A simple estimation method and finite-sample inference for a stochastic volatility model

Jean-Marie Dufour* Pascale Valery†

12th February 2002

Abstract

The aim of the paper is to fulfill the gap for testing hypotheses on parameters of the Stochastic Volatility model, more precisely, to propose finite sample exact tests in the sense that the tests have correct levels in small samples. To do this, we examine method-of-moments-based tests and it is interesting to note that explicit expressions are available for the moments and the estimators, simplifying highly the test procedures. We then state the asymptotic distribution of the estimator as well as that of the proposed test statistics for testing the null hypothesis of no persistence in the volatility. We then compare in a study of level and power these asymptotic techniques to the technique of Monte Carlo tests which is valid in finite samples and allow for test statistics whose null distribution may depend on nuisance parameters. In particular maximized Monte Carlo tests (MMC) introduced by Dufour (1995) have the exact level in finite samples when the p-value function is maximized over the entire set of nuisance parameters. In contrast to MMC tests which are highly computer intensive, simplified (asymptotically justified) approximate versions of Monte Carlo tests provide a halfway solution which achieves to control the level of the tests while alleviating the use of computers.

Key words: exact test; Monte Carlo test; finite sample test; stochastic volatility; Method-of-moments .
J.E.L classifications: C1, C13, C12, C32, C15

*CRDE,CIRANO,Université de Montreal,email: jean.marie.dufour@umontreal.ca
†CRDE,CIRANO,Université de Montreal,email: valeryp@magellan.umontreal.ca

Contents

1 Introduction

Evaluating the likelihood function of ARCH models is relatively easy compared to Stochastic Volatility models (SV) for which it is impossible to get an explicit closed-form expression for the likelihood function [see Shephard (1996), Mahieu and Schotman (1998)]. This is a generic feature common to almost all nonlinear latent variable models due to the curse of the high dimensionality of the integral appearing in the likelihood function of the stochastic volatility model. This is the reason why econometricians were reluctant to use this kind of models in their applications for a long time since in this setting, maximum likelihood methods are computationnaly intensive. But recently progress has been made regarding the estimation of nonlinear latent variable models in general and stochastic volatility models in particular. Among these methods to estimate the SV model, we can mention the Quasi Maximum Likelihood (QML) approach suggested by Nelson (1988) and Harvey, Ruiz and Shephard (1994) but also a Generalized Method of Moments(GMM) procedure proposed by Melino and Turnbull (1990). On the other hand, increased computer power has made simulation-based estimation methods more attractive among which we can mention the Simulated Method of Moments (SMM) proposed by Duffie and Singleton (1993), the indirect inference approach of Gouriéroux, Monfort and Renault (1993) and the moment matching methods of Gallant and Tauchen (1994). But computer intensive Markov Chain Monte Carlo methods applied to SV models by Jacquier, Polson and Rossi (1994) and Kim and Shephard (1994), Kim, Shephard and Chib (1998), and simulation-based Maximum Likelihood (SML) method proposed by Danielsson and Richard (1993), Danielsson (1994), are the most efficient methods to estimate this kind of models. In particular, Danielsson (1994), Danielsson and Richard (1993) developp an importance sampling technique to estimate the integral appearing in the likelihood function of the SV model. In a Bayesian setting, Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998) combine a Gibbs sampler with the Metropolis-Hastings algorithm to obtain the marginal posterior densities of the parameters of the SV model.

In addition to the previous estimation methods the aim of this paper is to estimate a stochastic volatility model with an autoregressive mean part. Indeed, the standard form as set forth, for instance, in Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994), Danielsson (1994), takes the form of an autoregression whose innovations are scaled by an unobservable volatility process, usually distributed as a lognormal autoregression but other distributions (student, mixture of normal distributions) can be considered [see Kim, Shephard and Chib (1998), Mahieu and Schotman (1998)]. The stochastic volatility specification we have chosen here comes from Gallant, Hsieh, Tauchen (1995), Tauchen (1996). Whereas all the authors quoted above, mainly focus on estimation procedures for the stochastic volatility model, often preoccupied by efficiency considerations [e.g. bayesian methods, Efficient Method of Moments], our paper unlike the others is mostly motivated by inference techniques applicable to the stochastic volatility model. Our concern for inference, in particular for simulation-based inference such as the technique of Monte Carlo tests introduced by Dwass (1957) for permutation tests, and later extended by Barnard (1963) and Birnbaum (1974), justifies an estimation method easy to implement. Thus, the estimation method used in this paper is mainly a method of moments in two steps which coincides with the GMM procedure in the particular case that the autoregressive mean part vanishes. Econometricians previously quoted mainly focused on efficient estimation procedures to estimate the SV model, they mostly examine specification tests such as the χ^2 tests for goodness of fit in Andersen and Sorensen (1996), Andersen, Chung and Sorensen (1997), specification tests with diagnostics in Gallant, Hsieh and Tauchen (1995), χ^2 specification tests through Indirect Inference criterion in Monfardini (1997), or likelihood ratio tests statistics for comparative fit in Kim, Shephard and Chib (1998). As a consequence, inference techniques for testing hypotheses on parameters remained underdeveloped, apart from standard t tests for individual parameters in Andersen and Sorensen (1996), in Andersen, Chung and Sorensen (1997) often performed with size distorsions.

In this setting, the aim of the paper is to fulfill the gap for testing hypotheses on parameters of the SV model, more precisely, to propose finite sample exact tests in the sense that the tests have correct levels in small samples. To do this, we examine method-of-moments-based tests and it is interesting to note that explicit expressions are available for the moments and the estimators, simplifying highly the test procedures. We then state the asymptotic distribution of the estimator as well as that of the proposed test statistics for testing the null hypothesis of no persistence in the volatility. We then compare in a study of level and power these asymptotic techniques to the technique of Monte Carlo tests which is valid in finite samples and allow for test statistics whose null distribution may depend on nuisance parameters. In particular maximized Monte Carlo tests (MMC) introduced by

Dufour (1995) have the exact level in finite samples when the p-value function is maximized over the entire set of nuisance parameters. In contrast to MMC tests which are highly computer intensive, simplified (asymptotically justified) approximate versions of Monte Carlo tests provide a halfway solution which achieves to control the level of the tests while alleviating the use of computers.

The paper is organized as follows. The second section sets the framework and the assumptions underlying the model. In the third section, we expose the estimation procedure used as well as the distributional results obtained for our estimator. Hypothesis testing is examined in the fourth section and the distribution of the test statistics is stated. The fifth section explicits the technique of Monte Carlo tests. The sixth section presents the data used in the empirical application while implementation results are discussed in the seventh section. All proofs are given in the appendix.

2 Framework

The basic form of the stochastic volatility model we study here for y_t comes from Gallant, Hsieh, Tauchen (1995). Let y_t denote the first difference over a short time interval, a day for instance, of the price of a financial asset traded on active speculative markets.

Assumption 1 *The process* $\{y_t,\ t \in \mathbb{N}\}$ *follows a stochastic volatility model of the form:*

$$y_t - \mu_y = \sum_{j=1}^{L_y} c_j (y_{t-j} - \mu_y) + exp(w_t/2) r_y z_t \ , \tag{1}$$

$$w_t - \mu_w = \sum_{j=1}^{L_w} a_j (w_{t-j} - \mu_w) + r_w v_t \ , \tag{2}$$

where μ_y, $\{c_j\}_{j=1}^{L_y}$, r_y, μ_w, $\{a_j\}_{j=1}^{L_w}$ *and* r_w *are the parameters of the two equations, called the mean and volatility equations respectively.* $s_t = (y_t, w_t)'$ *is initialized from its stationary distribution.*

The lag lengths of the autoregressive specifications used in the literature are typically short, e.g. $L_w = 1$, and $L_y = 1$, or $L_y = 0$ [see e.g. Andersen and Sorensen (1996), Andersen, Chung and Sorensen (1997) Gallant, Hsieh, Tauchen (1995)].

Assumption 2 *The vectors* $(z_t, v_t)'$, $t \in \mathbb{N}$ *are i.i.d. according to a* $N(0, I_2)$ *distribution.*

The process $\{s_t\ , t \in \mathbb{N}\}$ where $s_t = (y_t, w_t)'$ can be rewritten in the form:

$$(1 - c_1 L - \ldots - c_j L^j - \ldots - c_{L_y} L^{L_y})(y_t - \mu_y) = \exp(w_t/2) r_y z_t\ ,$$

$$(1 - a_1 L - \ldots - a_j L^j - \ldots - a_{L_w} L^{L_w})(w_t - \mu_w) = r_w v_t\ ,$$

and the roots of the lag operators

$$C(z) = 1 - c_1 z - \ldots - c_j z^j - \ldots - c_{L_y} z^{L_y} = 0$$

and

$$A(z) = 1 - a_1 z - \ldots - a_j z^j - \ldots - a_{L_w} z^{L_w} = 0$$

outside of the unit circle, is covariance-stationary [see Hamilton, (1994 p.57)] .

Assumption 3 *The process* $s_t = (y_t, w_t)'$ *is strictly stationary.*

The process is Markovian of order $L_s = max(L_y, L_w)$ with conditional density $p_s(s_t | s_{t-L_s}, \ldots, s_{t-1}, \rho)$ given by the stochastic volatility model,where

$$\rho = (\mu_y, c_1, \ldots, c_{L_y}, r_y, \mu_w, a_1, \ldots, a_{L_w}, r_w)' \tag{3}$$

is a vector which contains the free parameters of the stochastic volatility model. The process $\{y_t\}$ is observed whereas $\{w_t\}$ is regarded as latent. Write $p_{y,J}(y_{t-J}, \ldots, y_t | \rho)$ for the implied joint density under the model of a stretch $y_{t-J}, \ldots, y_t$. No general closed-form expressions are available for the moments of y_t, but they can be approximated by Monte Carlo integration.

3 Method-of-moments estimation of an AR(1)-SV model

In this section, we derive the analytic expressions of the moments and of the estimator of $\theta = (a, r_y, r_w)'$ as well as its distributional properties. Let us consider in a first step a simplified version of model (1)-(2) with $\mu_y = \mu_w = 0$ and $c_j = a_j = 0$, $\forall j \geq 2$. We then have:

$$y_t = cy_{t-1} + \exp(w_t/2) r_y z_t \ , \tag{4}$$

$$w_t = a w_{t-1} + r_w v_t \ . \tag{5}$$

We shall call the model represented by equations (4)-(5) the stochastic volatility model with an autoregressive mean part of order one [AR(1)-SV for short]. This specification of the stochastic volatility model comes from Gallant, Hsieh and Tauchen (1995). Let $u_t = exp(w_t/2) r_y z_t$ denote the error term in the mean equation. As the stochastic volatility model belongs to the general class of models characterized by strong conditional heteroscedasticity and correlation between the sample moments, working with $u_t^2 = exp(w_t) r_y^2 z_t^2$ will therefore be more appropriate. To estimate this model, we consider a two-step method whose first step consists in applying ordinary least squares (OLS) to the mean equation which yields a consistent estimate of the parameter c denoted by $\hat{c}_T$ and the adjusted residuals $\hat{u}_t$. Then, we apply in a second step a method of moments to the residuals $\hat{u}_t$ to get the estimate of the parameter $\theta' = (a, r_y, r_w)$ of the mean and volatility equations. The moments of interest used in the estimation procedure are of the following form.

Proposition 1 *Under Assumptions 1,2,3, with $\mu_y = \mu_w = 0$ and $c_j = a_j = 0$, $\forall\ j \geq 2$, let $u_t = exp(w_t/2) r_y z_t$. Then u_t has the following moments for even values of k:*

$$\mu_k(\theta) \equiv E(u_t^k) = r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp[\frac{k^2}{8} r_w^2/(1-a^2)] \ , \tag{6}$$

$$\begin{aligned} \mu_{k,l}(m|\theta) &\equiv E(u_t^k u_{t+m}^l) \\ &= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp[\frac{r_w^2}{8(1-a^2)}(k^2 + l^2 + 2kla^m)] \end{aligned} \tag{7}$$

The odd moments are equal to zero.

The proofs of the propositions are given in the Appendix. In particular, for $k = 2$, $k = 4$ and $k = l = 2$ and $m = 1$, we get:

$$\mu_2(\theta) = E(u_t^2) = r_y^2 \exp[r_w^2/2(1-a^2)] \ , \tag{8}$$

$$\mu_4(\theta) = E(u_t^4) = 3r_y^4 \exp[2r_w^2/(1-a^2)] \ , \tag{9}$$

and

$$\mu_{2,2}(1|\theta) = E[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1-a)] \ . \tag{10}$$

Solving the above moment equations corresponding to $k = 2$, $k = 4$ and $m = 1$ yields the following proposition.

Proposition 2 *Under the assumptions of Proposition 1, we have:*

$$a = \frac{[\log(\mu_{2,2}(1|\theta)) - \log(3) - 4\log(\mu_2) + \log(\mu_4)]}{\log(\frac{\mu_4}{3(\mu_2)^2})} - 1 \ , \tag{11}$$

$$r_y = \frac{3^{1/4}\mu_2}{\mu_4^{1/4}} \ , \tag{12}$$

$$r_w = \left(\log(\frac{\mu_4}{3(\mu_2)^2})(1-a^2)\right)^{1/2} \ . \tag{13}$$

Given the latter proposition, it is easy to compute a method-of-moments estimator for $\theta = (a, r_y, r_w)'$ replacing the theoretical moments by sample counterparts based on the estimated residuals $\hat{u}_t$, where $\hat{u}_t \equiv u_t(\hat{c}_T)$,$\forall\ t$ with $u_t(c) \stackrel{def}{=} y_t - cy_{t-1}$. Let $\hat{\theta}_T$ denote the method-of-moments estimator of θ. More precisely, $E(u_t^2)$, $E(u_t^4)$ and $E(u_t^2 u_{t-1}^2)$ are approximated by:

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

and

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

respectively. Now we introduce some useful notation. First, let:

$$\begin{aligned} u_t(c) &\equiv y_t - cy_{t-1} \\ &= \exp(\frac{aw_{t-1} + r_w v_t}{2}) r_y z_t \\ &\equiv v_t(\theta), \ \forall\ t \ . \end{aligned} \tag{14}$$

with $v_t(\theta) = \exp(\frac{aw_{t-1}+r_w v_t}{2}) r_y z_t$. We recall that $\hat{u}_t \equiv u_t(\hat{c}_T)\ \forall\ t$. Let c_0 and θ_0 denote the true value of the parameters c and θ respectively.

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

$$\bar{g}_T(\hat{U}) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 \\ \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^4 \\ \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2\hat{u}_{t-1}^2 \end{pmatrix}, \tag{15}$$

$\bar{g}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$ with $g_t(\theta)' = (v_t^2(\theta), v_t^4(\theta), v_t^2(\theta)v_{t-1}^2(\theta))$, i.e.

$$\bar{g}_T(\theta) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} v_t^2(\theta) \\ \frac{1}{T}\sum_{t=1}^{T} v_t^4(\theta) \\ \frac{1}{T}\sum_{t=1}^{T} v_t^2(\theta)v_{t-1}^2(\theta) \end{pmatrix}, \tag{16}$$

with $\mu(\theta)' = (\mu_2(\theta), \mu_4(\theta), \mu_{2,2}(1|\theta))$. We can now state the following proposition.

Proposition 3 *Under the assumptions of Proposition 1, the process* $\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0))$ *is asymptotically equivalent to* $\sqrt{T}(\bar{g}_T(\theta_0) - \mu(\theta_0))$.

The latter proposition will be useful in deriving the asymptotic distribution of the method-of-moments estimator of θ, but could have been stated with a faster rate of convergence as it is formulated in the corollary below.

Corollary 1 *Under the assumptions of Proposition 1, the process* $(\bar{g}_T(\hat{U}) - \mu(\theta_0))$ *is asymptotically equivalent to* $(\bar{g}_T(\theta_0) - \mu(\theta_0))$.

Its proof is similar to the proof of Proposition 3 and is based on L.L.N arguments and $p\lim_{T\to\infty}(\hat{c}_T - c_0) = 0$. Before deriving the asymptotic distribution of the method-of-moments estimator of θ we need to state the following proposition.

Proposition 4 *Under the assumptions of Proposition 1, the process* $\sqrt{T}(\bar{g}_T(\theta_0) - \mu(\theta_0))$ *is asymptotically distributed as a* $N(0, \Omega^*)$ *variable where* Ω^* *is a symetric positive definite matrix.*

Next proposition states the asymptotic normality of the method-of-moments estimator of $\theta = (a, r_y, r_w)'$ a subvector of $\rho' = (c, \theta')$.

Proposition 5 *Under the assumptions of Proposition 1, then the method-of-moments estimator $\hat{\theta}_T(\Omega)$ is such that:*

$$\sqrt{T}(\hat{\theta}_T(\Omega) - \theta_0) \xrightarrow{D} N(0, W(\Omega)) \tag{17}$$

where

$$\begin{aligned} W(\Omega) &= \left(\frac{\partial \mu'}{\partial \theta}(\theta_0) \Omega \frac{\partial \mu}{\partial \theta'}(\theta_0) \right)^{-1} \frac{\partial \mu'}{\partial \theta}(\theta_0) \Omega \Omega^* \Omega \frac{\partial \mu}{\partial \theta'}(\theta_0) \\ & \left(\frac{\partial \mu'}{\partial \theta}(\theta_0) \Omega \frac{\partial \mu}{\partial \theta'}(\theta_0) \right)^{-1} \end{aligned} \tag{18}$$

As usual, there is an optimal choice of this matrix, i.e. a choice which minimizes $W(\Omega)$.

Proposition 6 *Under the assumptions of Proposition 1, the optimal choice of the Ω matrix is: $\Omega = \Omega^{*-1}$ and*

$$W^* = W(\Omega^{*-1}) = \left(\frac{\partial \mu'}{\partial \theta}(\theta_0) \Omega^{*-1} \frac{\partial \mu}{\partial \theta'}(\theta_0) \right)^{-1} . \tag{19}$$

The optimal estimator thus obtained is denoted by $\hat{\theta}_T$. When the dimensions of μ and θ are the same, we have $W(\Omega) = W^*$, $\forall\, \Omega$ and

$$W^* = \left(\frac{\partial \mu'}{\partial \theta}(\theta_0) \right)^{-1} \Omega^* \left(\frac{\partial \mu}{\partial \theta'}(\theta_0) \right)^{-1} .$$

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

$$\bar{g}_T(\hat{U}) = \mu(\theta) .$$

A consistent estimator of W^* is obtained as soon as we have a consistent estimator of Ω^*. A consistent estimator of Ω^* can be easily obtained [see Newey and West (1987)] as a fixed-bandwith Bartlett kernel estimator, i.e.:

$$\hat{\Omega}^*(\theta_0) = \Gamma_0 + \sum_{k=1}^{K} (1 - \frac{k}{K+1})(\Gamma_k + \Gamma'_k) \tag{20}$$

with

$$\Gamma_k = \frac{1}{T} \sum_{t=k+1}^{T} [g_{t-k}(\theta_0) - \bar{g}_T(\theta_0)][g_t(\theta_0) - \bar{g}_T(\theta_0)]' \tag{21}$$

with θ_0 replaced by any consistent estimator $\tilde{\theta}_T$ of θ_0, i.e., $\hat{\theta}_T$,

$$\hat{\Omega}^*(\tilde{\theta}_T) = \hat{\Gamma}_0 + \sum_{k=1}^{K}(1 - \frac{k}{K+1})(\hat{\Gamma}_k + \hat{\Gamma}'_k) \tag{22}$$

with

$$\hat{\Gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}[g_{t-k}(\tilde{\theta}_T) - \bar{g}_T(\tilde{\theta}_T)][g_t(\tilde{\theta}_T) - \bar{g}_T(\tilde{\theta}_T)]' \tag{23}$$

since

$$\bar{g}_T(\tilde{\theta}_T) = \frac{1}{T}\sum_{t=1}^{T} g_t(\tilde{\theta}_T) \xrightarrow{p} \mu(\theta_0) = E[g_t(\theta_0)]$$

and

$$\frac{1}{T}\sum_{t=1}^{T} g_{t-k}(\tilde{\theta}_T) g_t(\tilde{\theta}_T)' \xrightarrow{p} E[g_{t-k}(\theta_0) g_t(\theta_0)']$$

since the perturbation vectors have been shown to be strictly stationary and ergodic [see proof of Proposition 4 in Appendix].

Therefore a consistent estimator of W^* is given by:

$$\hat{W}^* = \left(\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}^{*-1}(\tilde{\theta}_T)\frac{\partial \mu}{\partial \theta'}(\hat{\theta}_T)\right)^{-1}. \tag{24}$$

4 Tests of hypotheses

We assume that the parameter $\theta = (a, r_y, r_w)'$ is partitioned into

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

with $\theta_1 \overset{def}{=} a$. We consider the null hypothesis $H_0 :\ a = 0$ which corresponds to test the absence of long memory in the volatility. To define these tests we have to introduce the optimal unconstrained estimator:

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

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

$$\begin{pmatrix} 0 \\ \hat{\theta}^c_{2T} \end{pmatrix} = \hat{\theta}^c_T \ .$$

The Wald statistic is defined as

$$\xi^W_T = T(\hat{\theta}_T)'\hat{W}^{*-1}_1(\hat{\theta}_T) \tag{25}$$

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

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:

$$\mathcal{D}_T = \frac{\partial \mu'}{\partial \theta_1}(\hat{\theta}^c_T)\Omega^{*-1}(\mu(\hat{\theta}^c_T) - \bar{g}_T(\hat{U})) \tag{26}$$

and the test statistic is

$$\xi^S_T = T\mathcal{D}'_T\mathcal{A}\mathcal{D}_T \ . \tag{27}$$

Finally, we can introduce the difference between the optimal values of the objective function that we will call the LR-type test in the simulations:

$$\xi^C_T = T[M^*_T(\hat{\theta}^c_T) - M^*_T(\hat{\theta}_T)] \tag{28}$$

where the criterion to be minimized is:

$$M^*_T(\theta) \stackrel{def}{=} [\bar{g}_T(\hat{U}) - \mu(\theta)]'\hat{\Omega}^{*-1}[\bar{g}_T(\hat{U}) - \mu(\theta)] \tag{29}$$

Proposition 7 *Under the assumptions of Proposition 1, the test statistics ξ^W_T, ξ^S_T, and ξ^C_T are asymptotically equivalent under the null hypothesis, and have the common distribution $\chi^2(1)$.*

5 The technique of Monte Carlo tests

The technique of Monte Carlo tests has originally been suggested by Dwass (1957) for implementing permutation tests, and did not involve nuisance parameters. This technique has been later extended by Barnard (1963) and Birnbaum (1974). This technique 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 study here the case where the distribution of the test statistic S depends on nuisance parameters. For the test statistics exposed in section 4, their asymptotic distribution is asymptotically pivotal (Chi-square distribution), but their finite sample distribution remains unknown. At this stage, we need to make an effort of formalization to clearly expose the procedure. We consider a family of probability spaces $\{(\mathcal{Z}, \mathcal{A}_{\mathcal{Z}}, P_\rho) : \rho \in \Omega\}$ and suppose that S is a real valued $\mathcal{A}_{\mathcal{Z}}$-measurable function whose distribution is determined by $P_{\bar{\rho}}$ where $\bar{\rho}$ is the "true" parameter vector. We wish to test the hypothesis

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

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

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

or equivalently,

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

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

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

If we define the distribution and p-value functions,

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

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

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

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

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

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

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

We consider a real random variable S_0 and random vectors of the form

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

all defined on a common probability space $(\mathcal{Z}, \mathcal{A}_{\mathcal{Z}}, P)$, such that the variables $S_0, S_1(\bar{\rho}), \ldots , S_N(\bar{\rho})$ are i.i.d. or exchangeable for some $\bar{\rho} \in \Omega$, each one with distribution function $F[x|\bar{\rho}] = P[S_0 \leq x]$. Typically, S_0 will refer to a test statistic computed from the observed data when the true parameter vector is $\bar{\rho}$ (i.e., $\rho = \bar{\rho}$), while $S_1(\rho), \ldots , S_N(\rho)$ 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 ρ (i.e., $P[S_i(\rho) \leq x] = F[x|\rho]$). In other words, the observed statistic S_0 is simulated by first generating an "observation" vector y according to

$$y = g(\rho, z, v) \tag{30}$$

where the function g is bivariate for our AR(1)-SV model, and corresponds to equations (4) and (5), with $\rho = (c, \theta)'$, $\theta = (a, r_y, r_w)'$. The perturbations z and v have known distributions, which can be simulated ($N(0, 1)$ or student, or mixtures, e.g.) and then computing

$$S(\rho) \equiv S[g(\rho, z, v)] \equiv g_S(\rho, z, v) \; . \tag{31}$$

The observed statistic S_0 is then computed as $S_0 = S[g(\bar{\rho}, z_0, v_0)]$ and the simulated statistics $S_i(\rho) = S[g(\rho, z_i, v_i)]$, $i = 1, \ldots , N$ where the random vectors $z_0, z_1, \ldots , z_N$ are i.i.d. (or exchangeable) and $v_0, v_1, \ldots , v_N$ are i.i.d. (or exchangeable) as well.

The technique of Monte Carlo tests provides a simple method allowing one to replace the theoretical distribution $F(x|\rho)$ by its sample analogue based on $S_1(\rho), \ldots , S_N(\rho)$:

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

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

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

and the p-value function

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

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

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

The central property which is exploited here is the following: to obtain critical values or compute p-values, the "theoretical" null distribution $F[x|\bar{\rho}]$ can be replaced by its simulation-based "estimate" $\hat{F}_N[x|\rho] \equiv \hat{F}_N[x; S(N,\rho)]$ in a way that will preserve the level of the test in *finite samples, irrespective of the number N of replications used.*

Thus, in this framework, Dufour (1995) states the finite sample validity of Monte Carlo tests when the p-value function is maximized over the entire set of the nuisance parameters as it is formulated in the reported proposition below [see Dufour (1995, p.13, Proposition 4.1].

Proposition 8 (Validity of MMC tests when ties have zero probability) *Under the above assumptions and notations, set* $S_0(\bar{\rho}) = S_0$ *and suppose that*

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

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

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

where

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

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

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

Following the latter proposition, if we choose α_1 and N so that

$$\alpha = \frac{I[\alpha_1 N] + 1}{N+1} \tag{32}$$

is the desired significance level, the critical region $\sup\{\hat{G}_N[S_0|\rho] : \rho \in \Omega_0\} \leq \alpha_1$ has level α irrespective of the presence of nuisance parameters in the distribution of the test statistic S under the null hypothesis $H_0 : \bar{\rho} \in \Omega_0$. The same also holds if we use the (almost) equivalent critical regions $\inf\{\hat{F}_N[S_0|\rho] : \rho \in \Omega_0\} \geq 1 - \alpha_1$ or $S_0 \geq \sup\{\hat{F}_N^{-1}[1-\alpha_1|\rho] : \rho \in \Omega_0\}$. Dufour call such tests maximized Monte Carlo (MMC) tests. The function $\hat{G}_N[S_0|\rho]$ (or $\hat{p}_N[S_0|\rho]$) is then maximized with respect to $\rho \in \Omega_0$, keeping the observed statistic S_0 and the simulated disturbance vectors $z_1, ..., z_N$ and $v_1, ..., v_N$ fixed. The function $\hat{G}_N[S_0|\rho]$ is a step-type function which typically has zero derivatives almost everywhere, except on isolated points (or manifolds) where it is not differentiable. So it cannot be maximized with usual derivative-based algorithms. However, the required maximizations can be performed by using appropriate optimization algorithms that do not require differentiability, such as *simulated annealing.* For further discussion of such algorithms, the reader may consult Goffe, Ferrier, and Rogers(1994).

On the other hand, Dufour (1995) also proposes simplified (asymptotically justified) approximate versions of the Monte Carlo tests where this time the p-value function is evaluated at a consistent point estimate, the Bootstrap version, or a consistent set estimate of ρ, asymptotic Monte Carlo tests based on consistent set estimators. The author shows [see Dufour, (1995, p.16, Proposition 5.1 and p.19, Proposition 6.3)] that both tests are asymptotically valid in the sense that they have the correct level α asymptotically and the estimated p-values converge to the true p-values. He also assesses the validity of the MMC tests and the asymptotic Monte Carlo tests based on consistent set estimators for general distributions , when ties have non-zero probability [see Dufour, (1995, p.14, Proposition 4.2 and p.17, Proposition 5.2)].

It is this technique of Monte Carlo tests in its maximized and Bootstrap versions which will be applied in section 7 for comparison issues with standard asymptotic tests of section 4.

6 Data

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. We use a long time series, because, among other things, we want to investigate the long-term properties of stock market volatility through a persistence test. 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 Implementation results

The results given here are preliminary and have to be improved and completed. The maximized Monte Carlo tests (MMC) remains to be performed to reinforce the validity of the tests in finite samples. Here we test the null hypothesis of no-persistence in the volatility, which takes the form $H_0 : a = 0$ and the alternative is of the form $H_1 : a = 0.9$. The nominal level of the tests has been set to $\alpha = 5\%$. M represents the number of replications to evaluate the frequence of rejection of both hypotheses, and N represents the number of simulated statistics used in the Monte Carlo tests. T is the sample size of the serie y_t. Implementation is performed with the GAUSS software.

7.1 Study of the level of the tests

Here we study the frequence of rejection of the null hypothesis $H_0 : a = 0$ and compare it to the nominal level fixed at $\alpha = 5\%$.

LEVELS in % (under H_0)				
	Asymptotic tests			
	M=100	M=100	M=1000	M=1000
	T=80	T=1000	T=80	T=1000
Wald	8	14	8.5	9
Score	50	54	48.1	54.1
LR $\hat{\Omega}$	34	34	31.8	34
LR $\hat{\Omega} = Id$	0	0	0	0

LEVELS in % (under H_0)				
	Bootstrap tests			
	M=100	M=100	M=100	M=100
	N=19	N=19	N=99	N=99
	T=80	T=1000	T=80	T=1000
Wald	5	9	3	5
Score	7	5	6	2
LR $\hat{\Omega}$	2	7	2	3
LR $\hat{\Omega} = Id$	1	6	0	3

At this stage of the implementation which still remains preliminary and has to be improved, the Bootstrap test which is a simplified (asymptotically justified) version of the Maximized Monte Carlo test, achieves to control most of the time the level of the tests fixed at $\alpha = 5\%$ while its asymptotic counterpart fails, in particular there are serious size distorsions for the score test. The Maximized Monte Carlo test (MMC) remains to be performed to reinforce the validity of the tests in finite samples in the sense that they have the correct levels.

7.2 Study of the power of the tests

Here we study the frequence of rejection of the alternative hypothesis H_1 : $a = 0.9$.

POWER in % (under H_1)				
	Asymptotic tests			
	M=100	M=100	M=1000	M=1000
	T=80	T=1000	T=80	T=1000
Wald	26	85	25.7	89.2
Score	66	98	66.1	99.6
LR $\hat{\Omega}$	54	91	48.8	93.8
LR $\hat{\Omega} = Id$	15	99	18.6	98

POWER in % (under H_1)				
	Bootstrap tests			
	M=100	M=100	M=100	M=100
	N=19	N=19	N=99	N=99
	T=80	T=1000	T=80	T=1000
Wald	13	45	16	46
Score	11	27	7	32
LR $\hat{\Omega}$	12	28	11	23
LR $\hat{\Omega} = Id$	7	93	3	98

We do not prescribe these two methods when the sample size is very small (e.g. $T = 80$), the results being in this case unreliable. Both test procedures have more power when the sample size grows which is intuitive since both tests are asymptotically justified. Additionally, we can note that the "LR-type" test with $\hat{\Omega} = Id$ has more power than the one with $\hat{\Omega}$ probably due to precision errors in the estimation of the weighting matrix which is estimated by a kernel estimator with a fixed-Bandwith Bartlett Kernel, where the lag truncation parameter K has been set to $K = 5$.

7.3 Empirical applications

In this subsection we test the null hypothesis of no-persistence in the volatility of the data (Standard and Poor's Composite Price Index (SP), 1928-87) which counts 16,127 daily observations. The critical regions used to perform the tests are of the form:

$$\mathcal{R}_c = \{S_0 > \chi^2_{1-\alpha}(1) = 3.84\}$$

for the standard asymptotic tests, where S_0 denotes the statistic computed from the observed data and of the form:

$$\mathcal{R}_c = \{\hat{p}_N[S_0|\hat{\rho}_T] \leq \alpha\}$$

with the p-value function

$$\hat{p}_N[S_0|\rho] = \frac{N\hat{G}_N[S_0|\rho] + 1}{N + 1}\ .$$

and the tail area function

$$\hat{G}_N[S_0; S(N, \rho)] = \frac{1}{N}\sum_{i=1}^{N} s(S_i(\rho) - S_0).$$

for the Bootstrap test with $\rho = (c, \theta')' = (c, a, r_y, r_w)'$. α has been set to $\alpha = 5\%$. The estimated value of ρ from the data is:

$$\hat{\rho}_T = (0.129, 0.926, 0.829, 0.427)'$$

where the method-of-moments estimated value of a corresponds to $\hat{a}_T = 0.926$. We may conjecture that there is some persistence in the data (1928-87) what is statistically checked by performing the tests below.

data			
	Asymptotic tests	*Bootstrap tests*	
	S_0	N=19	N=99
Wald	48.11	0.10	0.04
Score	1481.4	0.05	0.04
LR $\hat{\Omega}$	18.3	0.15	0.17
LR $\hat{\Omega} = Id$	247463	0.05	0.01

All asymptotic tests reject indeed the null hypothesis of no-persistence in the volatility since $S_0 > \chi^2_{1-\alpha}(1) = 3.84$ as well as the Bootstrap tests for the Score statistic and the "LR-type" statistic with $\hat{\Omega} = Id$. Once again we can note that the results are different whether the criterion on which is based the "LR-type" statistic is evaluated at the Kernel estimator of $\hat{\Omega}$ or at $\hat{\Omega} = Id$ since we are in the just-identified case $(dim(\theta) = dim(\mu(\theta) = 3)$. Thus, the estimation of the weighting matrix Ω seems to poorly affect the "LR-type" statistic in the sense that the latter is strongly underevaluated when the weighting matrix Ω is estimated. We have further performed the Bootstrap test for the "LR-type" statistic with the estimation of Ω for $N = 199$: $p-value = 0.095$, for $N = 599 : p-value = 0.111$ and $N = 999 : p-value = 0.099$ and they still do not reject the null hypothesis of no-persistence since the p-values functions are still greater than $\alpha = 5\%$. The Wald version of the Bootstrap test does reject the null hypothesis with $N = 99$ but not with $N = 19$ for a nominal level of 5%, but could have rejected the null hypothesis at a level of 10%.

A Proof of proposition 1

First, we recall that if $U \sim N(0,1)$ then $E(U^{2p+1}) = 0$, $\forall\ p \in \mathbb{N}$ and $E(U^{2p}) = (2p)!/[2^p p!]$ $\forall\ p \in \mathbb{N}$ [see Gouriéroux, Monfort, p.518, vol.2]. Under Assumptions 1,2,3 with a stationary AR(1) specification for w_t, using the definition of u_t, we get:

$$\begin{aligned} E(u_t^k) &= r_y^k E(z_t^k) E[\exp(k w_t/2)] \\ &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp(\frac{k^2}{4} r_w^2/2(1-a^2)) \\ &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp(\frac{k^2}{8} r_w^2/(1-a^2)) \end{aligned} \tag{33}$$

where the second equality makes use of the definition of the gaussian Laplace transform of $w_t \sim N(0, \frac{r_w^2}{1-a^2})$ and of the moments of the $N(0,1)$ z_t variable. Let us now calculate the cross-product:

$$\begin{aligned} E[u_t^k u_{t+m}^l] &= E[r_y^{k+l} z_t^k z_{t+m}^l \exp(k\frac{w_t}{2} + l\frac{w_{t+m}}{2})] \\ &= r_y^{k+l} E(z_t^k) E(z_{t+m}^l) E[\exp(k\frac{w_t}{2} + l\frac{w_{t+m}}{2})] \\ &= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp\left(\frac{r_w^2}{8(1-a^2)}(k^2+l^2+2kla^m)\right) \end{aligned}$$

where $E(w_t) = 0$, $Var(w_t) = \frac{r_w^2}{1-a^2}$ and

$$\begin{aligned} Var(k\frac{w_t}{2} + l\frac{w_{t+m}}{2}) &= \frac{k^2}{4} Var(w_t) + \frac{l^2}{4} Var(w_{t+m}) + 2\frac{k}{2}\frac{l}{2} Cov(w_t, w_{t+m}) \\ &= \frac{r_w^2}{4(1-a^2)}(k^2+l^2+2kla^m)\ . \end{aligned} \tag{34}$$

B Proof of proposition 2

Taking the ratio of equation (9) on equation (8) to the square produces

$$\frac{E(u_t^4)}{(E(u_t^2))^2} = 3\exp(r_w^2/(1-a^2))\ , \tag{35}$$

i.e.

$$Q \equiv (r_w^2/(1-a^2)) = \log\left(\frac{E(u_t^4)}{3(E(u_t^2))^2}\right). \tag{36}$$

Inserting $Q \equiv (r_w^2/(1-a^2))$ in equation (8) yields

$$r_y = \left(\frac{E(u_t^2)}{\exp(Q/2)}\right)^{1/2} = \frac{3^{1/4}E(u_t^2)}{E(u_t^4)^{1/4}}. \tag{37}$$

From equation (10) we have

$$\exp(\frac{r_w^2}{(1-a)}) = \frac{E[u_t^2 u_{t-1}^2]}{r_y^4} \tag{38}$$

which, after a few manipulations, yields

$$1 + a = \frac{[\log(E[u_t^2 u_{t-1}^2]) - 4\log(r_y)]}{Q} \tag{39}$$

or either

$$a = \frac{[\log(E[u_t^2 u_{t-1}^2]) - \log(3) - 4\log(E[u_t^2]) + \log(E[u_t^4])]}{\log\left(\frac{E[u_t^4]}{3(E[u_t^2])^2}\right)} - 1. \tag{40}$$

From the expressions of $Q \equiv r_w^2/(1-a^2)$ at equation (36) and that of a above we can deduce

$$r_w = \left(\log\left(\frac{E[u_t^4]}{3(E[u_t^2])^2}\right).(1-a^2)\right)^{1/2}. \tag{41}$$

C Proof of Proposition 3

We are going to make an expansion of order one of the function:

$$\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T(\hat{u}_t^2 - \mu_2(\theta_0)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T(\hat{u}_t^4 - \mu_4(\theta_0)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T(\hat{u}_t^2\hat{u}_{t-1}^2 - \mu_{2,2}(1)(\theta_0)) \end{pmatrix}, \tag{42}$$

in order to show the asymptotic equivalence between $\hat{u}_t^r = u_t(\hat{c}_T)^r$ and $u_t^r = u_t(c_0)^r$.

C.1 The component $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2 - \mu_2(\theta_0))$

Recall that $u_t^2(c) \stackrel{def}{=} (y_t - cy_{t-1})^2$ and $\hat{u}_t^2 \stackrel{def}{=} u_t^2(\hat{c}_T)$. We shall make a Taylor's expansion of the function $u_t^2(\hat{c}_T)$ with $\hat{c}_T = c_0 + h$, i.e.,

$$u_t^2(\hat{c}_T) = u_t^2(c_0) + \frac{\partial u_t^2}{\partial c}|_{c=c_0} h + R_t$$

with $R_t = \frac{1}{2!}\frac{\partial^2 u_t^2}{\partial c^2}|_{c=\bar{c}} h^2 = y_{t-1}^2 h^2$ where $\bar{c} = c_0 + \alpha h$ with $0 < \alpha < 1$. We note that $\lim_{h\to 0}\frac{R_t}{h} = \lim_{h\to 0} y_{t-1}^2 h = 0$. Substracting $\mu_2(\theta_0)$ from both sides yields

$$u_t^2(\hat{c}_T) - \mu_2(\theta_0) = u_t^2(c_0) - \mu_2(\theta_0) - 2y_{t-1}u_t(c_0)h + y_{t-1}^2 h^2$$

and aggregating it over the sample size and standardizing by $\sqrt{T}$ yields

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(\hat{c}_T) - \mu_2(\theta_0)) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(c_0) - \mu_2(\theta_0)) - \frac{2}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}u_t(c_0)h + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}^2 h^2 ,$$

where $R_T \stackrel{def}{=} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}^2 h^2$ with $\lim_{h\to 0}\frac{R_T}{h} = 0$. Replacing then $h = \hat{c}_T - c_0$ we have:

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(\hat{c}_T) - \mu_2(\theta_0)) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(c_0) - \mu_2(\theta_0)) - \frac{2}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}u_t(c_0)(\hat{c}_T - c_0) \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}^2(\hat{c}_T - c_0)^2 , \qquad (43)
\end{aligned}$$

and

$$\begin{aligned}
R_T &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}^2(\hat{c}_T - c_0)^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \sqrt{T}(\hat{c}_T - c_0)(\hat{c}_T - c_0) \qquad (44)
\end{aligned}$$

where it is shown at equations (66), (68) and (69) respectively that $\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \stackrel{p}{\to} 0$, $\sqrt{T}(\hat{c}_T - c_0) \stackrel{\mathcal{L}}{\to} N(0, 1-c^2)$ and $(\hat{c}_T - c_0) \stackrel{p}{\to} 0$ yielding an $o_p(1)-$variable for

R_T. Let us now study the term $\frac{2}{T}\sum_{t=1}^{T} u_t(c_0)y_{t-1}\sqrt{T}(\hat{c}_T - c_0)$. $u_t(c_0)y_{t-1}$ is a martingale difference sequence (m.d.s.) w.r.t. the subfields $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$ where $X_t = (y_t, u_t, z_t, v_t)'$ since

$$
\begin{aligned}
E[u_t(c_0)y_{t-1}|\mathcal{F}_{t-1}] &= y_{t-1}E[u_t(c_0)|\mathcal{F}_{t-1}] \\
&= y_{t-1}E[v_t(\theta_0)|\mathcal{F}_{t-1}] \\
&= y_{t-1}E[\exp(w_t/2)r_{y0}z_t|\mathcal{F}_{t-1}] \\
&= y_{t-1}r_{y0}E[z_t|\mathcal{F}_{t-1}]E[\exp(w_t/2)|\mathcal{F}_{t-1}] \\
&= r_{y0}y_{t-1}E[z_t]E[\exp(w_t/2)|\mathcal{F}_{t-1}] \\
&= 0 \, , \qquad (45)
\end{aligned}
$$

where the second and third lines exploit equation (14), and the fourth, fifth and sixth lines make use of the assumptions 1 and 2. Moreover, from equation (45) we deduce that $E[u_t(c_0)y_{t-1} = 0$, we have besides that $E(u_t^2 y_{t-1}^2) = E(u_t^2)E(y_{t-1}^2) = \frac{\mu_2^2}{1-c^2}$.

Before going further, we will need to show that $E|u_t y_{t-1}|^r < \infty$ for some $r > 1$, which will be useful for the uniform integrability. Thus,

$$
\begin{aligned}
E|u_t y_{t-1}|^r &= E|r_y z_t \exp(w_t/2) y_{t-1}|^r \\
&= E(|z_t|^r |r_y|^r \exp(\frac{r}{2}w_t)|y_{t-1}|^r) \\
&= E|z_t|^r |r_y|^r E(\exp(\frac{r}{2}w_t))E|y_{t-1}|^r \, . \qquad (46)
\end{aligned}
$$

In order to compute equation (46), we need to compute first $E|z_t|^r$ and $E|y_{t-1}|^r$.

For $z_t \overset{i.i.d.}{\sim} N(0,1)$ we have:

$$E|z_t|^r = 2(2\pi)^{-1/2}\int_0^\infty z_t^r e^{-\frac{z_t^2}{2}} \, dz_t \, . \qquad (47)$$

- For $r = 2n$: it is known that [see Gradshteyn and Ryzhik (1980, p.337, formula 3.461.2)]:

  $$\int_0^\infty z_t^{2n} e^{-\frac{z_t^2}{2}} \, dz_t = \frac{(2n-1)!!}{2}\sqrt{2\pi} \, , \qquad (48)$$

  hence,

  $$E|z_t|^{2n} = (2n-1)!! \qquad (49)$$

  where $(2n-1)!! = 1 \times 3 \times 5 \times \ldots \times (2n-1)$.

- For $r = 2n+1$: it is known that [see Gradshteyn and Ryzhik (1980, p.337, formula 3.461.3)]:

$$\int_0^\infty z_t^{2n+1} e^{-\frac{z_t^2}{2}}\, dz_t = \frac{n!}{2(2)^{-(n+1)}} \;, \tag{50}$$

  hence,

$$E|z_t|^{2n+1} = \sqrt{\frac{2}{\pi}} 2^n n! \;. \tag{51}$$

We also need to compute $E|y_{t-1}|^r$. Thus,

$$y_{t-1} = c y_{t-2} + \exp(w_{t-1}/2) r_y z_{t-1} \;,$$

hence:

$$|y_{t-1}|^r = |c y_{t-2} + \exp(w_{t-1}/2) r_y z_{t-1}|^r \;. \tag{52}$$

Exploiting then the $\tilde{c}_r-$inequality, we can say for $r > 1$ that:

$$\begin{aligned} E|y_{t-1}|^r &= E|c y_{t-2} + \exp(w_{t-1}/2) r_y z_{t-1}|^r \\ &\leq \tilde{c}_r \left[E|c y_{t-2}|^r + E|\exp(w_{t-1}/2) r_y z_{t-1}|^r \right] \\ &= \tilde{c}_r \left[|c|^r E|y_{t-2}|^r + |r_y|^r E|z_{t-1}|^r E \exp(\frac{r}{2} w_{t-1}) \right] \;. \end{aligned} \tag{53}$$

Under the assumption of stationarity, we have:

$$E|y_{t-1}|^r \;\leq\; \frac{\tilde{c}_r}{1 - \tilde{c}_r |c|^r} |r_y|^r E|z_{t-1}|^r \exp(\frac{r^2}{8} \frac{r_w^2}{1-a^2}) \stackrel{def}{=} K_r < \infty \tag{54}$$

for $1 - \tilde{c}_r |c|^r \neq 0$, with

$$\begin{aligned} \tilde{c}_r &= 1 \quad \text{for } 0 < r \leq 1 \;, \\ \tilde{c}_r &= 2^{r-1} \quad \text{for } r > 1 \;, \end{aligned}$$

and where $E|z_t|^r$ is computed at equations (49) and (51). In particular, we have for $r = 1$,

$$E|y_{t-1}| \leq \frac{1}{1 - |c|} |r_y| \sqrt{\frac{2}{\pi}} \exp(\frac{1}{8} \frac{r_w^2}{1-a^2}) \stackrel{def}{=} K_1$$

Thus, we can say that:

$$E|y_{t-1}|^r \leq K_r < \infty \ ,$$

for r finite. At this step, it is now possible to compute equation (46) which becomes

$$E|u_t y_{t-1}|^r \leq E|z_t|^r |r_y|^r \exp(\frac{r^2}{8}\frac{r_w^2}{1-a^2})K_r \overset{def}{=} B_r < \infty \tag{55}$$

and this holds for any r finite. $E|u_t y_{t-1}|^r$ can be decomposed as $E|u_t y_{t-1}|^{1+\theta}$ with $\theta = (r-1) > 0$ i.e. $r > 1$ then we can state that [see Davidson, (1994, p.190 Theorem 12.10)]:

$$\lim_{M\to\infty} E(|u_t(c_0)y_{t-1}| 1_{|u_t(c_0)y_{t-1}|\geq M}) = 0 \ .$$

And this holds for all $t \in \mathbb{N}\backslash\{0\}$. Thus, the collection $\{u_t(c_0)y_{t-1},\ t \in \mathbb{N}\backslash\{0\}\}$ is uniformly integrable.

On the other hand, let $c_t = E|u_t(c_0)y_{t-1}|$ which corresponds to equation (55) with $r = 1$ from which we can deduce that $c_t \leq B_1 < \infty$ therefore we have that:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} c_t \leq \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} B_1 = B_1 < \infty \ .$$

Thus, the process $\{u_t(c_0)y_{t-1},\ t \in \mathbb{N}\backslash\{0\}\}$ which is a m.d.s. w.r.t. $\mathcal{F}_t$ could be described as a L_1-mixingale w.r.t. $\mathcal{F}_t$ with $\xi_0 = 1$ and $\xi_m = 0\ , m \geq 1$. We can then apply the Law of Large Numbers (L.L.N.) for L_1-mixingale that [see Hamilton (1994, p.191, Proposition 7.6)]:

$$\frac{1}{T}\sum_{t=1}^{T}(u_t(c_0)y_{t-1}) \overset{P}{\to} E(u_t(c_0)y_{t-1}) = 0 \ . \tag{56}$$

We shall show below that the process $\{y_t^2 - \mu_{Y^2}\ , t \in \mathbb{N}\}$ is a L_1-mixingale w.r.t. the subfields $\Omega_t = \sigma(Y_t, Y_{t-1}, \dots)$ where $Y_t = (y_t, w_t, z_t, v_t)'$. We know that:

$$\begin{aligned} y_t^2 &= (cy_{t-1} + \exp(w_t/2)r_y z_t)^2 \\ &= c^2 y_{t-1}^2 + \exp(w_t) r_y^2 z_t^2 + 2cy_{t-1}\exp(w_t/2)r_y z_t \end{aligned} \tag{57}$$

and iterating backwards on y_t^2 we have:

$$\begin{aligned} y_t^2 &= (c^2)^m y_{t-m}^2 + (c^2)^{m-1}[\exp(w_{t-m+1})r_y^2 z_{t-m+1}^2 + 2cy_{t-m}\exp(\frac{w_{t-m+1}}{2})r_y z_{t-m+1}] \\ &\quad + \dots + (c^2)^0[\exp(w_t)r_y^2 z_t^2 + 2cy_{t-1}\exp(w_t/2)r_y z_t] \ . \end{aligned} \tag{58}$$

Besides,

$$\mu_{Y2} \stackrel{def}{=} Ey_t^2 = \frac{\mu_2}{1-c^2} \ .$$

Therefore,

$$\begin{aligned} E(y_t^2 - \mu_{Y2}|\Omega_{t-m}) &= (c^2)^m y_{t-m}^2 - \mu_{Y2} \\ &\quad + \sum_{j=0}^{m-1}(c^2)^j \Big\{ E\Big[\exp(w_{t-j}) r_y^2 z_{t-j}^2 + 2c y_{t-1-j}\exp(w_{t-j}/2) r_y z_{t-j} |\Omega_{t-m}\Big]\Big\} \\ &= (c^2)^m y_{t-m}^2 - \mu_{Y2} + \sum_{j=0}^{m-1}(c^2)^j \Big\{ r_y^2 E(z_{t-j}^2|\Omega_{t-m}) \\ &\quad E\Big[\exp(a^{m-j} w_{t-m} + \sum_{l=0}^{m-j-1} a^l r_w v_{t-j-l})|\Omega_{t-m}\Big]\Big\} \\ &= (c^2)^m y_{t-m}^2 - \mu_{Y2} + \sum_{j=0}^{m-1}(c^2)^j \Big\{ r_y^2 E(z_{t-j}^2|\Omega_{t-m}) \exp(a^{m-j} w_{t-m}) \\ &\quad E\Big[\prod_{l=0}^{m-j-1} \exp(a^l r_w v_{t-j-l})|\Omega_{t-m}\Big]\Big\} \ . \end{aligned} \tag{59}$$

Thus,

$$\begin{aligned} E(y_t^2 - \mu_{Y2}|\Omega_{t-m}) &= (c^2)^m y_{t-m}^2 - \mu_{Y2} + \sum_{j=0}^{m-1}(c^2)^j \Big\{ r_y^2 \exp(a^{m-j} w_{t-m}) \\ &\quad \prod_{l=0}^{m-j-1} E\exp(a^l r_w v_{t-j-l})\Big\} \end{aligned} \tag{60}$$

because $z_t \stackrel{i.i.d.}{\sim} N(0,1)$ and $v_t \stackrel{i.i.d.}{\sim} N(0,1)$. Hence,

$$\begin{aligned} E(y_t^2 - \mu_{Y2}|\Omega_{t-m}) &= (c^2)^m y_{t-m}^2 - \mu_{Y2} + \sum_{j=0}^{m-1}(c^2)^j \Big\{ r_y^2 \exp(a^{m-j} w_{t-m}) \\ &\quad \prod_{l=0}^{m-j-1} \exp(\frac{1}{2} a^{2l} r_w^2)\Big\} \ . \end{aligned} \tag{61}$$

Therefore,

$$
\begin{aligned}
E|E(y_t^2 - \mu_{Y2}|\Omega_{t-m})| \;\leq\; & E\Bigg(|(c^2)^m y_{t-m}^2 - \mu_{Y2}| \\
& + |\sum_{j=0}^{m-1}(c^2)^j \Bigg\{ r_y^2 \exp(a^{m-j} w_{t-m}) \exp(\frac{1}{2}\sum_{l=0}^{m-j-1} a^{2l} r_w^2) \Bigg\} |\Bigg) \\
\leq\; & E\Bigg(|(c^2)^m y_{t-m}^2| + |\mu_{Y2}| \\
& + \sum_{j=0}^{m-1}(c^2)^j |r_y^2 \exp(a^{m-j} w_{t-m}) \exp(\frac{r_w^2}{2}\sum_{l=0}^{m-j-1} a^{2l})|\Bigg) \\
\leq\; & (c^2)^m E y_{t-m}^2 + \mu_{Y2} \\
& + \sum_{j=0}^{m-1}(c^2)^j r_y^2 \exp(\frac{a^{2(m-j)}}{2}\frac{r_w^2}{1-a^2}) \exp(\frac{r_w^2}{2}\sum_{l=0}^{m-j-1} a^{2l}) \; .
\end{aligned}
\tag{62}
$$

Since $\mu_{Y2} > 0$ we also have:

$$
\begin{aligned}
E|E(y_t^2 - \mu_{Y2}|\Omega_{t-m})| \;\leq\; & (c^2)^m E y_{t-m}^2 - \mu_{Y2} \\
& + \sum_{j=0}^{m-1}(c^2)^j r_y^2 \exp(\frac{a^{2(m-j)}}{2}\frac{r_w^2}{1-a^2}) \exp(\frac{r_w^2}{2}\sum_{l=0}^{m-j-1} a^{2l}) \; .
\end{aligned}
\tag{63}
$$

We have:

$$
\begin{aligned}
& \lim_{m\to\infty} r_y^2 \sum_{j=0}^{m-1}(c^2)^j \exp(\frac{a^{2(m-j)}}{2}\frac{r_w^2}{1-a^2}) \exp(\frac{r_w^2}{2}\sum_{l=0}^{m-j-1} a^{2l}) \\
& \qquad = \frac{1}{1-c^2} r_y^2 \exp(\frac{r_w^2}{2}\frac{1}{1-c^2}) = \frac{\mu_2}{1-c^2} = \mu_{Y2} \; .
\end{aligned}
$$

Therefore,

$$E|E(y_t^2 - \mu_{Y2}|\Omega_{t-m})| \leq c_t \xi_m$$

with $c_t = 1, \ \forall t$ and

$$\begin{aligned}
\xi_m &= (c^2)^m \mu_{Y2} - \mu_{Y2} \\
&\quad + r_y^2 \sum_{j=0}^{m-1} (c^2)^j \exp\left(\frac{a^{2(m-j)}}{2} \frac{r_w^2}{1-a^2}\right) \exp\left(\frac{r_w^2}{2} \sum_{l=0}^{m-j-1} a^{2l}\right) \quad \forall\, m \qquad (64)
\end{aligned}$$

with $\lim_{m\to\infty} \xi_m = 0$ and $\lim_{m\to\infty} \frac{1}{T}\sum_{t=1}^{T} c_t = 1 < \infty$. Thus, the process $\{y_t^2 - \mu_{Y2}\ , t \in \mathbb{N}\}$ is a L_1-mixingale w.r.t. the subfields $\Omega_t = \sigma(Y_t, Y_{t-1}, \dots)\ , t \in N$. Besides, exploiting once again the $\tilde{c}_r-$inequality we have:

$$\begin{aligned}
E|y_t^2 - \mu_{Y2}|^2 &\leq 2(E|y_t^2|^2 + E|\mu_{Y2}|^2) \\
&= 2\mu_{Y4} + 2\mu_{Y2}^2\ , \qquad (65)
\end{aligned}$$

where some algebra yields for μ_{Y4}:

$$\mu_{Y4} \stackrel{def}{=} Ey_t^4 = \frac{1}{1-c^4}\left[3r_y^4 \exp\left(\frac{2r_w^2}{1-a^2}\right) + 6c^2 r_y^2 \mu_{Y2} \exp\left(\frac{r_w^2}{2(1-a^2)}\right)\right] .$$

Hence, $E|y_t^2 - \mu_{Y2}|^2$ above is finite and can be decomposed as: $E|y_t^2 - \mu_{Y2}|^{1+\theta} < \infty$, with $\theta = 1$, then we can conclude [see James Davidson (1994), p.190, Theorem 12.10] that $\lim_{M\to\infty} E(|y_t^2 - \mu_{Y2}| 1_{|y_t^2 - \mu_{Y2}| \geq M}) = 0$. And this holds for all $t \in \mathbb{N}$. We can then say that the process $\{y_t^2 - \mu_{Y2}\ , t \in \mathbb{N}\}$ is uniformly integrable. We can then apply the Law of Large Numbers (L.L.N.) for L_1-mixingales [see Hamilton (1994, p.191, proposition 7.6)] on the process $\{y_t^2 - \mu_{Y2}\ , t \in \mathbb{N}\}$ to get

$$\frac{1}{T}\sum_{t=1}^{T} (y_t^2 - \mu_{Y2}) \stackrel{P}{\to} 0$$

that is

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 \stackrel{P}{\to} \mu_{Y2}\ . \qquad (66)$$

Then, using the L.L.N. for L_1-mixingales and the Central Limit Theorem (C.L.T.) for m.d.s. [see Hamilton (1994, p.193, Proposition 7.8)], we have

$$\begin{aligned}
p\lim_{T\to\infty} \sqrt{T}(\hat{c}_T - c_0) &= \left(p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} p\lim_{T\to\infty} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t-1} u_t \\
&= (\mu_{Y2})^{-1} Z \qquad (67)
\end{aligned}$$

with $Z \sim N(0, \frac{\mu_2^2}{1-c^2})$, hence,

$$\sqrt{T}(\hat{c}_T - c_0) \xrightarrow{D} N(0, 1 - c^2) \ . \tag{68}$$

And using equation (56) we can also say that:

$$\begin{aligned} p\lim_{T\to\infty}(\hat{c}_T - c_0) &= \left(p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_t \\ &= (\mu_{Y2})^{-1} \times 0 \ , \\ &= 0 \ . \end{aligned} \tag{69}$$

Consequently, the term $\frac{2}{T}\sum_{t=1}^{T}(u_t(c_0)y_{t-1})\sqrt{T}(\hat{c}_T - c_0)$ is $o_p(1)$ by equations (56), (67) and (68). Therefore, equation (43) is equivalent to

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2 - \mu_2(\theta_0)) \# \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_0)^2 - \mu_2(\theta_0))$$

asymptotically.

C.2 The component $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4 - \mu_4(\theta_0))$

Recall that $u_t(c)^4 \stackrel{def}{=} (y_t - cy_{t-1})^4$ and $\hat{u}_t^4 \stackrel{def}{=} u_t^4(\hat{c}_T)$, then making a Taylor's expansion of order one of the function $u_t(c)^4$ with $\hat{c}_T = c_0 + h$ we have:

$$u_t^4(\hat{c}_T) = u_t(c_0)^4 - 4y_{t-1}u_t(c_0)^3 h + \frac{12}{2!}y_{t-1}^2 u_t(\bar{c})^2 h^2 \tag{70}$$

where $R_t \stackrel{def}{=} 6y_{t-1}^2 u_t(\bar{c})^2 h^2$ where we note that $\lim_{h\to 0} \frac{R_t}{h} = 0$. Aggregating equation (71) over the sample size and replacing h by $\hat{c}_T - c_0$, we have

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^4(\hat{c}_T) - \mu_4(\theta_0)) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_0)^4 - \mu_4(\theta_0)) - \frac{4}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}u_t(c_0)^3(\hat{c}_T - c_0) \\ &\quad + \frac{6}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}^2 u_t(\bar{c})^2(\hat{c}_T - c_0)^2 \ , \end{aligned} \tag{71}$$

where $R_T \stackrel{def}{=} \frac{6}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}^2 u_t(\bar{c})^2(\hat{c}_T - c_0)^2$ with $\bar{c} = c_0 + \alpha h$ with $0 < \alpha < 1$. Rewritting R_T as

$$R_T = \frac{6}{T}\sum_{t=1}^{T} y_{t-1}^2 u_t(\bar{c})^2 \sqrt{T}(\hat{c}_T - c_0)(\hat{c}_T - c_0)$$

shows by equations (66), (68) and (69) respectively that R_T is an $o_p(1)$-variable. Equation (71) is equivalent to

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^4(\hat{c}_T)-\mu_4(\theta_0)) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_0)^4-\mu_4(\theta_0))-\frac{4}{T}\sum_{t=1}^{T}y_{t-1}u_t(c_0)^3\sqrt{T}(\hat{c}_T-c_0) + \frac{6}{\sqrt{T}}\sum_{t=1}^{T}y_{t-1}^2u_t(\bar{c})^2(\hat{c}_T-c_0)^2 \ . \quad (72)$$

We show that the process $\{u_t(c_0)^3y_{t-1}\}=\{v_t(\theta_0)^3y_{t-1}\}$ is a m.d.s. w.r.t. $\mathcal{F}_t$ since:

$$\begin{aligned} E[u_t(c_0)^3y_{t-1}|\mathcal{F}_{t-1}] &= y_{t-1}E[u_t(c_0)^3|\mathcal{F}_{t-1}] \\ &= y_{t-1}E[v_t(\theta_0)^3|\mathcal{F}_{t-1}] \\ &= y_{t-1}E[\exp(\frac{3}{2}w_t)r_{y0}^3z_t^3|\mathcal{F}_{t-1}] \\ &= y_{t-1}r_{y0}^3E(z_t^3|\mathcal{F}_{t-1})E[\exp(\frac{3}{2}w_t)|\mathcal{F}_{t-1}] \\ &= y_{t-1}r_{y0}^3E(z_t^3)E[\exp(\frac{3}{2}w_t)] \\ &= 0 \ , \end{aligned} \quad (73)$$

from which we can deduce that $E[u_t(c_0)^3y_{t-1}]=0 \ ,\forall t\in\mathbb{N}\backslash\{0\}$.

Besides, a little algebra yields for r a finite positive integer that:

$$\begin{aligned} E|u_t(c_0)^3y_{t-1}|^r &= E|z_t|^{3r}|r_{y0}|^{3r}E[\exp(\frac{3r}{2}w_t)]E|y_{t-1}|^r \\ &\leq E|z_t|^{3r}|r_{y0}|^{3r}\exp(\frac{9r^2}{8}\frac{r_w^2}{1-a^2})K_r \stackrel{def}{=} B_r<\infty \end{aligned} \quad (74)$$

according to equations (54), (49) and (51). Setting $c_t=E|u_t(c_0)^3y_{t-1}|$, we can see at the light of equation (74) with $r=1$ that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}c_t\leq\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}B_1=B_1<\infty \ .$$

Additionnaly, $E|u_t(c_0)^3y_{t-1}|^r$ can be decomposed as $E|u_t(c_0)^3y_{t-1}|^{1+\theta}<\infty$ for $\theta=r-1>0$ which yields [see James Davidson, (1994), p.190, Theorem 12.10] that

$$\lim_{M\to\infty}E\left(|u_t(c_0)^3y_{t-1}|1_{|u_t(c_0)^3y_{t-1}|\geq M}\right)=0 \ .$$

And this holds for all $t \in \mathbb{N}\backslash\{0\}$. So the collection $\{u_t(c_0)^3 y_{t-1}\ , t \in \mathbb{N}\backslash\{0\}\}$ is uniformly integrable. Thus, the latter which is a m.d.s. w.r.t. $\mathcal{F}_t$ could be described as a L_1-mixingale w.r.t. $\mathcal{F}_t$ with $\xi_0 = 1$, and $\xi_m = 0$ for $m \geq 1$. Then, we can apply the L.L.N. for L_1-mixingale to establish that:

$$\frac{1}{T}\sum_{t=1}^{T} u_t(c_0)^3 y_{t-1} \stackrel{P}{\to} E(u_t(c_0)^3 y_{t-1}) = 0\ . \tag{75}$$

Therefore, the term $\frac{1}{T}\sum_{t=1}^{T} u_t(c_0)^3 y_{t-1}\sqrt{T}(\hat{c}_T - c_0)$ is $o_p(1)$ since $\sqrt{T}(\hat{c}_T - c_0) \stackrel{D}{\to} N(0, 1-c^2)$ by equation (68).

Consequently, equation (72) is equivalent to

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4 - \mu_4(\theta_0))\#\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_0)^4 - \mu_4(\theta_0))$$

asymptotically.

C.3 The component $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2\hat{u}_{t-1}^2 - \mu_{2,2}(1|\theta_0))$

Finally, a Taylor's expansion of order one of $u_t(\hat{c}_T)^2 u_{t-1}^2(\hat{c}_T)$ with $\hat{c}_T = c_0 + h$ yields

$$\begin{aligned} u_t(\hat{c}_T)^2 u_{t-1}^2(\hat{c}_T) &= u_t(c_0)^2 u_{t-1}(c_0)^2 + [-2y_{t-1}u_t(c_0)u_{t-1}(c_0)^2 - 2y_{t-2}u_t(c_0)^2 u_{t-1}(c_0)]h \\ &\quad + \frac{1}{2!}[2y_{t-1}^2 u_{t-1}(\bar{c})^2 + 2y_{t-2}^2 u_t(\bar{c})^2 + 8y_{t-1}y_{t-2}u_t(\bar{c})u_{t-1}(\bar{c})]h^2 \end{aligned} \tag{76}$$

where $R_t \stackrel{def}{=} \frac{1}{2!}[2y_{t-1}^2 u_{t-1}(\bar{c})^2 + 2y_{t-2}^2 u_t(\bar{c})^2 + 8y_{t-1}y_{t-2}u_t(\bar{c})u_{t-1}(\bar{c})]h^2$ with $\lim_{h\to 0}\frac{R_t}{h} = 0$. After a few manipulations we have:

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(\hat{c}_T)u_{t-1}^2(\hat{c}_T) - \mu_{2,2}(1|\theta_0)) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_0)^2 u_{t-1}(c_0)^2 - \mu_{2,2}(1|\theta_0)) \\ &\quad - \frac{2}{\sqrt{T}}\sum_{t=1}^{T}[u_t(c_0)u_{t-1}^2(c_0)y_{t-1} + u_t^2(c_0)u_{t-1}(c_0)y_{t-2}]h \\ &\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[y_{t-1}^2 u_{t-1}(\bar{c})^2 + y_{t-2}^2 u_t(\bar{c})^2 \\ &\quad + 4y_{t-1}y_{t-2}u_t(\bar{c})u_{t-1}(\bar{c})]h^2 \end{aligned} \tag{77}$$

where the last term corresponds to R_T with $\lim_{h\to 0}\frac{R_T}{h}=0$. Replacing then h by $\hat{c}_T - c_0$ we have the equivalent form:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(\hat{c}_T)u_{t-1}^2(\hat{c}_T)-\mu_{2,2}(1|\theta_0)) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_0)^2u_{t-1}(c_0)^2-\mu_{2,2}(1|\theta_0)) \\ &\quad -\frac{2}{T}\sum_{t=1}^{T}[u_t(c_0)u_{t-1}^2(c_0)y_{t-1}+u_t^2(c_0)u_{t-1}(c_0)y_{t-2}] \\ &\quad \times\sqrt{T}(\hat{c}_T-c_0)+\frac{1}{T}\sum_{t=1}^{T}[y_{t-1}^2u_{t-1}(\bar{c})^2+y_{t-2}^2u_t(\bar{c})^2 \\ &\quad +4y_{t-1}y_{t-2}u_t(\bar{c})u_{t-1}(\bar{c})]\sqrt{T}(\hat{c}_T-c_0)(\hat{c}_T-c_0) \end{aligned} \quad (78)$$

where the remainder is $o_p(1)$ since $(\hat{c}_T - c_0)$ is $o_p(1)$.

As $E(u_t^2(c_0)u_{t-1}(c_0)y_{t-2}) = E(y_{t-2})E(u_t^2(c_0)u_{t-1}) = 0$ since $E(y_{t-2}) = 0$ and $E(u_t(c_0)u_{t-1}^2(c_0)y_{t-1}) = E[u_{t-1}^2(c_0)y_{t-1}E(u_t(c_0)|\mathcal{F}_{t-1})] = 0$ since

$$\begin{aligned} E[u_t(c_0)|\mathcal{F}_{t-1}] &= E[r_y z_t \exp(\frac{w_t}{2})|\mathcal{F}_{t-1}] \\ &= r_y E(z_t|\mathcal{F}_{t-1})E[\exp(\frac{w_t}{2})|\mathcal{F}_{t-1}] \\ &= 0 \end{aligned} \quad (79)$$

under Assumption 2, and we recall that $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$ with $X_t = (y_t, u_t, z_t, v_t)'$. At this step, we need to show that $u_t^2(c_0)u_{t-1}(c_0)y_{t-2}$ is a L_1-mixingale in order to allow for the L.L.N. for L_1-mixingales for the second term of the right-hand side of equation (**??**). To do so, we need to show that

$$E|E[u_t^2(c_0)u_{t-1}(c_0)y_{t-2}|\mathcal{F}_{t-m}]| \le c_t\xi_m$$

with $\lim_{m\to\infty}\xi_m = 0$. Let us first compute $E[u_t^2(c_0)u_{t-1}(c_0)y_{t-2}|\mathcal{F}_{t-m}]$ for $m \ge 3$, i.e.

$$\begin{aligned} E[u_t^2(c_0)u_{t-1}(c_0)y_{t-2}|\mathcal{F}_{t-m}] &= E[u_t^2u_{t-1}(c^{m-2}y_{t-m}+c^{m-3}u_{t-m+1}+c^{m-4}u_{t-m+2}+\dots \\ &\quad +cu_{t-3}+u_{t-2})|\mathcal{F}_{t-m}] \\ &= E[c^{m-2}y_{t-m}u_t^2u_{t-1}+c^{m-3}u_t^2u_{t-1}u_{t-m+1}+c^{m-4}u_t^2u_{t-1}u_{t-m+2} \\ &\quad +\dots+cu_t^2u_{t-1}u_{t-3}+u_t^2u_{t-1}u_{t-2}|\mathcal{F}_{t-m}] \\ &= 0 \end{aligned} \quad (80)$$

because for $m \geq 3$, the terms of the form:

$$\begin{aligned} E[u_t^2 u_{t-1} u_{t-m+1} | \mathcal{F}_{t-m}] &= E\Big[r_y^2 z_t^2 \exp(a^m w_{t-m} + a^{m-1} r_w v_{t-m+1} + \ldots + a r_w v_{t-1} + r_w v_t) \\ &\quad r_y z_{t-1} \exp[\frac{1}{2}(a^{m-1} w_{t-m} + a^{m-2} r_w v_{t-m+1} + \ldots + a r_w v_{t-2} + r_w v_{t-1})] \\ &\quad r_y z_{t-m+1} \exp[\frac{1}{2}(a w_{t-m} + r_w v_{t-m+1})] | \mathcal{F}_{t-m}\Big] \\ &= E(z_{t-1}|\mathcal{F}_{t-m}) E(z_{t-m+1}|\mathcal{F}_{t-m}) E\Big[r_y^4 z_t^2 \exp[(a^m + \frac{1}{2}a^{m-1} + \frac{1}{2}a) w_{t-m}] \\ &\quad \exp[(a^{m-1} + \frac{1}{2}a^{m-2} + \frac{1}{2}) r_w v_{t-m+1}] \prod_{j=1}^{m-2} \exp[(a^j + \frac{1}{2}a^{j-1}) r_w v_{t-j}] \\ &\quad \exp(r_w v_t) | \mathcal{F}_{t-m}\Big] \\ &= 0 \end{aligned} \tag{81}$$

because $E(z_{t-1}|\mathcal{F}_{t-m}) = E(z_{t-1}) = 0$ according to Assumption 2, and for the same reasons we also have:

$$\begin{aligned} E[y_{t-m} u_t^2 u_{t-1} | \mathcal{F}_{t-m}] &= y_{t-m} r_y^3 E(z_t^2|\mathcal{F}_{t-m}) E(z_{t-1}|\mathcal{F}_{t-m}) \\ &\quad E[\exp(a^m w_{t-m} + a^{m-1} r_w v_{t-m+1} + \ldots + a r_w v_{t-1} + r_w v_t) \\ &\quad \exp[\frac{1}{2}(a^{m-1} w_{t-m} + a^{m-2} r_w v_{t-m+1} + \ldots + a r_w v_{t-2} + r_w v_{t-1})] | \mathcal{F}_{t-m}] \\ &= 0\ . \end{aligned} \tag{82}$$

So we have for $m \geq 3$ that $E[u_t^2(c_0) u_{t-1}(c_0) y_{t-2} | \mathcal{F}_{t-m}] = 0$. Similarly, we also have that $E[u_t^2(c_0) u_{t-1}(c_0) y_{t-2} | \mathcal{F}_{t-2}] = 0$. Therefore,

$$E|E[u_t^2(c_0) u_{t-1}(c_0) y_{t-2} | \mathcal{F}_{t-m}]| = 0\ , \quad m \geq 2$$

Now for $m = 1$ we have:

$$\begin{aligned} E[u_t^2 u_{t-1} y_{t-2} | \mathcal{F}_{t-1}] &= u_{t-1} y_{t-2} E[r_y^2 z_t^2 \exp(a w_{t-1} + r_w v_t) | \mathcal{F}_{t-1}] \\ &= u_{t-1} y_{t-2} r_y^2 E(z_t^2 | \mathcal{F}_{t-1}) \exp(a w_{t-1}) E[\exp(r_w v_t) | \mathcal{F}_{t-1}] \\ &= u_{t-1} y_{t-2} r_y^2 \exp(a w_{t-1}) E(z_t^2) E[\exp(r_w v_t)] \\ &= u_{t-1} y_{t-2} r_y^2 \exp(a w_{t-1}) \exp(\frac{1}{2} r_w^2)\ . \end{aligned} \tag{83}$$

Then, we have:

$$\begin{aligned}
E|E[u_t^2 u_{t-1} y_{t-2}|\mathcal{F}_{t-1}]| &= E|r_y z_{t-1} \exp(w_{t-1}/2) y_{t-2} r_y^2 \exp(a w_{t-1}) \exp(\frac{1}{2} r_w^2)| \\
&= E\left(\exp[(a+\frac{1}{2}) w_{t-1}] \exp(\frac{1}{2} r_w^2) |r_y^3||z_{t-1}||y_{t-2}| \right) \\
&= |r_y^3| \exp(\frac{1}{2} r_w^2) \sqrt{\frac{2}{\pi}} E|y_{t-2}| \exp[\frac{(a+\frac{1}{2})^2}{2} \frac{r_w^2}{1-a^2}] \\
&\leq |r_y^3| \exp(\frac{1}{2} r_w^2) \sqrt{\frac{2}{\pi}} K_1 \exp[\frac{(a+\frac{1}{2})^2}{2} \frac{r_w^2}{1-a^2}] \overset{def}{=} B
\end{aligned} \tag{84}$$

recalling that $E|y_{t-2}| \leq K_1$ and $E|z_{t-1}| = \sqrt{\frac{2}{\pi}}$. A similar calculation yields also that

$$E|E[u_t^2 u_{t-1} y_{t-2}|\mathcal{F}_t]| = E|u_t^2 u_{t-1} y_{t-2}| \leq B \; .$$

Then, the process $\{u_t^2 u_{t-1} y_{t-2}\}$ is a L_1-mixingale with $c_t = B$, $\forall t$, $\xi_m = 1$ for $m = 0, 1$, and $\xi_m = 0$ for $m \geq 2$, and $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^T c_t = B < \infty$. Moreover, let us show now that the process $\{u_t^2 u_{t-1} y_{t-2}\}$ is uniformly integrable. To do this, we shall compute $E|u_t^2 u_{t-1} y_{t-2}|^r$ for $r = 1, 2, 3, \ldots$.

$$\begin{aligned}
E|u_t^2 u_{t-1} y_{t-2}|^r &= E|r_y^2 z_t^2 \exp(w_t) r_y z_{t-1} \exp(w_{t-1}/2) y_{t-2}|^r \\
&= E|r_y^3 z_t^2 z_{t-1} y_{t-2} \exp(a w_{t-1} + r_w v_t) \exp(w_{t-1}/2)|^r \\
&= E\left(|r_y^3||z_t^2||z_{t-1}||y_{t-2}| \exp[(a+\frac{1}{2}) w_{t-1}] \exp(r_w v_t) \right)^r \\
&= |r_y^3|^r E z_t^{2r} E|z_{t-1}|^r E|y_{t-2}|^r \exp[\frac{r^2(a+\frac{1}{2})^2}{2} \frac{r_w^2}{1-a^2}] \exp(\frac{r^2 r_w^2}{2}) \\
&= |r_y^3|^r \frac{(2r)!}{2^r r!} E|z_{t-1}|^r E|y_{t-2}|^r \exp[\frac{r^2(a+\frac{1}{2})^2}{2} \frac{r_w^2}{1-a^2}] \exp(\frac{r^2 r_w^2}{2}) \\
&\leq |r_y^3|^r \frac{(2r)!}{2^r r!} E|z_{t-1}|^r K_r \exp[\frac{r^2(a+\frac{1}{2})^2}{2} \frac{r_w^2}{1-a^2}] \exp(\frac{r^2 r_w^2}{2}) \\
&< \infty \quad \text{for } r < \infty \; .
\end{aligned} \tag{85}$$

Decomposing:

$$E|u_t^2 u_{t-1} y_{t-2}|^r = E|u_t^2 u_{t-1} y_{t-2}|^{1+\theta} < \infty$$

for $\theta = (r-1) > 0$ i.e. $r > 1$, r a finite integer, then

$$\lim_{M\to\infty} E(|u_t^2 u_{t-1} y_{t-2}| 1_{|u_t^2 u_{t-1} y_{t-2}| \geq M}) = 0 \ ,$$

[see Davidson (1994, p.190, Theorem 12.10)]. And this holds for any $t \in N\backslash\{0,1\}$. Then, the collection $\{u_t^2 u_{t-1} y_{t-2} \ , t \in \mathbb{N}\backslash\{0,1\}\}$ is a uniformly integrable L_1-mixingale w.r.t. $\mathcal{F}_t$. Finally, we can apply the L.L.N. for L_1-mixingales to have that:

$$\frac{1}{T}\sum_{t=1}^{T} u_t^2 u_{t-1} y_{t-2} \xrightarrow{P} E u_t^2 u_{t-1} y_{t-2} = 0 \ . \tag{86}$$

We shall show now that the process $\{u_t u_{t-1}^2 y_{t-1}\}$ is a L_1-mixingale w.r.t. $\mathcal{F}_t$. More precisely, it is a m.d.s. w.r.t. $\mathcal{F}_t$ since:

$$\begin{aligned} E[u_t(c_0)u_{t-1}^2(c_0)y_{t-1}|\mathcal{F}_{t-1}] &= u_{t-1}^2(c_0)y_{t-1}E[u_t(c_0)|\mathcal{F}_{t-1}] \\ &= u_{t-1}^2(c_0)y_{t-1}E[r_{y0}z_t\exp(w_t/2)|\mathcal{F}_{t-1}] \\ &= r_{y0}u_{t-1}^2(c_0)y_{t-1}E(z_t)E(\exp(w_t/2)) \\ &= 0 \ , \end{aligned} \tag{87}$$

from which we can deduce that $E(u_t(c_0)u_{t-1}^2(c_0)y_{t-1}) = 0$. Moreover, by the $\tilde{c}_r$-inequality we can state that:

$$\begin{aligned} E|u_t(c_0)u_{t-1}^2(c_0)y_{t-1}|^r &\leq \tilde{c}_r\Bigg\{|c|^r E|y_{t-2}|^r |r_y^3|^r E|z_t|^r E|z_{t-1}^2|^r \exp\left[\frac{r^2}{2}(1+a/2)^2\frac{r_w^2}{1-a^2}\right]\exp(\frac{r^2r_w^2}{8}) \\ &\quad + |r_y^4|^r E|z_t|^r E|z_{t-1}^3|^r \exp\left[\frac{r^2}{2}(1/2+1+a/2)^2\frac{r_w^2}{1-a^2}\right]\exp(\frac{r^2r_w^2}{8})\Bigg\} \\ &\leq \tilde{c}_r\Bigg\{|c|^r K_r |r_y|^{3r}\gamma_r\frac{(2r)!}{2^r r!}\exp\left[\frac{r^2}{2}(1+a/2)^2\frac{r_w^2}{1-a^2}\right]\exp(\frac{r^2r_w^2}{8}) \\ &\quad + |r_y|^{4r}\gamma_r\gamma_{3r}\exp\left[\frac{r^2}{2}(1/2+1+a/2)^2\frac{r_w^2}{1-a^2}\right]\exp(\frac{r^2r_w^2}{8})\Bigg\} \\ &\overset{def}{=} B_r < \infty \end{aligned} \tag{88}$$

where it has been shown earlier that $E|y_{t-2}|^r \leq K_r$ by equation (54), that $E|z_t|^r \equiv \gamma_r$, $E|z_{t-1}^3|^r = E|z_{t-1}|^{3r} \equiv \gamma_{3r}$, and $E|z_{t-1}^2|^r = E z_{t-1}^{2r} = \frac{(2r)!}{2^r r!}$ since $z_t \sim N(0,1)$, and γ_r and γ_{3r} have been computed at equations (49), (51). We recall that

$$\tilde{c}_r = 1 \quad \text{for } 0 < r \leq 1 \ ,$$
$$\tilde{c}_r = 2^{r-1} \quad \text{for } r > 1 \ .$$

Taking $c_t = E|u_t(c_0)u_{t-1}^2(c_0)y_{t-1}|$ which corresponds to equation (**??**) with $r = 1$ yields

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} c_t \leq \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} B_1 = B_1 < \infty \ .$$

On the other hand, decomposing $E|u_t(c_0)u_{t-1}^2(c_0)y_{t-1}|^r$ as $E|u_t(c_0)u_{t-1}^2(c_0)y_{t-1}|^{1+\theta} < \infty$ with $\theta = r-1 > 0$, i.e. $1 < r < \infty$, then

$$\lim_{M\to\infty} E\left(|u_t(c_0)u_{t-1}^2(c_0)y_{t-1}| 1_{|u_t(c_0)u_{t-1}^2(c_0)y_{t-1}|\geq M} \right) = 0 \ .$$

And this holds for all $t \in \mathbb{N}\backslash\{0\}$. Hence, the collection $\{u_t(c_0)u_{t-1}^2(c_0)y_{t-1} \ , \forall\ t \in \mathbb{N}\backslash\{0\}\}$ is uniformly integrable. Thus, the latter which is a m.d.s. w.r.t. $\mathcal{F}_t$ could ne described as a specific L_1-mixingale w.r.t. $\mathcal{F}_t$ with $\xi_0 = 1$ and $\xi_m = 0$ for $m \geq 1$. Then, we can apply the L.L.N. for L_1-mixingales to assess that:

$$\frac{1}{T}\sum_{t=1}^{T} u_t(c_0)u_{t-1}^2(c_0)y_{t-1} \xrightarrow{P} E(u_t(c_0)u_{t-1}^2(c_0)y_{t-1}) = 0 \ . \tag{89}$$

Finally, given that the terms $\frac{1}{T}\sum_{t=1}^{T} u_t^2 u_{t-1}y_{t-2}$ and $\frac{1}{T}\sum_{t=1}^{T} u_t(c_0)u_{t-1}^2(c_0)y_{t-1}$ are both $o_p(1)$ [see equations (**??**) and (**??**)] and that $\sqrt{T}(\hat{c}_T - c_0) \xrightarrow{D} N(0, 1-c^2)$ by equation (68), then the term $\frac{2}{T}\sum_{t=1}^{T}(u_t u_{t-1}^2 y_{t-1} + u_t^2 u_{t-1} y_{t-2})\sqrt{T}(\hat{c}_T - c_0)$ is also $o_p(1)$, yielding the asymptotic equivalence below.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2\hat{u}_{t-1}^2 - \mu_{2,2}(1)(\theta_0)) \# \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_0)^2 u_{t-1}(c_0)^2 - \mu_{2,2}(1)(\theta_0)) \ .$$

Thus,

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2 - \mu_2(\theta_0)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4 - \mu_4(\theta_0)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2\hat{u}_{t-1}^2 - \mu_{2,2}(1)(\theta_0)) \end{pmatrix} \overset{asy}{\#} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(c_0) - \mu_2(\theta_0)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^4(c_0) - \mu_4(\theta_0)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(c_0)u_{t-1}^2(c_0) - \mu_{2,2}(1)(\theta_0)) \end{pmatrix} \tag{90}$$

and from equation (14) we know that $u_t(c_0)^r = v_t(\theta_0)^r\ \forall\ t$ then we have the asymptotic equivalence

$$\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) \overset{asy}{\#} \sqrt{T}(\bar{g}_T(\theta_0) - \mu(\theta_0)) \ ,$$

with $\bar{g}_T(\theta)$ defined as in equation (16)

D Proof of Proposition 4

In order to get asymptotic normality for $\sqrt{T}(\bar{g}_T(\theta_0) - \mu(\theta_0))$ we shall use a Central Limit Theorem (C.L.T) for dependent processes [see Davidson (1994, p.385, Theorem 24.5)]. In this aim, we need to verify first the conditions under which this C.L.T holds. To do this, we will apply propositions 5 and 17 from Carrasco,Chen (1999) to conclude that

i) if $\{w_t\}$ is geometrically ergodic, then $\{(w_t, \ln|v_t|)\}$ is Markov geometrically ergodic with the same decay rate as that of $\{w_t\}$;

ii) if $\{w_t\}$ is stationary β-mixing with a certain decay rate, then $\{\ln|v_t|\}$ is β-mixing with a decay rate at least as fast as that of $\{w_t\}$.

If initialized from its stationary distribution, then $\{\ln|v_t|\}$ is strictly stationary β-mixing with an exponential decay rate. Since this property is preserved by any continuous transformation, $\{v_t\}$ and hence $\{v_t^k\}$ and $\{v_t^k v_{t-1}^k\}$ are strictly stationary and exponential β-mixing where we recall that $v_t = v_t(\theta_0) = \exp(\frac{w_t}{2}) r_{y0} z_t = \exp(\frac{a_0 w_{t-1} + r_{w0} v_t}{2}) r_{y0} z_t$. Then, the process

$$X_t \stackrel{def}{=} \begin{pmatrix} v_t^2(\theta_0) - \mu_2(\theta_0) \\ v_t^4(\theta_0) - \mu_4(\theta_0) \\ v_t^2(\theta_0) v_{t-1}^2(\theta_0) - \mu_{2,2}(1|\theta_0) \end{pmatrix} = g_t(\theta_0) - \mu(\theta_0) \tag{91}$$

is strictly stationary and exponential β-mixing. Moreover, application of theorem 14.2, p.211 [see Davidson (1994)] shows that a mixing zero-mean process is an adapted L_1-mixingale with respect to the subfields $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$ provided it is bounded in the L_1-norm. Indeed $\{X_t\}$ is bounded in the L_1-norm as we show below.

i)

$$\begin{aligned} E|v_t^2 - \mu_2(\theta_0)| &\leq E(|v_t^2| + |\mu_2(\theta_0)|) \\ &= 2\mu_2(\theta_0) \\ &= 2r_y^2 \exp(\frac{r_w^2}{2(1-a^2)}) \,. \end{aligned}$$

There exists $M < \infty$ such that: $E|v_t^2 - \mu_2(\theta_0)| < M < \infty$.

ii) In the same way,

$$E|v_t^4 - \mu_4(\theta_0)| \leq 2\mu_4(\theta_0) < M < \infty.$$

iii) Similarly,

$$E|v_t^2 v_{t-1}^2 - \mu_{2,2}(1|\theta_0)| \leq 2\mu_{2,2}(1|\theta_0) < M < \infty.$$

The process $\{X_t\}$ is L_1-bounded, therefore it is an adapted L_1-mixingale w.r.t. $\mathcal{F}_t$.

The process $\{X_t\}$ which is a L_1-mixingale w.r.t. $\mathcal{F}_t$, with mixing coefficients of the form $\beta_n = c\rho^n$, $c > 0$, , $0 < \rho < 1$, is shown below to be of size -1. To show that the process $\{X_t\}$ is of size -1, we need to show that its mixing coefficients β_n are such that: $\beta_n = O(n^{-\phi})$, with $\phi > 1$. Indeed,

$$\begin{aligned}\frac{\rho^n}{n^{-\phi}} &= n^{\phi} \exp(n \log \rho) \\ &= \exp(\phi \log n) \exp(n \log \rho) \\ &= \exp(\phi \log n + n \log \rho) \ .\end{aligned}$$

It is known [see Rudin (1976, p.57, Theorem 3.20d)] that $\lim_{n\to\infty} \phi \log n + n \log \rho = -\infty$ which yields

$$\lim_{n\to\infty} \exp(\phi \log n + n \log \rho) = 0 \ .$$

And this holds with $\phi > 1$.

Now, the last condition to verify before applying the Central Limit Theorem for dependent processes is to show that

$$\limsup_{T\to\infty} T^{-1/2} E|S_T| < \infty$$

where

$$S_T \stackrel{def}{=} \sum_{t=1}^{T} X_t = \sum_{t=1}^{T} \begin{pmatrix} v_t^2(\theta_0) - \mu_2(\theta_0) \\ v_t^4(\theta_0) - \mu_4(\theta_0) \\ v_t^2(\theta_0) v_{t-1}^2(\theta_0) - \mu_{2,2}(1|\theta_0) \end{pmatrix}$$

We will write the proof only for the two last components of X_t that is:

$$\begin{aligned}T^{-1/2} E|\sum_{t=1}^{T}(v_t^4(\theta_0) - \mu_4(\theta_0))| &\leq E\sum_{t=1}^{T} |v_t^4(\theta_0) - \mu_4(\theta_0)| \\ &< T^{-1/2} E\sum_{t=1}^{T}(|v_t^4(\theta_0)| + |\mu_4(\theta_0)|)\end{aligned}$$

since $v_t^4(\theta_0) > 0$ and $\mu_4(\theta_0) > 0$ since we work with real-valued processes. Then, we have

$$\begin{aligned} T^{-1/2}E|\sum_{t=1}^{T}(v_t^4(\theta_0) - \mu_4(\theta_0))| &< T^{-1/2}E\sum_{t=1}^{T}(|v_t^4(\theta_0)| + |\mu_4(\theta_0)|) \\ &= T^{-1/2}E\sum_{t=1}^{T}(v_t^4(\theta_0) + \mu_4(\theta_0)) \\ &= T^{-1/2}2\mu_4(\theta_0) \end{aligned}$$

And replacing $\mu_4(\theta_0)$ by its value we have that

$$\limsup_{T\to\infty} T^{-1/2}E|\sum_{t=1}^{T}(v_t^4(\theta_0) - \mu_4(\theta_0))| < \lim_{T\to\infty} 6T^{-1/2}r_y^4\exp(\frac{2r_w^2}{1-a^2}) = +\infty \ .$$

Likewise, we have

$$\begin{aligned} T^{-1/2}E|\sum_{t=1}^{T}(v_t^2(\theta_0)v_{t-1}^2(\theta_0) - \mu_{2,2}(1|\theta_0))| &\leq E\sum_{t=1}^{T}|v_t^2(\theta_0)v_{t-1}^2(\theta_0) - \mu_{2,2}(1|\theta_0)| \\ &< T^{-1/2}E\sum_{t=1}^{T}(|v_t^2(\theta_0)v_{t-1}^2(\theta_0)| + |\mu_{2,2}(1|\theta_0)|) \end{aligned}$$

since $v_t^2(\theta_0)v_{t-1}^2(\theta_0) > 0$ and $\mu_{2,2}(1)(\theta_0) > 0$. Thus, we have

$$\begin{aligned} T^{-1/2}E|\sum_{t=1}^{T}(v_t^2(\theta_0)v_{t-1}^2(\theta_0) - \mu_{2,2}(1|\theta_0))| &< T^{-1/2}E\sum_{t=1}^{T}(|v_t^2(\theta_0)v_{t-1}^2(\theta_0)| + |\mu_{2,2}(1|\theta_0)|) \\ &= T^{-1/2}E\sum_{t=1}^{T}(v_t^2(\theta_0)v_{t-1}^2(\theta_0) + \mu_{2,2}(1|\theta_0)) \\ &= 2T^{-1/2}\mu_{2,2}(1|\theta_0) \\ &= 2T^{-1/2}r_y^4\exp(\frac{r_w^2}{1-a}) \underset{T\to\infty}{\to} +\infty \ . \end{aligned}$$

Hence, we have

$$\limsup_{T\to\infty} T^{-1/2}E|\sum_{t=1}^{T}(v_t^2(\theta_0)v_{t-1}^2(\theta_0) - \mu_{2,2}(1|\theta_0))| < \infty \ .$$

Since the proof is similar for the first component of X_t we can assess that $\limsup_{T\to\infty} T^{-1/2}E|\sum_{t=1}^{T} X_t| < \infty$. Therefore we can apply Theorem 24.5, p.385 [see Davidson (1994)] to the process $\{X_t, \mathcal{F}_t\}$ defined in equation (**??**) with $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$, and establish that

$$T^{-1/2}S_T = T^{-1/2}\sum_{t=1}^{T} X_t = \sqrt{T}(\bar{g}_T(\theta_0) - \mu(\theta_0)) \xrightarrow{D} N(0, \Omega^*) \ .$$

E Proof of Proposition 5

The method-of-moments estimator $\hat{\theta}_T(\Omega)$ is solution of the following optimization problem:

$$\min_{\theta} M_T(\theta) = \min_{\theta}(\mu(\theta) - \bar{g}_T(\hat{U}))'\hat{\Omega}(\mu(\theta) - \bar{g}_T(\hat{U})) \tag{92}$$

where we recall that

$$\bar{g}_T(\hat{U}) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \\ \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4 \\ \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2 \end{pmatrix} \ , \tag{93}$$

or else $\bar{g}_T(\hat{U}) = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{U})$ and $\hat{u}_t = u_t(\hat{c}_T)$, and $\mu(\theta)' = (\mu_2(\theta), \mu_4(\theta), \mu_{2,2}(1|\theta))$. The first order conditions (F.O.C) associated with this problem are:

$$\frac{\partial\mu'}{\partial\theta}(\hat{\theta}_T)\hat{\Omega}(\mu(\hat{\theta}_T) - \bar{g}_T(\hat{U})) = 0 \ .$$

An expansion of the F.O.C above around the true value θ_0 yields

$$\frac{\partial\mu'}{\partial\theta}(\hat{\theta}_T)\hat{\Omega}\bigg(\mu(\theta_0) + \frac{\partial\mu}{\partial\theta'}(\theta_0)(\hat{\theta}_T - \theta_0) - \bar{g}_T(\hat{U})\bigg) \simeq 0$$

after rearranging the equation we have

$$\sqrt{T}(\hat{\theta}_T(\Omega) - \theta_0) \simeq \left(\frac{\partial\mu'}{\partial\theta}(\theta_0)\Omega\frac{\partial\mu}{\partial\theta'}(\theta_0)\right)^{-1}\frac{\partial\mu'}{\partial\theta}(\theta_0)\Omega\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) \ .$$

Using then, propositions 3 and 4 we get the asymptotic normality of $\hat{\theta}_T(\Omega)$ with asymptotic covariance matrix $W(\Omega)$ as specified in proposition 6.

F Proof of Proposition 7

The proofs derived here follow the lines of Gourieroux,Monfort,Renault (1993). The parameter θ such that $\theta' = (a, r_y, r_w)$ is partitioned into two subvectors

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

with $\theta_1 = a$ and $\theta_2' = (r_y, r_w)$. The null hypothesis is defined by $H_0 = \{\theta_1 = 0\}$ which corresponds to test the absence of long memory in the model, i.e. $a = 0$.

The expansion given earlier can be rewritten under the null hypothesis with the optimal metric Ω^{*-1}:

$$\sqrt{T}\begin{bmatrix} \hat{\theta}_{1T} \\ \hat{\theta}_{2T} - \theta_{20} \end{bmatrix} \sim \left[\begin{pmatrix} \frac{\partial \mu'}{\partial \theta_1} \\ \frac{\partial \mu'}{\partial \theta_2}, \end{pmatrix} \Omega^{*-1}(\frac{\partial \mu}{\partial \theta_1'}, \frac{\partial \mu}{\partial \theta_2'})\right]^{-1} \begin{pmatrix} \frac{\partial \mu'}{\partial \theta_1} \\ \frac{\partial \mu'}{\partial \theta_2}, \end{pmatrix} \Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0))$$

$$\sqrt{T}\hat{\theta}_{1T} \simeq (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}(\frac{\partial \mu'}{\partial \theta_1} - A_{12}A_{22}^{-1}\frac{\partial \mu'}{\partial \theta_2})\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) .$$

We note that:

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

where

$$M_2 = \frac{\partial \mu}{\partial \theta_2'}\left[\frac{\partial \mu'}{\partial \theta_2}\Omega^{*-1}\frac{\partial \mu}{\partial \theta_2'}\right]^{-1}\frac{\partial \mu'}{\partial \theta_2}\Omega^{*-1} \tag{94}$$

and $A_{ij} = \frac{\partial \mu'}{\partial \theta_i}\Omega^{*-1}\frac{\partial \mu}{\partial \theta_j}$ which yields

$$\sqrt{T}\hat{\theta}_{1T} \simeq (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}\frac{\partial \mu'}{\partial \theta_1}[Id - M_2]'\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) . \tag{95}$$

Thus,

$$Var_{as}(\sqrt{T}\hat{\theta}_{1T}) = (A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}\frac{\partial \mu'}{\partial \theta_1}[Id-M_2]'\Omega^{*-1}[Id-M_2]\frac{\partial \mu}{\partial \theta_1'}(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}$$

and we show that $\frac{\partial\mu'}{\partial\theta_1}[Id-M_2]'\Omega^{*-1}[Id-M_2]\frac{\partial\mu}{\partial\theta_1'} = (A_{11}-A_{12}A_{22}^{-1}A_{21})$ yielding

$$Var_{as}(\sqrt{T}\hat{\theta}_{1T}) = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} \ . \tag{96}$$

Indeed,

$$\begin{aligned}
\frac{\partial\mu'}{\partial\theta_1}[Id - M_2]'\Omega^{*-1}[Id - M_2]\frac{\partial\mu}{\partial\theta_1'} &= \frac{\partial\mu'}{\partial\theta_1}\left(Id - \frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\right)'\Omega^{*-1} \\
&\quad \left(Id - \frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\right)\frac{\partial\mu}{\partial\theta_1'} \\
&= \frac{\partial\mu'}{\partial\theta_1}\left(Id - \Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\right)\Omega^{*-1} \\
&\quad \left(Id - \frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\right)\frac{\partial\mu}{\partial\theta_1'} \\
&= \left(\frac{\partial\mu'}{\partial\theta_1}\Omega^{*-1} - \frac{\partial\mu'}{\partial\theta_1}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_1'}A_{22}^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\right) \\
&\quad \left(\frac{\partial\mu}{\partial\theta_1'} - \frac{\partial\mu}{\partial\theta_2'}A_{22}^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_1'}\right) \\
&= \frac{\partial\mu'}{\partial\theta_1}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_1'} - \frac{\partial\mu'}{\partial\theta_1}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}A_{22}^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_1'} \\
&= A_{11} - A_{12}A_{22}^{-1}A_{21} \ .
\end{aligned}$$

Thus, the Wald statistic

$$\xi_T^W = T\hat{\theta}_{1T}'\hat{W}_1^{-1}\hat{\theta}_{1T}$$

is asymptotically equivalent to:

$$\begin{aligned}
\xi_T^W &= T(\bar{g}_T(\hat{U}) - \mu(\theta_0))'\Omega^{*-1}[Id - M_2]\frac{\partial\mu}{\partial\theta_1'}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} \\
&\quad (A_{11} - A_{12}A_{22}^{-1}A_{21})\left\{\frac{\partial\mu'}{\partial\theta_1}[Id - M_2]'\Omega^{*-1}[Id - M_2]\frac{\partial\mu}{\partial\theta_1'}\right\}^{-1} \\
&\quad (A_{11} - A_{12}A_{22}^{-1}A_{21})(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}\frac{\partial\mu'}{\partial\theta_1}[Id - M_2]'\Omega^{*-1}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) \ ,
\end{aligned}$$

that is

$$\xi_T^W = T(\bar{g}_T(\hat{U}) - \mu(\theta_0))'\Omega^{*-1}[Id - M_2]\frac{\partial \mu}{\partial \theta_1'}$$
$$\left\{\frac{\partial \mu'}{\partial \theta_1}[Id - M_2]'\Omega^{*-1}[Id - M_2]\frac{\partial \mu}{\partial \theta_1'}\right\}^{-1}\frac{\partial \mu'}{\partial \theta_1}$$
$$[Id - M_2]'\Omega^{*-1}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) \ .$$

The score statistic is based on the gradient of the objective function with respect to θ_1 evaluated at the constrained estimator $\hat{\theta}_T'^c = (0, \hat{r}_y^c, \hat{r}_w^c)$ i.e.

$$\begin{aligned}
\mathcal{D}_T &= \frac{\partial \mu'}{\partial \theta_1}(\hat{\theta}_T^c)\Omega^{*-1}(\mu(\hat{\theta}_T^c) - \bar{g}_T(\hat{U})) \\
&\simeq \frac{\partial \mu'}{\partial \theta_1}(\theta_0)\Omega^{*-1}\left(\mu(\theta_0) + \frac{\partial \mu}{\partial \theta_2'}(\theta_0)(\hat{\theta}_{2T}^c - \theta_{20}) - \bar{g}_T(\hat{U})\right) \\
&\simeq -\frac{\partial \mu'}{\partial \theta_1}(\theta_0)\Omega^{*-1}\left(\bar{g}_T(\hat{U}) - \mu(\theta_0) - \frac{\partial \mu}{\partial \theta_2'}(\theta_0)(\hat{\theta}_{2T}^c - \theta_{20})\right) \\
&\simeq -\frac{1}{\sqrt{T}}\frac{\partial \mu'}{\partial \theta_1}(\theta_0)\Omega^{*-1}\left(\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) - \frac{\partial \mu}{\partial \theta_2'}(\theta_0)\sqrt{T}(\hat{\theta}_{2T}^c - \theta_{20})\right)
\end{aligned}$$

Given that

$$\sqrt{T}(\hat{\theta}_{2T}^c - \theta_{20}) \simeq \left(\frac{\partial \mu'}{\partial \theta_2}(\theta_0)\Omega^{*-1}\frac{\partial \mu}{\partial \theta_2'}(\theta_0)\right)^{-1}\frac{\partial \mu'}{\partial \theta_2}(\theta_0)\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0))$$

we have

$$\begin{aligned}
\mathcal{D}_T &\simeq -\frac{1}{\sqrt{T}}\frac{\partial \mu'}{\partial \theta_1}(\theta_0)\Omega^{*-1}\left(\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) - \frac{\partial \mu}{\partial \theta_2'}(\theta_0)A_{22}^{-1}\frac{\partial \mu'}{\partial \theta_2}(\theta_0)\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0))\right) \\
&\simeq -\frac{1}{\sqrt{T}}\frac{\partial \mu'}{\partial \theta_1}(\theta_0)(Id - M_2)'\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0))
\end{aligned}$$

where M_2 has been defined at equation (**??**). Finally, from equation (**??**) we have

$$\mathcal{D}_T \simeq -(A_{11} - A_{12}A_{22}^{-1}A_{21})\hat{\theta}_{1T} \ . \tag{97}$$

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

$$Var_{as}(\mathcal{D}_T) = A_{11} - A_{12}A_{22}^{-1}A_{21} \ . \tag{98}$$

On the other hand, the difference between the two optimal values of the objective function (the constrained minus the unconstrained one) is such that

$$\begin{aligned}\xi_T^C &\simeq T(\bar{g}_T(\hat{U})-\mu(\theta))'[Id-M_2]'\Omega^{*-1}[Id-M_2](\bar{g}_T(\hat{U})-\mu(\theta))\\ &\quad -(\bar{g}_T(\hat{U})-\mu(\theta))'[Id-M]'\Omega^{*-1}[Id-M](\bar{g}_T(\hat{U})-\mu(\theta))\end{aligned}$$

where $M_2 = \frac{\partial\mu}{\partial\theta_2'}A_{22}^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}$ and $M = \frac{\partial\mu}{\partial\theta_2'}\left(\frac{\partial\mu'}{\partial\theta}\Omega^{*-1}\frac{\partial\mu}{\partial\theta'}\right)^{-1}\frac{\partial\mu'}{\partial\theta}\Omega^{*-1}$. Thus,

$$\begin{aligned}\xi_T^C &\simeq T(\bar{g}_T(\hat{U})-\mu(\theta))'\Omega^{*-1}[Id-M_2](\bar{g}_T(\hat{U})-\mu(\theta))\\ &\quad -(\bar{g}_T(\hat{U})-\mu(\theta))'\Omega^{*-1}[Id-M](\bar{g}_T(\hat{U})-\mu(\theta))\\ &\simeq T(\bar{g}_T(\hat{U})-\mu(\theta))'\Omega^{*-1}[M-M_2](\bar{g}_T(\hat{U})-\mu(\theta))\ .\end{aligned}$$

A classical argument of block inverse gives

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

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

References

Andersen, T., G. (1994, April). Stochastic autoregressive volatility: A framework for volatility modelling. *Mathematical Finance 4*(2), 75–102.

Andersen, T. and B. Sorensen (1996). Gmm estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Economics and Business 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.

Baillie, R. and H. Chung (2001). Estimation of garch models from the autocorrelations of the squares of a process. *Journal of Time Series mAnalysis 22*(6), 631–650.

Bansal, R., A. R. Gallant, R. Hussey, and G. E. Tauchen (1995). Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics 66*, 251–287.

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.

Birnbaum, Z. W. (1974). Computers and unconventional test-statistics. In F. Proschan and R. J. Serfling (Eds.), *Reliability and Biometry*, pp. 441–458. Philadelphia, PA: SIAM.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 51*, 307–327.

Carrasco, M. and X. Chen (1999). β-mixing and moment properties of various garch, stochastic volatility and acd models. Discussion paper, London School of Economics.

Chen, M. and G. Chen (2000). Geometric ergodicity of nonlinear autoregressive models with changing conditional variances. *The Canadian Journal of Statistics 28*(3), 605–613.

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.

Davidson, J. (1994). *Stochastic Limit Theory*. Advanced Texts in Econometrics. New York: Oxford University Press Inc.

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

Dufour, J.-M. (1995). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M., L. Khalaf, J.-T. Bernard, and I. Genest (2001). Simulation-based finite-sample tests for heteroscedasticity and arch effects. Discussion Paper 08-2001, Universitè de Montrèal, C.R.D.E.

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 (1995). Estimation of stochastic volatility models with diagnostics. Discussion paper, Duke University.

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

Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. In C. Rao and G. Maddala (Eds.), *Statistical Methods in Finance*, Holland. North-Holland.

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. New York: Springer-Verlag.

Gouriéroux, C. and A. Monfort (1995). *Simulation Based Econometric Methods*. CORE Lecture Series. Louvain-la-Neuve: CORE Foundation.

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

Gradshteyn, I. and I. Ryzhik (1980). *Table of integrals, series, and products. Corrected and enlarged edition*. Academic Press inc.

Hamilton, J. (1994). *Time Series Analysis*. New Jersey: Princeton University Press Inc.

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.

Kim, S. and N. Shephard (1994). Stochastic volatility: Optimal likelihood inference and comparison with arch models. Technical report, Nuffield College, Oxford. unpublished paper.

Kim, S., N. Shephard, and S. Chib (1996). Stochastic volatility: Likelihood inference and comparison with arch models. Technical report, Nuffield College, Oxford. unpublished paper.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility : Likelihood inference and comparison with ARCH models. *Review of economic studies 65*, 361–393.

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.

Melino, A. and S. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics 45*, 239–265.

Monfardini, C. (1997). Estimating stochastic volatility models through indirect inference. Technical report, European University Institute.

Newey, W. K. and K. D. West (1987a). Hypothesis testing with efficient method of moments estimators. *International Economic Review 28*, 777–787.

Newey, W. K. and K. D. West (1987b). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, Inc.

Shephard, N. (1996). Statistical aspects of arch and stochastic volatility. In D. Cox, O. Barndorff-Nielsen, and D. Hinkley (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*. Chapman & Hall. London.

Tauchen, G. (1996, January). New minimum chi-square methods in empirical finance. Technical report, Department of Economics, Duke University.

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

Taylor, S. (1994). Modelling stochastic volatility. *Mathematical Finance 4*, 183–204.