Monte Carlo Tests and Regularized Indirect Inference for a Stochastic Volatility Model

Jean-Marie Dufour* Pascale Valéry†

9th December 2004

ABSTRACT

In this paper we assess the finite sample properties of the standard asymptotic tests built on the indirect estimator in a specific context when non-identification problems are encountered. More specifically, the auxiliary estimator used in the indirect estimation is based on moment conditions which become non-identifying under the null hypothesis of homoscedasticity of the volatility process. Under such non-identification issues, the covariance matrix of the auxiliary estimator and that of the Wald statistic become occasionally singular. As a result, the usual invertibility technique breaks down making the statistics non implementable. To remedy this problem, we resort to two competitive regularization techniques: the first one has originally been proposed by Lutkepohl and Burda (1997)(LB), for multi-step noncausality and amounts to estimating a reduced rank covariance matrix, taking its Moore-Penrose generalized inverse and then modifying the Wald statistic accordingly. Alternatively, we propose a slightly different regularization technique which consists in keeping the eigenvalues of the estimated covariance matrix which are greater than a predetermined threshold and setting the smaller ones to the threshold. Then we can still be proceeding as usual to invert the covariance matrix thus regularized. To account for the lack of identification, we implement the proposed regularization techniques at two distinct levels: one to overcome the singularity problem of the covariance matrix of the auxiliary estimator appearing in the indirect criterion as a way to improve the explanation power of such a *weak* instrument and improve thereby the performance of all the test statistics together whereas the second-step regularization only helps the Wald statistic as it was originally suggested by LB. Unlike the nonregularized test statistics, the modified statistics can always be computed. They also demonstrate more power than their nonregularized counterparts. We further combine the asymptotic modified tests with simulation-based inference techniques such as the technique of Monte Carlo tests [see Dwass (1957), Barnard (1963), Dufour (2002)] to control for the size. We provide some simulation results before illustrating the methodology on the Standard and Poor's Composite Price Index (SP),daily, 1928-87.

*CIREQ,CIRANO,Université de Montreal,email: jean.marie.dufour@umontreal.ca
†HEC-Montréal,IFM^2,email: pascale.valery@hec.ca

Key words: Indirect inference; Exact tests; non-identification, Moore-Penrose inverse.
JEL classification: C1, C13, C12, C32, C15

Contents

1. Introduction

Indirect estimation was proposed by Gouriéroux, Monfort and Renault (1993) [henceforth GMR] as an estimation and inference procedure for models having complex formulations or untractable likelihood functions. Basically, it consists in optimizing an auxiliary criterion that does not directly provide a consistent estimator of the parameter of interest. A consistent estimator is then obtained by means of simulations. Indirect inference techniques belong to the class of modern statistical procedures which exploit Monte Carlo methods to derive powerful estimator and tests for complex models. Bootstrap and Monte Carlo Markov chain methods belong to this class and more generally any simulation-based inference technique is a potential candidate. The only requirement for implementing simulation-based procedures is that the model or the statistic can be simulated. In this framework, various test statistics have been proposed to make inference on the parameters of interest of the structural model, among which the Wald-type statistic and the likelihood ratio-type statistic [see GMR (1993)]. 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 [see Dufour (1997)]. Further under nonregular conditions, asymptotic tests are known to have incorrect size even asymptotically in part of the parameter space [see Andrews (1987), Gregory and Veal (1985), Breusch and Schmidt (1988), Lutkepohl and Burda (1997), (henceforth LB)]. More specifically, LB examined the behavior of the Wald statistic for multi-step causality for finite order vector autoregressive (VAR) processes. In such a setup, multi-step noncausality implies a set of highly nonlinear restrictions on the VAR coefficient matrices. For this type of nonlinear restrictions standard Wald tests fail to have limiting χ^2-distributions in general. In this respect, LB proposed modifications to the Wald statistic which ensure an asymptotic χ^2-distribution under the null hypothesis. In this paper, we examine the behavior of the standard asymptotic indirect tests as proposed by GMR in a situation close to non-identification. Under such non-identification issues, the covariance matrix of the auxiliary estimator and that of the Wald statistic become occasionally singular. As a result, the usual invertibility technique breaks down making the statistics non implementable. To remedy this problem, we resort to two competitive regularization techniques: the first one has originally been proposed by LB (1997) for multi-step noncausality and amounts to estimating a reduced rank covariance matrix and then modifying the Wald statistic accordingly. When the covariance matrix becomes singular, LB replace the usual inverse by its Moore-Penrose generalized inverse, by setting to zero the inverse of eigenvalues of the estimated covariance matrix when they drop below a threshold. Alternatively, we propose a slightly different regularization technique which consists in keeping the eigenvalues of the estimated covariance matrix which are greater than a predetermined threshold and setting the smaller ones to the threshold instead of zero. Then we can still be proceeding as usual to invert the covariance matrix thus regularized. Unlike LB who did study multi-step noncausality in VAR processes, we implement these regularization techniques in a completely different setup by investigating the behavior of the Wald-type statistic and the likelihood ratio-type (LR) statistic under nonidentification conditions for testing a null hypothesis of homoskedasticity in the volatility process of a lognormal stochastic volatility (SV) model. Indeed, the auxiliary estimator which enters the second step objective criterion in the indirect estimation procedure is based on moment conditions which become non-identifying under the null hypothesis of homoscedasticity of the volatility process. To account for the lack of identification, we implement the proposed regularization techniques at two distinct levels: one to overcome the singularity problem of the covariance matrix of the auxiliary estimator appearing in the indirect criterion as a way to improve the explanation power of such a *weak* instrument and improve thereby the performance of all the test statistics together whereas the second-step regularization only helps the Wald statistic as it was originally suggested by LB. Unlike the nonregularized test statistics, the modified statistics can always be computed. They also demonstrate more power than their nonregularized counterparts. Therefore the regularization techniques appear very useful in two ways, by keeping the statistics computable in nonregular conditions and further by increasing power performances when compared with their nonregularized counterparts.

The main contribution of the paper is to assess the finite sample properties of the modified statistics in a setup of non-identification. We further combine the asymptotic modified tests with simulation-based inference techniques to control for the size. In this scope, Monte Carlo tests (MC, henceforth)[see Dwass (1957), Barnard (1963), Birnbaum (1974)], more specifically *maximized* Monte Carlo (MMC, henceforth), [see Dufour (2002)] tests are mainly a simulation-based-inference technique which provides exact tests irrespective of the presence of nuisance parameters in the distribution of the test statistic.

The paper is organized as follows. In Section 2 we describe the structural model which is assumed to be the true data generating process and takes the form of a log-normal stochastic volatility model with an autoregressive conditional mean part of order one. Section 3 describes the auxiliary estimator chosen to implement the indirect estimator while the testing problem as well as the regularization techniques are exposed in Section 4. In Section 5 we review briefly the methodology of Monte Carlo tests which still provides reliable inference for distributions which are not pivotal even asymptotically. We provide some simulation results in Section 6 before illustrating the methodology on the Standard and Poor's Composite Price Index (SP),daily, 1928-87 in Section 7. Finally, we conclude in Section 8.

2. Framework

The data generating process we consider in this paper takes the form of an autoregression whose innovations are scaled by an unobservable volatility process, usually distributed as a lognormal autoregression [see Gallant, Hsieh and Tauchen (1997), Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994), Danielsson (1994)]. More specifically, let the structural model be described as follows.

$$y_t = \alpha + c y_{t-1} + \exp(w_t/2) r_y z_t \ , \quad |c| < 1 \tag{2.1}$$

$$w_t = aw_{t-1} + r_w v_t \ , \quad |a| < 1 \ . \tag{2.2}$$

Let $\theta = (c, r_y, a, r_w)'$ denote the parameter of interest of the structural model above. Let $\alpha = \mu_y(1-c)$ where μ_y is the conditional mean of y. We shall call the model represented by equations (2.1)-(2.2) the stochastic volatility model with an autoregressive mean part of order one [AR(1)-SV for short].

Assumption 2.1 *The vectors* $(z_t, v_t)'$, $t \in \mathbb{N}$ *are i.i.d. according to a* $N(0, I_2)$ *distribution.*

Assumption 2.2 *The process* $s_t = (y_t, w_t)'$ *is strictly stationary.*

The process is Markovian of order 1. The process $\{y_t\}$ is observed whereas $\{w_t\}$ is regarded as latent. Accordingly, the joint density of the vector of observations $\bar{y} = (y_1, \ldots, y_T)$ is not available in closed-form since it requires evaluating an integral with dimension equal to the whole path of the latent volatilities. Let $F_T(\theta) = F(\bar{y}, \theta) = P[Y_1 \le y_1, \ldots, Y_T \le y_T|\theta]$ denote its unknown distribution function.

3. Estimation by Indirect Inference

In this section we describe the indirect estimation procedure chosen to estimate the parameter of interest θ of the AR(1)-SV model given at equations (2.1)-(2.2). For a more exhaustive description of the method see Gouriéroux, Monfort and Renault (1993). Usually the likelihood function of the structural model is unknown or untractable as it is the case here for our AR(1)-SV model and therefore requires to resort on an approximated model called the auxiliary model which is simpler to estimate. The auxiliary model should closely approximate the distribution of the observed data but is not required to nest it. However, if the auxiliary model nests the structural model then the estimator is as efficient as maximum likelihood [see Gallant and Tauchen (1996)].

The auxiliary estimator we have chosen is a moment estimator available in closed form which rules out the cumbersome task of optimizing the first step auxiliary criterion. We recall the expression of this moment estimator which has been developed in Dufour and Valéry (2004). It is essentially a two-step estimation method whose first step consists in providing a consistent estimate of the parameter of the mean equation and thereby of the residuals $\hat{u}_T = u(\hat{c}_T)$. Then in a second step we apply a method-of-moments to the residuals $\hat{u}_T$ to get the estimate of the parameter of interest. More specifically, let us recall the equations defining the estimator in the proposition below based on the three following moments:

$$\mu_2 = E(u_t^2) = r_y^2 \exp[r_w^2/2(1-a^2)] \ , \tag{3.3}$$

$$\mu_4 = E(u_t^4) = 3r_y^4 \exp[2r_w^2/(1-a^2)] \ , \tag{3.4}$$

and

$$\mu_{2,2}(1) = E[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1-a)] \ , \tag{3.5}$$

where $\beta = (c, r_y, a, r_w)'$. Solving the above moment equations yields the proposition.

Proposition 3.1 Estimating equations.
Under Assumptions **2.1**-**2.2**, *we have:*

$$a = \frac{[\log(\mu_{2,2}(1)) - \log(3) - 4\log(\mu_2) + \log(\mu_4)]}{\log(\frac{\mu_4}{3(\mu_2)^2})} - 1 \ , \tag{3.6}$$

$$r_y = \frac{3^{1/4}\mu_2}{\mu_4^{1/4}} \ , \tag{3.7}$$

$$r_w = \left(\log(\frac{\mu_4}{3(\mu_2)^2})(1-a^2)\right)^{1/2} \ . \tag{3.8}$$

Given the latter proposition, it is easy to compute a method-of-moments estimator for $\beta = (c, r_y, a, r_w)'$ replacing the theoretical moments by sample counterparts based on the residuals $\hat{u}_t$. Let $\hat{\beta}_T$ denote the method-of-moments estimator of β. Typically, $E(u_t^2)$, $E(u_t^4)$ and $E(u_t^2 u_{t-1}^2)$ are approximated by:

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \quad \hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4 \ , \quad \hat{\mu}_2(1) = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2$$

respectively. Let $\bar{g}_T(\hat{U}) = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{U})$ with $g_t(\hat{U}) = (y_t, \hat{u}_t^2, \hat{u}_t^4, \hat{u}_t^2\hat{u}_{t-1}^2)'$. Then

$$\bar{g}_T(\hat{U}) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} y_t \\ \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 \\ \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^4 \\ \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2\hat{u}_{t-1}^2 \end{pmatrix} \ . \tag{3.9}$$

Let $M_T(\beta)$ denote the auxiliary criterion :

$$M_T(\beta) \equiv [\bar{g}_T(\hat{U}) - \mu(\beta)]'\hat{\Omega}_1[\bar{g}_T(\hat{U}) - \mu(\beta)] \ , \tag{3.10}$$

with $\mu(\beta) = (\mu_y, \mu_2, \mu_4, \mu_{2,2}(1))'$. We denote by $\hat{\beta}_T$ the solution to this problem:

$$\hat{\beta}_T = \arg\max_{\beta \in B} M_T(\beta) \ . \tag{3.11}$$

Since the moment conditions $[\bar{g}_T(\hat{U}) - \mu(\beta)]$ and β have the same dimensions, then we can take $\hat{\Omega}_1 = I_4$ and hence, $\hat{\beta}_T$ is easily obtained by solving the equation: $\bar{g}_T(\hat{U}) = \mu(\beta)$. On the other hand, as the moments used to compute the estimator are not the true population moments but their sample counterparts computed on the residuals, we can correct for the approximation error by simulating the true unknown binding function $b(F_T(\theta), \theta)$. Then in a second step we can obtain the indirect estimator $\hat{\theta}_T$ by minimizing the second step criterion:

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} [\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_T^{(s)}(\theta)]' \hat{\Omega}_2 [\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_T^{(s)} \hat{\theta}_T] \tag{3.12}$$

where $\hat{\Omega}_2$ is a positive definite matrix defining the metric. $\hat{\beta}_T$ denotes the estimate of the auxiliary parameter based on the observed data whereas $\tilde{\beta}_T^{(s)}(\theta)$ denotes the corresponding estimate for a data set simulated under the structural model for a value θ. A consistent estimator of the metric is given by:

$$\Omega_2 = J(\theta)' I(\theta)^{-1} J(\theta) \tag{3.13}$$

where

$$J(\theta) = -\frac{\partial^2 M_T(\beta)}{\partial\beta\partial\beta'}(y_T(\theta), \hat{\beta}_T) \; , \tag{3.14}$$

and

$$I(\theta) = \Gamma_0(\theta) + \sum_{k=1}^{K}(1 - \frac{k}{K+1})(\Gamma_k(\theta) + \Gamma_k'(\theta)) \; , \tag{3.15}$$

with

$$\Gamma_k(\theta) = \frac{1}{T}\sum_{t=k+1}^{T} \frac{\partial M_{t-k}(\beta)}{\partial \beta'}(y_T(\theta), \hat{\beta}_T) \frac{\partial M_t'(\beta)}{\partial \beta}(y_T(\theta), \hat{\beta}_T) \; . \tag{3.16}$$

Replacing θ by any root-n consistent estimator yields a consistent estimate of Ω. The metric Ω_2 defined at equation (3.13) is the metric which minimizes the asymptotic variance-covariance matrix of the indirect estimator, yielding the optimal estimator. This asymptotic variance-covariance matrix is given by:

$$W_S^* = (1 + \frac{1}{S})\left(\frac{\partial^2 M_\infty}{\partial\theta\partial\beta'}(F(\theta_0), \theta_0, \beta_0) I(\theta_0)^{-1} \frac{\partial^2 M_\infty}{\partial\beta\partial\theta'}(F(\theta_0), \theta_0, \beta_0)\right)^{-1} \; . \tag{3.17}$$

A consistent estimator of W_S^* is obtained as:

$$W_S^* = (1 + \frac{1}{S})\left(\frac{\partial^2 M_T}{\partial\theta\partial\beta'}(\hat{\theta}_T, \hat{\beta}_T) I(\hat{\theta}_T)^{-1} \frac{\partial^2 M_T}{\partial\beta\partial\theta'}(\hat{\theta}_T, \hat{\beta}_T)\right)^{-1} \; . \tag{3.18}$$

as soon as we can compute the derivative of $\frac{\partial M_T}{\partial \beta}$ with respect to θ. The computation of such a derivative has to be made numerically.

4. Hypothesis tests

In this section we are interested in testing general hypotheses such as $H_0 : F \in \mathcal{H}_0$, where $\mathcal{H}_0$ is a subset of all possible distributions for the stochastic volatility model (2.1)- (2.2), that is,

$$\mathcal{H}_0 \equiv \{F(.)\ : F(\bar{y}) = F_0(\bar{y}|\psi(\theta))\ and\ \psi(\theta) = 0\}\ , \tag{4.19}$$

where $\psi(\theta)$ is a $p \times 1$ continuously differentiable function of θ. H_0 is usually abbreviated as: $H_0 :\ \psi(\theta) = 0$. The derivative of the constraints $P(\theta) = \frac{\partial\psi}{\partial\theta'}$ has full row rank. Let $\hat{\theta}_T$ be the unrestricted indirect estimator and $\hat{\theta}^c_T$ the constrained estimator obtained by minimizing the second step indirect criterion under H_0:

$$M_T(\theta) \equiv \min_{\theta\in\Theta}[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}^{(s)}_T(\theta)]'\hat{\Omega}_2[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}^{(s)}_T\hat{\theta}_T]\ . \tag{4.20}$$

To test the null hypothesis we are interested in two types of statistics, the Wald statistic and the likelihood ratio statistic. The Wald statistic is defined as

$$\lambda_1 = T\psi(\hat{\theta}_T)'[\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']^{-1}\psi(\hat{\theta}_T) \tag{4.21}$$

where $\hat{P} = P(\hat{\theta}_T)$, $\hat{I} = I(\hat{\theta}_T)$, $\hat{J} = J(\hat{\theta}_T)$.

The likelihood ratio statistic is basically the difference between the optimal values of the objective function as defined below.

$$\lambda_2 = \frac{TS}{1+S}[M_T(\hat{\theta}^c_T) - M_T(\hat{\theta}_T)]\ . \tag{4.22}$$

In the simulations we shall focus on a particular form of the constraint, i.e. $\psi(\theta) = (0,1)\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \theta_2$ and the null hypothesis $H_0 :\ \psi(\theta) = 0$ simplifies to $H_0 :\ \theta_2 = 0$, (e.g. $\theta_1 \equiv (a, r_w)'$). This specific form $H_0 :\ (a, r_w)' = \underline{0}$ of the constraint corresponds to testing no heteroscedasticity in the volatility process against an alternative of stochastic volatility. However, when implementing the null hypothesis of homoscedasticity, a severe identification issue arises since under the null, the moment conditions which define the auxiliary estimator are no more identifying. Indeed, the three moment conditions (3.3), (3.4) and (3.5) reduce to only one relevant moment condition under H_0. This lack of identification creates some singularity problems for the covariance matrix of $\sqrt{T}[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}^{(s)}_T(\theta)] = (1+1/S)\Omega_2^{-1}$ through the non-invertibility of $I(\theta)$ defined at equations (3.15) and (3.16). Hence, the usual invertibility of the matrix fails occasionally. To remedy this problem, we implement two regularization techniques among which the Moore-Penrose generalized inverse of the corresponding matrix. The idea comes from LB(1997) to use the principal components associated with the largest eigenvalues of the estimated covariance matrix. To do so, let $\hat{\Sigma}$ be a suitable reduced rank consistent estimator of a covariance matrix Σ with eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_J$, and $\hat{V}$ an orthogonal matrix consisted of the associated eigenvectors, such that $\hat{\Sigma} = \hat{V}\hat{\Lambda}\hat{V}'$, where $\hat{\Lambda} = diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_J)$. For some $c > 0$, define $\hat{J}_c$ to be the number of $\hat{\lambda}_j > c$ and let $\hat{\Lambda}_c = diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_{\hat{J}_c}, 0, \ldots, 0)$. Moreover, define $\hat{\Lambda}^+_c = diag(\hat{\lambda}_1^{-1}, \ldots, \hat{\lambda}^{-1}_{\hat{J}_c}, 0, \ldots, 0)$. Then, the Moore-Penrose generalized inverse of $\hat{\Sigma}$ denoted by $\hat{\Sigma}^+$ is obtained as:

$$\hat{\Sigma}^+ = \hat{V}\hat{\Lambda}_c^+\hat{V}' \ . \tag{4.23}$$

$\hat{\Sigma}^+$ will alternately denote $\hat{\Sigma}^+ = I(\theta)^+$, or $\hat{\Sigma}^+ = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']^+$. When regularizing the estimated covariance matrices by taking their Moore-Penrose generalized inverse as proposed by LB (1997), the modified statistics will be referred to as λ_1^+ for the modified Wald statistic and λ_2^+ for the modified LR statistic. Alternately, to regularize the estimated covariance matrix $\hat{\Sigma} = \hat{V}\hat{\Lambda}\hat{V}'$, we propose instead to keep the estimated eigenvalues $\hat{\lambda}_j > c$ and set $\hat{\lambda}_j = c$ whenever they drop below the threshold c. For $c > 0$, let $\tilde{J}_c$ be the number of eigenvalues for which $\hat{\lambda}_j > c$. Let $\tilde{\Lambda}_c = diag(\hat{\lambda}_1, \dots, \hat{\lambda}_{\tilde{J}_c}, c, \dots, c)$. Then, the regularized covariance matrix is obtained as the sum of the non-regularized initial matrix and a regularizing matrix such that:

$$\tilde{\Sigma} = \frac{1}{2}[\hat{V}\hat{\Lambda}\hat{V}' + \hat{V}\tilde{\Lambda}_c\hat{V}'] \ . \tag{4.24}$$

Finally, the inverse of $\tilde{\Sigma}$ is obtained by taking a usual inverse defined for positive definite matrices. In particular, note that when all eigenvalues are greater than the threshold, the regularizing matrix coincides with the original matrix, that is: $\tilde{\Lambda}_c = \hat{\Lambda}$. $\tilde{\Sigma}$ will sequentially denote $\tilde{\Sigma} = I(\theta)$ and/or $\tilde{\Sigma} = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']$. Likewise, when using this regularization technique, the modified Wald statistic will be referred to as $\tilde{\lambda}_1$ and the modified LR statistic $\tilde{\lambda}_2$. Finally we also consider a combination of the two kinds of regularization by regularizing the covariance matrix $I(\theta)$ of the auxiliary estimator $\hat{\beta}_T$ through $\tilde{\Sigma}^{-1} = I(\theta)^{-1}$ and that of the Wald statistic by choosing the LB technique for $\hat{\Sigma}^+ = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']^+$. The modified statistics according to the combined regularization techniques will be denoted λ_1^C and λ_2^C. These generalized inverses will be built sequentially if necessary. The first one will help in regularizing the indirect criterion to account for "weak instruments", and thereby will benefit to all the statistics together whereas at the opposite only the Wald statistic will take advantage of the two inverses jointly when the covariance matrices become singular. In the remaining of the paper, we will compare the modified statistics with the original statistics proposed by GMR (1993). We will further compare the finite sample properties (in terms of size and power) of inference techniques based on asymptotic approximations to simulation-based inference techniques such as the technique of Monte Carlo tests (and its maximized version), as described in the following section.

5. Monte Carlo testing

The technique of Monte Carlo tests has originally been proposed by Dwass (1957) for implementing permutation tests and did not involve nuisance parameters. This technique has been extended by Barnard (1963) and Birnbaum (1974). It has the great attraction of providing *exact* (randomized) tests based on any statistic whose finite sample distribution may be intractable but can be simulated. We briefly review the methodology of Monte Carlo tests covering both cases, first without nuisance parameters and then with nuisance parameters as it is proposed in Dufour (2002). The technique of Monte Carlo tests provides a simple method allowing one to replace the unknown or untractable theoretical distribution $F(x|\theta)$ by its sample analogue based on the statistics $S_1(\theta), \ldots, S_N(\theta)$ simulated under the null hypothesis. The procedure can be designed as follows.

First we present the case without nuisance parameters which provides an *exact* test.

- STEP 1: Using the observed sample, we calculate the relevant statistic denoted by S_0.
- STEP 2: Using draws under H_0, we generate N simulated samples: $S_1, \ldots, S_N$.
- STEP 3: Then we compute the estimated survival function:

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

  and the associated p-value function

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

  If N is chosen so that $\alpha(N+1)$ is an integer, under H_0:

  $$P(\hat{p}_N[S_0] \leq \alpha) = \alpha \; ,$$

  yielding an *exact* test.

Second, in presence of nuisance parameters, Dufour (2002) proposes to maximize the nuisance parameters over the parameter space conformable with the null hypothesis. In this case the procedure is the following.

- STEP 1: To test the null hypothesis

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

  we use first the observed sample to calculate the relevant statistic denoted by S_0.
- STEP 2: For each $\theta \in \Omega_0$, we generate N replications of S: $S_1(\theta), \ldots, S_N(\theta)$.
- STEP 3: Using these simulations we compute the corresponding simulated p-value function:

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

Finally the p-value function $\hat{p}_N[S_0|\theta]$ as a function of θ is maximized over the parameter space. If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:

$$P[\sup\{\hat{p}_N(S_0|\theta) : \theta \in \Omega_0\} \leq \alpha] \leq \alpha \; ,$$

that is we control for the size. Such a technique which provides an *exact* test irrespective of the presence of nuisance parameters under the null hypothesis is called a *Maximized Monte Carlo* test (henceforth MMC) by Dufour (2002). A proof of this assertion can be found in Dufour (2002). In the simulation exercises below we will implement the test in two forms, one in a local maximized version we call (MMC) and another one when the nuisance parameters are evaluated at a consistent point estimate yielding a form of parametric Bootstrap we shall call (MC) tests. For the MMC version the nuisance parameters are maximized over a fine grid since there are only two nuisance parameters. When the nuisance parameters are numerous one can use simulated annealing [see Goffe, Ferrier and Rogers (1994)] an appropriate optimization algorithm which does not require differentiability. Indeed $\hat{G}_N[S_0|\theta]$ is step-type function which typically has zero derivatives almost everywhere, except on isolated points where it is not differentiable.

6. Simulation results

In this section, we implement the Wald test and the Likelihood ratio test for testing the null hypothesis of homoscedasticity in the volatility process, say, $H_0 : a = 0, r_w = 0$. The tests are performed in three ways. The first one uses the asymptotic χ^2 critical point $(\chi^2_{1-\alpha}(2))$, while the other ones are based on the simulated p-values. For the Monte Carlo test (hereafter, MC), the p-value is evaluated at a consistent restricted point estimate of the nuisance parameters. Concerning the maximized Monte Carlo test (hereafter, MMC), the p-value function is maximized over a neighborhood of the restricted estimate of the nuisance parameters and the null hypothesis is rejected each time the maximized p-value is less than the nominal level fixed at 5%. We assess the actual sizes of the tests averaged on 100 replications. The Monte Carlo tests are performed with $N = 19$ statistics simulated under the null hypothesis. The nuisance parameters have been set to $r_y = 0.4$ and $c = 0.95$ to produce a high level of persistence in the mean equation. In the simulations the drift parameter α has been fixed at 0.5 throughout the experiment. The simulations are run on the GAUSS software, 3.2.37 version. Concerning the regularization techniques, all the thresholds have been set to $c = 0.001$ to facilitate comparisons across methods. The LR statistic modified according to LB regularization and denoted by λ_2^+, will only benefit from the first regularization for computing $\hat{\Sigma}_1^+ = I(\hat{\theta})^+ = \hat{V}\hat{\Lambda}_{c_1}^+\hat{V}'$, with $\hat{\Lambda}_{c_1}^+ = diag(\hat{\lambda}_1^{-1}, \ldots, \hat{\lambda}_{\hat{J}_{c_1}}^{-1})$ and $\hat{J}_{c_1}$ denotes the number of $\hat{\lambda}_j$ of the estimated covariance matrix $I(\hat{\theta})$ which satisfy: $\hat{\lambda}_j > c_1$. The modified Wald statistic (denoted by λ_1^+), will benefit not only from $\hat{\Sigma}_1^+ = \hat{I}^+$, but also from $\hat{\Sigma}_2^+ = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']^+ = \hat{V}\hat{\Lambda}_{c_2}^+\hat{V}'$. Likewise, the LR statistic and the Wald statistic modified by our regularization technique are denoted $\tilde{\lambda}_2$ and $\tilde{\lambda}_1$ respectively, in the simulation experiment whereas the statistics using a combination of both techniques are denoted λ_1^C for the Wald and λ_2^C for the LR statistic.

6.1. Size analysis

First of all, we can see in Table 1 that the frequency at which the non-regularized Wald statistic fails is around 10% in small sample and it diminishes when the sample size increases. At the opposite the non-regularized LR statistic nearly never fails, less than 3% of failure. The rejection frequencies for the non-regularized procedures have been computed after excluding the cases when the usual inverses crash. Overall, the size distortions displayed by the non-regularized Wald statistic is not severe and varies between 7 and 9% but do not diminish when the sample size increases. Indeed, it is well-known [see Dufour (1997)], that the Wald test does not provide reliable inference under non-identification conditions. Dufour (1997) shows that the distribution of the Wald test cannot be bounded by any finite set of distribution functions when "weak instruments" are involved in IV regressions. This is precisely the case encountered here since the auxiliary estimator is built on weak instruments. Under $H_0 : a = 0, r_w = 0$, the moment conditions defining the auxiliary estimator are no more identifying. However, the attempts of regularization of the covariance matrices performed at two levels, at the estimation step when regularizing the indirect criterion and at the testing step when regularizing the covariance matrix in the Wald statistic seem to help the latter. Resorting to other kinds of inverse prevents the statistic from breaking down but also help in reducing the large standard errors. Our regularization technique better control for the size of the Wald test as well as the combined version when compared with that of LB. Indeed, the Wald statistic regularized with the technique proposed by LB still slightly overrejects in small samples and more than all other statistics including the non-regularized one. However, when the sample size increases, its overrejection is getting less severe. In such situations, simulation-based inference techniques such as Monte Carlo tests [see Dufour (2002)] help controlling for the size mostly for the modified Wald statistic λ_1^+ in small samples. At the opposite, the non-regularized LR statistic tends to underreject. For $T = 1000$, the non-regularized LR statistic never rejects the null hypothesis. The results support earlier works that both finite sample and asymptotic distributions of the LR test may also be modified when identifiability conditions are not satisfied [see Sargan (1983), Phillips (1989), Staiger and Stock (1994), and other references in Dufour (1997)]. However, Dufour (1997) shows that LR statistics have null distributions which can be bounded by a nuisance-parameter-free distribution (possibly derived from the Wilks Λ distribution), hence inference methods based on such statistics are more reliable. Further, the LR statistic is known to be robust to non-invariance problems unlike the Wald statistics [see Breusch and Schmidt (1988), Nelson and Savin (1990), Dagenais and Dufour (1991)]. Concerning the regularized LR statistics, there size performances are quite similar and help correcting for the underrejection. Moreover, we observe for the LR tests at $T = 500$ and $T = 2000$ in Table 1 that when MC tests whose distribution also depends on strong regularity conditions [see Dufour (2002)], cannot achieve in correcting for some over-rejections, the *maximized* MC test usually solves over-rejection problems.

6.2. Power analysis

We also study in Table 2 the power properties of the tests for an alternative hypothesis of stochastic volatility with a quite high persistence feature in the volatility process, namely $H_1: \ a = 0.9, r_w = 0.9$. The asymptotic tests have been corrected for size distortions. Monte Carlo tests are implemented with $N = 99$ simulated statistics since for power considerations, the number of simulated statistics may have an impact on gains in power. As expected, the Wald test is not consistent at all. When the sample size increases, the gains in power for the Wald statistic for the three procedures (Asy, MC, MMC), are not significant and are even diminishing for $\tilde{\lambda}_1$. This observation carries out a crucial message concerning the behavior of the Wald statistic in a context of (almost) non-identification when using "poor instruments". It is impossible to build a valid test based on the Wald statistic despite the various technical tools in hand, such as regularization techniques which may also contributes in correcting for "poor" standard errors. The Wald statistic is not *reformable* in situations close to non-identification [see Dufour (1997)]. As for the LR test, note first of all that $\tilde{\lambda}_2 = \lambda_2^C$, since the same regularization technique is used for the covariance matrix of the auxiliary estimator which is the only relevant regularization taken into account by the LR statistic. Note the very erratic behavior of the non-regularized LR statistic which support the fact that the regularization techniques help increasing the power performances significantly. This observation is particularly outstanding in large samples for $T = 1000, 2000$ where for instance $P[\lambda_2 > \chi^2_{1-\alpha(2)}|H_1] = 0.17$ compared with $P[\lambda_2^+ > \chi^2_{1-\alpha(2)}|H_1] = 0.57$ and $P[\tilde{\lambda}_2 > \chi^2_{1-\alpha(2)}|H_1] = 0.50$. We further observe that λ_2^+ demonstrates more power than $\tilde{\lambda}_2$ in small samples ($T = 200, 500$) but in large samples ($T = 1000, 2000$) their power performances become quite equivalent around 68% for λ_2^+ and 69% for $\tilde{\lambda}_2$. We further observe in Table 2, a loss in power for both versions of Monte Carlo tests w.r.t. their asymptotic counterparts. Indeed, there is always a loss of "power" of the simulated tests compared with the asymptotic ones due to the noise introduced by the simulations. In this respect, one has to be aware that the asymptotic tests remain infeasible and are considered as a benchmark useful for comparisons purposes. Indeed, implementing the asymptotic tests requires the prior knowledge of the nuisance parameters which is not available in practice. More specifically, the *maximized* Monte Carlo tests are always feasible in any contexts dealing with nuisance parameters and controls for size distortions. The only requirement of the procedure is that the test statistic can be simulated.

7. Empirical application

In this subsection we test the null hypothesis of no-persistence in the volatility from real data (Standard and Poor's Composite Price Index (SP), 1928-87).

7.1. Data

The data have been provided by Tauchen where Efficient Method of Moments have been used by Gallant, Hsieh and Tauchen to fit a standard stochastic volatility model. The data to which we fit the univariate stochastic volatility model is a long time series comprised of 16,127 daily observations, $\{\tilde{y}_t\}_{t=1}^{16,127}$, on adjusted movements of the Standard and poor's Composite Price Index, 1928-87. The raw series is the Standard and Poor's Composite Price Index (SP),daily, 1928-87. The raw series is converted to a price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and then adjusted for systematic calendar effects, that is, systematic shifts in location and scale due to different trading patterns across days of the week, holidays, and year-end tax trading. This yields a variable we shall denote y_t.

7.2. Results

To conduct the asymptotic tests, we use the asymptotic critical value of a $\chi^2_{1-\alpha}(2) = 5.99$ for a $\alpha = 5\%$ significance level. In Table 3, we observe that λ_1^+ and λ_1^C both reject the null hypothesis $H_0 : \ a = 0 \ \ r_w = 0$ of homoscedasticity in the volatility process whereas two other ones, that are $\tilde{\lambda}_1$ and λ_1 do not reject the null hypothesis. The same observation holds for simulated tests where this time λ_1 and λ_1^+ cannot reject H_0 at both level whereas Monte Carlo tests based on $\tilde{\lambda}_1$ and λ_1^C statistics do reject H_0 at $\alpha = 5\%$ and $\alpha = 1\%$. Once again, these controversial results obtained with the Wald statistic highlight the unreliable feature of the latter when making inference with *weak* instruments. As predicted by Dufour (1997), whatever powerful tools in hand, the Wald statistic is not reformable. Such a statistic cannot produce valid inference in situations close-to-non-identification. By contrast, the LR statistic still provides reliable inference in presence of *weak* instruments, even though its finite and asymptotic distribution may be modified. Our results reported in Table 3 for the LR statistics give evidence on this statement. Whatever LR statistic one considers, modified or non-modified, whatever tests one chooses, asymptotic or simulated, they all reject H_0 at $\alpha = 5\%$ and $\alpha = 1\%$. Therefore we can infer that the null hypothesis of homoscedasticity in the volatility observed on the Standard and Poor's Composite Price Index (SP),daily, 1928-87 can be rejected at both level of significance. Indeed, it is well-known that high-frequency financial data are time-varying and displays strong volatility clustering effects [see Engle (1982)].

8. Concluding remarks

To summarize, we provide regularization techniques of covariance matrices when these ones become singular and non invertible by resorting to some specific generalized inverses. As a result, the tests statistics remain computable in any situations. It is worth noting that the regularization techniques implemented here in the context of a stochastic volatility model estimated by indirect inference is not restricted to this particular framework but could be employed in more general models. On the other hand, despite the attempts to regularize the covariance matrix of the Wald statistic, it still provides invalid inference when the estimator of the parameter of interest is based on "weak instruments". Indeed, the distribution of the Wald statistic cannot be bounded by any finite set of distribution functions under non-identification problems. In such situations, *maximized* Monte Carlo tests can control for the size but at the cost of no power at all under the alternative. By contrast, the likelihood ratio test behave much better (both in size and power) in such situations even though its finite and asymptotic distributions may be modified.

Table 1: Actual size

LEVEL in % ($H_0 : a = 0, r_w = 0$)								
$c = 0.95, r_y = 0.4$								
	T=200				*T=500*			
	Asy NON reg.	**Asy**	**MC**	**MMC**	**Asy NON reg.**	**Asy**	**MC**	**MMC**
λ_1^+	8.8	11	6	3	7.9	11	6	1
$\tilde{\lambda}_1$	8.8	3	3	3	7.9	5	7	4
λ_1^C	8.8	2	4	4	7.9	5	4	3
failure	10	0	0	0	11	0	0	0
λ_2^+	1	2	5	2	4	4	7	0
$\tilde{\lambda}_2$	1	2	3	2	4	5	7	4
λ_2^C	1	2	3	2	4	5	7	4
failure	3	0	0	0	1	0	0	0
	T=1000				*T=2000*			
	Asy NON reg.	**Asy**	**MC**	**MMC**	**Asy NON reg.**	**Asy**	**MC**	**MMC**
λ_1^+	7.4	8	2	1	7.9	7	2	1
$\tilde{\lambda}_1$	7.4	6	2	2	7.9	6	5	4
λ_1^C	7.4	5	2	2	7.9	6	5	5
failure	6	0	0	0	4	0	0	0
λ_2^+	0	2	1	1	2	4	9	1
$\tilde{\lambda}_2$	0	1	1	1	2	4	7	4
λ_2^C	0	1	1	1	2	4	7	4
failure	30	0	0	0	0	0	0	0

References

ANDERSEN, T., AND B. SØRENSEN (1996): "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business and Economic Statistics*, 14(3), 328–352.

ANDERSEN, T. G., H.-J. CHUNG, AND B. E. SØRENSEN (1999): "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics*, 91, 61–87.

ANDREWS, D. (1987): "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348–358.

BARNARD, G. A. (1963): "Comment on "The Spectral Analysis of Point Processes" by M. S. Bartlett," *Journal of the Royal Statistical Society, Series B*, 25, 294.

Table 2: Actual power

POWER in % ($H_1 : a = 0.9, r_w = 0.9$)								
$c = 0.95, r_y = 0.4$								
	T=200				*T=500*			
	Asy NON reg.	**Asy**	**MC**	**MMC**	**Asy NON reg.**	**Asy**	**MC**	**MMC**
λ_1^+	10	5	8	5	1	9	11	3
$\tilde{\lambda}_1$	10	7	8	3	1	4	5	4
λ_1^C	10	6	11	6	1	8	15	10
failure	0	0	0	0	0	0	0	0
λ_2^+	44	45	26	20	39	48	28	21
$\tilde{\lambda}_2$	44	32	25	20	39	38	30	28
λ_2^C	44	32	25	20	39	38	30	28
	T=1000				*T=2000*			
	Asy NON reg.	**Asy**	**MC**	**MMC**	**Asy NON reg.**	**Asy**	**MC**	**MMC**
λ_1^+	0	6	5	2	0	5	4	1
$\tilde{\lambda}_1$	0	0	2	2	0	0	1	0
λ_1^C	0	8	11	4	0	10	8	5
failure	0	0	0	0	0	0	0	0
λ_2^+	17	57	51	46	42	68	61	50
$\tilde{\lambda}_2$	17	50	49	40	42	69	63	51
λ_2^C	17	50	49	40	42	69	63	51

Table 3: Empirical application

Standard and Poor's Composite Price index			
$H_0: \ a = 0 \ \ r_w = 0$			
	Asymptotic tests	*Monte Carlo tests*	
	S_0	N=19	N=99
λ_1	0.000772	0.249	0.23
λ_1^+	6.70	0.30	0.30
$\tilde{\lambda}_1$	5.20	0.05	0.01
λ_1^C	32155.84	0.05	0.01
λ_2	111.91	0.05	0.01
λ_2^+	18.94	0.05	0.01
$\tilde{\lambda}_2$	111.91	0.05	0.01

Birnbaum, Z. W. (1974): "Computers and Unconventional Test-Statistics," in *Reliability and Biometry*, ed. by F. Proschan, and R. J. Serfling, pp. 441–458. SIAM, Philadelphia, PA.

Breusch, T. S., and P. Schmidt (1988): "Alternative forms of the Wald test: How long is a piece of string ?," *Communications in statistics, Theory and Methods*, 17, 2789–2795.

Dagenais, M., and J.-M. Dufour (1991): "Invariance, Nonlinear Models and Asymptotic Tests," *Econometrica*, 59, 1601–1615.

Danielsson, J. (1994): "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics*, 61, 375–400.

Danielsson, J., and J.-F. Richard (1993): "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models," *Journal of Applied Econometrics*, 8, S153–S173.

Dufour, J.-M. (1995): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," *Journal of Econmetrics, forthcoming.*

——— (1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1389.

Dufour, J.-M., L. Khalaf, J.-T. Bernard, and I. Genest (2001): "Simulation-Based Finite-Sample Tests for Heteroscedasticity and ARCH Effects," *Journal of Econometrics, forthcoming.*

Dufour, J.-M., and M. Taamouti (2000): "Statistical Inference and Projection techniques in Simultaneous Equations Models," *Econometrica, forthcoming.*

Dufour, J.-M., and P. Valéry (2004): "On a simple closed-form estimator for a stochastic volatility model," Discussion paper, Université de Montréal.

Dwass, M. (1957): "Modified Randomization Tests for Nonparametric Hypotheses," *Annals of Mathematical Statistics*, 28, 181–187.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Foutz, R. V. (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., D. Hsieh, and G. Tauchen (1997): "Estimation of Stochastic Volatility Models with Diagnostics," *Journal of Econometrics*, 81(1), 159–192.

Gallant, A. R., and G. Tauchen (1996): "Which Moments to Match?," *Econometric Theory*, 12, 657–681.

Goffe, W. L., G. D. Ferrier, and J. Rogers (1994): "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65–99.

Gouriéroux, C. (1997): *ARCH Models and Financial Applications*, Springer Series in Statistics. Springer-Verlag, New York.

Gouriéroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, S8, 85–118.

Gregory, A. V., and M. R. Veal (1985): "Formulating Wald tests of nonlinear restrictions," *Econometrica*, 53, 1465–1468.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Jacquier, E., N. Polson, and P. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models (with discussion)," *Journal of Economics and Business Statistics*, 12, 371–417.

Lutkepohl, H., and M. M. Burda (1997): "Modified Wald tests under nonregular conditions," *Journal of Econometrics*, 78, 315–332.

Mahieu, R., and P. Schotman (1998): "An Empirical application of stochastic volatility models," *Journal of Applied Econometrics*, 13, 333–360.

Marriott, F. H. C. (1979): "Barnard's Monte Carlo Tests: How Many Simulations?," *Applied Statistics*, 28, 75–77.

Monfardini, C. (1998): "Estimating Stochastic volatility Models through Indirect Inference," *The Econometrics journal*, 1(1).

Nelson, F. D., and N. Savin (1990): "The Danger of Extrapolating Asymptotic Local Power," *Econometrica*, 58, 977–981.

Newey, W. K., and K. D. West (1987a): "Hypothesis Testing with Efficient Method of Moments Estimators," *International Economic Review*, 28, 777–787.

——— (1987b): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Phillips, P. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

Sargan, J. (1983): "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633.

Staiger, D., and J. Stock (1997): "Instrumental Variable Regressions with Weak Instruments," *Econometrica*, 65(3), 557–586.

Tauchen, G. (1997): "New Minimum Chi-Square Methods in Empirical Finance," *Advances in Econometrics, Seventh World Congress, ed. by David Kreps, and Kenneth F. Wallis. Cambridge University Press.*

Taylor, S. (1986): *Modelling Financial Time Series.* John Wiley, Chichester.

——— (1994): "Modelling Stochastic Volatility," *Mathematical Finance*, 4, 183–204.