Exact Monte Carlo Tests applied to models estimated by Indirect Inference and by Efficient Method of Moments

Jean-Marie Dufour* Christian Gouriéroux† Pascale Valéry‡

31st January 2000

Abstract

The aim of this paper is to provide exact inference in finite sample for econometric models whose likelihood function is intractable and require thereby simulation-based estimation method like Indirect Inference Method or Efficient Method of Moments. To do so, we resort to the technique of Monte Carlo Tests which naturally applies to any simulable model. In particular, maximized Monte Carlo tests allow for test statistics whose distribution depends on nuisance parameters. This technique of Monte Carlo tests is applied here to a stochastic differential equation.

*Université de Montréal,CRDE
†CREST,CEPREMAP
‡Université de Montréal,CRDE

Contents

1 Introduction

Econometrics models often lead to complex formulations for the conditional distribution of the endogenous variables given the exogenous variables and the lagged endogenous values. Econometricians have therefore tried to approximate the structural model by an auxiliary model chosen for its simplicity of estimation. Indirecte Inference estimation method [Gouriéroux. Monfort et Renault (1993)] and Efficient Method of Moments [Gallant et Tauchen (1996)] come within this scope. They both proceed by simulation and require the structural model to be simulated. In this framework, various test statistics have also been proposed to make inference on the parameters of interest of the structural model. The latter include among others Wald-type statistics, score-type statistics or statistics based on the difference of optimized objective functions: see Gouriéroux, Monfort and Renault (1993), Gallant and Tauchen (1996), Gallant (1987), Andersen, Chung and Sorensen (1997). However, the distributional theory associated with those statistics is asymptotic and the choice of the existing statistics importantly depends on the possibility to obtain an asymptotic nuisance-parameter free distribution under the null hypothesis. This opens up the way for errors of approximation of any magnitude [voir Dufour (1997)]. The aim of the paper is to develop exact tests as well as more reliable asymptotic procedures in such models. To do so, we resort to the technique of Monte Carlo tests which naturally applies to any simulable model. In particular, maximized Monte Carlo tests allow for test statistics whose distribution depends on nuisance parameters [Dufour (1996)]. Thus the technique of Monte Carlo tests naturally matches with simulated estimation methods.

In the second section we explain the Indirect Inference estimation method [Gouriéroux, Monfort and Renault (1993)] together with tests of hypotheses which allow for correct inference from an incorrect criterion. The third section summarizes Efficient Method of Moments (EMM) [Gallant and Tauchen (1996)] where a specific criterion based on the quasi-scores is used for the estimation of the parameter of interest. The EMM estimator is then used in inference methods. The fourth section extends the previous standard tests to finite sample Monte Carlo tests insofar as the structural model can be simulated. Monte Carlo tests with or without nuisance parameters are analyzed. The asymptotic validity of Monte Carlo tests based on consistent set estimators of the nuisance parameters is then stated in a fith section. In a sixth section we propose new test statistics straightforwardly built from the objective function of the auxiliary model. The last section applies Monte Carlo tests to a geometric brownian motion estimated by Indirect Inference method.

2 Indirect Inference method

Econometric models often lead to complex formulations yielding intractable formulation of the likelihood function and thereby the impossibility to efficiently estimate the parameters of interest. A natural procedure consists in replacing the likelihood function by another criterion (an approximation of the exact likelihood function or either the likelihood function of an auxiliary model): this is the Indirect Inference approach [Gouriéroux, Monfort and Renault (1993)]

2.1 The estimation method

We consider the following dynamic model :

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

where the x_t' are observable exogenous variables whereas the u_t and ϵ_t are not observed. We assume moreover that:

Hypothèse 1 *$\{x_t\}$ is an homogeneous Markov process independent of $\{\epsilon_t\}$ and $\{u_t\}$, the process $\{\epsilon_t\}$ is a white noise whose distribution G_0 is known, and the process $\{y_t, x_t\}$ is stationary.*

With such a parametric model it is theoretically possible to compute the conditional density function of $y_1, \dots, y_T$ given $z_0, x_1, \dots, x_T$, where z_0 is a vector of initial values 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. The Indirect Inference method constitutes an alternative two-step estimation method, in which all that is required from model (1) is to be easily simulated. We can summarize this method in the following way.

First an observation-dependent criterion and an auxiliary parameter $\beta \in B \subset R^q$ are introduced. Let $\mathbf{y_T^1} = (y_1, \dots, y_T)$ and $\mathbf{x_T^1} = (x_1, \dots, x_T)$ be the observation vectors. Let $\hat{\beta}_T$ be the solution to this problem:

$$\hat{\beta}_T = \arg\max_{\beta \in B} \mathcal{Q}_T(\mathbf{y_T^1}, \mathbf{x_T^1}, \beta)$$

where the criterion $\mathcal{Q}_T$ is defined by

$$\mathcal{Q}_T(\mathbf{y_T^1}, \mathbf{x_T^1}, \beta) = \frac{1}{T} \sum_{t=1}^{T} \psi(y_t, x_t, \beta) \ ,$$

which suggests an M-estimation procedure for the auxiliary parameter β.

Let us assume that the criterion $\mathcal{Q}_T$ tends asymptotically (and uniformly almost certainly) to a non-stochastic limit and that the limit criterion is continuous in β with a unique maximum.

Hypothêse 2

$$\lim_{T\to\infty} \mathcal{Q}_T(\mathbf{y_T^1}, \mathbf{x_T^1}, \beta) = \mathcal{Q}_\infty(F_0, G_0, \theta_0, \beta) \quad a.s.$$

Hypothêse 3

$$\beta_0 = \arg\max_{\beta \in B} \mathcal{Q}_\infty(F_0, G_0, \theta_0, \beta) \ .$$

According to the usual asymptotic theory [see Gallant and White (1988), ch. 3) and under the assumptions 1, 2 and 3, 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 the unknown transition distribution F_0 of the exogenous variables. We can then define the binding function:

$$b(F, G, \theta) = \arg\max_{\beta \in B} \mathcal{Q}_\infty(F, G, \theta, \beta) \ .$$

In particular,

$$\beta_0 = b(F_0, G_0, \theta_0) \ .$$

If the binding function

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

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 $\tilde{\theta}_T$ of $\hat{\beta}_T = b(F_0, G_0, \tilde{\theta}_T)$.

It is clear that the estimator $\hat{\beta}_T$ 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 distributions of x and ϵ. Therefore if the finite sample binding function $b(\hat{F}_T, \hat{G}_T, \cdot)$ was known and one to one we could deduce from $\hat{\beta}_T$ the *exact value* θ_0 of the unknown parameters while the knowledge of the true binding function $b(F_0, G_0, \cdot)$ the solution of $\hat{\beta}_T = b(F_0, G_0, \tilde{\theta}_T)$ only provides a consistent estimator $\tilde{\theta}_T$. This is the reason why the second step of the 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 $y's$. The following assumption is required for identifying the parameter θ.

Hypothèse 4 *$b(F_0, G_0, \cdot)$ is one to one and $\frac{\partial}{\partial \theta'} b(F_0, G_0, \theta_0)$ is of full-column rank.*

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 $\epsilon_t, (\tilde{\epsilon}_1^h, \dots, \tilde{\epsilon}_T^h)$, and on initial values z_0^h, $h = 1, \dots, H$. For each of these paths, we can also consider the optimization problem:

$$\max_{\beta \in B} \mathcal{Q}_T((\mathbf{\tilde{y}^h})_\mathbf{T}^\mathbf{1}, \mathbf{x}_\mathbf{T}^\mathbf{1}, \beta)$$

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

$$\tilde{\beta}_T^h(\theta, z_0^h) = \arg \max_{\beta \in B} \mathcal{Q}_T((\mathbf{\tilde{y}^h}(\theta, \mathbf{z_0^h}))_\mathbf{T}^\mathbf{1}, \mathbf{x}_\mathbf{T}^\mathbf{1}, \beta) \ .$$

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

$$\max_{\beta \in B} \mathcal{Q}_\infty(F_0, G_0, \theta, \beta)$$

i.e.

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

$\tilde{\beta}_T^h(\cdot, z_0^h)$ is therefore a consistent functional estimator of $b(F_0, G_0, \theta)$. It is now possible to define the indirect estimator of θ. The idea is simply to calibrate the value of θ in order to make

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

as close as possible 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:*

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

where $\hat{\Omega}_T$ is a positive definite matrix converging to a deterministic positive definite matrix Ω. Under the assumptions 1, 2, 3 et 4, the indirect estimator $\tilde{\theta}_T^H(\Omega)$ is a consistent estimator of θ_0.

Proposition 2 *Under the assumptions 1, 2, 3 et 4, the assumptions A5-A8 of Gouriéroux and al (1993) and under some usual regularity conditions, the indirect estimator is asymptotically normal, when H is fixed and T tends to infinity:*

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

where

$$\begin{aligned}\mathcal{W}(H,\Omega) = &\left(1 + \frac{1}{H}\right)\left(\frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0)\Omega\frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0)\right)^{-1} \\ &\frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0)\Omega\mathcal{J}_0^{-1}(\mathcal{I}_0 - \mathcal{K}_0)\mathcal{J}_0^{-1}\Omega\frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0) \\ &\left(\frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0)\Omega\frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0)\right)^{-1}\end{aligned}$$

The asymptotic variance-covariance matrix depends on the metric Ω and there is an optimal choice of this matrix in the sense that it minimizes $\mathcal{W}(H,\Omega)$. The optimal choice of the Ω matrix is: $\Omega^* = \mathcal{J}_0(\mathcal{I}_0 - \mathcal{K}_0)^{-1}\mathcal{J}_0$ and

$$\mathcal{W}_H^* = \mathcal{W}(H,\Omega) = \left(1+\frac{1}{H}\right)\left(\frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0)\mathcal{J}_0(\mathcal{I}_0 - \mathcal{K}_0)^{-1}\mathcal{J}_0\frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0)\right)^{-1}$$

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

2.2 Specification tests

We will now describe the different test criteria which have been proposed by Gouriéroux, Monfort and Renault (1993) in order to test hypotheses on models estimated by indirect inference method.

A specification test for the model may be based on the optimal value of the objective function used in the second step of the indirect estimation method.

Proposition 3 *The statistic*

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

where $\hat{\Omega}_T^*$ is a consistent estimator of Ω^*, is asymptotically distributed as a chi-square with $q-p$ degrees of freedom, with $q = \dim \beta$ and $p = \dim \theta$, when the reduced-form (1) or either the stuctural model is well spesified. The specification test of asymptotic level α is associated with the critical region $\mathcal{W} = \{\varepsilon_T > \chi^2_{1-\alpha}(q-p)\}$.

2.3 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

Let

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

be the optimal unconstrained indirect estimator and let

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

be the optimal constrained indirect estimator obtained by optimizing the criterion submitted to $\theta_1 = 0$.

The Wald statistic is defined as:

$$\varepsilon_T^{\mathcal{W}} = T(\hat{\theta}_{1T}^H)'\hat{\mathcal{W}}_1^{*-1}(\hat{\theta}_{1T}^H)$$

where $\hat{\mathcal{W}}_1^*$ is a consistent estimator of the variance-covariance 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 constrained estimator. This gradient is given by

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

and the test statistic is

$$\varepsilon_T^{\mathcal{S}} = T\mathcal{D}'_T\mathcal{A}\mathcal{D}_T \ ,$$

where $\mathcal{A}$ is a consistent estimator of $(V_{as}(\sqrt{T}\mathcal{D}_T))^{-1} = (1+\frac{1}{H})^{-1}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$ Finally, we can introduce the difference between the optimal values of the objective function:

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

Proposition 4 *The test statistics $\varepsilon_T^{\mathcal{W}}$, $\varepsilon_T^{\mathcal{S}}$ and $\varepsilon_T^{\mathcal{C}}$ are asymptotically equivalent under the null hypothesis, and have the common distribution $\chi^2(p_1)$.*

3 Efficient Method of Moments

The Efficient Method of Moments (EMM) proposed by Bansal, Gallant, Hussey and Tauchen (1993, 1995) and Gallant and Tauchen (1996) relies on a specific form of the auxiliary criterion $\mathcal{Q}_T(\mathbf{y_T^1}, \mathbf{x_T^1}, \theta)$ optimized in the indirect inference approach. This one takes here the form of a quasilikelihood. But while the indirect inference approach uses the parameters of the auxiliary model to define the GMM criterion, efficient method of moments builds the GMM criterion from the score function of the auxiliary model.

3.1 The method

The efficient method of moments (EMM) consists in taking the expectation under the structural model of the score from an auxiliary model as the vector of moment conditions. The structural model assumed to have generated the data $\{\tilde{y}_t, \tilde{x}_t\}_{t=1}^n$ is defined by:

$$\left\{p_1(x_1|\rho^0), \left\{p_t(y_t|x_t, \rho^0)\right\}_{t=1}^n\right\}$$

where ρ^0 indicates the true value of ρ in the model

$$\{p_1(x_1|\rho), \{p_t(y_t|x_t, \rho)\}_{t=1}^\infty\}_{\rho \in R} \ .$$

Moreover, we can consider an auxiliary model

$$\{f_1(x_1 \mid \theta), \{f_t(y_t|x_t, \theta)\}_{t=1}^\infty\}_{\theta \in \Theta}$$

which constitutes an approximation of the structural model. The former is called the "score generator" model.

Definition 1 *The model $\{p_1(x_1|\rho), \{p_t(y_t|x_t, \rho)\}_{t=1}^\infty\}_{\rho \in R}$ is said to be smoothly embedded within the score generator $\{f_1(x_1|\theta), \{f_t(y_t|x_t, \theta)\}_{t=1}^\infty\}_{\theta \in \Theta}$ if for some open neighborhood R^0 of ρ^0 there is a twice continuously differentiable mapping $g : R^0 \to \Theta$ such that*

$$p_t(y_t|x_t, \rho) = f_t[y_t|x_t, g(\rho)] \ t = 1, 2, ...,$$

for every $\rho \in R^0$ and $p_1(x_1|\rho) = f_1[x_1|g(\rho)]$ for every $\rho \in R^0$.

The score generator need not encompass the structural model. If it does, then the estimator is as efficient as the maximim likelihood estimator. If the score generator closely approximates the actual distribution of the data, even though it does not encompass it, then the estimator is nearly fully efficient.

Hypothèse 5 *There is a* θ_0 *such that* $p_t(y_t|x_t,\rho^0) = f_t(y_t|x_t,\theta_0)$ *for* $t = 1,2,\ldots,$ *and* $p_1(x_1|\rho^0) = f_1(x_1|\theta_0)$.

The method consists of two steps. First, the idea is to use the scores

$$\frac{\partial}{\partial\theta}\ln f_t(y_t|x_t,\theta)$$

evaluated at the quasimaximum likelihood estimate

$$\tilde{\theta}_n = \arg\max_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\ln f_t(\tilde{y}_t|\tilde{x_t},\theta)\ ,$$

to generate GMM moments conditions. In the case that the data $\{\tilde{y}_t,\tilde{x}_t\}_{t=1}^n$ are a sample from $\prod_{t=1}^n p(y_t|x_t,\rho^0)p(x_1|\rho^0)$ the moment conditions are :

$$\begin{aligned} m(\rho,\tilde{\theta}_n) = \quad & E_\rho[s_f(Y_t,\tilde{\theta}_n)] \\ & = \int\int \tfrac{\partial}{\partial\theta}\ln f(y|x,\tilde{\theta}_n)p(y|x,\rho)\,dyp(x|\rho)\,dx\ . \end{aligned}$$

These are the moment conditions that define the EMM estimator ($\hat{\rho}_{EMM}$). It is because the expected score of the auxiliary model is defined under the probability measure induced by the structural model that the moments depend on parameters of both models. In most applications, analytic expressions for the integrals will not be available due to their high complexity, therefore the sample moments will be computed through Monte-Carlo integration following the simulated method of moments of Duffie and Singleton (1993).

Second, a simulated serie $\{\hat{y}_t,\hat{x}_t\}_{t=1}^n$ is generated from the density $\prod_{t=1}^n p(y_t|x_t,\rho)p(x_1|\rho)$ for a given value of ρ and used to evaluate the sample moments at the QML estimate $\tilde{\theta}_n$:

$$m_n(\rho,\tilde{\theta}_n) = \frac{1}{n}\sum_{\tau=1}^{n}\frac{\partial}{\partial\theta}\ln f(\hat{y}_\tau|\hat{x}_\tau,\tilde{\theta}_n)$$

where $\hat{x}_\tau$ is a vector of exogenous variables with possibly lagged dependent variables . When $n\to\infty$, $m_n(\rho,\tilde{\theta}_n)\to m(\rho,\tilde{\theta}_n)$. For n sufficiently large, the Monte-Carlo error is negligible. Identification conditions are on the other hand required to the identification of the parameter of interest. Thus the number of moment conditions has to be greater or equal to the number of structural parameters: $\dim\theta \geq \dim\rho$. The rank condition $m_n(\rho, \theta_n^0) = 0 \Rightarrow \rho = \rho^0$ for $n \geq n^0$ is more difficult to verify.

A GMM criterion function can then be constructed from the sample moment conditions of the auxiliary model. The EMM estimator of ρ is defined by

$$\hat{\rho}_n = \arg\min_{\rho}[m_n(\rho, \tilde{\theta}_n)'(\tilde{\mathcal{I}}_n)^{-1}m_n(\rho, \tilde{\theta}_n)]$$

where $\tilde{\mathcal{I}}_n$ is a consistent estimator of the asymptotic variance-covariance matrix of the scores of the auxiliary model $\mathcal{I}_n^0$,

$$\sqrt{n}m_n(\rho^0, \tilde{\theta}_n) \approx \mathcal{N}(0, \mathcal{I}_n^0)$$

where

$$\mathcal{I}_n^0 = Var\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln f_t(\tilde{y}_t|\tilde{x}_t, \theta_n^0)\right]$$

or either

$$\mathcal{I}_n^0 = p\lim_{n\to\infty}\left(\frac{1}{n}\sum_{t=1}^{n}\sum_{\tau=1}^{n}s_f(Y_t, \tilde{\theta}_n)s_f(Y_\tau, \tilde{\theta}_n)'\right).$$

If the auxiliary model fits well the systematic features of the data in the sense that the quasi-scores constitute a martingale difference sequence, then from the standard likelihood theory a convenient estimator of $\mathcal{I}_n^0$ is given by

$$\tilde{\mathcal{I}}_n = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta}\ln f_t(\tilde{y}_t|\tilde{x}_t, \tilde{\theta}_n)\right]\left[\frac{\partial}{\partial\theta}\ln f_t(\tilde{y}_t|\tilde{x}_t, \tilde{\theta}_n)\right]'.$$

3.2 Asymptotic distribution of the EMM estimator

Gallant et Tauchen (1996) show under general regularity conditions (conditions 8-11 in Gallant 1987, ch.7), that the EMM estimator is consistent and asymptotically normal.

Theorem 1

$$\begin{aligned}
\lim_{n\to\infty} \hat{\rho}_n &= \rho_0 \text{ a.s}, \\
\sqrt{n}(\hat{\rho}_n - \rho_0) &\overset{a.s}{\longrightarrow} \mathcal{N}\left\{0, \left[(\mathcal{M}_n^0)'(\mathcal{I}_n^0)^{-1}(\mathcal{M}_n^0)\right]^{-1}\right\}, \\
\lim_{n\to\infty}(\hat{\mathcal{M}}_n - \mathcal{M}_n^0) &= 0 \ a.s,
\end{aligned}$$

where $\hat{\mathcal{M}}_n = \mathcal{M}_n(\hat{\rho}_n, \tilde{\theta}_n)$, $\mathcal{M}_n^0 = \mathcal{M}_n(\rho^0, \theta_n^0)$ *and* $\mathcal{M}_n(\rho, \theta) = \frac{\partial}{\partial \rho'} m_n(\rho, \theta)$.

3.3 Specification tests

The two following subsections describe test statistics which have been proposed by Gallant and Tauchen (1996) and Gallant (1987) within the Efficient Method of Moments estimation framework. These authors mostly focused on the asymptotic behaviour of these test procedures. We will rather analyze their exact behaviour.

Under correct specification hypothesis, n times the GMM optimized criterion is distributed as a $\chi^2(n_\eta - n_\rho)$:

$$n \min_{\rho \in R} \left[m_n(\rho, \tilde{\theta}_n)'(\tilde{\mathcal{I}}_n)^{-1} m_n(\rho, \tilde{\theta}_n) \right] .$$

This statistic may constitute a test of the overidentifying restrictions of the structural model (Hansen 1982). The p-values are asymptotically uniform under the null hypothesis.

3.4 Tests of hypotheses

Direct use of the above theorem for setting confidence intervals on the elements of ρ or testing hypotheses with the Wald test requires computation of $\mathcal{M}_n(\rho, \theta)$. Computation of $\mathcal{M}_n(\rho, \theta)$ can be avoided by testing hypotheses using the criterion difference test statistic (Gallant, 1987, ch. 7, theorem 15) and setting confidence intervals by inverting it. Let

$$d_n(\rho) = m_n(\rho, \tilde{\theta}_n)'(\tilde{\mathcal{I}}_n)^{-1} m_n(\rho, \tilde{\theta}_n)$$

We will exploit here the asymptotic normality of the EMM estimator $\hat{\rho}_n$ as established by Gallant and Tauchen in theorem 1 as well as the asymptotic normality of the score of the objective function $d_n(\rho)$, i.e.:

$$\sqrt{n}\frac{\partial}{\partial \rho'}d_n(\rho) \overset{\mathcal{L}}{\sim>} \mathcal{N}_p(0, \bar{\mathcal{I}}). \tag{1}$$

where

$$\bar{\mathcal{I}} = \mathcal{M}_n^{0'}\mathcal{D}_n^0\mathcal{I}_n^0\mathcal{D}_n^0\mathcal{M}_n^0$$

and where $\mathcal{D}_n^0 = \frac{\partial^2}{\partial m_n \partial m_n'}d_n(\rho^0)$, to construct Wald-type statistics, likelihood ratio statistics and Lagrange multiplier-type statistics. We will use usual asymptotic derivations in order to derive the asymptotic chi-square distribution of these three statistics. The hypothesis to be tested is : H_0: $h(\rho_n^0) = h_n^*$ against H_1: $h(\rho_n^0) \neq h_n^*$,
where $h(\rho)$ is a twice continuously differentiable mapping from $\Lambda \subset R^p$ onto R^q . The first statistic considered is a Wald-type statistic :

$$W = n(\hat{h} - h_n^*)'(\hat{H}\hat{V}\hat{H}')^{-1}(\hat{h} - h_n^*)$$

where $\hat{h} = h(\hat{\rho})$, $h_n^* = h(\rho_n^0)$, $\hat{H} = \frac{\partial h}{\partial \rho}(\hat{\rho})$. $\hat{V} = V\hat{a}r(\hat{\rho}_n) = (\hat{\mathcal{M}}_n'\tilde{\mathcal{I}}_n^{-1}\hat{\mathcal{M}}_n)^{-1}$ is a consistent estimator of $((\mathcal{M}_n^0)'(\mathcal{I}_n^0)^{-1}(\mathcal{M}_n^0))^{-1}$ Given asymptotic normality of the estimator $\hat{\rho}_n$, this quadratic form will asymptotically be distributed as a χ^2 with q degrees of freedom. The second statistic "likelihood ratio" type statistic (LR):

$$L = 2n[d_n(\tilde{\rho}_n) - d_n(\hat{\rho}_n)]$$

The LR test requires that :

$$H_n^*(\mathcal{M}_n^{*'}(\mathcal{I}_n^*)^{-1}\mathcal{M}_n^*)^{-1}H_n^{*'} = H_n^*(\mathcal{M}_n^*)^{-1}H_n^{*'} + o(1)$$

where $H_n^* = \frac{\partial h}{\partial \rho}(\rho_n^0)$ in order to obtain the usual asymptotic distribution. $\mathcal{M}_n^* = \mathcal{I}_n^*$ if the score generator is a good approximation to the true data generating process. This LR-type statistic is asymptotically distributed as a chi-square with q degrees of freedom.

The last statistic to be considered is a Lagrange multiplier-type statistic:

$$R = n\left(\frac{\partial}{\partial \rho}d_n(\tilde{\rho_n})\right)' Var\left[\sqrt{n}\frac{\partial}{\partial \rho}d_n(\tilde{\rho_n})\right]^{-1}\frac{\partial}{\partial \rho}d_n(\tilde{\rho_n}) .$$

This one derives from the Lagrangian

$$\frac{\partial}{\partial \rho}\mathcal{L}(\tilde{\rho}_n, \tilde{\lambda}) = \frac{\partial}{\partial \rho}\left\{ d_n(\rho) + \lambda'[h(\tilde{\rho}_n) - h_n^*] \right\} = 0$$

corresponding to the minimization of the criterion $d_n(\rho)$ submitted to the restriction: $h(\rho) = h_n^*$. This statistic follows also a chi-square distribution due to the asymptotic normality of the score function of the objective function $d_n(\rho)$.

For these three tests, the hypothesis: $H_0 : h(\rho_n^0) = h_n^*$ is rejected when the test statistic exceeds the critical values $\alpha \times 100\%$ of a chi-square variable with q degrees of freedom to achieve a test of asymptotic level α.

4 Finite sample Monte Carlo tests

The technique of Monte Carlo tests has been suggested for the first time by Dwass (1957) to implement permutation tests, then independently by Barnard (1963) and Birnbaum (1974) for statistics with continuous distributions. The advantage of this technique is to provide (randomized) exact tests based on test statistics whose finite sample distribution may be intractable but can be easily simulated. The validity of the tests so obtained does not dependat all on the number of replications made (which can be small). Only the power of the procedure is influenced by the number of replications, but the power gains associated with lenghty simulations are typically rather small. For further discussion of Monte Carlo tests, see Dufour (1996), Dufour and Kiviet (1996,1997),Kiviet and Dufour (1996),Dufour and Khalaf (1996a,1996b,1996c) Edgington (1987), Foutz (1980), Jöckel (1986) and Marriott (1979) .

Only standard test statistics have been proposed up to now in the literature as these authors only focus on the asymptotic distributional properties of these statistics. The technique of Monte Carlo tests allow for general test statistics. In particular, maximized Monte Carlo tests (Dufour, 1996) allow for test statistics whose distribution depends on nuisance parameters. In a first step we will consider the case without nuisance parameters in order to better analyze the case with nuisance parameters.

4.1 Monte Carlo tests without nuisance parameters

Consider now a situation where the distribution of S under H_0 may not be easy to compute analytically but can be simulated. Let $S_1, \cdots, S_N$ be a sample of independent and identically distributed (i.i.d) real random variables with the same distribution as S. The technique of Monte Carlo tests provides a simple method allowing one to replace the theoretical distribution $F(x)$ by its sample analogue based on $S_1, \ldots, S_N$:

$$\hat{F}_N[x; S(N)] = \frac{1}{N} \sum_{i=1}^{N} s(x - S_i) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{[0,\infty]}(x - S_i)$$

where $S(N) = (S_1, \ldots, S_N)'$, $s(x) = 1_{[0,\infty]}(x)$ and $1_A(x)$ is the indicator function associated with the set A. We also consider the corresponding sample tail area function:

$$\hat{G}_N[x; S(N)] = \frac{1}{N} \sum_{i=1}^{N} s(S_i - x) \cdot$$

The sample distribution function is related to the ranks $R_1, \cdots, R_N$ of the variables $S_1, \ldots, S_N$ (when put in ascending order) by the expression:

$$R_j = N\hat{F}_N[S_j; S(N)] = \sum_{i=1}^{N} s(S_j - S_i), \; j = 1, \ldots, N.$$

The central property we shall exploit here is the following: to obtain critical values or compute p-values, the "theoretical" null distribution $F(x)$ can be replaced by its simulation-based "estimate" $\hat{F}_N$ in a way that will preserve the level of the test in *finite samples, irrespective of the number N of replications used.* For continuous distributions, this property is expressed by proposition 5 below which is easily proved by using the following simple lemma.

Lemme 1 *Let $(y_1, \ldots, y_N)'$ be a $N \times 1$ vector of exchangeable real random variables such that*

$$P[y_i = y_j] = 0 \;\; \textit{for } i \neq j, \; i,j = 1, \ldots, N$$

and let $R_j = \sum_{i=1}^{N} s(y_j - y_i)$ be the rank of y_j when $(y_1, \ldots, y_N)$ are ranked in nondecreasing order. Then, for $j = 1, \ldots, N$,

$$P[R_j/N \leq x] = \frac{I[xN]}{N} \qquad , \textit{for} \quad 0 \leq x \leq 1$$

and

$$\begin{aligned} P[R_j/N \geq x] &= 1 & &, \textit{ if } x \leq 0 \\ &= \frac{I[(1-x)N] + 1}{N} & &, \textit{ if } 0 < x \leq 1 \\ &= 0 & &, \textit{ if } x > 1, \end{aligned}$$

where $I[x]$ is the largest integer less than or equal to x.

Proposition 5 *(Validity of Monte Carlo tests when ties have zero probability). Let* $(S_0, S_1, \ldots, S_N)'$ *be a* $(N+1)\times 1$ *vector of exchangeable real random variables such that*

$$P[S_i = S_j] = 0 \ \ \textit{for } i \neq j, \ i,j = 0, 1, \ldots, N$$

let $\hat{F}_N \equiv \hat{F}_N[x; S(N)], \hat{G}_N(x) = \hat{G}_N[x; S(N)]$ *and* $\hat{F}_N^{-1}(x)$ *be the quantile function defined by:*

$$\begin{aligned} \hat{F}_N^{-1}(x) &= \inf\{y : \hat{F}_N(y) \geq x\} \ , \ \textit{if } 0 < x < 1, \\ &= \inf\{y : \hat{F}_N(y) > 0\} \ , \ \textit{if } x = 0, \\ &= \sup\{y : \hat{F}_N(y) < 1\}, \ \textit{if } x = 1, \end{aligned}$$

and

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N + 1}.$$

Then

$$\begin{aligned} P[\hat{G}_N(S_0) \leq \alpha_1] &= P[\hat{F}_N(S_0) \geq 1 - \alpha_1] \\ &= \frac{I[\alpha_1 N] + 1}{N + 1}, \ \textit{for } 0 \leq \alpha_1 \leq 1, \end{aligned} \tag{2}$$

$$P[S_0 \geq \hat{F}_N^{-1}(1 - \alpha_1)] = \frac{I[\alpha_1 N] + 1}{N + 1}, \textit{for } 0 < \alpha_1 < 1, \tag{3}$$

and

$$P[\hat{p}_N(S_0) \leq \alpha] = \frac{I[\alpha(N + 1)]}{N + 1}, \textit{for } 0 \leq \alpha \leq 1. \tag{4}$$

The latter proposition can be used as follows: choose α_1 and N so that

$$\alpha = \frac{I[\alpha_1 N] + 1}{N + 1} \tag{5}$$

is the desired significance level. Provided N is reasonably large, α_1 will be very close to α; in particular, if $\alpha(N + 1)$ is an integer, we can take

$$\alpha_1 = \alpha - \frac{(1 - \alpha)}{N},$$

in which case we see easily that the critical region $\hat{G}_N(S_0) \leq \alpha_1$ is equivalent to $\hat{G}_N(S_0) \leq \alpha$. For $0 < \alpha_1 < 1$, the randomized critical region $S_0 \geq \hat{F}_N^{-1}(1-\alpha_1)$ has the same level α as the nonrandomized critical region $S_0 \geq F^{-1}(1-\alpha)$, or equivalently the critical regions $\hat{p}_N(S_0) \leq \alpha$ and $\hat{G}_N(S_0) \leq \alpha_1$ have the same level as the critical region $G(S_0) \equiv 1 - F(S_0) \leq \alpha$.

4.2 Monte Carlo tests with nuisance parameters

We will now study the case where the distribution of the test statistic S depends on nuisance parameters. We consider a family of probability spaces $\{(\mathcal{L}, \mathcal{A}_{\mathcal{L}}, P_\theta) : \theta \in \Omega\}$ and suppose that S is a real valued $\mathcal{A}_{\mathcal{L}}$-measurable function whose distribution is determined by $P_{\bar{\theta}}$ where $\bar{\theta}$ is the "true" parameter vector. We wish to test the hypothesis

$$H_0 : \bar{\theta} \in \Omega_0,$$

where Ω_0 is a nonempty subset of Ω. Again we take a critical region of the form $S \geq c$, where c is a constant which does not depend on θ. The critical region $S \geq c$ has *level* α if and only if

$$P_\theta[S \geq c] \leq \alpha, \forall \theta \in \Omega_0,$$

or equivalently,

$$\sup_{\theta \in \Omega_0} P_\theta[S \geq c] \leq \alpha.$$

Firthermore, $S \geq c$ has *size* α when

$$\sup_{\theta \in \Omega_0} P_\theta[S \geq c] = \alpha.$$

If we define the distribution and p-value functions,

$$F[x|\theta] = P_\theta[S \leq x], x \in \bar{R},$$

$$G[x|\theta] = P_\theta[S \geq x], x \in \bar{R},$$

where $\theta \in \Omega$, it is again easy to see that the critical regions

$$\sup_{\theta \in \Omega_0} G[S|\theta] \leq \alpha(c),$$

where $\alpha(c) \equiv \sup_{\theta \in \Omega_0} G[c|\theta]$, and

$$S \geq \sup_{\theta \in \Omega_0} F^{-1}[(1 - G[c|\theta])^{+}|\theta] \equiv \bar{c}$$

are equivalent to $S \geq c$ in the sense that $c \leq \bar{c}$.

We shall now extend proposition 5 in order to allow for the presence of nuisance parameters. For that purpose, we consider a real random variable S_0 and random vectors of the form

$$S(N, \theta) = (S_1(\theta), \ldots, S_N(\theta))', \theta \in \Omega,$$

all defined on a common probability space $(\mathcal{L}, \mathcal{A}_{\mathcal{L}}, P)$, such that the variables $S_0, S_1(\bar{\theta}), \ldots, S_N(\bar{\theta})$ are exchangeable for some $\bar{\theta} \in \Omega$, each one with distribution function $F[x|\bar{\theta}] = P[S_0 \leq x]$. Typically, S_0 will refer to a test statistic computed from observed data when the true parameter vector is $\bar{\theta}$ (i.e., $\theta = \bar{\theta}$), while $S_1(\theta), \ldots, S_N(\theta)$ will refer to i.i.d replications of the test statistic obtained independently (e.g., by simulation) under the assumption that the parameter vector is θ (i.e., $P[S_i(\theta) \leq x] = F[x|\theta]$). The notation $S_i(\theta)$ does not mean that the value of θ is required for computing the test statistic $S_i(\theta)$: it simply indicates that the distribution function of $S_i(\theta)$ is $F[x|\theta]$. Let also

$$\hat{F}_N[x|\theta] \equiv \hat{F}_N[x; S(N, \theta)],\ \hat{G}_N[x|\theta] \equiv \hat{G}_N[x; S(N, \theta)],$$

$$\hat{p}_N[x|\theta] = \frac{N\hat{G}_N[x|\theta]}{N+1}$$

and $\hat{F}_N^{-1}[x|\theta]$ be defined as in proposition 5 and suppose the variables $\sup\{\hat{G}_N[S_0|\theta] : \theta \in \Omega_0\}$ and $\inf\{\hat{F}_N[S_0|\theta] : \theta \in \Omega_0\}$ are $\mathcal{A}_{\mathcal{L}}$-measurable (where $\emptyset \neq \Omega_0 \subseteq \Omega$). We then get the following proposition.

Proposition 6 *Under the above assumptions and notations, set $S_0(\bar{\theta}) = S_0$ and suppose that*

$$P[S_i(\bar{\theta}) = S_j(\bar{\theta})] = 0, \quad \textit{for } i \neq j,\ i, j = 0, 1, \ldots, N.$$

If $\bar{\theta} \in \Omega_0$, then for $0 \leq \alpha_1 \leq 1$,

$$P[\sup\{\hat{G}_N[S_0|\theta] : \theta \in \Omega_0\} \leq \alpha_1] \leq P[\inf\{\hat{F}_N[S_0|\theta] : \theta \in \Omega_0\} \geq 1 - \alpha_1] \leq \frac{I[\alpha_1 N] + 1}{N+1}$$

where

$$P[\inf\{\hat{F}_N[S_0|\theta] : \theta \in \Omega_0\} \geq 1 - \alpha_1] = P[S_0 \geq \sup\{\hat{F}_N^{-1}[1 - \alpha_1|\theta] : \theta \in \Omega_0\}]$$

for $0 < \alpha_1 < 1$, *and*

$$P[\sup\{\hat{p}_N[S_0|\theta] : \theta \in \Omega_0\} \leq \alpha] \leq \frac{I[\alpha(N+1)]}{N+1}, \textit{for } 0 \leq \alpha \leq 1.$$

Following the latter proposition, if we choose α_1 and N so that equation 5 holds, the critical region $\sup\{\hat{G}_N[S_0|\theta] : \theta \in \Omega_0\} \leq \alpha_1$ has level α irrespective of the presence of nuisance parameters in the distribution of the test statistic S under the null hypothesis $H_0 : \bar{\theta} \in \Omega_0$. The same also holds if we use the (almost) equivalent critical regions $\inf\{\hat{F}_N[S_0|\theta] : \theta \in \Omega_0\} \geq 1 - \alpha_1$ or $S_0 \geq \sup\{\hat{F}_N^{-1}[1 - \alpha_1|\theta] : \theta \in \Omega_0\}$. We shall call such tests maximized Monte Carlo (MMC) tests.

We will now describe how to make exact the standard tests developped in sections 2 and 3. We consider for instance the Wald statistic established in section 2.3:

$$S = \epsilon_T^W = T(\hat{\theta}_{1T}^H)'\hat{W}_1^{*-1}(\hat{\theta}_{1T}^H)$$

We want to generate by simulation N different values of this test statistic denoted by S. Each simulated statistic simulée S_i $(i = 1, \ldots, N)$ requires H simulated series each one of length T generated under the structural model for a given value of θ: $[\tilde{y}_t^h(\theta; z_0^h)\ \ t = 0, \ldots, T]$ for $h = 1, \ldots, H$. Moreover, inorder to obtain N independent replications of the statistic S, it will be necessary to simulate also N samples of quasi-true data for some value échantillons de pseudo-vraies données pour une certaine valeur θ^0, i.e. $[\hat{y}_t^{(i)}(\theta^0; z_0^{(i)})\ \ t = 0, \ldots, T]_{i=1,\ldots,N}$ based on independent drawings of ϵ_t, $(\hat{\epsilon}_1^{(i)}, ..., \hat{\epsilon}_T^{(i)})$ and on initial values $z_0^{(i)}$ such that:

$$\hat{\beta}_T^{(i)}(\theta^0) = \arg\max_{\beta \in B} \mathcal{Q}_T((\hat{y}_T^1(\theta^0))^{(i)}, x_T^1, \beta), i = 1, ..., N.$$

For each simulated statistic S_i the following optimization problem will need to be solved:

$$\min_{\theta \in \Theta}\left[\hat{\beta}_T^{(i)}(\theta^0) - \frac{1}{H}\sum_{h=1}^{H}\tilde{\beta}_T^h(\theta, z_0^h)\right]'\hat{\Omega}_T^{(i)}(\theta^0)\left[\hat{\beta}_T^{(i)}(\theta^0) - \frac{1}{H}\sum_{h=1}^{H}\tilde{\beta}_T^h(\theta, z_0^h)\right],\ i = 1, \ldots, N$$

which yields $(\hat{\theta}_{1T}^H)^{(i)}(\theta^0)$ for $i = 1, \ldots, N$. This yields in total N optimization problems of this form and $N \times H$ auxiliary optimization problems to obtain

$\tilde{\beta}_T^h$. We then get the vector $S(N, \theta^0) = (S_1(\theta^0), \ldots,$ $S_N(\theta^0))$ where each $S_i(\theta^0)$ is computed from this estimator $(\hat{\theta}_{1T}^H)^{(i)}(\theta^0)$. In the Wald statistic considered here:

$$S_i(\theta^0) = T(\hat{\theta}_{1T}^H)'^{(i)}(\theta^0)\hat{W}_{1H(i)}^{*-1}(\theta^0)(\hat{\theta}_{1T}^H)^{(i)}(\theta^0) \qquad i = 1, \ldots, N.$$

The sample p-value function $\hat{G}_N[\ .\ |S(N, \theta^0)]$ is then built from this vector $S(N, \theta^0) = (S_1(\theta^0), \ldots, S_N(\theta^0))$ for each $\theta^0 \in \Omega_0$. The function $\hat{G}_N[\ .\ |S(N, \theta^0)]$ is a function of θ^0 through $S(N, \theta^0)$ the p-values will therefore be maximized over the set Ω_0 of the admissible values of θ^0 i.e.

$$\sup\{\hat{G}_N[\ S_0\ |S(N, \theta^0)] : \theta^0 \in \Omega_0\}.$$

But as $\hat{G}_N[\ .\ |S(N, \theta^0)]$ is a piece-wise function, maximizing $\sup\{\hat{G}_N[S_0|S(N, \theta^0)] : \theta^0 \in \Omega_0\}$ requires specific algorithms capable of optimizing nondifferentiable functions. We will resort to "simulated annealing" algorithm which optimizes from a stochastic search of the values of θ^0 in the set Ω_0 (Goffe, Ferrier and Rogers, 1994). The best thing to do would be to maximize the p-values over the larger possible set of admissible values of the nuisance parameter θ^0. But this raises numerical difficulties mostly in this context of computationnally intensive simulated methods as we have just seen. Thus a serious efficiency problem raises from simulated methods. We will search for that to restrict the set of admissible values for θ^0 to only consistent estimators.

5 Tests de Monte Carlo asymptotiques

In this section, we propose simplified approximate versions of the procedures proposed in the previous section when a consistent point or set estimate of θ is available. To do this, we shall need to reformulate the setup used previously in order to allow for an increasing sample size.

5.1 Monte Carlo tests based on consistent set estimators

Consider

$$S_{T0}, S_{T1}, \ldots, S_{TN}, T \geq I_0, \theta \in \Omega,$$

real random variables all defined on a common probability space $(\mathcal{L}, \mathcal{A}_{\mathcal{L}}, P)$ and set

$$S_T(N, \theta) = (S_{T1}(\theta), \ldots, S_{TN}(\theta)), \quad T \geq I_0.$$

We will be primarily interested by situations where the variables $S_{T0}, S_{T1}(\bar{\theta}), \ldots, S_{TN}(\bar{\theta})$ are exchangeable for some $\bar{\theta} \in \Omega$ each one with a distribution function $F_T[x|\bar{\theta}] = P[S_{T0} \leq x]$. Here S_{T0} will normally refer to a test statistic with distribution $F_T[.|\theta]$ based on a sample of size T, while $S_{T1}(\theta), \ldots, S_{TN}(\theta)$ i.i.d. replications of the same test statistic obtained independently under the assumption that the parameter vector is $\theta : P[S_{Ti}(\theta) \leq x] = F_T[x|\theta],\ i = 1, \ldots, N$. Let also

$$\hat{F}_{TN}[x|\theta] = \hat{F}_N[x; S_T(N, \theta)],\ \hat{G}_{TN}[x|\theta] = \hat{G}_N[x; S_T(N, \theta)],$$

$$\hat{p}_{TN}[x|\theta] = \frac{N\hat{F}_{TN}[x|\theta] + 1}{N + 1},$$

and let $\hat{F}_{TN}^{-1}[x|\theta]$ be defined as in proposition 5. We consider first the situation p-values are maximized over a subset C_T of Ω (e.g., a confidence set for θ) instead of Ω_0. Consequently, we introduce the following assumption: $C_T, T \geq I_0$, is a sequence of (possibly random) subsets of Ω such that $\sup\{\hat{G}_{TN}[S_{T0}|\theta] : \theta \in C_T\}$ and $\inf\{\hat{F}_{TN}[S_{T0}|\theta] : \theta \in C_T\}$ are $\mathcal{A}_{\mathcal{L}}$-measurable for all $T \geq I_0$ where $\emptyset \neq \Omega_0 \subseteq \Omega$. Then we have the following proposition.

Proposition 7 *(Asymptotic validity of confidence-set restricted maximized Monte Carlo tests. Continuous distributions). Under the previous assumptions and notations, set $S_{T0}(\bar{\theta}) = S_{T0}$ and suppose*

$$P[S_{Ti}(\bar{\theta}) = S_{Tj}(\bar{\theta})] = 0 \ \text{ for } i \neq j \text{ and } i, j = 0, 1, \ldots, N,$$

for all $T \geq I_0$, *and let* $C_T, T \geq I_0$ *be a sequence of (possibly random) subsets of* Ω *such that*

$$\lim_{T\to\infty} P[\bar{\theta} \in C_T] = 1. \tag{6}$$

If $\bar{\theta} \in \Omega_0$, *then*

$$\begin{aligned}
\lim_{T\to\infty} P[\sup\{\hat{G}_{TN}[S_{T0}|\theta] : \theta \in C_T\} \leq \alpha_1] &\leq \lim_{T\to\infty} P[\inf\{\hat{F}_{TN}[S_{T0}|\theta] : \theta \in C_T\} \geq 1-\alpha_1] \\
&= \lim_{T\to\infty} P\left[S_{T0} \geq \sup\{\hat{F}_{TN}^{-1}[1-\alpha_1|\theta] : \theta \in C_T\}\right] \\
&\leq \frac{I[\alpha_1 N]+1}{N+1}
\end{aligned} \tag{7}$$

and

$$\lim_{T\to\infty} P[\sup\{\hat{p}_{TN}[S_{T0}|\theta] : \theta \in C_T\} \leq \alpha] \leq \frac{I[\alpha(N+1)]}{N+1}, \quad \textit{pour } 0 \leq \alpha \leq 1.$$

It is quite easy to find a consistent set estimate of $\bar{\theta}$ whenever a consistent point estimate $\hat{\theta}_T$ of $\bar{\theta}$ is available. For instance, if $\hat{\theta}$ is a consistent estimator of $\bar{\theta}$, with $\bar{\theta} \in \Omega_0$, a possible confidence set for $\bar{\theta}$ could be a ball of fixed radius centred on $\hat{\theta}_T$ satisfying the condition

$$\lim_{T\to\infty} P[\|\hat{\theta}_T - \bar{\theta}\| < \epsilon] = 1, \forall \epsilon > 0, \tag{8}$$

where $\Omega \subseteq R^k$ and where $\|.\|$ is the Euclidean norm in R^k. Then any set of the form $C_T = \{\theta \in \Omega : \|\hat{\theta}_T - \theta\| < c\}$ where c is a fixed positive constant, which does not depend on T, satisfies (6). We can also consider balls whose radius decreases with sample size T. More generally, if there is a sequence of (possibly random) matrices A_T and a non-negative exponent δ such that

$$\lim_{T\to\infty} P[T^\delta \|A_T(\hat{\theta}_T - \bar{\theta})\|^2 < c] = 1, \forall c > 0,$$

then any set of the form

$$\begin{aligned}
C_T &= \{\theta \in \Omega : (\hat{\theta}_T - \theta)' A_T' A_T (\hat{\theta}_T - \theta) < c/T^\delta\} \\
&= \{\theta \in \Omega : \|A_T(\hat{\theta}_T - \theta)\|^2 < c/T^\delta\}, c > 0
\end{aligned}$$

satisfies (6) since in this case,

$$\begin{aligned}
P[\bar{\theta} \in C_T] &= P_{\bar{\theta}}[(\hat{\theta}_T - \bar{\theta})' A_T' A_T (\hat{\theta}_T - \bar{\theta}) < c/T^\delta] \\
&= P[T^\delta (\hat{\theta}_T - \bar{\theta})' A_T' A_T (\hat{\theta}_T - \bar{\theta}) < c] \underset{T\to\infty}{\longrightarrow} 1.
\end{aligned}$$

6 Empirical applications

We will apply in this section Monte Carlo tests procedures to a stochastic volatility model and to a brownian motion with drift.

6.1 Geometric Brownian motion: Indirect Inference estimation and Monte Carlo tests

6.1.1 Geometric Brownian motion with drift estimated by indirect inference

The price y_t of the underlying asset is assumed to satisfy the stochastic differential equation:

$$\frac{dy_t}{y_t} = \mu dt + \sigma dW_t \ , \tag{9}$$

where W_t is a standard brownian motion. μ and σ are the drift and volatility parameters respectively. Let

$$\theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \tag{10}$$

denote the parameter of interest. The equation (9) will be the structural model under which we will perform the simulations. Let (M) denote it. By applying Ito's formula, we get the equivalent form:

$$d(\log y_t) = (\mu - \frac{\sigma^2}{2})dt + \sigma dW_t \ . \tag{11}$$

We deduce from (11) the exact discretized version of (9), which corresponds to a random walk with drift in the log-price:

$$\log y_t = \log y_{t-1} + (\mu - \frac{\sigma^2}{2}) + \sigma \epsilon_t \ , \quad (\epsilon_t) \sim IIN(0,1), \tag{12}$$

and to a lognormal distribution for the price. We may also introduce the direct Euler approximation of equation (9):

$$\begin{aligned} y_t &= y_{t-1} + \mu^* y_{t-1} + \sigma^* y_{t-1} \epsilon_t^* \\ &= (1+\mu^*) y_{t-1} + \sigma^* y_{t-1} \epsilon_t^* \ , \quad (\epsilon_t^*) \sim IIN(0,1). \end{aligned} \tag{13}$$

It gives an autoregressive form for (y_t) with some conditional heteroscedasticity. This will be the instrumental model which replaces the structural model in the estimation procedure. Let (M^*) denote it and

$$\beta = \begin{pmatrix} \mu^* \\ \sigma^* \end{pmatrix} \tag{14}$$

the auxiliary parameter. The auxiliary model criterion to optimize will be denoted by $L_T^*(y_1^T, \beta)$ which corresponds to the log-likekihood function based on the instrumental model (M^*) but evaluated in the data simulated under the true model (M).

$$\begin{aligned}
\tilde{\beta}_T^h(\theta, y_0^h, \tilde{\epsilon}^h) &= \arg\max_{\beta \in B} L_T^*(\tilde{y}_t^h(\theta, y_0^h, \tilde{\epsilon}^h)_1^T, \beta),\ h = 1, \ldots, H \\
&= \arg\max_{\beta \in B} \Big[-\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T} \log(\sigma^{*2} \tilde{y}_{t-1}^h(\theta, y_0^h, \tilde{\epsilon}^h)^2) \\
&\quad - \frac{1}{2}\sum_{t=1}^{T} \frac{(\tilde{y}_t^h(\theta, y_0^h, \tilde{\epsilon}^h) - (1+\mu^*)\tilde{y}_{t-1}^h(\theta, y_0^h, \tilde{\epsilon}^h))^2}{\sigma^{*2}(\tilde{y}_{t-1}^h(\theta, y_0^h, \tilde{\epsilon}^h))^2} \Big]
\end{aligned} \tag{15}$$

Then the simulated data come from the initial model (M) for a small time unit δ. Let us rewrite model (M) as :

$$dy_t = \mu y_t dt + \sigma y_t dW_t \tag{16}$$

So by integrating (16) we obtain

$$\begin{aligned}
\int_t^{t+\delta} dy_s &= \mu \int_t^{t+\delta} y_s ds + \sigma \int_t^{t+\delta} y_s dW_s \\
y_{t+\delta} - y_t &\simeq \mu y_t \delta + \sigma y_t \sqrt{\delta}\epsilon_{t+\delta}
\end{aligned} \tag{17}$$

More precisely, we define the process $(y_t^{(\delta)})$ such that:

$$y_t^{(\delta)} = y_{k\delta}^{(\delta)} \text{ if } \quad k\delta \leq t < (k+1)\delta$$

where

$$y_{(k+1)\delta}^{(\delta)} = y_{k\delta}^{(\delta)} + (\mu y_{k\delta}^{(\delta)})\delta + (\sigma y_{k\delta}^{(\delta)})\sqrt{\delta}\epsilon_{(k+1)\delta}^{(\delta)} \tag{18}$$

and $(\epsilon_k^{(\delta)},\ k$ varying) is a Gaussian white noise with unit variance. Then for each parameter value θ, we may simulate using (18) some values $\tilde{y}_{k\delta}^{(\delta)h}(\theta, y_0^h, \tilde{\epsilon}_k^{(\delta)h})$, $k =$

$1, \ldots, [T/\delta]$, $h = 1, \ldots, H$ and deduce simulated values for the observation dates by just selecting the values corresponding to integer indexes:

$$\tilde{y}_t^h(\theta, y_0^h, \tilde{\epsilon}^h) = \tilde{y}_t^{(\delta)h}(\theta, y_0^h, \tilde{\epsilon}_k^{(\delta)h}) \quad [i.e.\text{with } k = t/\delta]. \tag{19}$$

δ will be taken equal to $1/10$ and $y_0^h = y_{k\delta}^{(\delta)h}$ with $k = 1$. We will have to simulate $H \times [T/\delta]$ perturbation vectors $\tilde{\epsilon}_k^{(\delta)h}$. For instance, for $T = 150$ and $\delta = 1/10$ we will need $150/\frac{1}{10}$ i.e. 1500 perturbation vectors for each simulation h.

So the objective function used to calibrate θ where

$$\hat{\theta}_T^H(y_0, \tilde{\epsilon}_T^H, \Omega) = \begin{pmatrix} \hat{\mu}_T^H(\Omega) \\ \hat{\sigma}_T^H(\Omega) \end{pmatrix}$$

is:

$$\begin{pmatrix} \hat{\mu}_T^H(\Omega) \\ \hat{\sigma}_T^H(\Omega) \end{pmatrix} = \arg\min \left[\hat{\beta}_T - \frac{1}{H}\sum_{h=1}^{H} \tilde{\beta}_T^H(\theta, y_0^h, \tilde{\epsilon}^h)\right]' \hat{\Omega}_T \left[\hat{\beta}_T - \frac{1}{H}\sum_{h=1}^{H} \tilde{\beta}_T^H(\theta, y_0^h, \tilde{\epsilon}^h)\right] \tag{20}$$

where $\tilde{\epsilon}_T^H = (\tilde{\epsilon}_t^h)_{t=1,\ldots,T}^{h=1,\ldots,H}$ and $\hat{\beta}_T = \hat{\beta}_T(\theta^0)$ is obtained from optimizing $L_T^*(y_1^T, \beta)$ based on the true observations $(y_1, \ldots, y_T)$. $\hat{\Omega}_T$ is a consistent estimator of the inverse of the asymptotic covariance matrix of $\sqrt{T}\left[\hat{\beta}_T - \frac{1}{H}\sum_{h=1}^{H} \tilde{\beta}_T^H(\theta, y_0^h, \tilde{\epsilon}^h)\right]$.

6.1.2 The estimators

The parameters'estimators The estimators which maximize the log-likelihood function (15) are solution of the following equations. Let

$$\tilde{\beta}_T^h(\theta, y_0^h, \tilde{\epsilon}^h) = \begin{pmatrix} \tilde{\mu}^*(\theta) \\ \tilde{\sigma}^*(\theta) \end{pmatrix}$$

be the quasi maximum likelihood estimators. Let $\tilde{y}_t^h$ and $\tilde{y}_{t-1}^h$ stand for $\tilde{y}_t^h(\theta)$ and $\tilde{y}_{t-1}^h(\theta)$ At first, let us consider the estimator of σ^* i.e.

$$\frac{\partial L_T^*}{\partial \sigma^*}(\tilde{\sigma}^*) = 0$$

$$\Leftrightarrow \frac{-1}{2}\sum_{t=1}^{T}\frac{2\sigma^*(\tilde{y}_{t-1}^h)^2}{\sigma^{*2}(\tilde{y}_{t-1}^h)^2} - \frac{1}{2}\sum_{t=1}^{T}\frac{-2\sigma^*}{\sigma^{*4}}.\left(\frac{\tilde{y}_t^h - (1+\mu^*)\tilde{y}_{t-1}^h}{\tilde{y}_{t-1}^h}\right)^2 = 0$$

$$\Leftrightarrow \frac{-T}{\sigma^*} + \frac{1}{\sigma^{*3}}\sum_{t=1}^{T}\left(\frac{\tilde{y}_t^h - (1+\mu^*)\tilde{y}_{t-1}^h}{\tilde{y}_{t-1}^h}\right)^2 = 0$$

$$\Leftrightarrow -T + \frac{1}{\sigma^{*2}}\sum_{t=1}^{T}\left(\frac{\tilde{y}_t^h - (1+\mu^*)\tilde{y}_{t-1}^h}{\tilde{y}_{t-1}^h}\right)^2 = 0\ . \tag{21}$$

So the estimator is

$$\begin{aligned} \tilde{\sigma}_{ML}^{*2} &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\tilde{y}_t^h - (1+\mu^*)\tilde{y}_{t-1}^h}{\tilde{y}_{t-1}^h}\right)^2 \\ &= T^{-1}\|(\tilde{y}_{-1}^h)^{-1}.(\tilde{y}^h - (1+\mu^*)\tilde{y}_{-1}^h)\|^2 \end{aligned} \tag{22}$$

The concentrated log-likelihood is then given by

$$\begin{aligned} L_T^*(\tilde{y}, \mu^*)^c &= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log[T^{-1}\|(\tilde{y}_{-1}^h)^{-1}.(\tilde{y}^h - (1+\mu^*)\tilde{y}_{-1}^h)\|^2] - \frac{T}{2} \\ &= -\frac{T}{2}\log(2\pi) + \frac{T}{2}\log(T) - \frac{T}{2}\log[\|(\tilde{y}_{-1}^h)^{-1}.(\tilde{y}^h - (1+\mu^*)\tilde{y}_{-1}^h)\|^2] - \frac{T}{2} \end{aligned} \tag{23}$$

The estimator of μ^* is then obtained by maximizing the concentrated log-likelyhood function, i.e.

$$\max_{\mu^*} L_T^{*c}(\tilde{y}, \mu^*)$$

$$\Leftrightarrow \min_{\mu^*} \log(\|(\tilde{y}_{-1}^h)^{-1}.(\tilde{y}^h - (1+\mu^*)\tilde{y}_{-1}^h)\|^2)$$

$$\Leftrightarrow \min_{\mu^*} \|(\tilde{y}_{-1}^h)^{-1}.(\tilde{y}^h - (1+\mu^*)\tilde{y}_{-1}^h)\|^2 \tag{24}$$

since the log-function is an increasing function of its arguments. So from the first-order conditions

$$\frac{\partial L_T^{*c}}{\partial \mu^*} = -2\sum_{t=1}^{T} \frac{\tilde{y}_{t-1}^h(\tilde{y}_t^h - (1+\tilde{\mu}_{ML}^*)\tilde{y}_{t-1}^h)}{(\tilde{y}_{t-1}^h)^2} = 0$$

$$\Leftrightarrow \sum_{t=1}^{T}(\tilde{y}_{t-1}^h)^{-1}\tilde{y}_t^h - \sum_{t=1}^{T}(1+\tilde{\mu}_{ML}^*) = 0$$

$$\Leftrightarrow \sum_{t=1}^{T}(1+\tilde{\mu}_{ML}^*) = \sum_{t=1}^{T}\left(\frac{\tilde{y}_t^h}{\tilde{y}_{t-1}^h}\right) \tag{25}$$

which yields

$$(1+\tilde{\mu}_{ML}^*) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\tilde{y}_t^h}{\tilde{y}_{t-1}^h}\right) \tag{26}$$

we obtain

$$\tilde{\mu}_{ML}^{*h}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\tilde{y}_t^h(\theta)}{\tilde{y}_{t-1}^h(\theta)}\right) - 1\ . \tag{27}$$

The estimator of σ^{*2} is then given by

$$\tilde{\sigma}_{ML}^{*2(h)}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\tilde{y}_t^h(\theta)}{\tilde{y}_{t-1}^h(\theta)} - \frac{1}{T}\sum_{t=1}^{T}\frac{\tilde{y}_t^h(\theta)}{\tilde{y}_{t-1}^h(\theta)}\right)^2 \tag{28}$$

A consistent estimator of Ω In order to calibrate θ, we need a consistent estimator of $\Omega^* = J_0(I_0 - K_0)^{-1}J_0$. J_0 can be consistently estimated by

$$-\frac{1}{T}\frac{\partial^2 L_T^*}{\partial\beta\partial\beta'}(y(\theta^0)_1^T, \hat{\beta}_T(\theta^0)) \tag{29}$$

where $y(\theta^0)_1^T$ denotes the true observations and $\hat{\beta}_T(\theta^0)$ the estimator based on the true observations. So the second derivative of L_T^* is

$$\begin{aligned}\frac{\partial^2 L_T^*}{\partial\sigma^*\partial\sigma^{*'}} &= \frac{T}{\hat{\sigma}^{*2}} - \frac{3\hat{\sigma}^{*2}}{\hat{\sigma}^{*6}}\sum_{t=1}^{T}\left(\frac{y_t(\theta^0) - (1+\hat{\mu}^*)y_{t-1}(\theta^0)}{y_{t-1}(\theta^0)}\right)^2 \\ &= \frac{T}{\hat{\sigma}^{*2}} - \frac{3}{\hat{\sigma}^{*4}}.T\hat{\sigma}^{*2} \\ &= -\frac{2T}{\hat{\sigma}^{*2}}\ . \end{aligned} \tag{30}$$

So

$$-\frac{1}{T}\frac{\partial^2 L_T^*}{\partial\sigma^*\partial\sigma^{*'}} = \frac{2}{\hat{\sigma}^{*2}}\ . \tag{31}$$

On the other hand,

$$\frac{\partial^2 L_T^*}{\partial\sigma^*\partial\mu^{*'}} = -\frac{T}{\hat{\sigma}^*} - \frac{2}{\hat{\sigma}^{*3}}\sum_{t=1}^{T}\left(y_{t-1}(\theta^0)\frac{y_t(\theta^0) - (1+\hat{\mu}^*)y_{t-1}(\theta^0)}{(y_{t-1}(\theta^0))^2}\right) \tag{32}$$

$$-\frac{1}{T}\frac{\partial^2 L_T^*}{\partial\sigma^*\partial\mu^{*'}} = 0 \tag{33}$$

since from the first-order conditions we have

$$\left.\frac{\partial L_T^*}{\partial\mu^*}\right|_{\substack{\hat{\mu}^* \\ \hat{\sigma}^*}} = \sum_{t=1}^{T}\left(y_{t-1}(\theta^0)\frac{y_t(\theta^0) - (1+\hat{\mu}^*(\theta^0))y_{t-1}(\theta^0)}{\sigma^{*2}(y_{t-1}(\theta^0))^2}\right) = 0. \tag{34}$$

The second-order derivative of L_T^* with respect to μ^* is given by

$$\begin{aligned}-\frac{1}{T}\frac{\partial^2 L_T^*}{\partial\mu^*\partial\mu^{*'}} &= -\frac{1}{T}\left(-\frac{1}{\hat{\sigma}^{*2}}\sum_{t=1}^{T}\frac{y_{t-1}(\theta^0)}{y_{t-1}(\theta^0)}\right) \\ &= \frac{1}{\hat{\sigma}^{*2}} \end{aligned} \tag{35}$$

So the estimator of J_0 is given by

$$\hat{J}(\theta^0) = \hat{J}_0 = \begin{pmatrix} \frac{1}{\hat{\sigma}^{*2}(\theta^0)} & 0 \\ 0 & \frac{2}{\hat{\sigma}^{*2}(\theta^0)} \end{pmatrix}. \tag{36}$$

On the other hand, a consistent estimator of $(I_0 - K_0)$ as $T \to \infty$ is given by

$$\frac{T}{H} \sum_{h=1}^{H} (W_h - \bar{W})(W_h - \bar{W})' \tag{37}$$

where

$$\begin{aligned} W_h &= \frac{\partial L_T^*}{\partial \beta}(y^h(\tilde{\theta})_1^T, \hat{\beta}_T) \\ &= \begin{pmatrix} \frac{\partial L_T^*}{\partial \mu^*}\left(y^h(\tilde{\theta})_1^T, \begin{pmatrix} \hat{\mu}^*(\theta^0) \\ \hat{\sigma}^*(\theta^0) \end{pmatrix}\right) \\ \frac{\partial L_T^*}{\partial \sigma^*}\left(y^h(\tilde{\theta})_1^T, \begin{pmatrix} \hat{\mu}^*(\theta^0) \\ \hat{\sigma}^*(\theta^0) \end{pmatrix}\right) \end{pmatrix} \end{aligned} \tag{38}$$

and

$$\bar{W} = \frac{1}{H} \sum_{h=1}^{H} W_h = \begin{pmatrix} \frac{1}{H} \sum_{h=1}^{H} \frac{\partial L_T^*}{\partial \mu^*}\left(y^h(\tilde{\theta})_1^T, \begin{pmatrix} \hat{\mu}^*(\theta^0) \\ \hat{\sigma}^*(\theta^0) \end{pmatrix}\right) \\ \frac{1}{H} \sum_{h=1}^{H} \frac{\partial L_T^*}{\partial \sigma^*}\left(y^h(\tilde{\theta})_1^T, \begin{pmatrix} \hat{\mu}^*(\theta^0) \\ \hat{\sigma}^*(\theta^0) \end{pmatrix}\right) \end{pmatrix} \tag{39}$$

where $\tilde{\theta}$ is any consistent estimator of θ (for instance $\tilde{\theta}_T^H(Id)$).

The case considered here does not include any exogenous variables, in this case $K_0 = 0$ and I_0 is equal to

$$\lim_{T \leftarrow \infty} V_0 \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial L^*}{\partial \beta} (\tilde{y}_t^h(\theta, y_0^h, \tilde{\epsilon}^h), \tilde{y}_{t-1}^h(\theta, y_0^h, \tilde{\epsilon}^h), \beta) \right] = \lim_{T \leftarrow \infty} V_0 \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial L_t^*}{\partial \beta}(\beta) \right]. \tag{40}$$

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

$$\hat{\Gamma}_0 + \sum_{k=1}^{K} \left(1 - \frac{1}{K+1}\right)(\hat{\Gamma}_k + \hat{\Gamma}_k') \tag{41}$$

with

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^{T} \frac{\partial L_{t-k}^*}{\partial \beta}(\hat{\beta}_T) \frac{\partial L_t^*}{\partial \beta'}(\hat{\beta}_T) \tag{42}$$

A consistent estimator of W_1^* Let W_1^* denote the inverse of the asymptotic covariance matrix of the indirect estimator $\sqrt{T}\hat{\theta}_{1T}^H = \sqrt{T}\hat{\mu}_T^H$.

$$V_{as}[\sqrt{T}\hat{\theta}_{1T}^H] = (1+\frac{1}{H})(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} \ , \tag{43}$$

so

$$\begin{aligned} W_1^* &= [V_{as}(\sqrt{T}\hat{\theta}_{1T}^H)]^{-1} \\ &= (1+\frac{1}{H})^{-1}(A_{11} - A_{12}A_{22}^{-1}A_{21}) \ , \end{aligned} \tag{44}$$

where

$$A_{ij} = \frac{\partial b'}{\partial \theta_i}\Omega^* \frac{\partial b}{\partial \theta_j} \ . \tag{45}$$

By using the fact that

$$\frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0) = J_0^{-1}\frac{\partial^2 L_\infty^*}{\partial \beta \partial \theta'}(F_0, G_0, \theta_0, \beta_0) \tag{46}$$

and that

$$\Omega^* = J_0(I_0 - K_0)^{-1}J_0 \ . \tag{47}$$

So a consistent estimator of A_{ij} will be given by

$$\hat{A}_{ij} = (\frac{\partial^2 L_T^*}{\partial \beta \partial \theta_i})'(\frac{T}{H}\sum_{h=1}^{H}(W_h - \bar{W})(W_h - \bar{W})')^{-1}(\frac{\partial^2 L_T^*}{\partial \beta \partial \theta_j}) \tag{48}$$

where $\frac{\partial^2 L_T^*}{\partial \beta \partial \theta'} = \frac{\partial^2 L_T^*}{\partial \beta \partial \theta'}(y^h(\theta)_T^1, x_T^1, \hat{\beta}_T(\theta^0))$ evaluated at the indirect estimator $\hat{\theta}_T^H$ where $y^h(\theta)_T^1$ is a simulated path of y based on the parameter θ. When $K_0 = 0$, $\Omega^* = J_0(I_0)^{-1}J_0$,

$$\hat{A}_{ij} = (\frac{\partial^2 L_T^*}{\partial \beta \partial \theta_i})'(\hat{I}_0)^{-1}(\frac{\partial^2 L_T^*}{\partial \beta \partial \theta_j}) \ , \tag{49}$$

and when $\Omega^* = Id$,

$$\hat{A}_{ij} = (\frac{\partial^2 L_T^*}{\partial \beta \partial \theta_i})'\hat{J}_0^{-1}\hat{J}_0^{-1}(\frac{\partial^2 L_T^*}{\partial \beta \partial \theta_j}) \ . \tag{50}$$

$$\frac{\partial^2 L_T^*}{\partial\beta\partial\theta'} = \begin{pmatrix} \frac{\partial}{\partial\theta'}[\frac{\partial L_T^*}{\partial\mu^*}(y^h(\theta)_T^1, x_T^1, \hat{\beta}_T(\theta^0))]_{|\hat{\theta}_T^H} \\ \frac{\partial}{\partial\theta'}[\frac{\partial L_T^*}{\partial\sigma^*}(y^h(\theta)_T^1, x_T^1, \hat{\beta}_T(\theta^0))]_{|\hat{\theta}_T^H} \end{pmatrix} \tag{51}$$

is a matrix 2 by 2. The following second derivatives call on numerical derivatives.

$$\frac{\partial^2 L_T^*}{\partial\mu^*\partial\theta'} = \frac{\partial}{\partial\theta'}\left[\sum_{t=1}^{T}\left(y_{t-1}^h(\theta).\frac{(y_t^h(\theta)-(1+\hat{\mu}^*(\theta^0))y_{t-1}^h(\theta))}{\hat{\sigma}^{*2}(\theta^0)(y_{t-1}^h(\theta))^2}\right)\right]_{|\hat{\theta}_T^H} \tag{52}$$

$$\frac{\partial^2 L_T^*}{\partial\sigma^*\partial\theta'} = \frac{\partial}{\partial\theta'}\left[\frac{-T}{\hat{\sigma}^*(\theta^0)} + \frac{1}{\hat{\sigma}^{*3}(\theta^0)}\sum_{t=1}^{T}\left(\frac{y_t^h(\theta)-(1+\hat{\mu}^*(\theta^0))y_{t-1}^h(\theta)}{y_{t-1}^h(\theta)}\right)^2\right]_{|\hat{\theta}_T^H} \tag{53}$$

Finally

$$\hat{W}_1^* = (1+\frac{1}{H})^{-1}(\hat{A}_{11}-\hat{A}_{12}\hat{A}_{22}^{-1}\hat{A}_{21})\ . \tag{54}$$

will denote a consistent estimator of the inverse of the asymptotic covariance matrix of the indirect inference estimator $\sqrt{T}\hat{\theta}_{1T}^H = \sqrt{T}\hat{\mu}_T^H$.

A consistent estimator of W_H^* A consistent estimator of

$$W_H^* = (1+\frac{1}{H})[\frac{\partial L_\infty^*}{\partial\theta\partial\beta'}(I_0-K_0)^{-1}\frac{\partial L_\infty^*}{\partial\beta\partial\theta'}]^{-1} \tag{55}$$

is

$$\hat{W}_H^* = (1+\frac{1}{H})[\frac{\partial L_T^*}{\partial\theta\partial\beta'}(I_0 \hat{-} K_0)^{-1}\frac{\partial L_T^*}{\partial\beta\partial\theta'}]^{-1} \tag{56}$$

obtained as soon as we have a consistent estimator of $\frac{\partial^2 L_\infty^*}{\partial\theta\partial\beta'}$ which can be obtained as aforementionned by numerical derivation of $\frac{\partial L_T^*}{\beta'}[(y^s(\theta)_T^1, x_T^1, \hat{\beta}_T(\theta^0)]$ 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.

6.1.3 Test Statistics

We are interested in testing a null hypothesis on the drift μ of the initial model (M) but the volatility parameter σ is a nuisance parameter for our test. The interest parameter is

$$\theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$$

and the instrumental parameter is:

$$\beta = \begin{pmatrix} \mu^* \\ \sigma^* \end{pmatrix}$$

The null hypothesis we are interested in is $H_0 : \mu = \mu_0$. We will denote here $\hat{\theta}_T^H(y_0, \tilde{\epsilon}_T^H, \Omega)$ by $\hat{\theta}_T^H$ for short. Let

$$(\hat{\theta}_T^H) = \begin{pmatrix} \hat{\theta}_{1T}^H \\ \hat{\theta}_{2T}^H \end{pmatrix} = \begin{pmatrix} \hat{\mu}_T^H \\ \hat{\sigma}_T^H \end{pmatrix} \tag{57}$$

denote the unrestricted indirect estimator and

$$(\hat{\theta}_T^{0H}) = \begin{pmatrix} \mu_0 \\ \hat{\sigma}_T^{0H} \end{pmatrix} \tag{58}$$

the restricted estimator.

The Wald statistic

$$\begin{aligned} \xi_T^W &= T(\hat{\theta}_{1T}^H)'\hat{W}_1^*(\hat{\theta}_{1T}^H) \\ &= T(\hat{\mu}_T^H - \mu_0)'\hat{W}_1^*(\hat{\mu}_T^H - \mu_0) \end{aligned} \tag{59}$$

where $\hat{W}_1^*$ is a consistent estimator of the inverse of the asymptotic covariance-variance matrix of $\sqrt{T}\hat{\mu}_T^H$.

The Score statistic The Score function is

$$\begin{aligned} \mathcal{D}_T &= \frac{\partial \tilde{\beta}'_{HT}}{\partial \theta_1}(\hat{\theta}_T^{0H})\hat{\Omega}^*[\hat{\beta}_T - \tilde{\beta}_{HT}(\hat{\theta}_T^{0H})] \\ &= \frac{\partial \tilde{\beta}'_{HT}}{\partial \mu}\begin{pmatrix} \mu_0 \\ \hat{\sigma}_T^{0H} \end{pmatrix}\hat{\Omega}^*\left[\hat{\beta}_T - \tilde{\beta}_{HT}\begin{pmatrix} \mu_0 \\ \hat{\sigma}_T^{0H} \end{pmatrix}\right] \end{aligned} \tag{60}$$

which yields for the score statistic

$$\xi_T^S = T\mathcal{D}_T\mathcal{A}\mathcal{D}_T \tag{61}$$

where $\mathcal{A}$ is a consistent estimator of $[Var(\sqrt{T}\mathcal{D}_T)]^{-1}$.

The statistic based on the difference between the optimal values of the objective function

$$\begin{aligned}\xi_T^C &= \frac{TH}{1+H}\left[\hat{\beta}_T - \tilde{\beta}_{HT}\begin{pmatrix}\mu_0 \\ \hat{\sigma}_T^{0H}\end{pmatrix}\right]'\hat{\Omega}^*\left[\hat{\beta}_T - \tilde{\beta}_{HT}\begin{pmatrix}\mu_0 \\ \hat{\sigma}_T^{0H}\end{pmatrix}\right] \\ &\quad -\frac{TH}{1+H}\left[\hat{\beta}_T - \tilde{\beta}_{HT}\begin{pmatrix}\hat{\mu}_T^H \\ \hat{\sigma}_T^H\end{pmatrix}\right]'\hat{\Omega}^*\left[\hat{\beta}_T - \tilde{\beta}_{HT}\begin{pmatrix}\hat{\mu}_T^H \\ \hat{\sigma}_T^H\end{pmatrix}\right] \qquad (62)\end{aligned}$$

6.1.4 Implementation of Monte Carlo Tests for the above Test Statistics

At first the implementation of the Wald Statistic and the Statistic based on the difference between the optimal values of the objective function, with the identity matrix for the metric Ω in the Indirect Inference optimization criterion together with the simulation number $H = 1$ is much less costly in computation time than the use of the optimal weight matrix Ω^* together with $H > 1$. Moreover the use of this optimal weight matrix Ω^* is mainly justified through asymptotic efficiency considerations and thereby not relevant in finite samples. The optimization of the p-value over a subset of the parameters space under the null hypothesis due to the presence of nuisance parameters is relatively easy to perform through the GAUSS version of Simulated Annealing optimization programm and yields a test with the exact level α. Here it has been performed at the levels of 5% and 1% for the Wald Statistic and for the Statistic based on the difference between the optimal values of the objective function under the null hypothesis $\mu = 0$ with the nuisance parameter corresponding to σ. The startup value for the nuisance parameter in the Simulated Annealing optimization programm has been set up to a consistent restricted estimator. The Maximized Monte Carlo Test to test the null hypothesis $\mu = 0$ together with only one nuisance parameter σ to optimize is thus theoretically and computationally feasible in this Indirect Inference framework for both of these statistics. The next step would be to investigate the way to reduce the computation time by picking up an intermediary estimator instead of the Indirect Inference estimator obtained through the use of the Optmum algorithm of the version GAUSS 3.2.18. On the other hand, an exchangeability property could be explored in this Indirect Inference setting in order to reduce the simulations length.

References

Andersen, T., H.-J. Chung, and B. Sorensen (1997). Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. Working paper, Northwestern University and Brown University.

Bansal, R., A. Gallant, R. Hussey, and G. Tauchen. Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics 66*, 251–287.

Bansal, R., A. Gallant, R. Hussey, and G. Tauchen (1993). *Computational Aspects of Nonparametric Simulation Estimation*. Boston Massachussets: Kluwer Academic Publishers. In D.A. Belsley (eds). Computationnal techniques for Econometrics and Economic Analysis.

Barnard, G. (1963). The spectral analysis of point processes. *Journal of the Royal Statistical Society B 25*, 294. Discussion on M.S. Bartlett.

Birnbaum, Z. (1974). *Computers and Unconventionnal Test-statistics*. Philadelphia : SIAM. In F. Proschan and R.J. Serfling.(eds). Reliablity and Biometry.

Duffie, D. and K. Singleton (1993). Simulated moments estimation of markov models of asset prices. *Econometrica 61*, 929–952.

Dufour, J.-M. (1996). Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics. Working paper, Université de Montréal.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics, with applications to structural and dynamic models. *Econometrica*. Forthcoming.

Dufour, J.-M. and L. Khalaf (1996a). Monte Carlo tests for contemporaneous correlation of disturbances in multi-equation models. Technical Report 16, Université de Montréal.

Dufour, J.-M. and L. Khalaf (1996b). Simulation based finite and large sample inference methods in seemingly unrelated regressions. Technical Report 60, Université de Montréal. C.R.D.E.

Dufour, J.-M. and L. Khalaf (1996c). Simulation based finite and large sample inference methods in simultaneous equations. Technical Report 20, Université de Montréal. C.R.D.E.

Dufour, J.-M. and J. Kiviet (1996). Exact tests for structural change in first-order dynamic models. *Journal of Econometrics 70*, 39–68.

Dufour, J.-M. and J. Kiviet (1997). Exact inference methods in first-order autoregressive distributed lag models. *Econometrica*. Forthcoming.

Dwass, M. Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics 28*, 181–187.

Edgington, E. (1987). *Randomization Tests* (2 Edition ed.). New York : Marcel Dekker.

Foutz, R. (1980). A method for constructing exact tests from test statistics that have unknown null distributions. *Journal of Statistical Computation and Simulation 10*, 187–193.

Gallant, A. (1987). *Nonlinear Statistical Models*. New York : John Wiley & Sons.

Gallant, A. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*, 657–681.

Gallant, A. and H. White (1988). *Estimation and Inference for Nonlinear Dynamic Models*. New York : Blackwell.

Goffe, W., G. Ferrier, and J. Rogers (1994). Global optimisation of statistical functions. *Journal of Econometrics 60*, 65–99.

Gouriéroux, C., A. Monfort, and E. Renault (1993). Indirecte inference. *Journal of Applied Econometrics 8*, 85–118.

Jöckel, K. (1986). Finite sample properties and asymptotic efficiency of Monte Carlo tests. *Annals of Statistics 14*, 336–347.

Kiviet, J. and J.-M. Dufour (1996). Exact tests in single equation autoregressive distributed lag models. *Journal of Econometrics*. Forthcoming.

Marriott, F. Barnarb's Monte Carlo tests: How many simulations? *Applied Statistics 28*, 75–77.