marquette, 31000 Toulouse, France

SUMMARY

In this paper we present inference methods which are based on an ‘incorrect’ criterion, in the sense that the optimization of this criterion does not directly provide a consistent estimator of the parameter of interest. Moreover, the argument of the criterion, called the auxiliary parameter, may have a larger dimension than that of the parameter of interest. A second step, based on simulations, provides a consistent and asymptotically normal estimator of the parameter of interest. Various testing procedures are also proposed. The methods described in this paper only require that the model can be simulated, therefore they should be useful for models whose complexity rules out a direct approach. Various fields of applications are suggested (microeconometrics, finance, macroeconometrics).

1. INTRODUCTION

Econometric models often lead to complex formulations for the conditional distribution of the endogenous variables given the exogenous variables and the lagged endogenous values. These formulations may even be such that, it is impossible to efficiently estimate the parameters of interest because of the intractability of the likelihood function. The examples of such situations are numerous in the literature: discrete choice models with a large number of alternatives, job search models, description of optimal dynamic behaviours, non-linear models with random coefficients, non-linear rational expectation models, continuous-time model with stochastic volatility, factor ARCH models. In such cases, a natural procedure is to replace the likelihood function by another criterion. This criterion could be, for instance, an approximation of the exact likelihood function, or the exact likelihood function of an approximated model, but we shall see that the choice is much larger.

The aim of this paper is to show that a correct inference can be based on this ‘incorrect’ criterion. The main steps of the estimation method, called the indirect estimation method, are the following. First, we consider an auxiliary parameter and an estimation method for this parameter. Then this method is applied to the observations and to simulated values drawn

0883–7252/93/0S0S85–34\$22.00

© 1993 by John Wiley & Sons, Ltd.

Received September 1992

Revised May 1993

from the true model and associated with a value θ of the parameter of interest. Finally, θ is calibrated in order to obtain close values for the two estimators of the auxiliary parameter. The resulting estimator of θ has rather nice asymptotic properties when the auxiliary parameter and the criterion are well chosen and this allows for the development of a complete theory of indirect estimation. This kind of method was first proposed by Smith (1990) and the extensions described here are threefold. First, we take explicitly into account the presence of exogenous variables and we derive the modified variance–covariance matrix of the estimator. Second, we consider more general criteria for the first-step estimator of the auxiliary parameter. Third, we introduce several specification tests of the initial model and a general theory of hypotheses testing based on indirect estimation. Moreover, we show how these methods can be applied to the microeconometrics, macroeconometrics, and econometrics of finance.

In Section 2 we describe the indirect estimation method with its two steps: (1) definition and estimation of the auxiliary parameter and (2) calibration of the parameter of interest by comparison of the estimators of the auxiliary parameter computed from the observations and from simulations, respectively.

The asymptotic distributional properties of indirect estimators are given in Section 3. They depend on the retained auxiliary parameter, on the first-step estimation procedure, on the metric used to compare the two estimators, and on the number of simulated values. We discuss the optimal choice of the metric. Moreover, we show that the indirect method contains, as a special case, the simulated method of moments (SMM) (McFadden, 1989; Pakes and Pollard, 1989; Duffie and Singleton, 1989; Smith, 1993; Gouriéroux and Monfort, 1993).

In Section 4 we show how indirect estimators may be used to construct specification tests. We also propose a consistency test for the estimator of the parameter of interest based on a proxy model.

In Section 5 we introduce a general test theory based on indirect estimators and we show, in particular, that the usual equivalence between the Wald test, the score test, and the test based on optimal values of the objective function remains valid despite the presence of simulations.

In Section 6 we consider a simple Monte Carlo illustration of the technique: the estimation of the parameters of a moving-average process by means of a preliminary estimation of an autoregressive representation. In this case the indirect estimation method may be seen as a way of correcting for the lag truncation effect in the AR representation. Some light on the finite sample properties of the method is obtained from this example.

In Section 7 we show how the indirect methods can be used for the estimation of the parameters appearing in stochastic differential equations. Monte Carlo results are given for the case of a geometric brownian motion and of an Ornstein–Uhlenbeck process.

In Section 8 we consider two applications to the econometrics of finance: estimation of stochastic volatility models and of factor ARCH models.

In Section 9 we describe potential applications to microeconometrics. In discrete choice models with a large number of alternatives, the log-likelihood function often contains high-dimensional integrals. To solve this problem we first exhibit a logit approximation of the c.d.f of the multivariate normal distribution. Then this approximation is used to base the indirect approach on an approximated maximum likelihood procedure.

In Section 10 we describe some examples in the domain of macroeconometrics: correction for the linearization of a non-linear model, correction for the use of the extended Kalman filter.

Section 11 concludes and the proofs are given in four appendices.

2. INDIRECT ESTIMATION METHODS

2.1. The Model

We consider the following dynamic model:

$$\begin{aligned} y_t &= r(y_{t-1}, x_t, u_t, \theta) \\ u_t &= \varphi(u_{t-1}, \varepsilon_t, \theta), \quad \theta \in \mathbb{H} \in \mathbb{R}^p \end{aligned} \quad (1)$$

where the x_t 's are observable exogenous variables whereas the u_t 's and the ε_t 's are not observed. We assume that: (A1) $\{x_t\}$ is an homogeneous Markov process, independent of $\{\varepsilon_t\}$ (and $\{u_t\}$); the process $\{\varepsilon_t\}$ is a white noise whose distribution G_0 is known, and the process $\{y_t, x_t\}$ is stationary.

Note that the knowledge of the distribution of ε_t is not a real assumption, in the parametric case, since ε_t can always be considered as a function of a white noise with a known distribution and of a parameter which can be incorporated into θ .

The conditional probability density function (p.d.f.) of x_t given its past is denoted by: $f_0(x_t/\mathbf{x}_{t-1}) = f_0(x_t/x_{t-1})$, where $\mathbf{x}_{t-1} = (x_{t-1}, x_{t-2}, \dots)$.

Models with processes of larger order or with reduced forms containing more than one lag in y , x , or u can be included in the previous formulation by increasing the dimension of the processes.

Under the above assumption it is possible to simulate values of y_1, \dots, y_T for a given initial condition $z_0 = (y_0, u_0)$ and a given value of the parameter θ , conditionally to an observed path of the exogenous variables x_0, \dots, x_T . This is done by drawing simulated values $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_T$ in G_0 , and by computing

$$\tilde{y}_t(\theta, z_0), \quad t = 0, \dots, T \quad (2)$$

where

$$\begin{aligned} \tilde{y}_0(\theta, z_0) &= y_0 \\ \tilde{y}_t(\theta, z_0) &= r[\tilde{y}_{t-1}(\theta, z_0), x_t, \tilde{u}_t(\theta, u_0), \theta] \\ \tilde{u}_t(\theta, u_0) &= \varphi[\tilde{u}_{t-1}(\theta, u_0), \tilde{\varepsilon}_t, \theta] \end{aligned}$$

Note that assumption (A1) implies that x_t is a strongly exogenous process. This means that we assume that potential problems due to non-strongly exogenous variables have been solved by considering them as functions of lagged endogenous variables. It is also worth noting that models in which non-strongly exogenous variables appear have the serious drawback of not being simulable.

With such a parametric model it is theoretically possible to compute the conditional density function of $y_1 \dots y_T$ given $z_0, x_1 \dots x_T$, and therefore to estimate the unknown true value θ_0 of θ by a conditional maximum likelihood approach. However, in practice this likelihood function may be computationally intractable. In the following subsections we describe two-step estimation methods, in which all that is required from model (1) is to be easily simulated.

2.2. The Auxiliary Parameter and its Estimator

The auxiliary parameter and its estimator are defined in the following way. We introduce a criterion which depends on the observations $\mathbf{y}_T^1 = (y_1, \dots, y_T)$, $\mathbf{x}_T^1 = (x_1, \dots, x_T)$ and on the

auxiliary parameter $\beta \in B \subset \mathbb{R}^q$. This parameter is estimated by maximizing the criterion

$$\text{Max}_{\beta \in B} Q_T(\mathbf{y}_T^1, \mathbf{x}_T^1, \beta) \quad (3)$$

We denote by $\hat{\beta}_T$ the solution of this problem:

$$\hat{\beta}_T = \text{Argmax}_{\beta \in B} Q_T(\mathbf{y}_T^1, \mathbf{x}_T^1, \beta) \quad (4)$$

Let us assume that the criterion tends asymptotically (and uniformly almost certainly) to a non-stochastic limit:

$$(A2) \quad \lim_{T \rightarrow \infty} Q_T(\mathbf{y}_T^1, \mathbf{x}_T^1, \beta) = Q_\infty(F_0, G_0, \theta_0, \beta) \quad (5)$$

This limit depends on the unknown auxiliary parameter β , on the characteristics of the true distribution (i.e. the transition distribution F_0 of $\{x_t\}$, which is unknown), on the marginal distribution G_0 of $\{e_t\}$ (which is known), and on the true parameter of interest θ_0 . It may also depend on the initial value z_0 , and equation (5) includes the assumption that the initial conditions have no asymptotic effects. Moreover, let us assume that this limit criterion is continuous in β and as a unique maximum:

$$(A3) \quad \beta_0 = \text{Argmax}_{\beta \in B} Q_\infty(F_0, G_0, \theta_0, \beta) \quad (6)$$

We know from the usual asymptotic theory (see Gallant and White, 1988, Chapter 3) that under assumptions (A1), (A2) and (A3) the estimator $\hat{\beta}_T$ is a consistent estimator of the auxiliary parameter β_0 . It is clear that the auxiliary parameter β_0 is unknown since it depends on the unknown parameter of interest θ_0 as well as on the unknown transition distribution F_0 of the exogenous variables. Moreover, it will be useful to introduce the so-called *binding function* (Gouriéroux and Monfort, 1992) defined by

$$b(F, G, \theta) = \text{Argmax}_{\beta \in B} Q_\infty(F, G, \theta, \beta) \quad (7)$$

We have

$$\beta_0 = b(F_0, G_0, \theta_0) \quad (8)$$

If the function

$$b(F_0, G_0, \cdot): \theta \rightarrow b(F_0, G_0, \theta)$$

was known and one to one, we could deduce from $\hat{\beta}_T$ a consistent estimator of the unknown parameter of interest θ_0 by considering the solution $\hat{\theta}_T$ of $\hat{\beta}_T = b(F_0, G_0, \hat{\theta}_T)$. However, if the model contains some exogenous variables, F_0 is unknown and the previous approach cannot be used.¹ Even for models without exogenous variables, i.e. for pure time-series models, the

¹ Moreover, even if F_0 was known, such an approach would be generally suboptimal since it does not efficiently use the information provided by observations \mathbf{x}_T^1 , of exogenous variables. For instance, in the case of i.i.d. random variables (x_t, y_t) , $t = 1, 2, \dots, T$ and of an auxiliary parameter β which is estimated by an M-estimation procedure:

$$Q_T(\mathbf{y}_T^1, \mathbf{x}_T^1, \beta) = \frac{1}{T} \sum_{t=1}^T \psi(y_t, x_t, \beta)$$

it is clear that this estimator satisfies $\hat{\beta}_T = b(\hat{F}_T, \hat{G}_T, \theta_0)$, where \hat{F}_T and \hat{G}_T are the empirical probability distribution of x and $u (= \varepsilon)$. Therefore if the 'finite sample binding function' $b(\hat{F}_T, \hat{G}_T, \cdot)$ (see Gouriéroux and Monfort, 1992) was known and one to one we could deduce from $\hat{\beta}_T$ the exact value θ_0 of unknown parameters while the knowledge of the true binding function $b(F_0, G_0, \cdot)$ and the solution of $b(F_0, G_0, \theta) = \hat{\beta}_T$ only provides a consistent estimator $\hat{\theta}_T$! This is the reason why the indirect estimation procedure of θ_0 follows the previous idea after replacement of the unknown function $b(\hat{F}_T, \hat{G}_T, \cdot)$ by a functional estimator based on simulations of the y 's. It is important to keep in mind that we try to estimate $b(\hat{F}_T, \hat{G}_T, \cdot)$ rather than $b(F_0, G_0, \cdot)$. In order to do this, it is better to perform only simulations of the u_t associated with observed values of the x_t .

binding function may be difficult to compute. The second step of the estimation procedure follows the previous idea after replacement of the unknown function $b(F_0, G_0, \cdot)$ by a functional estimator based on simulations of the y 's. The following assumption will be necessary:

$$(A4) \quad b(F_0, G_0, \cdot) \text{ is one to one and } \frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0) \text{ is of full-column rank} \quad (9)$$

2.3. The Second Step

For a given value of θ , we can consider H simulated paths $[\tilde{y}_t^h(\theta, z_0^h), t = 0, \dots, T], h = 1 \dots H$ based on independent drawings of ε_t , $(\tilde{\varepsilon}_1^h, \dots, \tilde{\varepsilon}_T^h)$, and on initial values $z_0^h, h = 1, \dots, H$.

For each of these paths, we can also consider the optimization problem:

$$\text{Max}_{\beta \in B} Q_T((\tilde{y}^h)_T^1, x_1^1, \beta) \quad (10)$$

in which the observed values are replaced by the simulated ones. This problem has a solution:

$$\tilde{\beta}_T^h(\theta, z_0^h) = \text{Argmax}_{\beta \in B} Q_T((\tilde{y}^h)_T^1, x_1^1, \beta) \quad (11)$$

When T tends to infinity this solution tends to a solution of the limit problem:

$$\text{Max}_{\beta \in B} Q_\infty(F_0, G_0, \theta, \beta)$$

i.e.

$$\lim_{T \rightarrow \infty} \tilde{\beta}_T^h(\theta, z_0^h) = b(F_0, G_0, \theta)$$

Therefore $\tilde{\beta}_T^h(\cdot, z_0^h)$ is a consistent functional estimator of $b(F_0, G_0, \cdot)$.

It is now possible to define the indirect estimator of θ . The idea is simply to calibrate the value of θ in order to have

$$\frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h)$$

close to $\hat{\beta}_T$.

Proposition 1: An indirect estimator of θ is defined as a solution $\tilde{\theta}_T^H(\Omega)$ of a minimum distance problem:

$$\text{Min}_{\theta \in \Theta} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right]' \hat{\Omega}_T \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right]$$

where $\hat{\Omega}_T$ is a positive definite matrix converging to a deterministic positive definite matrix Ω . Under assumptions (A1), (A2), (A3) and (A4) the indirect estimator $\tilde{\theta}_T^H(\Omega)$ is a consistent estimator of θ_0 .

The previous approach necessitates H optimizations of the simulated criterion for each value of θ involved in the minimization algorithm. It is possible to replace these H optimizations by only one. Let us first consider the TH values of the x variables obtained by repeating H times the values $x_1 \dots x_T$:

$$\tilde{x}_1 = x_1, \dots, \tilde{x}_T = x_T, \quad \tilde{x}_{T+1} = x_1, \dots, \tilde{x}_{TH} = x_T.$$

Then we compute

$$\tilde{y}_t(\theta, z_0), \quad t = 0, \dots, HT$$

where

$$\begin{aligned}\tilde{y}_0(\theta, z_0) &= y_0 \\ \tilde{y}_t(\theta, z_0) &= r[\tilde{y}_{t-1}(\theta, z_0), \tilde{x}_t, \tilde{u}_t(\theta, z_0), \theta] \\ \tilde{u}_t(\theta, u_0) &= \varphi[\tilde{u}_{t-1}(\theta, u_0), \tilde{\varepsilon}_t, \theta]\end{aligned}$$

and $\tilde{\varepsilon}_t, t = 1, \dots, TH$ are TH drawings of the white noise process $\{\varepsilon_t\}$.

This approach is equivalent to H drawings of values $\tilde{y}_1^h \dots \tilde{y}_T^h$, with different initial values $z_0, z_T = (y_T, \tilde{u}_T), \dots, z_{T(H-1)} = (y_{T(H-1)}, \tilde{u}_{T(H-1)})$. From these simulated data we may deduce a simulated criterion $Q_T[\tilde{y}_{1H}^1, \tilde{x}_{1H}^1, \beta]$ and the associated estimator

$$\tilde{\beta}_{HT}[\theta, z_0] = \underset{\beta \in B}{\text{Argmax}} \quad Q_T[\tilde{y}_{1H}^1, \tilde{x}_{1H}^1, \beta] \quad (12)$$

Proposition 2: Another version of the indirect estimator of θ is defined as the solution of

$$\underset{\theta \in \Theta}{\text{Min}} \quad (\hat{\beta}_T - \tilde{\beta}_{HT}[\theta, y_0])' \hat{\Omega}_T[\hat{\beta}_T - \tilde{\beta}_{HT}(\theta, y_0)]$$

It is a consistent estimator of θ_0 under the same conditions as in proposition 1.

It will be shown in Appendix 1 that both versions of the indirect estimator have the same asymptotic properties.

This kind of estimator is considered by Smith (1990) for specific criteria based on misspecified log-likelihood functions and when no exogenous variables are present.

2.4. Numerical Aspects

It should be emphasized that the indirect estimator is obtained without evaluating the complete functions $\tilde{\beta}_T^h(\cdot, z_0^h)$ or $\tilde{\beta}_{HT}(\cdot, z_0)$, but only their values for the θ 's appearing in the optimization algorithm. Moreover, the same simulated values $\tilde{\varepsilon}_1^h, \dots, \tilde{\varepsilon}_T^h, h = 1, \dots, H$ are used for *all* the values of θ ; these values can be either stored in memory or redrawn with the same seed values.

In several applications (see Sections 7–10) the auxiliary parameter and the criterion are defined in one of the following ways. First, we can replace the initial reduced-form equation (1) by another:

$$y_t = r^*(y_{t-1}, x_t, u_t, \beta) \quad (13)$$

where r^* is an approximation of r for which the conditional log-likelihood function $L_T^*(\beta)$ may be easily derived and we take $Q_T = (1/T)L_T^*(\beta)$. Second, we can take for Q_T an approximation of $(1/T)L_T(\beta)$, where $L_T(\beta)$ is the exact log-likelihood function. In such cases the estimation method based on Q_T is an approximated log-likelihood method and the β parameter often has the same dimension as θ (and a similar interpretation). In these important cases, the indirect estimator is independent of the choice of $\hat{\Omega}_T$ and is the solution of

$$\hat{\beta}_T = \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h)$$

or of

$$\hat{\beta}_T = \tilde{\beta}_{HT}(\theta, z_0)$$

This system may be solved numerically by using as initial value of θ : $\theta^0 = \hat{\beta}_T$.

2.5. An Equivalent Method

Following Gallant and Tauchen (1992), we can propose a third version of the indirect estimator of θ obtained by

$$\tilde{\theta}^H(\Sigma) = \underset{\theta}{\text{Argmin}} \frac{\partial Q_T}{\partial \beta'} [\tilde{y}_{TH}^1(\theta), \tilde{x}_{TH}^1, \hat{\beta}_T] \Sigma \frac{\partial Q_T}{\partial \beta} [\tilde{y}_{TH}^1(\theta), \tilde{x}_{TH}^1, \hat{\beta}_T]$$

It is shown in Appendix 1 that $\tilde{\theta}(\Sigma)$ is asymptotically equivalent to $\tilde{\theta}_T^H(J_0 \Sigma J_0)$ [see (17) below]. Therefore the family of estimators $\tilde{\theta}^H(\Sigma)$ is asymptotically the same as the family $\tilde{\theta}_T^H(\Omega)$; the optimal value of Σ is obviously $\Sigma^* = J_0^{-1} \Omega^* J_0^{-1} = (I_0 - K_0)^{-1}$ [see (19) below].

This version may be computationally interesting if the gradient $\partial Q_T / \partial \beta$ has a closed form, because it necessitates only one optimization in θ . This kind of approach was first proposed by Gallant and Tauchen (1992) in a special case where: the criterion function is a likelihood function, the number of simulations is infinite, there are no exogenous variables, and the model corresponding to the 'pseudo-likelihood' function is asymptotically well specified. In this special case K_0 is equal to zero, I_0 is the information matrix based on the 'pseudo-score', and the term $1 + (1/H)$ appearing in proportions 4 and 5 is equal to 1.

This third version of the indirect estimators is different from the estimator generalizing the Simulated Quasi Maximum Likelihood estimator (SQML) proposed by Smith (1993). This generalized estimator is obtained by maximizing the constrained criterion:

$$\underset{\theta}{\text{Max}} Q_T[y_T^1, x_T^1, \hat{\beta}_{HT}(\theta)]$$

It is shown in Appendix 1 that this estimator is consistent but not asymptotically efficient, since it has the same asymptotic behaviour as $\tilde{\theta}_T^H(J_0)$ and, in general, $J_0 \neq \Omega^* = J_0(I_0 - K_0)^{-1}J_0$.

3. ASYMPTOTIC DISTRIBUTION OF INDIRECT ESTIMATORS

In this section we deduce the main asymptotic results, whose proofs are given in Appendix 1. We make the following assumptions:

$$(A5) \quad \xi_T = \sqrt{T} \frac{\partial Q_T}{\partial \beta} (y_T^1, x_T^1, \beta_0) - \sqrt{T} \frac{1}{H} \sum_{h=1}^H \frac{\partial Q_T}{\partial \beta} ([\tilde{y}_T^h(\theta_0, z_0^h)]_T^1, x_T^1, \beta_0) \quad (14)$$

is asymptotically normal with zero mean, with an asymptotic variance-covariance matrix given by

$$W = \lim_{T \rightarrow \infty} V(\xi_T)$$

and independent of the initial values z_0^h , $h = 1, \dots, H$:

$$(A6) \quad \lim_{T \rightarrow \infty} V \left\{ \sqrt{T} \frac{\partial Q_T}{\partial \beta} ([\tilde{y}_T^h(\theta_0, z_0^h)]_T^1, x_T^1, \beta_0) \right\} = I_0 \quad (15)$$

independent of z_0^h :

$$(A7) \quad \lim_{T \rightarrow \infty} \text{Cov} \left\{ \sqrt{T} \frac{\partial Q_T}{\partial \beta} ([\tilde{y}_T^h(\theta_0, z_0^h)]_T^1, x_T^1, \beta_0), \sqrt{T} \frac{\partial Q_T}{\partial \beta} ([\tilde{y}_T^l(\theta_0, z_0^l)]_T^1, x_T^1, \beta_0) \right\} \\ = K_0, \text{ independent of } z_0^h, z_0^l, \text{ for } h \neq l: \quad (16)$$

$$(A8) \quad \text{plim}_{T \rightarrow \infty} - \frac{\partial^2 Q_T}{\partial \beta \partial \beta'} [[\tilde{y}_T^h(\theta_0, z_0^h)]_T^1, x_T^1, \beta_0] = - \frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'} (F_0, G_0, \theta_0, \beta_0) = J_0 \quad (17)$$

independent of z_0^h .

Since $\sqrt{T} \frac{\partial Q_T}{\partial \beta} ([\tilde{y}_T^H(\theta_0, z^H)]_T^1, x_T^1, \beta_0)$ and $\sqrt{T} \frac{\partial Q_T}{\partial \beta} ([\tilde{y}_T^L(\theta_0, z^L)]_T^1, x_T^1, \beta_0)$

are independent conditionally to the exogenous variables, with the same asymptotic distribution, we deduce that

$$K_0 = \lim_{T \rightarrow \infty} \frac{1}{T} E \left[\sqrt{T} \frac{\partial Q_T}{\partial \beta'} [y_T^1, x_T^1, \beta_0] / x_T^1 \right] \quad (18)$$

$$I_0 - K_0 = \lim_{T \rightarrow \infty} \frac{1}{T} \left[\sqrt{T} \frac{\partial Q_T}{\partial \beta} [y_T^1, x_T^1, \beta_0] - E \left[\sqrt{T} \frac{\partial Q_T}{\partial \beta} [y_T^1, x_T^1, \beta_0] / x_T^1 \right] \right] \quad (19)$$

Proposition 3: Under assumptions (A1)–(A8) and the usual regularity conditions, the indirect estimator is asymptotically normal, when H is fixed and T goes to infinity:

$$\sqrt{T} (\hat{\theta}_T^H(\Omega) - \theta_0) \xrightarrow[T \rightarrow \infty]{d} N[0, W(H, \Omega)]$$

where

$$\begin{aligned} W(H, \Omega) &= \left(1 + \frac{1}{H} \right) \left(\frac{\partial b'}{\partial \theta} (F_0, G_0, \theta_0) \Omega \frac{\partial b}{\partial \theta'} (F_0, G_0, \theta_0) \right)^{-1} \\ &\quad \frac{\partial b'}{\partial \theta} (F_0, G_0, \theta_0) \Omega J_0^{-1} (I_0 - K_0) J_0^{-1} \Omega \frac{\partial b}{\partial \theta'} (F_0, G_0, \theta_0) \\ &\quad \left(\frac{\partial b'}{\partial \theta} (F_0, G_0, \theta_0) \Omega \frac{\partial b}{\partial \theta'} (F_0, G_0, \theta_0) \right)^{-1} \end{aligned}$$

The asymptotic variance–covariance matrix depends on the metric Ω and, as usual, there is an optimal choice of this matrix, i.e. a choice which minimizes $W(H, \Omega)$.

Proposition 4. The optimal choice of the Ω matrix is: $\Omega^* = J_0(I_0 - K_0)^{-1}J_0$ (assuming that $I_0 - K_0$ is invertible) and

$$W_H^* = W(H, \Omega^*) = \left(1 + \frac{1}{H} \right) \left(\frac{\partial b'}{\partial \theta} (F_0, G_0, \theta_0) J_0 (I_0 - K_0)^{-1} J_0 \frac{\partial b}{\partial \theta'} (F_0, G_0, \theta_0) \right)^{-1}$$

The optimal estimator thus obtained is denoted by $\hat{\theta}_T^H$.

This bound W_H^* has another equivalent form. Indeed, the binding function satisfies the first-order condition:

$$\frac{\partial Q_\infty}{\partial \beta} [F_0, G_0, \theta, b(F_0, G_0, \theta)] = 0, \quad \forall \theta \in \mathbb{H}$$

Therefore a second derivation with respect to θ gives

$$\frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} (F_0, G_0, \theta_0, \beta_0) + \frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'} (F_0, G_0, \theta_0, \beta_0) \frac{\partial b}{\partial \theta'} (F_0, G_0, \theta_0) = 0$$

and from (17):

$$\frac{\partial b}{\partial \theta'} (F_0, G_0, \theta_0) = J_0^{-1} \frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} (F_0, G_0, \theta_0, \beta_0)$$

This implies the following proposition.

Proposition 5. We have

$$W_H^* = \left(1 + \frac{1}{H}\right) \left(\frac{\partial^2 Q_\infty}{\partial \theta \partial \beta'} (F_0, G_0, \theta_0, \beta_0) (I_0 - K_0)^{-1} \frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} (F_0, G_0, \theta_0, \beta_0) \right)^{-1}$$

When the dimensions of β and θ are the same, we have $W(H, \Omega) = W_H^*$, $\forall \Omega$, and

$$W_H^* = \left(1 + \frac{1}{H}\right) \left(\frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} (F_0, G_0, \theta_0, \beta_0) \right)^{-1} (I_0 - K_0) \left(\frac{\partial^2 Q_\infty}{\partial \theta \partial \beta'} (F_0, G_0, \theta_0, \beta_0) \right)^{-1}$$

It is the asymptotic variance–covariance matrix of the estimator solution of

$$\hat{\beta}_T = \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h)$$

or of

$$\hat{\beta}_T = \tilde{\beta}_{HT}(\theta, z_0).$$

The number of replications appears in a multiplicative factor, since

$$W(H, \Omega) = \left(1 + \frac{1}{H}\right) W(\infty, \Omega)$$

and

$$W_H^* = \left(1 + \frac{1}{H}\right) W_\infty^* \quad (20)$$

where $W(\infty, \Omega)$ and W_∞^* are implicitly defined by the formulae in propositions 3 and 4. It is worth noting that these matrices are not, in general, the asymptotic variance–covariance matrices of the indirect estimators that would be based on the binding function.² It is only true when there is no exogenous variable (see Gouriéroux and Monfort, 1992).

The computation of the optimal indirect estimator $\hat{\theta}^H$ necessitates a preliminary consistent estimator of Ω^* . In some cases, such an estimator may be directly based on the observations and in others it has to be based on a first-stage consistent estimator of θ (for instance, $\hat{\theta}^H(Id)$ or on the value

$$\text{Argmax}_{\theta} Q_T \left(y_1^1, x_1^1, \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right)$$

(see Smith, 1993). These issues are discussed in Appendix 2, as well as the estimation of W_H^* .

The previous results are sufficiently general to cover a large number of estimation methods (for instance, the Simulated Generalized Method of Moments).

Let us consider a multidimensional function of the observable variables

$$k[y_t, \dots, y_{t-r}, x_t, \dots, x_{t-r}] \text{ (also denoted by } k_t)$$

the associated empirical moments:

$$\bar{k}_T = \frac{1}{T} \sum_{t=1}^T k[y_t, \dots, y_{t-r}, x_t, \dots, x_{t-r}]$$

² The asymptotic variance–covariance matrix of the estimator based on the true binding function (if we knew it) would be

$$W_H^{**} = \left(1 + \frac{1}{H}\right) \left[\frac{\partial b'}{\partial \theta} (F_0, G_0, \theta_0) J_0 I_0^{-1} J_0 \frac{\partial b}{\partial \theta'} (F_0, G_0, \theta_0) \right]$$

as if there were no observable exogenous variables ($J_0 - K_0$ becomes J_0). It is clear that the variance–covariance matrix W_H^{**} is generally greater than W_H^* . As explained in footnote 1, the accuracy of the indirect estimator is improved by the observation of the x_t 's (which have not to be simulated).

and the criterion:

$$Q_T = -\frac{1}{2}(\bar{k}_T - \beta)'(\bar{k}_T - \beta)$$

Then $\hat{\beta}_T = \bar{k}_T$, and an indirect estimator $\tilde{\theta}_T^H(\Omega)$ of θ is obtained by minimizing:

$$\left[\bar{k}_T - \frac{1}{TH} \sum_{t=1}^{TH} k(\tilde{y}_t(\theta, z_0), \dots, \tilde{y}_{t-r}(\theta, z_0), x_t, \dots, x_{t-r}) \right]' \\ \Omega \left[\bar{k}_T - \frac{1}{TH} \sum_{t=1}^{TH} k(\tilde{y}_t(\theta, z_0), \dots, \tilde{y}_{t-r}(\theta, z_0), x_t, \dots, x_{t-r}) \right]$$

with the usual convention for $x_\tau, \tau > T$.

This estimator generalizes the SMM estimator proposed by Duffie and Singleton (1993) and becomes identical when there is no exogenous variables. Moreover, we have

$$Q_\infty = -\frac{1}{2} \left(E_0 k - \beta \right)' \left(E_0 k - \beta \right), \quad b(F_0, G_0, \theta_0) = E_{\theta_0} k, \quad J_0 = Id$$

$$I_0 - K_0 = V_{as} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^T [k_t - E_0(k_t | \mathbf{x}_t^1)] \right\}$$

Therefore the asymptotic variance-covariance of $\sqrt{T}(\tilde{\theta}_T^H(\Omega) - \theta_0)$ is

$$W(H, \Omega) = \left(1 + \frac{1}{H} \right) \left[E_0 \left(\frac{\partial k'}{\partial \theta} \right) \Omega E_0 \left(\frac{\partial k}{\partial \theta'} \right) \right]^{-1} E_0 \left(\frac{\partial k'}{\partial \theta} \right) \\ \times \Omega (I_0 - K_0) \Omega E_0 \left(\frac{\partial k}{\partial \theta'} \right) \left[E_0 \left(\frac{\partial k'}{\partial \theta} \right) \Omega E_0 \left(\frac{\partial k}{\partial \theta'} \right) \right]^{-1}$$

The optimal choice of Ω is $\Omega^* = (I_0 - K_0)^{-1}$, and, in this case, $W(H, \Omega)$ becomes

$$W_H^* = \left(1 + \frac{1}{H} \right) \left[E_0 \left(\frac{\partial k'}{\partial \theta} \right) (I_0 - K_0)^{-1} E_0 \left(\frac{\partial k}{\partial \theta'} \right) \right]^{-1}$$

In the pure time series cases K_0 is equal to 0 and $I_0 - K_0$ reduces to

$$I_0 = V_{as} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T k_t \right]$$

which provides the Duffie and Singleton (1993) result.

4. SPECIFICATION TESTS

The indirect estimation model may be used in a modelling strategy, based on nested hypotheses or on the encompassing principle as shown in Gouriéroux-Monfort (1992), (see also Gouriéroux *et al.*, 1983; Mizon and Richard, 1986; Wooldridge, 1990; Hendry and Richard, 1990; Pesaran and Pesaran, 1991). This section is in the same spirit since it deals with the detection of possible misspecifications.

4.1. Global Specification Test

A specification test for model (1) may be based on the optimal value of the objective function used in the second step of the indirect estimation method. The following proposition is proved in Appendix 3.

Proposition 6: The statistics:

$$\xi_T = \frac{TH}{1+H} \min_{\theta \in \Theta} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right]' \hat{\Omega}_T^* \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right]$$

and

$$\tilde{\xi}_T = \frac{TH}{1+H} \min_{\theta \in \Theta} [\hat{\beta}_T - \tilde{\beta}_{HT}(\theta, z_0)]' \hat{\Omega}_T^* [\hat{\beta}_T - \tilde{\beta}_{HT}(\theta, z_0)]$$

where $\hat{\Omega}_T^*$ is a consistent estimator of Ω^* , are asymptotically distributed as a chi-square with $q - p$ degrees of freedom, where $q = \dim \beta$, $p = \dim \theta$, when the reduced-form equation (1) is well specified. Specification tests of asymptotic level α are associated with the critical regions:

$$\mathcal{W} = \{\xi_T > \chi^2_{1-\alpha}(q-p)\}, \quad \tilde{\mathcal{W}} = \{\tilde{\xi}_T > \chi^2_{1-\alpha}(q-p)\}$$

4.2. A Consistency Test for the Proxy Model

When the initial complex model is replaced by another, in which the parameters β have the same dimension as θ and similar interpretations, it is natural to see if the proxy model is really a good approximation of the initial one. For this purpose it is useful to compare:

- (1) The estimator $\hat{\beta}_T$ of β computed from the observations;
- (2) The estimator $\tilde{\beta}_{HT}(\hat{\beta}_T, z_0)$ computed from the simulations with the value $\hat{\beta}_T$ as the value of θ parameter.

Such an approach does not require the knowledge of the observations, but simply the forms of the two models and the knowledge of the estimation under the proxy model. It is a tool for checking *a posteriori* the accuracy of the approximation.

The implicit null hypothesis is defined by the constraint

$$\begin{aligned} b(F_0, G_0, \theta_0) &= b(F_0, G_0, b(F_0, G_0, \theta_0)) \\ \Leftrightarrow \theta_0 &= b(F_0, G_0, \theta_0) \text{ (because of the injectivity of } b) \\ \Leftrightarrow \hat{\beta}_T &\text{ consistent estimator of } \theta_0 \end{aligned}$$

This implicit null hypothesis appears as a constraint on θ_0 (see next section), where the function b defining the constraint has to be estimated. If, for instance, the auxiliary parameter is estimated by a maximum likelihood method on the proxy model, i.e. by a pseudo-maximum likelihood method, we are simply testing for the hypothesis of consistency of the PML estimator.

The asymptotic distribution of $\sqrt{T} [\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\beta}_T, z_0)]$ is derived in Appendix 3. Other specifications tests based on different criteria could be proposed (see Gouriéroux *et al.*, 1992).

5. INDIRECT TESTS OF HYPOTHESES ON THE PARAMETER OF INTEREST

The indirect estimation approach can be used to test hypotheses on parameter θ . We assume that the parameter θ is partitioned into

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

where θ_1 and θ_2 have dimensions p_1 and p_2 , respectively. We consider the null hypothesis $H_0 = (\theta_1 = 0)$.

Despite the use of simulated values the usual equivalence between the Wald test, the score

test, and the test based on the comparison of the constrained and unconstrained values of the objective function used in the second step remains valid, as shown in Appendix 4.

To define these tests we have to introduce the optimal unconstrained indirect estimator:

$$\begin{pmatrix} \hat{\theta}_{1T}^H \\ \hat{\theta}_{2T}^H \end{pmatrix} = (\hat{\theta}_T^H)$$

and the optimal constrained indirect estimator obtained by optimizing the criterion submitted to $\theta_1 = 0$. This estimator is denoted by

$$\begin{bmatrix} 0 \\ \hat{\theta}_{2T}^{0H} \end{bmatrix} = \hat{\theta}_T^{0H}$$

The Wald statistic is defined as

$$\xi_T^W = T(\hat{\theta}_{1T}^H)' \hat{W}_1^{*-1} (\hat{\theta}_{1T}^H) \quad (21)$$

where \hat{W}_1^* is a consistent estimator of the asymptotic covariance–variance matrix of $\sqrt{T} \hat{\theta}_{1T}^H$.

The score statistic is defined from the gradient of the objective function with respect to θ_1 evaluated at the unconstrained estimator. This gradient is given by:

$$\mathcal{D}_T = \frac{\partial \tilde{\beta}_{HT}}{\partial \theta_1} (\hat{\theta}_T^{0H}) \hat{\Omega}_T^* [\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\theta}_T^{0H})] \quad (22)$$

and the test statistic is

$$\xi_T^S = T \mathcal{D}_T' A W_1^* \mathcal{D} \quad (23)$$

Finally, we can introduce the difference between the optimal values of the objective function:

$$\begin{aligned} \xi_T^C &= \frac{TH}{1+H} \left[\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\theta}_T^{0H}) \right]' \hat{\Omega}_T^* [\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\theta}_T^{0H})] \\ &\quad - \frac{TH}{1+H} [\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\theta}_T^H)]' \hat{\Omega}_T^* [\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\theta}_T^H)] \end{aligned} \quad (24)$$

Proposition 7. The test statistics ξ_T^W , ξ_T^S , and ξ_T^C are asymptotically equivalent under the null hypothesis, and have the common distribution $\chi^2(p_1)$.

Proof. See Appendix 4.

As usual, the test based on the optimal values of the objective function requires both constrained and unconstrained estimation (and simulations). In contrast, the Wald or score test procedures only necessitate, respectively, unconstrained and constrained estimations (and simulations).

6. A SIMPLE EXAMPLE

In this section we consider the estimation of the parameter θ of a Gaussian moving-average process:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1} \quad t = 1, \dots, T$$

where the true value of θ is 0.5 and the true variance of ε_t is 1. The aim is to compare the

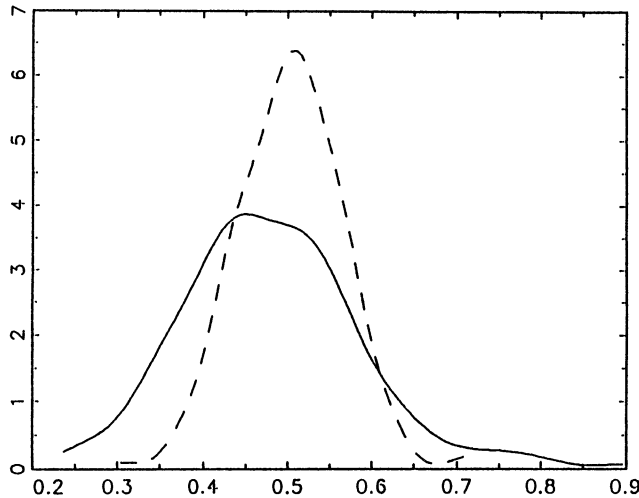


Figure 1. Estimated p.d.f. of the indirect estimator based on $Qf^{(1)}$ (——) and of the ML estimator (---)

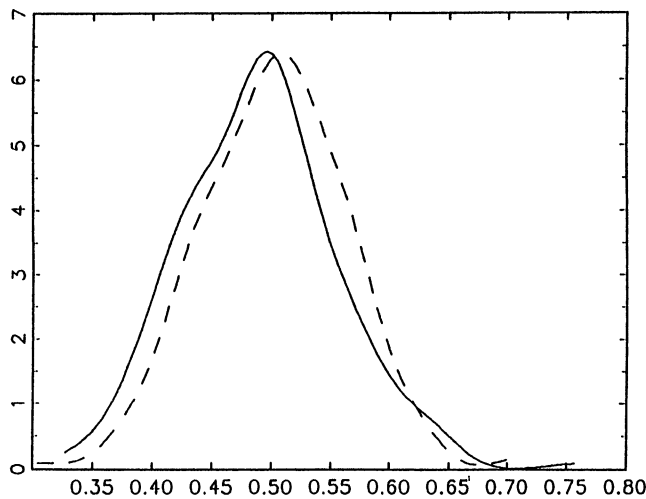


Figure 2. Estimated p.d.f. of the indirect estimator based on $Qf^{(2)}$ (——) and of the ML estimator (---)

properties of the exact maximum likelihood estimator of θ , based on the Kalman filter algorithm, with those of several indirect estimators of θ based on criteria of the form:

$$Qf^{(r)} = \sum_{t=r+1}^T (y_t - \beta_1 y_{t-1} \dots - \beta_r y_{t-r})^2$$

where r takes various values.

For each experiment, the number of observations is $T = 250$ and the number of replications is 200. In these experiments we only use one simulated path ($H = 1$). Moreover, since the

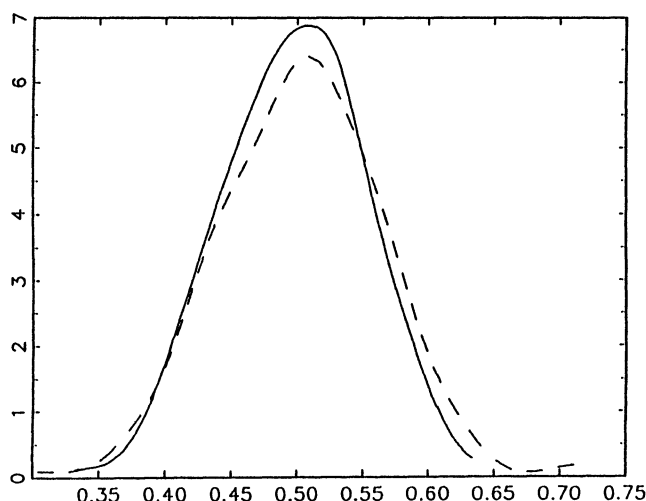


Figure 3. Estimated p.d.f. of the indirect estimator based on $Qf^{(3)}$ (——) and of the ML estimator (---)

Table I.

Estimator	Mean	Standard deviation	Root mean square error
Based on $Qf^{(1)}$	481	105	106
Based on $Qf^{(2)}$	491	065	066
Based on $Qf^{(3)}$	497	053	053
ML	504	061	061

efficiency gain obtained by using the optimal estimator is negligible, we only consider the estimators based on $\Omega = Id$.

In Figures 1–3 we show the estimated p.d.f. of the indirect estimators based on $Qf^{(1)}$, $Qf^{(2)}$, $Qf^{(3)}$, respectively, and the estimated p.d.f. of the ML estimator. Table I shows the empirical mean, the standard deviation, and the square root of the mean square error for each of these estimators. For all the methods the bias is small. For one lag ($r = 1$) in the criterion, the efficiency loss is relatively important. However, for $r = 2$ the efficiency is almost reached and for $r = 3$ it is completely reached and even overshoot in this particular example.

It is also worth noting that, in our experiment, the computation of the indirect estimator based on $Qf^{(3)}$ was about eighteen times faster than the computation of the ML estimator.

7. APPLICATIONS TO STOCHASTIC DIFFERENTIAL EQUATIONS

7.1. Description of the Method

Let us consider a diffusion process defined by

$$dy_t = g(\theta, y_t) dt + h(\theta, y_t) dw_t \quad (25)$$

where $\{w_t\}$ is a Brownian motion and the functions g, h satisfy some Lipschitz conditions to

ensure the stationarity of process $\{y_t\}$. For simplicity, we consider the case of a unidimensional process, but the results can be extended to the multidimensional case.

We assume that the process $\{y_t\}$ is observed at some discrete points y_1, \dots, y_T . It is generally impossible to deduce from equation (25) the form of the distribution of y_1, \dots, y_T , and a usual practice consists in a discretization of the differential equation and in the estimation of the parameters from this discretized version. Existing estimators generally are approximated ML estimators, whose properties are studied when the interval between successive observations tends to zero (Lipster and Shirayev, 1977, 1978; Ibragimov and Has'minskii, 1981; Genon and Catalot, 1990), or simulation based estimators (Duffie and Singleton, 1993). Two papers (Hansen and Scheinkman, 1991 and Aït-Sahalia, 1993) consider the case of fixed sampling period and a GMM approach based on the infinitesimal operator. The approach described in this section provides an alternative to these papers.

In fact the discretized version of the model may be seen as an approximated model whose likelihood function is used as the criterion of the indirect method. The second step requires simulations of the continuous process (25), i.e. discrete simulations with a very small time interval.

A large number of discretized versions of equations (25) exists. Indeed, if $k(y)$ is a twice-differentiable function, invertible with respect to y , we deduce from Ito's formula:

$$dk(y_t) = \left[\frac{\partial k}{\partial y}(y_t)g(\theta, y_t) + \frac{1}{2} \frac{\partial^2 k}{\partial y^2}(y_t)h^2(\theta, y_t) \right] dt + \frac{\partial k}{\partial y}(y_t)h(\theta, y_t) dw_t$$

The associated discretized version is

$$k(y_t) - k(y_{t-1}) = \frac{\partial k}{\partial y}(y_{t-1})g(\beta, y_{t-1}) + \frac{1}{2} \frac{\partial^2 k}{\partial y^2}(y_{t-1})h^2(\beta, y_{t-1}) + \frac{\partial k}{\partial y}(y_{t-1})h(\beta, y_{t-1})\varepsilon_t \quad (26)$$

where $\varepsilon_t \sim \text{IIN}(0, 1)$.

β is estimated by maximizing the associated log-likelihood or, equivalently:

$$\begin{aligned} Q_{T,k} = & \frac{1}{T} \sum_{t=2}^T -\text{Log } h(\beta, y_{t-1}) \\ & - \frac{1}{2T} \sum_{t=2}^T \frac{\{k(y_t) - k(y_{t-1}) - (\partial k / \partial y)(y_{t-1})g(\beta, y_{t-1}) - \frac{1}{2}(\partial^2 k / \partial y^2)(y_{t-1})h^2(\beta, y_{t-1})\}^2}{(\partial k^2 / \partial y)(y_{t-1})h^2(\beta, y_{t-1})} \end{aligned} \quad (27)$$

The second step of the indirect estimation method may be seen as a simple tool for eliminating the bias due to the discretization, i.e. the so-called convexity effect.

Let us now apply the indirect methods in two cases in which the exact likelihood function of y_1, \dots, y_T is known in order to evaluate the efficiency of the method. In both cases we shall also evaluate the importance of the implicit correction for the discretization which is performed by the indirect estimator, compared to the naive estimator based on the discretized model.³

7.2. Estimation of a Geometric Brownian Motion with Drift

Let us consider the geometric Brownian motion with drift:

$$dy_t = \mu y_t dt + \sigma y_t dw_t \quad (28)$$

³ We are grateful to Emmanuelle Clement for helpful assistance in these Monte Carlo experiments.

where $\{w_t\}$ is a Brownian motion of variance 1. We want to estimate $\theta = (\mu, \sigma)$ from discrete observations y_1, \dots, y_T .

In this particular case, there exists an exact discretization of equation (28), namely:

$$\text{Log } y_t = \text{Log } y_{t-1} + \mu - \frac{\sigma^2}{2} + \sigma \varepsilon_t \tag{29}$$

from which the maximum likelihood estimators of μ and σ are easily deduced and will be used

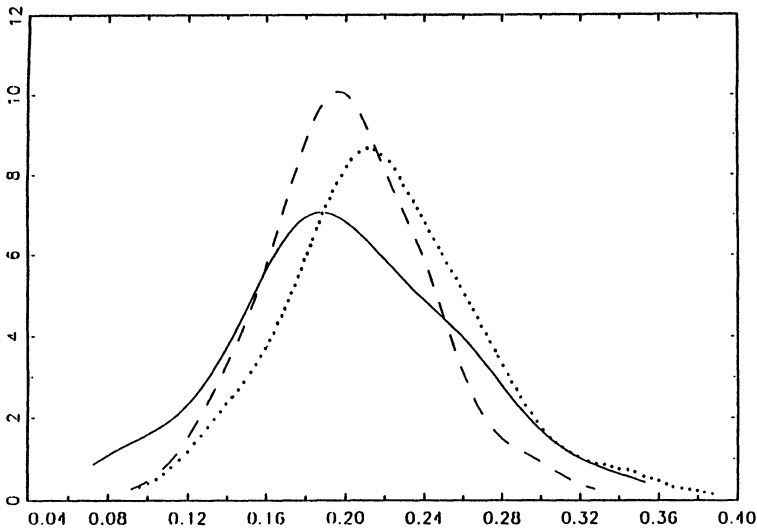


Figure 4. Geometric Brownian motion: estimation of μ . — Indirect estimator; --- ML estimator; naïve estimator

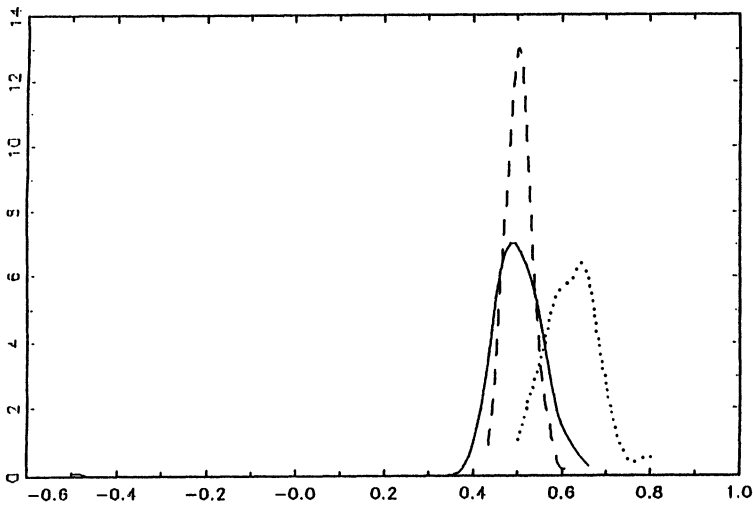


Figure 5. Geometric Brownian motion: estimation of σ . — Indirect estimator; --- ML estimator; naïve estimator

Table II. Gometric Brownian motion

Estimator		Mean	Bias	Standard deviation	Root mean square error
ML	μ	0.201	0.001	0.040	0.040
	σ	0.503	0.003	0.030	0.030
Indirect	μ	0.201	0.001	0.057	0.057
	σ	0.499	-0.001	0.087	0.087
Naive	μ	0.220	0.020	0.049	0.053
	σ	0.624	0.124	0.061	0.138

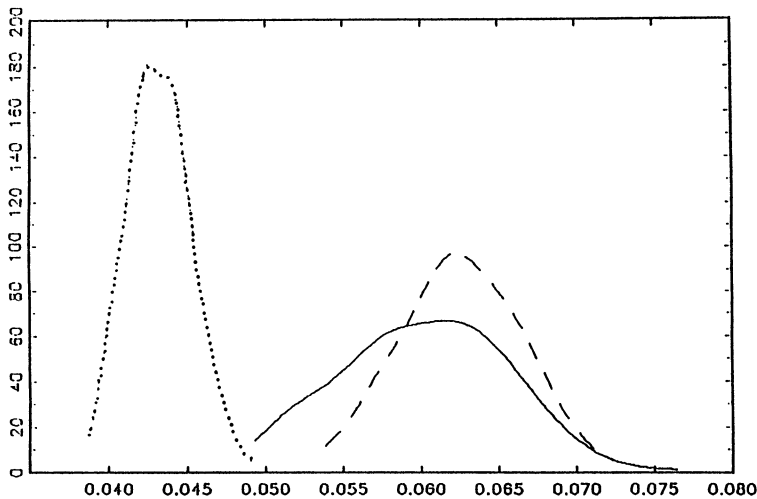


Figure 6. Ornstein-Uhlenbeck process: estimation of σ . — Indirect estimator; ---- ML estimator; naive estimator

as a reference for the evaluation of the other estimators. The other estimator considered is the 'naive' estimator based on the discretization of equation (28):

$$y_t = (1 + \mu)y_{t-1} + \sigma y_{t-1} \varepsilon_t \quad (30)$$

Finally, we consider a simple indirect estimator based on simulations of the naive discretization (30) with a time unit equal to $1/n$. The numerical values used are $\mu = 0.2$, $\sigma = 0.5$, $T = 150$, $y_0 = 10$, and, in the indirection estimation, we use $n = 10$ and $H = 1$; the number of replications is 200. The results are summarized in Figures 4 and 5, and Table II.

It is clear that the indirect estimator is much less biased than the naive estimator. In fact, the bias of the indirect estimator is comparable (or even smaller for σ) with that of the ML estimator. The root mean square error of the indirect estimator is also smaller than that of the naive estimator (for σ) or comparable (for μ).

7.3. Estimation of an Ornstein-Uhlenbeck Process

The model considered is

$$dy_t = k(a - y_t) dt + \sigma dw_t \quad (31)$$

and we want to estimate $\theta = (k, a, \sigma)$. Again, in this particular model, there exists an exact discretization:

$$y_t = a(1 - e^{-k}) + e^{-k}y_{t-1} + \sigma \left[\frac{(1 - e^{-2k})}{2k} \right]^{1/2} \varepsilon_t \quad (32)$$

We also consider the naive discretization:

$$y_t = (1 - k)y_{t-1} + ka + \sigma \varepsilon_t \quad (33)$$

and the naive and indirect estimators based on this discretization.

The numerical values used are $k = 0.8$, $a = 0.1$, $\sigma = 0.06$, $y_0 = 0.1$, $T = 250$, $n = 10$, $H = 1$, and the number of replications is 200. The results are given in Figures 6–8 and Table III.

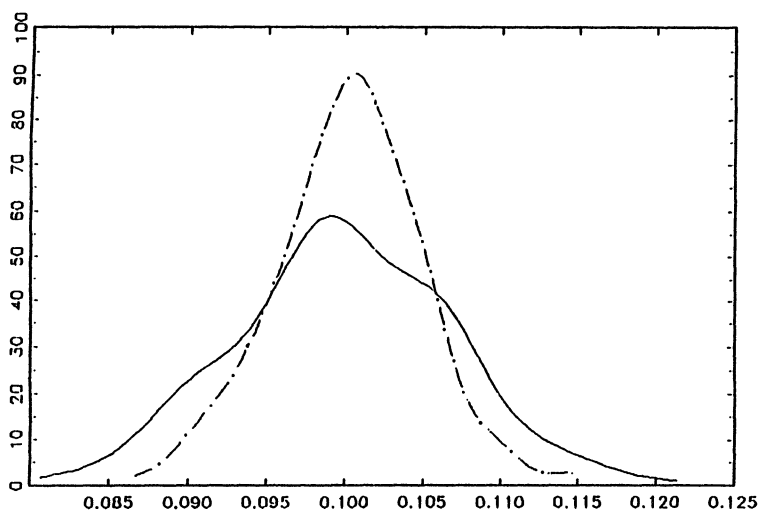


Figure 7. Ornstein-Uhlenbeck process: estimation of a . — Indirect estimator; ---- ML estimator; naive estimator

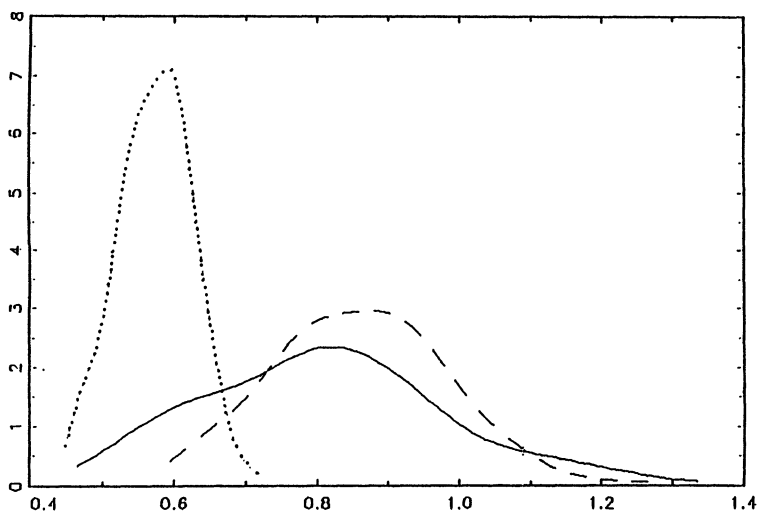


Figure 8. Ornstein-Uhlenbeck process: estimation of k . — Indirect estimator; ---- ML estimator; naive estimator

Table III. The Ornstein–Uhlenbeck process

Estimator		Mean	Bias	Standard deviation	Root mean square error
ML	k	0.859	0.059	0.122	0.135
	a	0.100	0.000	0.005	0.005
	σ	0.063	0.003	0.004	0.005
Indirect	k	0.811	0.011	0.170	0.170
	a	0.100	0.000	0.007	0.007
	σ	0.060	0.000	0.005	0.005
Naive	k	0.574	-0.226	0.051	0.232
	a	0.100	0.000	0.005	0.005
	σ	0.043	-0.017	0.002	0.017

It turns out that, by chance, the naive and the ML estimators of a are identical in this particular model. This implies that the indirect estimator of a , although perfectly unbiased as the other estimators, is slightly less efficient. As far as the estimation of k and σ is concerned, the indirect estimator is clearly much better, specially for the bias, than the naive estimator, and the efficiency loss, compared to the ML estimator, is not high even in this simple case where $H = 1$; moreover the naive estimator shows a misleading accuracy around a biased value.

8. APPLICATIONS TO FINANCIAL MODELS

8.1. Estimation of Stochastic Volatility Models

This kind of model appears in the literature in relation with the option-valuation problem and as the natural continuous time analogue of the ARCH processes (Nelson, 1990). In the basic model, the stock price is defined by

$$dy_t = \mu y_t dt + \sigma_t y_t dw_{1t} \quad (34)$$

where the logarithm of the volatility follows an Ornstein–Uhlenbeck process:

$$d\text{Log } \sigma_t = a(b - \text{Log } \sigma_t) dt + c dw_{2t} \quad (35)$$

and (w_{1t}) , (w_{2t}) are two correlated Brownian motions.

As in the previous section, it is possible to replace models (34) and (35) by a discrete time version. However, since only prices are observed and not the volatility, the discrete time model appears as a non-linear state space model, for which, in general, no simple filtering algorithm exists.

However, it may be possible to introduce slight modifications of this discrete time model in order to be able to apply the standard Kalman filter for the computation of a pseudo-likelihood function. Such an idea is described in Harvey *et al.* (1992) for the proxy model

$$\begin{cases} \text{Log } y_t = \text{Log } y_{t-1} + \sigma_t \varepsilon_{1t} \\ \text{Log } \sigma_t = d_0 + d_1 \text{Log } \sigma_{t-1} + \gamma \varepsilon_{2t} \end{cases} \quad (36)$$

where $(\varepsilon_{1t}, \varepsilon_{2t})$ is a Gaussian process with zero mean, unit variance, and an unknown correlation. This proxy model is not exactly a discrete version of equations (34) and (35) since it is implicitly assumed that the parameters μ , a , b , c are such that there is no drift on the logarithm of the price. It implies that the proxy model (36) has fewer parameters than the initial one (equations (34) and (35)), and the binding function is not injective. With such a

proxy model it is not possible to identify all the parameters by the indirect estimation approach. Nevertheless, if the proxy model is modified into

$$\begin{cases} \text{Log } y_t = \lambda + \text{Log } y_{t-1} + \sigma_t \varepsilon_{1t} \\ \text{Log } \sigma_t = d_0 + d_1 \text{Log } \sigma_{t-1} + \gamma \varepsilon_{2t} \end{cases} \quad (37)$$

and even if the Kalman filter does not apply to the above equation, we could use it inside a two-step procedure, as soon as $\{\sigma_t\}$ is stationary. Indeed, a consistent estimator of λ is

$$\hat{\lambda}_T = \frac{1}{T} \sum_{t=1}^T (\text{Log } y_t - \text{Log } y_{t-1})$$

and the Kalman filter may be applied to the variable $\text{Log}(\text{Log } y_t - \text{Log } y_{t-1} - \hat{\lambda}_T)^2$ instead of $\text{Log}(\text{Log } y_t / (y_{t-1}))^2$ in order to estimate the other auxiliary parameters d_0 , d_1 , γ .

Therefore it is possible to look for values of the parameters μ , a , b , c , for which these two-step estimators of the auxiliary parameters determined from the observations and from the simulations, respectively, are close together. Of course, the general results on the asymptotic variance-covariance matrix are not valid for this two-step estimation procedure of the auxiliary parameter, but they can be extended to this case.

8.2. Factor ARCH Models

Multivariate ARCH models naturally contain a large number of parameters, and it is necessary to introduce constraints to make this number smaller. A usual approach, compatible with the needs of financial theory, leads to the introduction of unobserved factors, which drive the whole dynamics of the system (Diebold and Nerlove, 1989; Engle *et al.*, 1989; King *et al.*, 1990; Gouriéroux *et al.*, 1991). Let us consider, for instance, a one-factor model of the Diebold–Nerlove type:

$$y_t = \lambda f_t + \varepsilon_t \quad (38)$$

where $\{y_t\}$ is the observable n -dimensional process, $\{\varepsilon_t\}$ is a Gaussian white noise with an unknown variance-covariance matrix Ω , $\{f_t\}$ is a unidimensional unobserved factor independent of $\{\varepsilon_t\}$, and λ gives the n sensitivity parameters of the components of y_t to the common factor f_t . We assume that the factor follows an ARCH(1) representation:

$$f_t / f_{t-1} \sim N[0, \alpha_0 + \alpha_1 f_{t-1}^2] \quad (\text{with the identifying restriction } \alpha_0 + \alpha_1 = 1) \quad (39)$$

The p.d.f. of $y_1, f_1, \dots, y_T, f_T$ conditional to f_0 is:

$$\begin{aligned} l(y_T, f_T, \dots, y_1, f_1 / f_0) &= \prod_{t=1}^T \\ &\times \left\{ \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det \Omega}} \exp -\frac{1}{2} (y_t - \lambda f_t)' \Omega^{-1} (y_t - \lambda f_t) \right. \\ &\times \left. \frac{1}{(2\pi)^{1/2}} \frac{1}{\sqrt{\alpha_0 + \alpha_1 f_{t-1}^2}} \exp -\frac{1}{2} \frac{f_t^2}{\alpha_0 + \alpha_1 f_{t-1}^2} \right\}. \end{aligned}$$

The likelihood function associated with the observable variables is

$$l(y_T, \dots, y_1 / f_0) = \int \dots \int l(y_T, f_T, \dots, y_1, f_1, f_0) \, df_T \dots df_1,$$

and has the form of an integral whose dimension is equal to the number of observations.

Use of the extended Kalman filter

The model defined by equations (38) and (39) can be put in a state-space form:

$$f_t = (\alpha_0 + \alpha_1 f_{t-1}^2)^{1/2} \eta_t \quad (\text{transition equation})$$

$$y_t = \lambda f_t + \varepsilon_t \quad (\text{measurement equation})$$

with $\eta_t \sim \text{IIN}(0, 1)$, independent of $\{\varepsilon_t\}$. These equations are non-linear with respect to the state f_t . However, the extended Kalman filter can be used and, in this case, this simply amounts to replacing f_{t-1}^2 in the transition equation by the square of the filtered value obtained at time $t-1$. If this extended filter is used to compute an approximate likelihood function, the estimators obtained are inconsistent, but this estimator can also be considered as the first-step estimator of the indirect estimation method, which leads to a consistent estimator.

State discretization

Another approach for approximating the exact log-likelihood function consists of approximating directly the transition equation. For this purpose it is interesting to replace the initial factor by a discretized factor (see Gouriéroux, 1992, Chapter 6). Let us consider a partition of the range of f_t into K given classes (a_k, a_{k+1}) , $k = 0, \dots, K-1$, with $a_0 = -\infty$, $a_k = +\infty$ and let us define the discretized factor:

$$\tilde{f}_t = b_k, \text{ if } f_t \in (a_k, a_{k+1}) \quad (40)$$

where b_k , $k = 0, \dots, K-1$, are given real numbers, such as the centres of the classes, except for the extreme ones. Then we get:

$$\begin{aligned} P[\tilde{f}_t = b_k | \tilde{f}_{t-1} = b_l] &= P(f_{t-1} \in (a_k, a_{k+1}) | f_{t-1} \in (a_l, a_{l+1})) \\ &\approx P(f_t \in (a_k, a_{k+1}) | f_{t-1} = b_l) \\ &= \Phi\left(\frac{a_{k+1}}{(\alpha_0 + \alpha_1 b_l^2)^{1/2}}\right) - \Phi\left(\frac{a_k}{(\alpha_0 + \alpha_1 b_l^2)^{1/2}}\right) \\ &= P_{kl}(\alpha_0, \alpha_1) \text{ (say).} \end{aligned}$$

The initial factor ARCH model can be replaced by the proxy model:

$$y_t = \lambda \tilde{f}_t + \varepsilon_t \quad (41)$$

where $\{\varepsilon_t\}$, $\{\tilde{f}_t\}$ are independent, $\{\varepsilon_t\} \sim \text{IIN}(0, \Omega)$, and $\{\tilde{f}_t\}$ is a qualitative Markov process with transition probabilities $P_{kl}(\alpha_0, \alpha_1)$.

In this auxiliary model a recursive evaluation of the likelihood function is available (see Hamilton, 1989). Moreover, using Kitagawa's (1987) smoothing formula, the EM algorithm provides explicit expressions both at the E and at the M stage for the ML estimation of the unrestricted parameters. Therefore these parameters can be chosen as the auxiliary parameters and the expression of P_{kl} above can provide relevant starting values for the estimators of these parameters based on simulations.

9. APPLICATION TO MICROECONOMETRICS

In this section we describe a potential application to discrete-choice models. When the independence of irrelevant alternatives (IIA) hypothesis is not satisfied and a probit formulation is used, the log-likelihood function contains integrals whose dimension is equal to the number of alternatives minus one (McFadden, 1976; Hausman and Wise, 1978). When

this dimension is small the integrals may be approximated by polynomial expansions: when it is reasonably large, they may be approximated by the Monte Carlo approach using, for instance, the simulated maximum likelihood method. The indirect estimation approach may also be used in the previous context for any dimension of the integrals. In this subsection we describe a logit approximation of the c.d.f. of a multivariate normal distribution. This approximation should be useful in various contexts and, in particular, for job-search models (see Gouriéroux *et al.*, 1992), or for the treatment of the serial correlation in limited dependent variables models (Robinson, 1982; Gouriéroux *et al.*, 1984).

Let Z be a Gaussian random vector, with zero mean and a variance-covariance matrix $Id + R$, where matrix R has a zero diagonal. We are interested in an approximation of the c.d.f.:

$$P[Z < z] = P(Z_1 < z_1, \dots, Z_K < z_K) \\ = \int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_K} \frac{1}{(2\pi)^{K/2}} \frac{1}{[\det(Id + R)]^{1/2}} \exp[-\frac{1}{2} z'(Id + R)^{-1} z] dz$$

To obtain such an approximation, we use the expansion of the probability around $R = 0$. A first-order expansion gives:

$$P[Z < z] \sim \int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_K} \frac{1}{(2\pi)^{K/2}} \exp[-\frac{1}{2} z'(Id - R)z] dz \\ \sim \int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_K} \prod_{k=1}^K \varphi(z_k) \left[1 + \sum_{l < k} \rho_{kl} z_k z_l \right] dz \\ = \left[\prod_{k=1}^K \Phi(z_k) \right] \left\{ 1 + \sum_{l < k} \rho_{kl} \frac{\varphi(z_k)}{\Phi(z_k)} \frac{\varphi(z_l)}{\Phi(z_l)} \right\}$$

where φ and Φ are the p.d.f. and the c.d.f. of the standard normal, and ρ_{kl} the correlation between Z_k and Z_l .

This approximation may be replaced by another, since:

$$P[Z < z] \sim \prod_{k=1}^K \Phi \left[z_k + \sum_{l < k} \rho_{kl} \frac{\varphi(z_l)}{\Phi(z_l)} \right] \quad (42)$$

Finally, we know that the logit distribution is close to the normal distribution $N[0, (\pi^2/3)]$. Therefore if $F(z) = 1/(1 + \exp(-z))$ is the c.d.f. of the logistic distribution, we have approximately:

$$P[Z < z] \sim \prod_{k=1}^K F \left(\frac{\pi}{\sqrt{3}} z_k + \sum_{l < k} \frac{\pi^2}{3} \rho_{kl} \left(1 - F \left(\frac{\pi}{\sqrt{3}} z_l \right) \right) \right) \quad (43)$$

The above two expansions have two advantages. They avoid the computation of the initial integral and they have a product form, which makes simpler the expression of the log-probability:

$$\text{Log } P[Z < z] \sim \sum_{k=1}^K \text{Log } F \left(\frac{\pi}{\sqrt{3}} z_k + \sum_{l < k} \frac{\pi^2}{3} \rho_{kl} \left(1 - F \left(\frac{\pi}{\sqrt{3}} z_l \right) \right) \right).$$

Approximation (43) seems better than (42) since it also defines a c.d.f. of a multivariate distribution (see e.g. Thelot, 1981, for $K = 2$). It is not the case in (42). Moreover, it is easily seen that the marginal distribution associated with approximation (43) are logistic distributions $F((\pi/\sqrt{3})z_k)$. In Gouriéroux *et al.* (1992) these results are used for job-search models.

A difficulty must be stressed in this kind of discrete-choice application. The simulated paths

$\tilde{y}_{TH}(\theta)$ are not continuous with respect to θ and, therefore, the classical asymptotic theory cannot be used. However, following Pakes and Pollard (1989), there is a way of building a generalized asymptotic theory.

10. APPLICATION TO MACROECONOMETRICS

Macroeconometrics is also a source of potential applications for indirect estimation methods. In this section we consider the practice of linearization in macroeconomic models without or with latent variables. Some other examples are related to dynamic optimization problems (the first application of indirect methods to such a problem is due to Smith, 1993).

10.1. Model without Latent Variables

We consider a dynamic model of the form

$$y_t = g(y_{t-1}, x_t, u_t, \theta) \quad (44)$$

where y and u have the same dimension. If g is a complicated non-linear function or if the u_t are correlated, equation (44) may be difficult to handle. It is often replaced by a linearized version with respect to u_t around zero:

$$y_t = g(y_{t-1}, x_t, 0, \beta) + \frac{\partial g}{\partial u'}(y_{t-1}, x_t, 0, \beta)u_t \quad (45)$$

or with respect to u_t and y_{t-1} , around zero and a long-run equilibrium value \bar{y} :

$$y_t = g(\bar{y}, x_t, 0, \gamma) + \frac{\partial g}{\partial y'}(\bar{y}, x_t, 0, \gamma)(y_{t-1} - \bar{y}) + \frac{\partial g}{\partial u'}(\bar{y}, x_t, 0, \gamma)u_t \quad (46)$$

In such a case the indirect approach corrects for the linearization bias. *A priori* there is no reason to expand around $u = 0$ rather than around another point \bar{u} :

$$y_t = g(y_{t-1}, x_t, \bar{u}, \beta) + \frac{\partial g}{\partial u}(y_{t-1}, x_t, \bar{u}, \beta)(u_t - \bar{u}) \quad (47)$$

For a given value \bar{u} we may maximize the approximated log-likelihood function (in the univariate case):

$$\begin{aligned} Q_{T, \bar{u}} = & \frac{1}{T} \sum_{t=1}^T -\text{Log} \left| \frac{\partial g}{\partial u}(y_{t-1}, x_t, \bar{u}, \beta) \right| \\ & - \frac{1}{2T} \sum_{t=1}^T \frac{(y_t - g(y_{t-1}, x_t, \bar{u}, \beta) + (\partial g / \partial u)(y_{t-1}, x_t, \bar{u}, \beta)\bar{u})^2}{(\partial^2 g / \partial u^2)(y_{t-1}, x_t, \bar{u}, \beta)} \end{aligned} \quad (48)$$

Thus we derive an indirect estimator whose asymptotic properties depend on the point \bar{u} . It may also be interesting to see if there exists a point \bar{u} for which the asymptotic variance-covariance matrix is minimal, since a part of the linearization bias can certainly be corrected by a suitable choice of the value \bar{u} .

10.2. Model with Latent Variables

Let us assume that the model can be put in the nonlinear state space form:

$$z_t = g_{1t}(z_{t-1}) + g_{2t}(z_{t-1})\eta_t \quad (49)$$

$$y_t = g_{3t}(z_t) + \varepsilon_t \quad (50)$$

where z_t is a state vector which is, in general (partially) unobserved, and $(\varepsilon'_t, \eta'_t)$ is a Gaussian white noise. In this case, the extended Kalman filter (Anderson and Moore, 1979) could be used to compute an approximate log-likelihood function. The estimator based on this approximate log-likelihood function would be inconsistent, but it could be used as a first-step estimator in the indirect estimation procedure, which, in its second step, gives a consistent and asymptotically normal estimator.

11. CONCLUSIONS

In this paper we have proposed general inference methods which should open the way to the treatment of complicated models for which the usual methods are intractable. Estimation methods, significance tests, and specification tests are based on simulations. Various potential applications are proposed and several Monte Carlo experiments show the feasibility and efficiency of the methods. These methods seem particularly promising when the criterion is based on approximations of the likelihood function, time discretizations, range discretizations, linearizations, etc. In this case the method is simpler (since the weighting matrix plays no role) and appears as an automatic correction for the asymptotic bias implied by the approximation. Moreover, the choice of the criterion function is natural and is likely to capture most features of the true distribution (more than a simulated moment type method) even in situations where direct simulated maximum likelihood methods are not available. It is clear, however, that there is room for further research in several directions: identification conditions, optimal choice of the criterion, use of different criteria to estimate different parts of the models, etc.

APPENDIX 1: ASYMPTOTIC DISTRIBUTION OF THE INDIRECT ESTIMATORS

The limit of the optimization problem:

$$\begin{aligned} \text{Min}_{\theta} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right]' \hat{\Omega}_T \left(\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right) \\ \text{Min}_{\theta} [b(\theta_0) - b(\theta)]' \Omega(b(\theta_0) - b(\theta)) \end{aligned}$$

where $b(\theta)$ is a notation for $b(F_0, G_0, \theta)$. The consistency of the indirect estimator of θ follows.

First-order Conditions for the Indirect Estimator

The first-order conditions corresponding to the optimization problem

$$\text{Min}_{\theta} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right]' \hat{\Omega}_T \left(\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta, z_0^h) \right)$$

are

$$\left[\frac{1}{H} \sum_{h=1}^H \frac{\partial \tilde{\beta}_T^h}{\partial \theta}(\tilde{\theta}_T^H, z_0^h) \right] \hat{\Omega}_T \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\tilde{\theta}_T^H, z_0^h) \right] = 0$$

The expansion of these first-order conditions gives

$$\begin{aligned}
& \left[\frac{1}{H} \sum_{h=1}^H \frac{\partial \tilde{\beta}_T^h}{\partial \theta} (\tilde{\theta}^H, z^h) \right] \\
& \times \hat{\Omega}_T \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta_0, z^h) - \frac{1}{H} \sum_{h=1}^H \frac{\partial \tilde{\beta}_T^h}{\partial \theta'} [\theta_0, z^h] [\tilde{\theta}^H - \theta_0] \right] \approx 0 \\
& \sqrt{T}(\tilde{\theta}^H - \theta_0) \approx \left[\frac{\partial b'}{\partial \theta} (F_0, G_0, \theta_0) \Omega \frac{\partial b}{\partial \theta'} (F_0, G_0, \theta_0) \right]^{-1} \\
& \times \frac{\partial b'}{\partial \theta} (F_0, G_0, \theta_0) \Omega \sqrt{T} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta_0, z^h) \right]
\end{aligned} \tag{51}$$

Expansions of $\hat{\beta}_T$ and $\tilde{\beta}_T^h(\theta_0, z^h)$

We get asymptotically

$$\begin{aligned}
& \frac{\partial Q_T}{\partial \beta} (\underline{y}_T^1, \underline{x}_T^1, \hat{\beta}_T) = 0 \\
& \sqrt{T} \frac{\partial Q_T}{\partial \beta} [\underline{y}_T^1, \underline{x}_T^1, \beta_0] + \frac{\partial^2 Q_T}{\partial \beta \partial \beta'} [\underline{y}_T^1, \underline{x}_T^1, \beta_0] \sqrt{T}(\hat{\beta}_T - \beta_0) \approx 0 \\
& \sqrt{T}(\hat{\beta}_T - \beta_0) \approx J_0^{-1} \sqrt{T} \frac{\partial Q_T}{\partial \beta} [\underline{y}_T^1, \underline{x}_T^1, \beta_0],
\end{aligned}$$

using (A8) and the fact that \underline{y}_T^1 may be seen as the ‘simulation of the nature’. Similarly, we get

$$\sqrt{T}(\tilde{\beta}_T^h(\theta_0, z^h) - \beta_0) \approx J_0^{-1} \sqrt{T} \frac{\partial Q_T}{\partial \beta} [(\tilde{y}^h(\theta_0, z^h))_T^1, \underline{x}_T^1, \beta_0]$$

Asymptotic Distribution of

$$\sqrt{T} \left(\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta_0, z^h) \right)$$

We have

$$\begin{aligned}
\sqrt{T} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta_0, z^h) \right] & \approx J_0^{-1} \left\{ \sqrt{T} \frac{\partial Q_T}{\partial \beta} \{ \underline{y}_T^1, \underline{x}_T^1, \beta_0 \} \right. \\
& \left. - \sqrt{T} \frac{1}{H} \sum_{h=1}^H \frac{\partial Q_T}{\partial \beta} ((\tilde{y}^h(\theta_0, z^h))_T^1, \underline{x}_T^1, \beta_0) \right\}.
\end{aligned}$$

Therefore under (A1)–(A8) this difference is asymptotically normal with zero mean and a variance–covariance matrix given by:

$$\begin{aligned}
V_{as} \sqrt{T} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^h(\theta_0, z^h) \right] & = J_0^{-1} \left[\left(1 + \frac{1}{H} \right) I_0 - 2K_0 + \frac{H(H-1)}{H^2} K_0 \right] J_0^{-1} \\
& = \left(1 + \frac{1}{H} \right) J_0^{-1} (I_0 - K_0) J_0^{-1} \left(1 + \frac{1}{H} \right) \Omega^{*-1}
\end{aligned}$$

Proposition 3 is a consequence of this form of the asymptotic distribution and of expansion of $\sqrt{T}(\tilde{\theta}^H - \theta_0)$ above.

The Indirect Estimator based on $\tilde{\beta}_{HT}(\theta, z_0)$

The proof is similar, since

$$\sqrt{T}(\tilde{\beta}_{HT}(\theta_0, z_0) - \beta_0) \simeq J_0^{-1} \sum_{h=1}^H \frac{\sqrt{T}}{H} \frac{\partial Q_T}{\partial \beta} [[\tilde{y}^h(\theta_0, z_0^h)]_1^1, \mathbf{x}_1^1, \beta_0]$$

where

$$\begin{aligned} z_0^1 &= [y_0, \tilde{u}_0] \\ z_0^h &= \{\tilde{y}_{TH}(\theta_0, z_0^1), \tilde{u}_{TH}\} \\ y_0^1 &= y_0, y_0^h = \tilde{y}_{TH}[\theta_0, z_0^h] \end{aligned}$$

The second version of the indirect estimator of θ discussed in proposition 2 has the same asymptotic properties as the first.

The Third Version of the Indirect Estimator

This is given by

$$\tilde{\theta}_T^H(\Sigma) = \underset{\theta}{\text{Argmin}} \frac{\partial Q_T}{\partial \beta'} [\tilde{y}^1_{TH}(\theta), \tilde{\mathbf{x}}^1_{TH}, \hat{\beta}_T] \Sigma \frac{\partial Q_T}{\partial \beta} [\tilde{y}^1_{TH}(\theta), \tilde{\mathbf{x}}^1_{TH}, \hat{\beta}_T]$$

The limit of the minimization problem is

$$\underset{\theta}{\text{Min}} \frac{\partial Q_\infty}{\partial \beta'} (\theta, b(\theta_0)) \Sigma \frac{\partial Q_\infty}{\partial \beta} (\theta, b(\theta_0))$$

where $b(\theta_0)$ is a notation for $b(F_0, G_0, \theta_0)$. Since $(\partial Q_\infty / \partial \beta)(\theta_0, b(\theta_0)) = 0$ the consistency of the estimator follows.

The first-order conditions are

$$\frac{\partial^2 Q_T}{\partial \beta \partial \beta'} [\tilde{y}^1_{TH}(\tilde{\theta}_T^H), \hat{\beta}_T] \Sigma \frac{\partial Q_T}{\partial \beta} [\tilde{y}^1_{TH}(\tilde{\theta}_T^H), \hat{\beta}_T] = 0$$

An expansion of these conditions gives

$$\begin{aligned} \frac{\partial^2 Q_\infty}{\partial \theta \partial \beta'} \Sigma \left[\frac{\partial Q_T}{\partial \beta} (\tilde{y}^1_{TH}(\theta_0), \beta_0) + \frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} (\tilde{\theta}_T^H - \theta_0) + \frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'} (\hat{\beta}_T - \beta_0) \right] &= 0 \\ \sqrt{T}(\tilde{\theta}_T^H - \theta_0) &\simeq - \left[\frac{\partial^2 Q_\infty}{\partial \theta \partial \beta'} \Sigma \frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} \right]^{-1} \frac{\partial^2 Q_T}{\partial \theta \partial \beta'} \\ &\times \Sigma \left[\sqrt{T} \frac{\partial Q_T}{\partial \beta} (\theta_0, \beta_0) + \frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'} \sqrt{T}(\hat{\beta}_T - \beta_0) \right] \end{aligned}$$

where $(\partial Q_T / \partial \beta)(\theta_0, \beta_0)$ is a notation for $(\partial Q_T / \partial \beta)(\tilde{y}^1_{TH}(\theta_0), \beta_0)$.

Moreover, we have

$$0 = \frac{\partial Q_T}{\partial \beta} (\theta_0, \tilde{\beta}_{TH}(\theta_0)) \simeq \frac{\partial Q_T}{\partial \beta} (\theta_0, \beta_0) + \frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'} [\tilde{\beta}_{TH}(\theta_0) - \beta_0]$$

Using this formula we get

$$\begin{aligned}\sqrt{T}(\tilde{\theta}_T^H - \theta_0) &\approx - \left[\frac{\partial^2 Q_\infty}{\partial \theta \partial \beta'} \Sigma \frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} \right]^{-1} \frac{\partial^2 Q_\infty}{\partial \theta \partial \beta'} \\ &\quad \times \Sigma \left[-\sqrt{T} \frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'} [\tilde{\beta}_{TH}(\theta_0) - \beta_0] + \sqrt{T} \frac{\partial^2 Q_\infty}{\partial \theta \partial \beta'} (\hat{\beta}_T - \beta_0) \right].\end{aligned}$$

or, since $(\partial b / \partial \theta') = J_0^{-1} [(\partial^2 Q_\infty) / (\partial \beta \partial \theta')]$ and $J_0 = - [(\partial^2 Q_\infty) / \partial \beta \partial \beta']$

$$\sqrt{T}(\tilde{\theta}_T^H - \theta_0) = \left[\frac{\partial b'}{\partial \theta} J_0 \Sigma J_0 \frac{\partial b}{\partial \theta'} \right]^{-1} \frac{\partial b'}{\partial \theta} J_0 \Sigma J_0 \sqrt{T} [\hat{\beta}_T - \tilde{\beta}_{TH}(\theta_0)]$$

Comparing with the expression of $\sqrt{T}(\tilde{\theta}_T^H - \theta_0)$ given above, we see immediately that these estimators are asymptotically equivalent if $\Omega = J_0 \Sigma J_0$.

The Estimator Based on the Constrained Maximization of the Criterion

This estimator θ_T^{*H} is obtained by the following maximization:

$$\text{Max } Q_T[\mathbf{y}_T^1, \mathbf{x}_T^1, \tilde{\beta}_{TH}(\theta)]$$

The limit problem is

$$\text{Max}_\theta Q_\infty(\theta_0, b(\theta))$$

whose maximum is reached for $\theta = \theta_0$ and the consistency of the estimator follows. The first-order conditions are

$$\frac{\partial Q_T}{\partial \theta} [\mathbf{y}_T^1, \mathbf{x}_T^1, \tilde{\beta}_{HT}(\theta_T^{*H})] = 0$$

or

$$\sqrt{T} \frac{\partial \tilde{\beta}_{TH}(\theta_T^{*H})}{\partial \theta} \frac{\partial Q_T}{\partial \beta} [\mathbf{y}_T^1, \mathbf{x}_T^1, \tilde{\beta}_{HT}(\theta_T^{*H})] = 0.$$

A first-order expansion gives, using $(\partial Q_\infty / \partial \beta)(\theta_0, b(\theta_0)) = 0$,

$$0 \approx \sqrt{T} \frac{\partial \tilde{\beta}_{TH}(\theta_0)}{\partial \theta} \frac{\partial Q_T}{\partial \beta} [\mathbf{y}_T^1, \mathbf{x}_T^1, \tilde{\beta}_{HT}(\theta_0)] + \frac{\partial b'(\theta_0)}{\partial \theta} \frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} (\theta_0, b(\theta_0)) \sqrt{T}(\theta_T^{*H} - \theta_0)$$

Using again $(\partial Q_\infty / \partial \beta)(\theta_0, b(\theta_0)) = 0$ and $(\partial^2 Q_\infty / \partial \beta \partial \theta') = J_0(\partial b / \partial \theta')$, we get

$$\begin{aligned}\sqrt{T}(\theta_T^{*H} - \theta_0) &\approx - \left[\frac{\partial b'}{\partial \theta} J_0 \frac{\partial b}{\partial \theta'} \right]^{-1} \frac{\partial b'}{\partial \theta} \sqrt{T} \frac{\partial Q_T}{\partial \beta} [\mathbf{y}_T^1, \mathbf{x}_T^1, \tilde{\beta}_{TH}(\theta_0)] \\ &\approx - \left[\frac{\partial b'}{\partial \theta} J_0 \frac{\partial b}{\partial \theta'} \right]^{-1} \frac{\partial b'}{\partial \theta} \sqrt{T} \left[\frac{\partial Q_T}{\partial \beta} [\mathbf{y}_T^1, \mathbf{x}_T^1, \beta_0] - J_0 [\tilde{\beta}_{TH}(\theta_0) - \beta_0] \right] \\ &\approx - \left[\frac{\partial b'}{\partial \theta} J_0 \frac{\partial b}{\partial \theta'} \right]^{-1} \frac{\partial b'}{\partial \theta} J_0 \sqrt{T} [\hat{\beta}_T - \tilde{\beta}_{TH}[\theta_0]]\end{aligned}$$

Therefore $\sqrt{T}(\theta_T^{*H} - \theta_0)$ has the same asymptotic distribution as $\sqrt{T}[\tilde{\theta}_T^H(J_0) - \theta_0]$ and is asymptotically less efficient than the optimal indirect estimator whose asymptotic distribution is that of $\sqrt{T}[\tilde{\theta}_T^H[J_0(I_0 - K_0)^{-1}J_0] - \theta_0]$.

APPENDIX 2: ESTIMATION OF Ω^* AND W_H^* **Estimation of Ω^* and Optimal Estimation of θ**

Ω^* is equal to $J_0(I_0 - K_0)^{-1}J_0$.

From (17) we that J_0 can be consistently estimated by

$$-\frac{\partial^2 Q_T}{\partial \beta \partial \beta'}(y_1^1, x_1^1, \hat{\beta}_T)$$

As far as the estimation of $I_0 - K_0$ is concerned, let us first consider the particular case where there is no exogenous variable and where Q_T is equal to

$$\frac{1}{T} \sum_{t=1}^T \psi(y_t, \dots, y_{t-r}, \beta) = \frac{1}{T} \sum_{t=1}^T \psi_t(\beta)$$

In this case $K_0 = 0$ and I_0 is equal to

$$\lim_{T \rightarrow \infty} V \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial \psi}{\partial \beta}(y_t, \dots, y_{t-r}, \beta) \right]$$

Therefore I_0 can be approximated (see Newey and West, 1987) by

$$\hat{\Gamma}_0 + \sum_{k=1}^K \left(1 - \frac{k}{K+1} \right) (\hat{\Gamma}_k + \hat{\Gamma}_k')$$

with

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^T \frac{\partial \psi_{t-k}}{\partial \beta}(\hat{\beta}_T) \frac{\partial \psi_t}{\partial \beta'}(\hat{\beta}_T)$$

In the general case, the estimation of $I_0 - K_0$ is slightly more complex and it is based on the following:

$$\begin{aligned} I_0 - K_0 &= \lim_{T \rightarrow \infty} T V_0 \left[\frac{\partial Q_T}{\partial \beta} - E_0 \left(\frac{\partial Q_T}{\partial \beta} \middle| x_1^1 \right) \right] \\ &= \lim_{T \rightarrow \infty} T E_0 V_0 \left(\frac{\partial Q_T}{\partial \beta} \middle| x_1^1 \right) \\ &= \lim_{T \rightarrow \infty} T V_0 \left(\frac{\partial Q_T}{\partial \beta} \middle| x_1^1 \right) \end{aligned}$$

This implies that a consistent estimator of $I_0 - K_0$ is

$$\frac{T}{S} \sum_{s=1}^S (W_s - \bar{W})(W_s - \bar{W})'$$

with

$$W_s = \frac{\partial Q_T}{\partial \beta} [(y^s(\tilde{\theta}))_T^1, x_1^1, \hat{\beta}_T]$$

$$\bar{W} = \frac{1}{S} \sum_{s=1}^S W_s$$

and $\tilde{\theta}$ is any consistent estimator of θ (for instance $\tilde{\theta}_T^H(Id)$). It is also worth noting that, once

a consistent estimator $\hat{\Omega}_T^*$ of Ω^* is available, an optimal estimator of θ is readily obtained from the linearized minimum distance estimator:

$$\text{Min}_{\theta} \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^H(\tilde{\theta}, z_0^h) - \frac{\partial b}{\partial \theta'}(\tilde{\theta})(\theta - \tilde{\theta}) \right]' \hat{\Omega}_T \left[\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^H \tilde{\beta}_T^H(\tilde{\theta}, z_0^h) - \frac{\partial b}{\partial \theta'}(\tilde{\theta})(\theta - \tilde{\theta}) \right]$$

which is just a GLS estimator; the derivative $(\partial b / \partial \theta')(\tilde{\theta})$ can be computed numerically by replacing the function $b(\cdot)$ by $\tilde{\beta}_{ST}(\cdot, z_0)$.

Estimator of W_H^*

Since

$$W_H^* = \left(1 + \frac{1}{H} \right) \left[\frac{\partial^2 Q_{\infty}}{\partial \theta \partial \beta'} (I_0 - K_0)^{-1} \frac{\partial^2 Q_{\infty}}{\partial \beta \partial \theta'} \right]^{-1}$$

a consistent estimator of W_H^* is obtained as soon as we have a consistent estimator of $(\partial^2 Q_{\infty} / \partial \theta \partial \beta')$. Such an estimator can be obtained by a numerical derivation of $(\partial Q_T / \partial \beta')[(y^s(\theta))_T^1, x_T^1, \hat{\beta}_T]$ with respect to θ evaluated at $\hat{\theta}_T^H$, where $(y^s(\theta))_T^1$ is a simulated path of y based on the parameter θ .

APPENDIX 3: SPECIFICATION TESTS

Global Specification Test

The optimal value of the objective function is (for instance, for ξ_T)

$$\xi_T = T[\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\theta}_T^H)]' \hat{\Omega}_T^* [\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\theta}_T^H)]$$

where $\hat{\theta}_T^H$ is the optimal indirect estimator. A first-order expansion gives

$$\begin{aligned} \xi_T &\approx T \left[\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0) - \frac{\partial b}{\partial \theta'}(\theta_0)(\hat{\theta}_T^H - \theta_0) \right]' \\ &\quad \Omega^* \left[\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0) - \frac{\partial b}{\partial \theta'}(\theta_0)(\hat{\theta}_T^H - \theta_0) \right] \end{aligned}$$

By using the asymptotic expansion of $\sqrt{T}(\hat{\theta}_T^H - \theta_0)$ given in Appendix 1 we get

$$\frac{\partial b}{\partial \theta'}(\theta_0) \sqrt{T}(\hat{\theta}_T^H - \theta_0) \approx \frac{\partial b}{\partial \theta'}(\theta_0) \left[\frac{\partial b'}{\partial \theta}(\theta_0) \Omega^* \frac{\partial b}{\partial \theta'}(\theta_0) \right]^{-1} \frac{\partial b'}{\partial \theta}(\theta_0) \Omega^* \sqrt{T}(\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0))$$

and therefore,

$$\sqrt{T}(\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0)) - \sqrt{T} \frac{\partial b}{\partial \theta'}(\theta_0)(\hat{\theta}_T^H - \theta_0) \approx (Id - M) \sqrt{T}(\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0))$$

where

$$M = \frac{\partial b}{\partial \theta'}(\theta_0) \left[\frac{\partial b'}{\partial \theta}(\theta_0) \Omega^* \frac{\partial b}{\partial \theta'}(\theta_0) \right]^{-1} \frac{\partial b'}{\partial \theta}(\theta_0) \Omega^*$$

is the orthogonal projector on the space spanned by the columns of $(\partial b / \partial \theta')(\theta_0)$ for the inner product Ω^* .

From Appendix 1 we know that

$$\sqrt{T} [\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0)] \xrightarrow{d} N\left(0, \left(1 + \frac{1}{H}\right) (\Omega^*)^{-1}\right)$$

Therefore a usual argument gives

$$\frac{TH}{1+H} (\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0))' (Id - M)' \Omega^* (Id - M) [\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0)] \rightarrow \chi^2(q-p)$$

since $(\partial b / \partial \theta')(\theta_0)$ is of full-column rank.

The Consistency Test

We have under the null hypothesis $\theta_0 = b[F_0, G_0, \theta_0]$

$$\begin{aligned} \sqrt{T} [\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\beta}_T)] &\approx \sqrt{T} \left(\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0) - \frac{\partial b}{\partial \theta'}(\theta_0) (\hat{\beta}_T - \theta_0) \right) \\ &\approx \sqrt{T} \left\{ \left[Id - \frac{\partial b}{\partial \theta'}(\theta_0) \right] (\hat{\beta}_T - \theta_0) - (\tilde{\beta}_{HT}(\theta_0) - \theta_0) \right\} \\ &\approx \left[Id - \frac{\partial b}{\partial \theta'}(\theta_0) \right] J_0^{-1} \sqrt{T} \frac{\partial Q_T}{\partial \beta}(\mathbf{y}_1^1, \mathbf{x}_1^1, \theta_0) \\ &\quad - J_0^{-1} \frac{1}{H} \sum_{h=1}^H \sqrt{T} \frac{\partial Q_T}{\partial \beta}[(\tilde{\mathbf{y}}^h)^1, \mathbf{x}_1^1, \theta_0] \end{aligned}$$

Therefore we deduce that $\sqrt{T} [\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\beta}_T)]$ is asymptotically normal, with zero mean, and an asymptotic variance-covariance matrix given by

$$\begin{aligned} \Sigma &= \left[Id - \frac{\partial b}{\partial \theta'}(\theta_0) \right] J_0^{-1} I_0 J_0^{-1} \left[Id - \frac{\partial b'}{\partial \theta}(\theta_0) \right] + \frac{1}{H} J_0^{-1} I_0 J_0^{-1} - \left[Id - \frac{\partial b}{\partial \theta'}(\theta_0) \right] J_0^{-1} K_0 J_0^{-1} \\ &\quad - J_0^{-1} K_0 J_0^{-1} \left(Id - \frac{\partial b'}{\partial \theta}(\theta_0) \right) + \left(1 - \frac{1}{H} \right) J_0^{-1} K_0 J_0^{-1} \end{aligned}$$

APPENDIX 4: THE TEST OF A NULL HYPOTHESIS ON THE PARAMETER OF INTEREST

This parameter θ is partitioned into two subvectors

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

By convention, the null hypothesis is defined by $H_0 = \{\theta_1 = 0\}$.

First-order Expansion of the Unconstrained Indirect Estimator

This expansion has been given in Appendix 1 and it may be rewritten, under the null

hypothesis:

$$\sqrt{T} \begin{bmatrix} \hat{\theta}_{1T}^H \\ \hat{\theta}_{2T}^H - \theta_{20} \end{bmatrix} \sim \left[\begin{pmatrix} \frac{\partial b'}{\partial \theta_1} \\ \frac{\partial b'}{\partial \theta_2} \end{pmatrix} \Omega^* \begin{pmatrix} \frac{\partial b}{\partial \theta_1} & \frac{\partial b}{\partial \theta_2} \end{pmatrix} \right]^{-1} \\ \begin{pmatrix} \frac{\partial b'}{\partial \theta_1} \\ \frac{\partial b'}{\partial \theta_2} \end{pmatrix} \Omega^* J_0^{-1} \left[\sqrt{T} \frac{\partial Q_T}{\partial \beta} - \sqrt{T} \frac{1}{H} \sum_{h=1}^H \frac{\partial Q_T^h}{\partial \beta} \right]$$

We deduce the expansion of the first subvector:

$$\sqrt{T} \hat{\theta}_{1T}^H \simeq [A_{11} - A_{12} A_{22}^{-1} A_{21}]^{-1} \left(\frac{\partial b'}{\partial \theta_1} - A_{12} A_{22}^{-1} \frac{\partial b'}{\partial \theta_2} \right) \Omega^* \sqrt{T} (\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0, y_0))$$

where

$$A_{ij} = \frac{\partial b'}{\partial \theta_i} \Omega^* \frac{\partial b}{\partial \theta_j}$$

We note that

$$\frac{\partial b'}{\partial \theta_1} - A_{12} A_{22}^{-1} \frac{\partial b'}{\partial \theta_2} = \frac{\partial b'}{\partial \theta_1} [Id - M_2]'$$

where

$$M_2 = \frac{\partial b}{\partial \theta_2} \left[\frac{\partial b'}{\partial \theta_2} \Omega^* \frac{\partial b}{\partial \theta_2} \right]^{-1} \frac{\partial b'}{\partial \theta_2} \Omega^*$$

is the orthogonal projector on the columns of $(\partial b / \partial \theta_2)$ for the inner product Ω^* . Therefore the Wald statistic is asymptotically equivalent to

$$\xi_T^W \simeq \frac{TH}{1+H} T(\hat{\beta}_T - \tilde{\beta}_{HT})' \Omega^* (Id - M_2) \frac{\partial b}{\partial \theta_1}$$

$$\left\{ \frac{\partial b'}{\partial \theta_1} (Id - M_2)' \Omega^* (Id - M_2) \frac{\partial b}{\partial \theta_1} \right\}^{-1} \frac{\partial b'}{\partial \theta_1} (Id - M_2)' \Omega^* (\hat{\beta}_T - \tilde{\beta}_{HT})$$

since the asymptotic variance-covariance matrix of $\sqrt{T}(\hat{\beta}_T - \tilde{\beta}_{HT})$ is $(1 + (1/H))(\Omega^*)^{-1}$.

First-order Expansion of the Constrained Estimator of θ_2

We get

$$\sqrt{T}(\hat{\theta}_{2T}^H - \theta_{20}) \\ \simeq \left(\frac{\partial b'}{\partial \theta_2} \Omega^* \frac{\partial b}{\partial \theta_2} \right)^{-1} \frac{\partial b'}{\partial \theta_2} \Omega^* \sqrt{T}(\hat{\beta}_T - \tilde{\beta}_{HT}) \\ \simeq A_{22}^{-1} \frac{\partial b'}{\partial \theta_2} \Omega^* \sqrt{T}(\hat{\beta}_T - \tilde{\beta}_{HT})$$

Difference of the Optimal Values of the Unconstrained and Constrained Objective Function

From Appendix 3 we deduce that the difference between the two optimal values of the objective function (the constrained minus the unconstrained one) is such that.

$$\begin{aligned}\xi_T^C &\approx \frac{TH}{1+H} (\hat{\beta}_T - \tilde{\beta}_{HT})' (Id - M_2)' \Omega^* (Id - M_2) (\hat{\beta}_T - \tilde{\beta}_{HT}) \\ &\quad - \frac{TH}{1+H} (\hat{\beta}_T - \tilde{\beta}_{HT})' (Id - M)' \Omega^* (Id - M) (\hat{\beta}_T - \tilde{\beta}_{HT})\end{aligned}$$

where

$$\begin{aligned}M_2 &= \frac{\partial b}{\partial \theta_2'} A_{22}^{-1} \frac{\partial b'}{\partial \theta_2} \Omega^* \\ M &= \frac{\partial b}{\partial \theta'} \left[\frac{\partial b}{\partial \theta'} \Omega^* \frac{\partial b'}{\partial \theta'} \right]^{-1} \frac{\partial b'}{\partial \theta} \Omega^* \\ \xi_T^C &\approx \frac{TH}{1+H} (\hat{\beta}_T - \tilde{\beta}_{HT})' \Omega^* (Id - M_2) (\hat{\beta}_T - \tilde{\beta}_{HT}) \\ &\quad - \frac{TH}{1+H} (\hat{\beta}_T - \tilde{\beta}_{HT})' \Omega^* (Id - M) (\hat{\beta}_T - \tilde{\beta}_{HT}) \\ &= \frac{TH}{1+H} (\hat{\beta}_T - \tilde{\beta}_{HT})' \Omega^* (M - M_2) (\hat{\beta}_T - \tilde{\beta}_{HT})\end{aligned}$$

A classical argument of block inverse gives

$$\begin{aligned}\Omega^* (M - M_2) &= \Omega^* (Id - M_2) \frac{\partial b}{\partial \theta_1'} \left(\frac{\partial b'}{\partial \theta_1} (Id - M_2)' \Omega^* (Id - M_2) \frac{\partial b}{\partial \theta_1'} \right)^{-1} \\ &\quad \times \frac{\partial b'}{\partial \theta_1} (Id - M_2)' \Omega^*\end{aligned}$$

and the asymptotic equivalence between ξ_T^C and ξ_T^W follows.

First-order Expansion of the Score Test Statistic

The score test statistic is based on the gradient of the objective function with respect to θ_2 evaluated under the null hypothesis. This gradient is proportional to

$$\begin{aligned}\mathcal{Q}_T &= \frac{\partial \tilde{\beta}_{HT}}{\partial \theta_1} \begin{bmatrix} 0 \\ \hat{\theta}_{2H}^0 \end{bmatrix} \Omega^* \left(\hat{\beta}_T - \tilde{\beta}_{HT} \begin{bmatrix} 0 \\ \hat{\theta}_{2H}^0 \end{bmatrix} \right) \\ &\approx \frac{\partial b'}{\partial \theta_1} (\theta_0) \Omega^* \left[\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0) - \frac{\partial b}{\partial \theta_2'} (\theta_0) (\hat{\theta}_{2H}^0 - \theta_{20}) \right] \\ &\approx \frac{1}{\sqrt{T}} \frac{\partial b'}{\partial \theta_1} \Omega^* \left[\sqrt{T} (\hat{\beta}_T - \tilde{\beta}_{HT}(\theta_0)) - \frac{\partial b}{\partial \theta_2'} A_{22}^{-1} \frac{\partial b'}{\partial \theta_2} \Omega^* \sqrt{T} (\hat{\beta}_T - \tilde{\beta}_{HT}) \right] \\ &\approx \frac{1}{\sqrt{T}} \frac{\partial b'}{\partial \theta_1} (Id - M_2)' \Omega^* \sqrt{T} (\hat{\beta}_T - \tilde{\beta}_{HT}) \\ &\approx [A_{11} - A_{12} A_{22}^{-1} A_{21}] \tilde{\theta}_{1T}^H\end{aligned}$$

There is asymptotically a one-to-one linear relationship between \mathcal{D} and $\hat{\theta}_{1T}^H$ and this shows that the score test is asymptotically equivalent to the Wald test.

Asymptotic variance-covariance matrix of the unconstrained estimator $\hat{\theta}_{1T}^H$ and of the score

We get

$$V_{as}(\sqrt{T} \hat{\theta}_{1T}^H) = \left(1 + \frac{1}{H}\right) (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}$$

and therefore we deduce

$$V_{as}(\sqrt{T} \mathcal{D}_T) = \left(1 + \frac{1}{H}\right) (A_{11} - A_{12} A_{22}^{-1} A_{21})$$

REFERENCES

- Aït-Sahalia, Y. (1993), 'Nonparametric pricing of interest rate derivative securities', MIT.
- Anderson, B. D. O. and J. B. Moore (1979), *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ.
- Diebold, F. and M. Nerlove (1989), 'The dynamic of exchange rate volatility: a multivariate latent factor ARCH model', *Journal of Applied Econometrics*, 4, 1–22.
- Duffie, D. and K. Singleton (1993), 'Simulated moments estimation of Markov models of asset prices'.
- Engle, R., V. Ng and M. Rothschild (1989), 'Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills' UCSD DP, 89–31.
- Gallant, R. and G. Tauchen (1992), 'Which moments to match ?' Mimeo.
- Gallant, R. and H. White (1988), *Estimation and Inference for Nonlinear Dynamic Models*, Blackwell, New York.
- Genon-Catalot, V. (1990), 'Maximum contrast estimation for diffusion processes from discrete observations', *Statistics*, 21, 99–116.
- Gouriéroux, C. (1992), *Modèles ARCH: Applications Financières et Monétaires*, Economica, Paris.
- Gouriéroux C. and A. Monfort (1993), 'Simulation based inference: a survey with special reference to panel data models', *Journal of Econometrics*.
- Gouriéroux, C. and A. Monfort (1992), 'Testing, encompassing and simulating dynamic econometric models, CREST-Département de la Recherche INSEE, DP No. 9214.
- Gouriéroux, C., A. Monfort and E. Renault (1991), 'Dynamic factor models', CREST-Département de la Recherche INSEE.
- Gouriéroux C., A. Monfort and E. Renault (1992), 'Indirect inference', CREST-Département de la Recherche INSEE, DP No. 9215.
- Gouriéroux, C., A. Monfort and A. Trognon (1983), 'Testing nested or non-nested hypotheses', *Journal of Econometrics*, 21, 83–115.
- Gouriéroux, C., A. Monfort and A. Trognon (1984), 'Estimation and test in probit models with serial correlation', in Florens *et al.* (eds), *Alternative Approaches to Time Series Analysis*, Public-Universits St Louis, Brussels.
- Hamilton, J. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, 57, 357–384.
- Hansen, L. P. and J. Scheinkman (1991), 'Back to the future: generating moment implications for continuous time Markov processes', University of Chicago.
- Hausman, J. and D. Wise (1978), 'A conditional probit model for qualitative choice: discrete decisions, recognizing interdependence and heterogeneous preferences', *Econometrica*, 46, 403–426.
- Harvey, A., E. Ruiz and N. Shephard (1992), 'Multivariate stochastic variance models', LSE DP No. 132.
- Hendry, D. F., and J. F. Richard (1990), 'Recent developments in the theory of encompassing, in B. Tulkens and H. Tulkens (eds), *Contributions to Operation Research and Econometrics, the Twentieth Anniversary of CORE, Cornet*, MIT Press, Cambridge, MA, 393–400.
- Ibragimov, I. A. and R. Z. Has'minskii (1981), *Statistical Estimation, Asymptotic Theory*, Springer-Verlag, Berlin.

- King, M., E. Sentana and S. Wadhvani (1990), 'A heteroscedastic model of assets returns and risk premia with time varying volatility: an application to sixteen world stock markets', Mimeo, LSE.
- Kitagawa, G. (1987), 'Non Gaussian state space modeling of nonstationary time series', *Journal of the American Statistical Association*, **82**, 1032–1041.
- Kooreman, P. and G. Ridder (1983), 'The effects of age and unemployment percentage on the duration of unemployment', *European Economic Review*, **20**, 41–57.
- Lipster R. S. and A. N. Shirayev (1977), *Statistics of Random Processes I, General Theory*, Springer-Verlag, Berlin.
- Lipster, R. S. and A. N. Shirayev (1978), *Statistics of Random Processes II, Applications*, Springer-Verlag, Berlin.
- McFadden, D. (1976), 'Quantal choice analysis: a survey', *Annals of Economic and Social Measurement*, **5**, 363–390.
- McFadden, D. (1989), 'A method of simulated moments for estimation of discrete response models without numerical integration', *Econometrica*, **57**, 995–1026.
- Mizon, G. E. and J. F. Richard (1986), 'The encompassing principle and its application to testing non nested hypotheses', *Econometrica*, **54**, 657–678.
- Nelson, D. B. (1990), 'ARCH models as diffusion approximations', *Journal of Econometrics*, **45**, 7–38.
- Newey, W. K. and K. D. West (1987), 'A simple positive definite heteroscedasticity and autocorrelation consistent covariance matrix', *Econometrica*, **55**, 703–708.
- Pakes, A. and D. Pollard (1989), 'Simulation and the asymptotics of optimization estimators', *Econometrica*, **57**, 1027–1058.
- Pesaran, H. and B. Pesaran (1991), 'A simulation approach to the problem of computing Cox's statistic for testing non nested models', DP University of California at Los Angeles.
- Robinson, P. (1982), 'On the asymptotic properties of estimators of models containing limited dependent variables', *Econometrica*, **50**, 27–41.
- Smith, A. (1993), 'Estimating nonlinear time series models using simulated vector autoregressions', *Journal of Applied Econometrics*, this issue.
- Thelot, C. (1981), 'Note sur la loi logistique et l'imitation', *Annales de l'INSEE*, **42**, 111–125.
- Wooldridge, J. M. (1990), 'An encompassing approach to conditional mean tests with applications to testing non nested hypotheses', *Journal of Econometrics*, **45**, 331–350.