Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

WORKING PAPER 2002/14

Volatility Estimation via Hidden Markov Models

Alessandro Rossi,
Giampiero M. Gallo

Università degli Studi
di Firenze

Econometrics

Volatility Estimation via Hidden Markov Models*

Alessandro ROSSI† Giampiero M. GALLO‡

June 17, 2002

Abstract

In this paper we suggest a convenient way to obtain parameter estimates of a discrete state hidden Markov volatility process within a framework consistent with observed option prices and stochastic volatility. Relative to similar proposals, we simplify the model estimation by resorting to some parametric approximation of the model in a maximum likelihood context. We show how correlation between returns and volatility innovations can be easily accommodated within this framework. Empirical applications illustrate model search strategies for the SP500 stock index, comparing the performances to a standard GARCH model.

Keywords: Stochastic volatility, Hidden Markov, GARCH, Smile-consistent option pricing, Forecasting.

JEL: C22, C53, G13

*We thank Christophe Planas and Stewart Hodges for valuable comments on a preliminary version of this paper. The usual disclaimer applies.

†European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen, Ispra (VA), Italy. Tel. +39 0332 789724 Fax +39 0332 785733 e-mail: alessandro.rossi@jrc.it

‡Department of Statistics "G. Parenti", University of Florence, Italy. Tel. +39 055 4237257 Fax +39 055 4223560 e-mail: gallog@ds.unifi.it

1 Introduction

In the vast array of econometric models of volatility, it is customary to assume that volatility is time–varying and can take over a continuous range of positive values. GARCH–type (Bollerslev *et al.*, 1994) and stochastic volatility models (Harvey *et al.*, 1996) fall in this category. In the theoretical literature on option pricing, though, it is not infrequent to encounter models where volatility is assumed to take on a finite number of states (Duan *et al.*, 2002).

Britten-Jones and Neuberger (BN, 2000) have recently shown how the classic Black and Scholes (1973) option pricing model can be extended to make it compatible with stochastic volatility and observed option prices. Their objective is to evaluate and hedge path-dependent derivatives. Specifically, given the arbitrage-free prices of European options, and allowing for a wide range of stochastic volatility dynamics, BN derive the class of price processes for the underlying asset which are consistent with the observed smile surface. They suggest a simple algorithm to implement the model in practice, in a discrete-time state-space framework (see also Rossi, 2002) where volatility is assumed to take on a finite number of states. However, their paper does not go as far as to deal with parameter estimation.

In what follows, we show that working with a finite number of states for volatility is not restrictive from an estimation point of view, relative to other models where volatility can take positive real values. We present an efficient way to recover estimates of asset price volatility parameters when a multiplicative model for the evolution of the price of a certain asset is considered in discrete-time, and volatility dynamics is consistent with the BN approach and is assumed to evolve as a hidden, discrete-state, Markov process. This model has good general properties: for example, we show that the negative correlation between returns and volatility – the so-called *leverage effect* – can be easily accommodated. We may stress that our approach is not an attempt at specifying the "best" model of volatility, since a many volatility models are capable of matching given stylized facts, but it has suitable properties when pricing and hedging of exotic derivatives is of interest.

We follow a maximum likelihood (ML) estimation framework adopting filtering techniques to estimate the set of parameters involved in the model specification, given the unobservability of the volatility process. Among several proposals, we favor the filtering approach introduced by Elliott, Aggoun and Moore (1995), which makes use of *reference probability methods* to draw inference about the unobserved state variable. The appealing feature of their

approach is that it is possible to simplify the filtering procedure using standard results available for independent and identically distributed random variables, introducing a change of measure. Their procedure is applied in Elliott, Hanter and Jamieson (1998) to the IBM and gold prices. We depart from their approach, since we suggest a parameterization which allows us to reduce the number of parameters to be estimated, to manage Markov chains of larger sizes, and to introduce a *leverage effect* in the model. Moreover, given the hill–shape of the likelihood function, we show the gains in computation times obtained by using the Simulated Annealing (SA) algorithm as an efficient method to find the global maxima. By comparison, the popular Expectation Maximization (EM) algorithm (Dempster, Laird and Rubin 1977) appears to be more demanding (higher number of parameters, a more cumbersome filtering procedure, and the need to use different initialization values).

The paper is structured as follows. Section 2 states the model and the main assumptions. In section 3, we show how the imposition of some restrictions on the parameter space allows us to reduce the number of parameters for model estimation. In section 4 we show how the model can be extended to take into account the leverage effect. Model estimation is discussed in section 5 whereas in section 6 we estimate volatility parameters for the SP500 stock index, showing that the in and out-of sample performance of this model comes very close to standard *GARCH* family models. Section 7 concludes.

2 A Markovian framework for volatility

Let S_t be the price of a certain asset at time t. We consider asset returns, observable at time $t+1$, as a random variable $Y_{t+1} = \ln\left(\frac{S_{t+1}}{S_t}\right)$, for a given time horizon $t = 0, 1, ..., \tau - 1$. Its volatility is driven by a discrete-time, finite-state Markov chain $\mathbf{Z}_t$ with N states. We denote with I_t the P-augmented increasing sigma-field generated by $\{\mathbf{Z}_s, Y_s : s \leq t\}$, whereas restrictions to I_t generated by the specific random variables are denoted by superscripts (e.g. $I_t^{\mathbf{Z}}$).

We will assume the distribution of $\mathbf{Z}_0$ and N as known. For convenience, the Markov chain is assumed to be one of the N-dimensional unit vectors, $\mathbf{e}_i$ $(i = 1, 2, ..., N)$ with a one in the i-th position and zeros elsewhere. The stochastic volatility model of interest here can be written in the state space

form as

$$\begin{aligned} Y_{t+1} &= \mu_t + \sigma(\mathbf{Z}_t) W_{t+1} \\ \mathbf{Z}_{t+1} &= \mathbf{M}\mathbf{Z}_t + \mathbf{V}_{t+1} \end{aligned} \tag{1}$$

where $\mu_t = E[Y_{t+1} | I_t]$ is the conditional mean of the observable process, W_t is i.i.d. Normal $(0,1)$ and $\sigma(\mathbf{Z}_t)$ is the volatility at time $t+1$ with $\sigma(\cdot)$ a positive valued scaling function.

The transition equation for $\mathbf{Z}_{t+1}$ highlights how the short term dynamics of the first-order Markov chain is fully described by the $N \times N$ one-step transition matrix $\mathbf{M}$, the generic element of which is

$$m_{ij} \equiv P(\mathbf{Z}_{t+1} = \mathbf{e}_i | \mathbf{Z}_t = \mathbf{e}_j) \tag{2}$$

describes the transition probability of the chain. The entries of $\mathbf{M}$ satisfy $m_{ij} \geq 0$ and $\sum_i m_{ij} = 1$, for each $1 \leq i, j \leq N$. With $\mathbf{Z}_t$ one of the unit vectors in $\mathbf{R}^N$ and using (2) we have $E\left[\mathbf{Z}_{t+1} \middle| I_t^{\mathbf{Z}}\right] = E[\mathbf{Z}_{t+1} | \mathbf{Z}_t] = \mathbf{M}\mathbf{Z}_t$. Hence defining $\mathbf{V}_{t+1} \equiv \mathbf{Z}_{t+1} - \mathbf{M}\mathbf{Z}_t$, we have $E\left[\mathbf{V}_{t+1} \middle| I_t^{\mathbf{Z}}\right] = \mathbf{0}$, which provides a semi-martingale representation for the transition equation. In the simplest case, the error terms W_t and $\mathbf{V}_t$ are assumed to be independent, but, as we will see, this hypothesis may be easily relaxed.

According to the model (1) and (2), we assume that volatility can take only a finite number of values, defined by the choice of the scaling function $\sigma(\cdot)$. When time elapses, the latent state variable $\mathbf{Z}$ switches from one regime to another according to the one step transition matrix $\mathbf{M}$. Under the assumptions of model, the conditional variance of the observation process is given by

$$Var\left[Y_{t+1} \middle| I_t^Y\right] = E\left[\sigma(\mathbf{Z}_t)^2 \middle| I_t^Y\right] \tag{3}$$

This model is consistent with many features shown by asset price innovations which are well documented by the *GARCH* (see for instance Bollerslev, Engle and Nelson, 1994) and Stochastic Volatility literature (a review is contained in Ghysels, Harvey and Renault, 1996). It allows for *fat tails*, since it enables different degrees of kurtosis in the unconditional distribution of asset returns, thanks to the randomness of the process $\mathbf{Z}$; *volatility clustering*, via the autoregressive behavior of the conditional expectation (3); *volatility persistence* and *mean-reversion*, via the stochastic features of the latent process $\mathbf{Z}$. *Leverage effect*, that is a negative correlation between returns and volatility innovations, can be accommodated by removing the independence assumption between the observation and state errors.

In a recent paper (Elliott et al., 1998), this model is applied to IBM stock and gold prices with $\mu_t = \mathbf{c}'\mathbf{Z}_t$ and $\sigma(\mathbf{Z}_t) = \boldsymbol{\sigma}'\mathbf{Z}_t$, being $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_N)$

and $\mathbf{c} = (c_1, c_2, ..., c_N)$. Note that this approach does not take into consideration the leverage effect, a characteristic often encountered with financial series. Hereafter, we refer to this structure as the Hidden Markov Unrestricted Model (HMU).[1]

Let $\boldsymbol{\theta} = \{\mu_i, \sigma_{i,} m_{ij} : i, j = 1, 2, ..., N\}$ denote the population parameters characterizing the probability distribution $P(\mathbf{Y}; \boldsymbol{\theta})$ of the observed data. The primary tasks are

- to estimate $\boldsymbol{\theta}$ subject to the constraints $m_{ij} \geq 0$, $\sum_{i=1}^{N} m_{ij} = 1$ and $\sigma_j > 0$, for each $1 \leq j \leq N$; and
- to make inference about the unobserved sequence $\mathbf{Z}$, given the available information up to time t (filtering).

3 A simple model parameterization

The estimation of the unrestricted model is far from being problem-free. First, the dimension d of the parameter vector $\boldsymbol{\theta}$ increases quadratically with the number of states of the Markov chain $\left(d = N^2 + N\right)$. There are no guidelines to establish the size of the Markov chain in practical applications; however, to make an example, a value of N equal to 7, (which could be justified in practice to get a good fit to the data) implies a number of parameters equal to 56. From a theoretical point of view, overparameterized models lead to non efficient estimators even in large samples (Harvey, 1990), whereas, from a computational point of view, it might not be straightforward to calibrate the EM algorithm with respect to the parameter starting values. In practice, this requires several hundreds of values from which to start the maximization procedure even when the size of the Markov chain is small (see Hamilton, 1990) making the approach heavily time consuming. Nevertheless, non concavity of the likelihood function could make the global maxima still go undetected.

A second point concerns the MLE standard errors when the EM is involved. Although resampling techniques, such as the bootstrap, may solve this issue, the presence of multiple local maxima of the likelihood function makes this approach practically unfeasible.

[1] For reference purposes we recall some technical aspects related to the HMU estimation in Appendix A. It shows how recursive filters for states, the number of jumps, the occupation time of the Markov chain, and processes related to the observations, can be derived pursuing the Elliott et al. (1995) approach. These quantities turns out to be useful for parameter estimation of the unrestricted model via ML, using the EM algorithm.

In light of these remarks, we propose some possible alternatives. In this section, we suggest how the model could be parsimoniously parameterized in a way which removes the dependence upon N. Later in the paper, a dependence structure between the Markov process and observations is also introduced, simply relaxing the homogeneity property of the Markov chain.

We start by specifying the conditional mean. Several alternatives are possible: for instance,

$$\mu_t = \mu + \sum_{k=1}^{p} \alpha_k Y_{t-k+1} + \lambda \sigma^2 (\mathbf{Z}_t) \tag{4}$$

allows for dependence on the past of Y and response to variations in the conditional variance; calendar or seasonal effects may also be considered.

As in the case of Elliott and al. (1998), we assume $\sigma (\mathbf{Z}_t) = \boldsymbol{\sigma}' \mathbf{Z}_t$, but we depart from their framework by constraining $\boldsymbol{\sigma}$

$$\sigma_i = \exp \{\alpha + \delta g (\mathbf{e}_i)\}, \quad i = 1, 2, ..., N \tag{5}$$

which specifies volatility regimes as a function of only two coefficients, α and δ, whichever the number of states. The function $g(\cdot)$, is defined as follows

$$g (\mathbf{e}_i) = \frac{2i - (N+1)}{N-1}$$

which has the effect of associating distinct values between -1 and 1 to each regime. When δ is assumed to be positive, we obtain the result of identifying by σ_1 the lowest volatility regime, and by σ_N the highest one. In addition, since we wish to capture the mean-reverting property of volatility, we parameterize the entries of the transition matrix $\mathbf{M}$ as a function of a constant parameter ϕ the magnitude of which determines the mean-reverting rate

$$m_{ij} = \begin{cases} 1 - \phi & i = j \\ \frac{1}{2}\phi [1 + g (\mathbf{e}_j)] & i = j - 1 \\ \frac{1}{2}\phi [1 - g (\mathbf{e}_j)] & i = j + 1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

To ensure $0 \leq m_{ij} \leq 1$, for each i and j, ϕ must belong to $[0, 1]$. From (5) and (6) it is easy to prove[2] that the (log) volatility process is mean-reverting,

[2]To do this, let us reason by induction. First let $k = 1$, and consider $E [\ln \sigma (\mathbf{Z}_{t+1}) | \mathbf{Z}_t]$. Simple calculations give

$$E [\ln \sigma (\mathbf{Z}_{t+1}) | \mathbf{Z}_t] = \ln \sigma (\mathbf{Z}_t) + \phi [\alpha - \ln \sigma (\mathbf{Z}_t)]$$

that is,

$$\begin{aligned} E\left[\ln\sigma\left(\mathbf{Z}_{t+k}\right)\left|I_t^{\mathbf{Z}}\right.\right] &= E\left[\ln\sigma\left(\mathbf{Z}_{t+k}\right)|\mathbf{Z}_t\right] \\ &= \ln\sigma\left(\mathbf{Z}_t\right)+\left[1-(1-\phi)^k\right]\left[\alpha-\ln\sigma\left(\mathbf{Z}_t\right)\right] \quad (7) \end{aligned}$$

Lower values of ϕ imply higher persistence of the mean-reverting volatility process. Note that equation (6) not only constrains the volatility being mean reverting, but also states that large sudden shifts of volatility are not possible. This property of continuity in the behavior of the volatility process is exploited by BN to derive their option pricing formula.

Taking the limit of the above expression we get

$$\lim_{k\to\infty} E\left[\ln\sigma\left(\mathbf{Z}_{t+k}\right)\left|I_t^{\mathbf{Z}}\right.\right]=\alpha$$

Hence, α is best viewed as the long-run (log) volatility level forecast, thus making δ in (5) interpretable as a scaling coefficient tied to the volatility of volatility. We refer to the structure (1), (4), (5), (6) as the Hidden Markov (hereafter *HM*) restricted model.

4 Leverage effect

The empirical evidence of volatility reacting differently according to the sign of return innovations (higher volatility as a consequence of negative return innovations) can be accommodated in this model by allowing volatility innovations to be dependent on contemporaneous asset returns. One way to do this is by making the transition probabilities m_{ij} dependent on some function of return innovations. To clarify, let us suppose a negative correlation between returns and volatility innovations. If at time t there is a positive return innovation, then the probability of the state vector to move toward a higher volatility level should be lower than in the case when the price exhibits a negative return.[3] This means that, for example, for $i>j$

$$P\left(\mathbf{Z}_t=\mathbf{e}_i\left|\mathbf{Z}_{t-1}=\mathbf{e}_j,Y_t>0\right.\right)<P\left(\mathbf{Z}_t=\mathbf{e}_i\left|\mathbf{Z}_{t-1}=\mathbf{e}_j,Y_t\leq 0\right.\right)$$

proving that (7) is satisfied when $k=1$. Now suppose (7) to be true when $k=n$. Setting $k=n+1$

$$\begin{aligned} E\left[\ln\sigma\left(\mathbf{Z}_{t+n+1}\right)|\mathbf{Z}_t\right] &= E\left[E\left[\ln\sigma\left(\mathbf{Z}_{t+n+1}\right)|\mathbf{Z}_{t+1}\right]|\mathbf{Z}_t\right] \\ &= E\left[\ln\sigma\left(\mathbf{Z}_{t+1}\right)+\left[1-(1-\phi)^n\right]\left[\alpha-\ln\sigma\left(\mathbf{Z}_{t+1}\right)\right]|\mathbf{Z}_t\right] \\ &= \ldots=\ln\sigma\left(\mathbf{Z}_t\right)+\left[1-(1-\phi)^{n+1}\right]\left[\alpha-\ln\sigma\left(\mathbf{Z}_t\right)\right] \end{aligned}$$

which concludes the proof.

[3]This is not restrictive, since we can easily extend the definition to a finer grid of values of Y_t.

the opposite occurring when $i < j$. This can be achieved by selecting two different transition matrices according to the sign assumed by Y_t. To this end, let us define a new sequence of random variables $\mathbf{U} = (\mathbf{U}_t : t = 1, 2, ..., \tau)$ such that

$$\mathbf{U}_t = \begin{cases} \mathbf{f}_1 = [1\ 0]' & \text{if} \quad Y_t \leq 0 \\ \mathbf{f}_2 = [0\ 1]' & \text{otherwise} \end{cases}$$

Clearly $\mathbf{U}_t$ is I_t^Y-measurable. In addition, we introduce the $N \times N \times 2$ tensor $\widetilde{\mathbf{M}}$, the generic element of which describes the transition probabilities of the chain

$$\widetilde{m}_{ijk} = P\left(\mathbf{Z}_{t+1} = \mathbf{e}_i \,|\mathbf{Z}_t = \mathbf{e}_j, \mathbf{U}_{t+1} = \mathbf{f}_k\ \right) \quad 1 \geq i, j \geq N,\ k = 1, 2 \tag{8}$$

The entries of $\widetilde{\mathbf{M}}$ satisfy $0 \leq \widetilde{m}_{ijk} \leq 1$ and $\sum_i \widetilde{m}_{ijk} = 1$. In this setup, the knowledge of the sign of Y_t gives extra information about $\mathbf{Z}_t$. In other words, we have two different transition matrices, which will be selected depending on the state assumed by $\mathbf{U}_t$. As above, we can note that

$$E\left[\mathbf{Z}_{t+1} \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right] = E\left[\mathbf{Z}_{t+1} \,|\mathbf{Z}_t, \mathbf{U}_{t+1}\right] = \left(\widetilde{\mathbf{M}}\mathbf{U}_{t+1}\right)\mathbf{Z}_t.$$

Hence, defining $\widetilde{\mathbf{V}}_{t+1} \equiv \mathbf{Z}_{t+1} - \left(\widetilde{\mathbf{M}}\mathbf{U}_{t+1}\right)\mathbf{Z}_t$, the state equation takes the form

$$\mathbf{Z}_{t+1} = \left(\widetilde{\mathbf{M}}\mathbf{U}_{t+1}\right)\mathbf{Z}_t + \widetilde{\mathbf{V}}_{t+1},$$

where $\left(\widetilde{\mathbf{V}}_t : t = 1, 2, .., \tau\right)$ is a sequence of martingale increments satisfying

$$E\left[\widetilde{\mathbf{V}}_{t+1} \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right] = \mathbf{0}.$$

Therefore, the model with leverage effects can be described in the following state space representation

$$\begin{aligned} Y_{t+1} &= \mu_t + \sigma\left(\mathbf{Z}_t\right) W_{t+1} \\ \mathbf{Z}_{t+1} &= \left(\widetilde{\mathbf{M}}\mathbf{U}_{t+1}\right)\mathbf{Z}_t + \widetilde{\mathbf{V}}_{t+1} \end{aligned} \tag{9}$$

Putting $\sigma\left(\mathbf{Z}_t\right)$ as in (5) and assuming the following parametric specification of $\widetilde{\mathbf{M}}$

$$\widetilde{m}_{i,j,k} = \begin{cases} m_{ij} & k = 1 \text{ for each } i,\ j \\ 0 & k = 2 \text{ and } |i - j| > 1 \\ \frac{1}{2}\phi\rho\left[1 + g\left(\mathbf{e}_j\right)\right] & k = 2 \text{ and } i = j - 1 \\ \frac{1}{2}\frac{\phi}{\rho}\left[1 - g\left(\mathbf{e}_j\right)\right] & k = 2 \text{ and } i = j + 1 \\ 1 - \sum_i m_{i,j,2} & \text{otherwise} \end{cases} \tag{10}$$

where m_{ij} is defined in (6), we have obtained a generalization of the restricted model. Here ρ is introduced to capture different impacts of positive and negative returns on volatility innovations. Values of ρ greater than one allow for negative correlation between returns and volatility innovations. Note that when $\rho > 1$ and $i > j$

$$\begin{aligned} P\left(\mathbf{Z}_t = \mathbf{e}_i \left| \mathbf{Z}_{t-1} = \mathbf{e}_j, Y_t > 0\right.\right) &= \frac{1}{2}\frac{\phi}{\rho}\left[1 - g\left(\mathbf{e}_j\right)\right] < \\ \frac{1}{2}\phi\left[1 - g\left(\mathbf{e}_j\right)\right] &= P\left(\mathbf{Z}_t = \mathbf{e}_i \left| \mathbf{Z}_{t-1} = \mathbf{e}_j, Y_t \leq 0\right.\right) \end{aligned}$$

as outlined above. This structure is denoted as Hidden Markov Dependent (HMD) model and it nests the restricted model (HM) when $\rho = 1$.

5 Maximum likelihood estimation

We are now in a position to describe a simple recursive algorithm which allows us to derive the MLE numerically without resorting to the EM algorithm. As shown below, the estimation of the parameter vector

$$\boldsymbol{\theta} = \{\mu, \alpha, \delta, \phi, \rho\},$$

given the information $\mathbf{Y}$, involves filtered estimates for the states. This problem is dealt with in Appendix A in the case of the restricted model, whereas in the case of correlation between returns and volatility innovations a similar result is derived in the Appendix B.

Let us start by writing the conditional distribution of Y_{t+1} given past observations. This exercise provides the sample likelihood function, which is defined as

$$L_N\left(\boldsymbol{\theta}; \mathbf{Y}\right) = \prod_{t=1}^{\tau} f\left(Y_t \left| I_{t-1}^Y; \boldsymbol{\theta}\right.\right) \tag{11}$$

where $f\left(Y_t \left| I_{t-1}^Y; \boldsymbol{\theta}\right.\right)$ is the conditional density of Y_t given the information up to time $t-1$. The subscript N refers to the number of states of the Markov chain.

Let $\xi \in R$. Standard probability considerations yield

$$P\left(Y_{t+1} \leq \xi \left| I_t^Y\right.\right) = \sum_{i=1}^{N} P\left(Y_{t+1} \leq \xi \left| \mathbf{Z}_t = \mathbf{e}_i, I_t^Y\right.\right) P\left(\mathbf{Z}_t = \mathbf{e}_i \left| I_t^Y\right.\right)$$

Now, let $\widehat{\mathbf{Z}}_t \equiv E\left[\mathbf{Z}_t \left| I_t^Y \right.\right]$ denote the filtered estimate of $\mathbf{Z}_t$ at time t, based on processing past and present observations. Since

$$P\left(\mathbf{Z}_t = \mathbf{e}_i \left| I_t^Y \right.\right) = E\left[\mathbb{1}_{\left\{\mathbf{e}_i' \mathbf{Z}_t\right\}} \left| I_t^Y \right.\right] = E\left[\mathbf{e}_i' \mathbf{Z}_t \left| I_t^Y \right.\right] = \mathbf{e}_i' \widehat{\mathbf{Z}}_t,$$

we have

$$P\left(Y_{t+1} \leq \xi \left| I_t^Y \right.\right) = \sum_{i=1}^{N} \mathbf{e}_i' \widehat{\mathbf{Z}}_t P\left(Y_{t+1} \leq \xi \left| \mathbf{Z}_t = \mathbf{e}_i \right.\right),$$

where $Y_{t+1} \left| \mathbf{Z}_t = \mathbf{e}_i, I_t^Y \right. \sim N\left(\mu_t, \sigma_i^2\right)$. Denoting with $\phi_i(\cdot)$ the $N\left(\mu_t, \sigma_i^2\right)$ probability density function, $f\left(Y_{t+1} \left| I_t^Y \right.\right)$ can be written as

$$f\left(Y_{t+1} \left| I_t^Y; \boldsymbol{\theta} \right.\right) = \sum_{i=1}^{N} \mathbf{e}_i' \widehat{\mathbf{Z}}_t \phi_i\left(Y_{t+1}\right) \tag{12}$$

Hence the sample log-likelihood function is

$$l_N(\boldsymbol{\theta}; \mathbf{Y}) \equiv \log L_N(\boldsymbol{\theta}; \mathbf{Y}) = \sum_{t=1}^{\tau} \log\left[\sum_{i=1}^{N} \mathbf{e}_i' \widehat{\mathbf{Z}}_{t-1} \phi_i\left(Y_t\right)\right]. \tag{13}$$

The MLE is the value of $\boldsymbol{\theta}$ which maximize (13). Clearly the sequence $\left(\widehat{\mathbf{Z}}_t : t = 1, 2, ..., \tau\right)$ depends on $\boldsymbol{\theta}$, hence the maximization of the log-likelihood can be accomplished by an iterative pass through the data (filtering) following the scheme

$$\boldsymbol{\theta} \rightarrow \left(\widehat{\mathbf{Z}}_t(\boldsymbol{\theta}) : t = 1, 2, ..., \tau\right) \rightarrow l_N(\boldsymbol{\theta}; \mathbf{Y})$$

that is adopting the algorithm

1. choose $\boldsymbol{\theta}_0$;
2. compute $\widehat{\mathbf{Z}}_t(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0}\left[\mathbf{Z}_t \left| I_t^Z \right.\right], \quad t = 1, 2, ..., \tau$;
3. compute $l_N(\boldsymbol{\theta}_0; \mathbf{Y})$ as in (13);
4. find $\boldsymbol{\theta}_1$ such that $l_N(\boldsymbol{\theta}_1; \mathbf{Y}) > l_N(\boldsymbol{\theta}_0; \mathbf{Y})$
5. set $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$ and go back to step 2 until a stopping criterion is satisfied.
6. on the basis of the MLE $\widehat{\boldsymbol{\theta}}$ a final pass through the filter allows to make inference about $\mathbf{Z}$.

Table 1: Sample statistics for daily returns.

Number of observations	1128	
Mean $(\times 10^{-4})$	9.46	
Standard Deviation	0.0097	
Skewness	-0.609	
Excess Kurtosis	7.054	
$Q(12)$ Returns	18.606	(0.046)
$Q(12)$ Squared Returns	175.30	(0.000)

Note: p-values within parenthesis.

6 SP500 stock index volatility

We have tested the model on real data choosing daily returns (based on closing–time values) on the SP500 composite index, for which a large number of corresponding European option prices are available. The period of interest covers 1303 observations ranging from January 3, 1995 to December 31, 1999. The first 1128 observations (until April 30, 1999) were exploited for model estimation, while the remaining were used for the out-of-sample analysis.

Some descriptive statistics are reported in Table 1: we notice, as usual, negative skewness and the presence of fat tails, some evidence of autocorrelation (Q(12) is the Ljung Box statistics with 12 lags with the probability value in parenthesis) in the returns and volatility clustering (represented by correlation in the squared returns).

It is worth noting that two days, 8/31/98 and 10/27/99, present abnormal returns since the SP500 fell by 65, respectively, 70 points. Regressing returns against a constant and a single dummy variable for these two days causes a decrease of 4.05 points (from 7.05 to 2.90) in the excess kurtosis. Interestingly, when the time-varying features of volatility were modelled, the dummy variable was no longer significant and hence it was dropped. By the same token, we did not find relevant day–of–the–week effects. For the sake of space we do not report graphical evidence of the behavior of the index, as it is widely documented elsewhere in the literature.

Table 2: Model HMU – ML estimates.

N	*GP*	*d*	*Log-Lik*	*Volatility levels*[a] (%)					μ^b	*AIC*
				st. 1	st. 2	st. 3	st. 4	st. 5		
2	45	5	3756.16	10.29	22.21				10.52	-6.6510
3	135	10	3799.82	8.38	16.24	37.88			10.39	-6.7195
5	1215	26	3812.43	6.48	9.29	14.45	18.33	38.32	10.41	-6.7135

Note. d: number of parameters, a: annualized by the factor $\sqrt{252}$. b: $\times 10^{-4}$

6.1 Model estimation

We begin by estimating the unrestricted model with a constant mean (*HMU*) using the EM approach.[4] Table 2 reports the estimation results with 2, 3 and 5 states (the first is a standard Switching-Markov model used as a benchmark, the latter two are models directly comparable with the restricted case). Greater values for N (e.g. 7 and 9, which will be used later) need substantial additional computational efforts to detect the global maxima, making the approach unfeasible in practice.

Table 2 contains the number of gridpoints (GP) from which we started the maximization procedure and the number (d) of estimated parameters. The values of the log-likelihood and of the estimated volatility levels by state are reported, as well as the estimated constant mean $\widehat{\mu}$ and the value of the Akaike's Information Criterion (AIC) (Akaike, 1974). As expected,

[4]To initialize the EM algorithm we have chosen the following grid-points (GP) of the parameter space

- m_{ii} ranging from 0.85 to 0.99 in steps of 0.007 (3 GP for each $i = 1, 2, ..., N$) and $m_{ij} = \frac{m_{ii}}{N-1}$ when $i \neq j$;
- σ ranging from $(0.004, ..., 0.016)$ equally spaced, to $(0.008, ..., 0.02)$ in steps of 0.001 (5 GP)
- μ constant

The overall number of GP for a given N is then $3^N \times 5$. The adopted maximum likelihood stopping criterium is given by $l_N(\theta; Y)^{(k+1)} - l_N(\theta; Y)^{(k)} \leq 10^{-5}$ and will be held the same throughout all the analysis.

the increase in the number of states brings about an increase in the spread between the lowest and the highest volatility level. Judging on a comparison among AIC values there is some evidence that the preferred model would be the one with three states.

Let us now turn to the estimation of the restricted model. Recall that the model of reference in its general form can be written as

$$Y_{t+1} = \mu_t + \sigma\left(\mathbf{Z}_t\right) W_{t+1}$$

$$\mathbf{Z}_{t+1} = \mathbf{M}\mathbf{Z}_t + \mathbf{V}_{t+1}$$

where W_t and $\mathbf{V}_t$ are mutually independent, and $\sigma\left(\mathbf{Z}_t\right) = \boldsymbol{\sigma}^{'}\mathbf{Z}_t$ (recall that the elements of the vector $\boldsymbol{\sigma}$ is specified in (5) and the entries of the transition matrix $\mathbf{M}$ are given in (6)). The three estimated models can be characterized by the common restrictions described earlier in the text and by the presence (or absence) of an autoregressive component in the mean $\mu_t = \mu + \beta Y_{t-5}$ ($\mu_t = \mu$) (*HM-AR*$_5$ versus *HM*) or for the presence (or absence) of dependency between returns and volatility (*HMD-AR*$_5$ versus *HM-AR*$_5$). The Maximum Likelihood (ML) estimation results are reported, in decreasing order of generality of the models, in Table 3. We give the parameter estimates (with standard errors[5] in parentheses) for each of the number of states N ($N = 3, 5, 7, 9$).

Several points deserve to be stressed. A comparison between *HMU* and *HM*, in correspondence to the same number of states (3 and 5), suggests a slight preference for the latter when the AIC is used: Likelihood Ratio tests confirm these results. As shown in Table 3, the log-likelihood, viewed as a function of the number of states, is almost flat for values of N greater than 3; this implies that an increase in the size of the Markov chain does not improve the ability of the model to capture the time-varying behavior of volatility. This feature is common to all estimated models.[6]

[5]Standard errors were computed numerically by inverting the observed information matrix.

[6]Note that *HM-AR*$_5$ and *HM* show a non-monotonic behavior of the likelihood as a function of the number of states. Since these model are restricted, note that an increase in the number of states does not necessarily entail a better fit.

The estimated values of μ and β do not depend on N and are throughout statistically significant at a 5% significance level. It is also worth noting that the estimated volatility parameters (ϕ, α, and δ) of the *HM-AR*$_5$ and *HM* models are almost the same (by number of states), suggesting that a misspecification of the mean equation does not seriously affect volatility estimation. The issue is investigated in and confirmed by a Monte Carlo analysis presented in Appendix C. For the restricted model with asymmetric effects and a time-varying mean (*HMD-AR*$_5$), it would appear that the coefficient of asymmetry $\widehat{\rho}$ is not statistically different from 1 for N equal to 3 and 5. Since the standard errors may be affected by the specific method we adopted in estimating them, we will return on the issue later on, using a Likelihood Ratio test (cf. Table 3).

To compare different models in terms of goodness-of-fit, we use the Standardized Residuals (SR). In our setting, SR_t can be defined as

$$SR_t = \frac{Y_t - \widehat{Y}_t}{\widehat{\sigma}_t}$$

where $\widehat{Y}_t$ is the model fitted value, at time t, and $\widehat{\sigma}_t$ is the estimated volatility level prevailing at time t,

$$\widehat{\sigma}_t \equiv \widehat{\boldsymbol{\sigma}}' \widehat{\mathbf{Z}}_{t-1}. \tag{14}$$

If the models were exactly specified, we would expect standardized residuals to be serially uncorrelated and normally distributed with zero mean and unit variance. Model diagnostics (available upon request) show an overall satisfactory performance, exceptions being the non-normality of residuals. Isolated problems arise with the *HM* model which is still plagued by residual autocorrelation.

As a comparison, we have also estimated the following models: *GARCH(1,1)-AR*$_5$, *TGARCH(1,1)-AR*$_5$ and *EGARCH(1,1)-AR*$_5$. The performance of hidden Markov models in terms of goodness-of-fit seems to be slightly worse than the *GARCH*-type approach, judging from the residuals properties. In particular, the *EGARCH(1,1)-AR*$_5$ shows the lowest excess of kurtosis and skewness of standardized residuals. This should not be seen as a general result, since a similar analysis carried out using daily returns for the FTSE100, for the same time period, showed the *HMD*(5) as the "best" model[7].

Table 4 summarizes the Likelihood Ratio specification tests (with p–values in parentheses) for the three possible pairs of model comparisons

[7]These results are also available on request from the authors.

Table 3: HMD and HM ML estimates.

N	*Log-Lik*[a]	ϕ	α	δ	$\mu\ (\times 10^{-4})$	β	ρ	AIC
HMD-AR$_5$								
3	3802.67	.0158	-4.576	0.720	12.60	-0.074	2.56[b]	-6.73168
		(.0068)	(.036)	(.044)	(2.10)	(.029)	(1.06)	
5	3809.16	.0349	-4.548	0.953	12.52	-0.080	2.05[b]	-6.74319
		(.0079)	(.081)	(.090)	(2.10)	(.031)	(.61)	
7	3811.27	.0503	-4.417	1.136	12.59	-0.077	2.27	-6.74693
		(.0108)	(.088)	(.101)	(2.14)	(.030)	(.56)	
9	3812.34	.0553	-4.060	1.484	12.42	-0.077	3.12	-6.74883
		(.0050)	(.133)	(.142)	(2.05)	(.022)	(.82)	
HM-AR$_5$								
3	3800.25	0.134	-4.585	0.713	12.61	-0.076		-6.72917
		(.0048)	(.035)	(.043)	(2.09)	(.019)		
5	3807.72	.0336	-4.736	0.848	12.55	-0.074		-6.74241
		(.0057)	(.054)	(.055)	(2.10)	(.030)		
7	3808.08	.0479	-4.813	1.106	12.60	-0.078		-6.74305
		(.0070)	(.063)	(.108)	(2.03)	(.030)		
9	3807.68	.0755	-4.743	1.211	12.81	-0.077		-6.74234
		(.0083)	(.103)	(.156)	(1.84)	(.037)		
HM								
3	3797.02	.0135	-4.585	0.711	11.70			-6.72521
		(.0081)	(.037)	(.044)	(2.61)			
5	3804.71	.0340	-4.737	0.839	11.64			-6.73885
		(.0045)	(.060)	(.064)	(2.11)			
7	3804.82	.0491	-4.800	1.104	11.70			-6.73903
		(.0059)	(.043)	(.118)	(1.92)			
9	3804.47	.0772	-4.801	1.208	11.81			-6.73844
		(.0073)	(.078)	(.130)	(1.88)			

Note. Standard errors in parentheses. a: the maximization does include the starting value for Z. b: not statistically different from 1 at 5% significance level using the t-statistics.

Table 4: LR specification tests.

N	*HMD-AR*$_5$ vs *HM-AR*$_5$		*HMD-AR*$_5$ vs *HM*		*HM-AR*$_5$ vs *HM*	
	$H_0 : \rho = 1$		$H_0 : \beta = 0;\ \rho = 1$		$H_0 : \beta = 0$	
3	4.84	(0.039)	11.3	(0.004)	6.46	(0.011)
5	2.88	(0.090)	8.90	(0.012)	6.02	(0.014)
7	6.38	(0.011)	12.90	(0.002)	7.52	(0.006)
9	9.32	(0.002)	15.74	(0.000)	6.42	(0.011)

(*HMD-AR*$_5$ versus *HM-AR*$_5$; *HMD-AR*$_5$ versus *HM*; *HM-AR*$_5$ versus *HM*.) The first column corresponds to testing the null hypothesis that there are no asymmetric effects: in this respect there is stronger evidence than shown before against symmetry (less so for $N = 5$). When the null of symmetry and absence of autocorrelation is tested (second column), the results strongly reject it for all states. At any rate, the restrictions imposed by the *HM* model are rejected.

What we have learned so far is that temporal dependence in the mean and correlation between return and volatility are supported by the empirical evidence across all estimations by number of states; but it seems not straightforward to select the most appropriate number of states. Different number of states in each class of model make the models different from one another and, as such, they cannot be seen as nested. We should resort, therefore, to other ways of comparing performances: a reasonable suggestion seems to be to focus on forecasting performance, which is done in the next subsection.

6.2 Measures of forecasting performance

This subsection compares the out-of-sample performance of some of the volatility models discussed earlier. The forecast exercise was made extending the time window used for the in-sample-analysis from April 30 to December 31, 1999 (175 additional daily returns). The one-step-ahead predictive ability of each model is measured by the Mean Square Prediction Error (MSPE). We adopt as a benchmark the performance of some GARCH–type models (namely, *GARCH*, *TGARCH* and *EGARCH*).

Denoting by P the number of out-of-sample observations available, with $\mu_{t,i} = E\left[Y_{t+1} \left| I_t^Y \right.\right]$ and $\sigma^2_{t+1,i} = Var\left[Y_{t+1} \left| I_t^Y \right.\right]$ the conditional mean and

variance for i-th model, by $\widehat{\mu}_{t,i}$ and $\widehat{\sigma}^2_{t+1,i}$ the one-step-ahead forecasts of the conditional mean and variance, then the MSPE for i-th model is given by

$$\widehat{\psi}_i = \frac{1}{P}\sum_{t=\tau}^{\tau+P-1}\left[\left(Y_{t+1}-\widehat{\mu}_{t,i}\right)^2 - \widehat{\sigma}^2_{t+1,i}\right]^2 \qquad i=1,2,...,m \tag{15}$$

The rationale behind (15) can easily be explained by recognizing that

$$\sigma^2_{t+1,i} = Var\left[Y_{t+1}\left|I_t^Y\right.\right] = E\left[\left(Y_{t+1}-\mu_{t,i}\right)^2\left|I_t^Y\right.\right].$$

To calculate $\widehat{\mu}_{t,i}$ and $\widehat{\sigma}^2_{t+1,i}$ for each t and i, we estimated the model parameters just once using the initial sample. For hidden Markov models, forecasts of variances were obtained first applying the filtering methodology described above, which gives $\widehat{\mathbf{Z}}_\tau, \widehat{\mathbf{Z}}_{\tau+1}, ..., \widehat{\mathbf{Z}}_{\tau+P-1,}$ then multiplying these estimates by the scaling vector $\widehat{\boldsymbol{\sigma}}^2$

$$\widehat{\sigma}^2_{t+1} = \left\langle \widehat{\boldsymbol{\sigma}}^2, \widehat{\mathbf{Z}}_t \right\rangle \qquad t=\tau,\tau+1,...,\tau+P-1.$$

For the *GARCH* model, the following recursive equation was exploited

$$\widehat{\sigma}^2_{t+1} = \widehat{\omega} + \widehat{a}\left(Y_t - \widehat{Y}_t\right)^2 + \widehat{b}\widehat{\sigma}^2_t \qquad t=\tau,\tau+1,...,\tau+P-1.$$

Similar formulae apply for the *TGARCH* and *EGARCH* models. The MSPE results show the following order from the smallest to largest: *TGARCH–*AR_5, *EGARCH–*AR_5, *GARCH–*AR_5, *HMD(7)–*AR_5, *HMD(5)–*AR_5, *HMD(3)–*AR_5 with MSPE's ranging from 2.65 to 2.79 $\left(\times 10^{-8}\right)$.

Although these results may suggest a slightly preference for the TGARCH, it is helpful to compare the predictive performance of each model using a statistical testing procedure. We made use of the approach developed by West and Cho (1995). The procedure is designed to test the null $\mathbf{B}\boldsymbol{\psi}=\mathbf{0}$ of equal predictive ability across (two or more) competitive models, where $\widehat{\boldsymbol{\psi}} = \left(\widehat{\psi}_1, \widehat{\psi}_2, .., \widehat{\psi}_m\right)'$ is the vector of (estimated) MSPE's and $\mathbf{B}$ is a matrix $(m-1)\times m$ obtained by putting a column equal to $(-1,-1,...,-1)'$ next to an $(m-1\times m-1)$ identity matrix. The test statistic is given by

$$P\left[\widehat{\boldsymbol{\psi}'}\mathbf{B}'\left(\mathbf{B}\widehat{\mathbf{S}}\mathbf{B}'\right)^{-1}\mathbf{B}\widehat{\boldsymbol{\psi}}\right]$$

and it is asymptotically distributed as $\chi^2_{(m-1)}$. Following West and Cho, $\widehat{\mathbf{S}}$ is computed as

$$\widehat{\mathbf{S}} = \widehat{\Gamma}_0 + \sum_{j=1}^{k}\left(1-\frac{j}{k+1}\right)\left(\widehat{\Gamma}_j + \widehat{\Gamma}'_j\right), \qquad \widehat{\Gamma}_j = \frac{1}{P}\sum_{t=\tau+j}^{\tau+P}\widehat{\lambda}_t\widehat{\lambda}'_{t+j}$$

Table 5: Testing forecasting ability among different model specification.

Model	*Test statistics*	*p-value*	*k*
Overall	3.765	0.191	-1
Benchmark TGARCH(1,1)–AR$_5$			
*HMD(*3*)-AR*$_5$	1.766	0.184	-1
*HMD(*5*)-AR*$_5$	1.889	0.169	0
*HMD(*7*)-AR*$_5$	1.804	0.179	-1
*GARCH(*1,1*)-AR*$_5$	1.318	0.251	-1
*EGARCH(*1,1*)-AR*$_5$	0.868	0.352	-1
Benchmark HMD(5)–AR$_5$			
*HMD(*3*)-AR*$_5$	0.344	0.557	0
*HMD(*7*)-AR*$_5$	1.032	0.301	0
*GARCH(*1,1*)-AR*$_5$	0.801	0.371	0
*EGARCH(*1,1*)-AR*$_5$	0.837	0.360	-1

where $\widehat{\lambda}_t$ the $(m \times 1)$ vector whose i^{th} component is $\left[\left(Y_t - \widehat{\mu}_{t-1,i}\right)^2 - \widehat{\sigma}^2_{t,i}\right] - \widehat{\psi}_i$. When k is greater than zero, this approach allows forecast errors to be serially correlated[8]. Adopting this procedure, we can test for:

1. $\psi_1 = \psi_2 = ... = \psi_6$, that is a null hypothesis of equal predictive ability for all models;

2. bivariate comparison, that is a null hypothesis of equal predictive ability between the best MSPE model (in our case *TGARCH(1,1)*–AR$_5$) and each of the other models;

3. equal predictive ability between the *HMD(5)*–*AR*$_5$ and each of the other models.

In the first panel of Table 5 we have reported the value of the test statistics, their p-values, and the value of k in each case. The null of equal predictive ability among all models considered in the analysis cannot be rejected. The outcome of equal predictive ability emerges also when a pairwise comparison is adopted, either taking the *TGARCH* (panel 2) or the *HMD*(5) (panel 3) as the benchmark model.

[8]See West and Cho (1995) for details on how the value of k is derived.

Table 6: Comparison between computational methods in terms of CPU time

N	d	*CPU Time (minutes)*
HMU – EM Algorithm		
2	5	7.53
3	10	44.87
5	26	1729.80
HMD–AR_5 – SA Algorithm		
2	6	0.67
3	6	1.45
5	6	3.55
7	6	7.05
9	6	17.33

Note: Fortran codes were run on a Pentium III, 266 Mhz.

6.3 Some remarks on computational time for model estimation

As a final remark of interest we report, in Table 6, the computational time required to estimate the unrestricted (*HMU*), and the restricted models *HMD-*AR_5 with asymmetric effect for different states. *HMU* was estimated using the EM algorithm, whilst the *HMD–*AR_5, were estimated using our proposed algorithm on the basis of the simulated annealing maximization routine.[9]

Our proposal outperforms the EM estimation scheme. Although the comparison is not completely fair, since in most cases the *HMU* has a greater number of parameters to be estimated, it is worth noting that the large number of filtered estimates (roughly one for each parameter), which

[9]All the estimation procedures were implemented using Fortran codes built *ad hoc* for each model specification. While any numerical maximization method could be used, in our applications we tested the performance of two different numerical schemes: an updated version of the Nelder and Mead (1965) simplex algorithm (SIMP) and the continuous simulated annealing global optimization algorithm (SA) described in Corana et al. (1987). SIMP is an optimization method which does not need any computation of the score function. Unfortunately, it is not able to escape from local maxima. Conversely, SA is one of most promising techniques when the objective function is characterized by a multiple local maxima. SA outperforms both the EM and SIMP method, drastically reducing the computational time.

are required to implement the EM approach, slows down this estimation procedure. Conversely, in our proposal, the model parameterization allows us to avoid the number of parameters being dependent on the number of states; furthermore, only filtered estimates of the state of the Markov chain are required, making the overall result faster to achieve.

7 Conclusions

In this paper we have proposed a procedure to estimate volatility parameters for a model which can be applied in the pricing and hedging of path–dependent derivatives as suggested by Britten–Jones and Neuberger (2000). Volatility dynamics is ruled by a hidden–Markov structure with a finite number of states: from an empirical point of view, the performance of the proposed volatility model is comparable to the GARCH-type family of models both in and out-of-sample (as shown by the West and Cho tests of predictive ability). Relative to similar proposals of estimation of volatility which can take a finite number of values, the proposed approach simplifies the estimation procedure in a ML context, avoiding problems related to the EM algorithm: furthermore, the computational burden can be further reduced by resorting to the Simulated Annealing algorithm. We have also shown how the leverage effect can be accommodated and how the recursive filter of the states proposed by Elliott et al. (1995) needed for model estimation are to be modified in this case.

Some issues are still open. In particular, we have not found conclusive evidence on how to define the choice of the number of states of the Markov chain in such models in a optimal way. The approach proposed by Otranto and Gallo (2001) in a nonparametric Bayesian framework could provide some insights in the matter.

References

Akaike (1974): "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, AC-19, pp.716-723.

Black F. and Scholes M. (1973): "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy,* 81, pp. 637-654.

Bollerslev T., Engle R., and Nelson D. (1994): "ARCH models", Ch. 49 in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics*, Vol. IV, Amsterdam: North Holland.

Britten-Jones M. and Neuberger A. (2000): "Option Prices, Implied Price Processes, and Stochastic Volatility", *Journal of Finance*, 55, No. 2, pp 839-866.

Corona A., Marchesini M., Martini C., and Ridella S. (1987): "Minimizing Multimodal Functions of Continuous Variables with the Simulated Annealing Algorithm", *ACM Trans. Math. Software*, March 1987, pp. 272-280.

Dempster A. P., Laird N. M., and Rubin D. B. (1977): "Maximum Likelihood Estimation from Incomplete Data via the EM algorithm", *Journal of Royal Statistical Society, Series B* 39, pp. 1-38.

Duan J., Popova I., and Ritchken P. (2002): "Option Pricing under Regime Switching", *Quantitative Finance*, 2, 1-17.

Elliott R. J., Aggoun L., and Moore J. B. (1995): "Hidden Markov Models: Estimation and Control", Applications of Mathematics 29, Springer-Verlag, New York.

Elliott R. J., Hanter W. C., and Jamieson B. M. (1998): "Drift and Volatility Estimation in Discrete Time", *Journal of Economic Dynamics and Control* 22, pp. 209-218.

Ghysels E., Harvey A., and Renault E. (1996): "Stochastic Volatility", *Handbook of Statistics* 14, pp. 119-191.

Hamilton J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica* 57, pp. 357-384.

Hamilton J. D. (1990): "Analysis of Time Series Subject to Changes in Regimes", *Journal of Econometrics* 45, pp. 39-70.

Hamilton J. D. (1996): "Specification Testing in Markov-Switching Time-Series Models", *Journal of Econometrics* 70, pp. 127-157.

Harvey A. (1990): "The Econometrics Analysis of Time Series". 2nd edition, Philip Allan, Hemel Hempstead and MIT Press, Boston, MA.

Nelder J. A. and Mead R. (1965): "A Simplex Method for Function Minimization", *Computer Journal*, 7, pp. 308-313.

Otranto, E. and G.M. Gallo (2001) "A Nonparametric Bayesian Approach to Detect the Number of Regimes in Markov Switching Models", Working Paper 2001/04 Dipartimento di Statistica "G.Parenti", Università di Firenze, (www.ds.unifi.it/ricerca/pubblicazioni/working_papers/2001/wp2001_04.pdf), forthcoming *Econometric Reviews*.

Rossi A. (2002): "The Britten-Jones and Neuberger Smile-consistent with Stochastic Volatility Option Pricing Model: a Further Analysis", *Journal of Theoretical and Applied Finance,* 5, pp.1-31.

West K. D. and Cho D. (1995): "The Predictive Ability of Several Models of Exchange Rate Volatility", *Journal of Econometrics*, 69, pp. 367-391.

A Hidden Markov Models: the Elliott, Aggoun and Moore (1995) framework

This appendix contains some technical aspects involving the estimation of *HMM*. Given the discrete time set-up of section 2, leading to model (1), Elliott et al. (1995) obtain optimal filtered estimates of the states and model parameters estimation using the EM algorithm: the main difference of the setup here is that we derive the expectation step when the mean of the process is a constant. Let us first illustrate how recursive filters for the states can be obtained. Filtered estimates of the number of jumps, occupation time, and processes related to the observations, are derived in a similar manner.

A.1 Filtered estimates for the states

Although the sequence $\mathbf{Z}$ is not directly observed, filtered estimates, i.e. the conditional expectations of $\mathbf{Z}_t$, given I_t^Y, provide inference about $\mathbf{Z}_t$, given past and present observations. A change of measure is introduced to avoid non-linear recursive filtering. The basic idea, to simplify the filtering procedure, is to introduce a new probability measure, P^*, defined on $(\Omega, \Im)$, under which $\mathbf{Y}$ is a sequence of $IIN(0,1)$ random variables, independent of $\mathbf{Z}$, and $\mathbf{Z}_{t+1} = \mathbf{M}\mathbf{Z}_t + \mathbf{V}_{t+1}$, where $\mathbf{V}_{t+1}$ satisfies $E^*\left[\mathbf{V}_{t+1} \left| I_t^{\mathbf{Z}} \right.\right] = \mathbf{0}$. Results obtained under P^* are later reinterpreted under the original probability measure P. The key result is the following theorem.

Theorem 1 *(Conditional Bayes) Let $(\Omega, \Im, P)$ be a probability space and ζ a sub-sigma-field of $\Im$. If P^* is a probability measure on $\Im$, absolutely continuous with respect to P, with Radon-Nikodym derivative $\frac{dP^*}{dP} = \Lambda$, then for any integrable random variable X*

$$E^*[X|\zeta] = \frac{E[\Lambda X|\zeta]}{E[\Lambda|\zeta]} \quad \text{if} \quad E[\Lambda|\zeta] > 0$$

and zero otherwise. E^* and E, denote the (conditional) expectation under P^* and P, respectively.

Consider the model defined in (1), let the parameters $\boldsymbol{\theta}$ be known, and denote the $N(0,1)$ probability density function by $\phi(\cdot)$. Let us write

$$\lambda_k = \frac{\sigma(\mathbf{Z}_{k-1})\,\phi(Y_k)}{\phi(W_k)}$$

and

$$\Lambda_0 = 1, \quad \Lambda_t = \prod_{k=1}^{t} \lambda_k \tag{16}$$

Define P^* by putting $\frac{dP^*}{dP}\,|_{I_t} = \Lambda_t$. Then we have

Theorem 2 *Under P^*, $\mathbf{Y}$ is a sequence of $IIN\,(0,1)$ random variables.*

Proof. See Elliott et al. (1995), page 60.

Since we will work under P^*, we need to construct a probability measure P (the original measure) such that, under P,

- $W_t \equiv \frac{Y_t - \mu_{t-1}}{\sigma(\mathbf{Z}_{t-1})}$ is an $IIN\,(0,1)$, sequence of random variables and
- $\mathbf{Z}_{t+1} = \mathbf{M}\mathbf{Z}_t + \mathbf{V}_{t+1}$, where $\mathbf{V}_{t+1}$ satisfies $E\left[\mathbf{V}_{t+1}\left|I_t^{\mathbf{Z}}\right.\right] = \mathbf{0}$.

The construction of P from P^* follows the same arguments shown above. To this end, assuming $\sigma\left(\mathbf{Z}_{k-1}\right) \neq 0$, let

$$\lambda_k^* = \frac{\phi\left[\left(Y_k - \mu_{k-1}\right)/\sigma\left(\mathbf{Z}_{k-1}\right)\right]}{\sigma\left(\mathbf{Z}_{k-1}\right)\phi\left(Y_k\right)}$$

and

$$\Lambda_0^* = 1, \quad \Lambda_t^* = \prod_{k=1}^{t} \lambda_k^*. \tag{17}$$

Then, defining $\frac{dP}{dP^*}\,|I_t = \Lambda_t^*$, we have the following result

Theorem 3 *Under P, $\mathbf{W}$ is a sequence of $IIN\,(0,1)$ random variables.*

Working under P^*, and using the conditional Bayes Theorem 1 above, one gets a recursive filter for states of the Markov chain from

$$\widehat{\mathbf{Z}}_t \equiv E\left[\mathbf{Z}_t\left|I_t^Y\right.\right] = \frac{E^*\left[\Lambda_t^*\mathbf{Z}_t\left|I_t^Y\right.\right]}{E^*\left[\Lambda_t^*\left|I_t^Y\right.\right]} \quad t = 1, 2, ..\tau. \tag{18}$$

The recursive unnormalized estimator for the state, under P^*

$$\widehat{\mathbf{Z}}_t^* \equiv E^*\left[\Lambda_t^*\mathbf{Z}_t\left|I_t^Y\right.\right]$$

is given by

$$\begin{aligned} E^*\left[\Lambda_t^*\mathbf{Z}_t\left|I_t^Y\right.\right] &= E^*\left[\Lambda_t^*\left(\mathbf{M}\mathbf{Z}_{t-1} + \mathbf{V}_t\right)\left|I_t^Y\right.\right] \\ &= E^*\left[\Lambda_t^* E^*\left[\mathbf{M}\mathbf{Z}_{t-1} + \mathbf{V}_t\left|I_{t-1}^{\mathbf{Z}}, I_t^Y\right.\right]\left|I_t^Y\right.\right] \end{aligned}$$

The last equality is due to following result

$$E^*\left[\mathbf{V}_t \left| I_{t-1}^{\mathbf{Z}}, I_t^Y \right.\right] = E^*\left[\mathbf{V}_t \left| I_{t-1}^{\mathbf{Z}} \right.\right] = 0.$$

Therefore,

$$\begin{aligned}\widehat{\mathbf{Z}}_t^* &= E^*\left[\Lambda_t^* \mathbf{M}\mathbf{Z}_{t-1} \left| I_t^Y \right.\right] \\ &= E^*\left[\Lambda_{t-1}^* \frac{\phi\left[\left(Y_t - \mu_{t-1}\right)/\sigma\left(\mathbf{Z}_{t-1}\right)\right]}{\sigma\left(\mathbf{Z}_{t-1}\right)\phi\left(Y_t\right)} \mathbf{M}\mathbf{Z}_{t-1} \left| I_t^Y \right.\right].\end{aligned}$$

Setting

$$\mathbf{p}_i\left(Y_t\right) \equiv \frac{\phi\left(\frac{Y_t - \mu_{t-1}}{\sigma_i}\right)}{\sigma_i \phi\left(Y_t\right)} \mathbf{e}_i \tag{19}$$

and summing over the number of the states $1 \leq i \leq N$, we get

$$\widehat{\mathbf{Z}}_t^* = \sum_{i=1}^{N} \left[\mathbf{p}_i\left(Y_t\right)' \widehat{\mathbf{Z}}_{t-1}^*\right] \mathbf{M}\mathbf{e}_i. \tag{20}$$

Furthermore, letting $\mathbf{1} \equiv (1, 1, .., 1)' \in \mathbf{R}^N$, we have $\mathbf{1}'\mathbf{Z}_t = 1$. Hence

$$\mathbf{1}' E^*\left[\Lambda_t^* \mathbf{Z}_t \left| I_t^Y \right.\right] = E^*\left[\Lambda_t^* \mathbf{1}'\mathbf{Z}_t \left| I_t^Y \right.\right] = E^*\left[\Lambda_t^* \left| I_t^Y \right.\right]$$

Then, applying (18), a recursive estimator for the state $\widehat{\mathbf{Z}}_t$, under P, is obtained by dividing the unnormalized estimator $\widehat{\mathbf{Z}}_t^*$ by the sum of its components

$$\widehat{\mathbf{Z}}_t = \frac{\widehat{\mathbf{Z}}_t^*}{\mathbf{1}'\widehat{\mathbf{Z}}_t^*} \tag{21}$$

A.2 Filtered estimates for the number of jumps, occupation time and observation processes

The following theorem gives filtered estimates of the number of jumps of the chain, occupation time and observation processes. These are later used to estimate the model parameters using the EM approach.

Theorem 4 *Let H_t be a scalar $I_t -$ measurable process of the form: H_0 is $I_0^Z -$ measurable and $H_{t+1} = H_t + \alpha_{t+1} + \boldsymbol{\beta}_{t+1}' \mathbf{V}_{t+1} + \delta_{t+1} f\left(Y_{t+1}\right)$,*

where f is a scalar valued real function, and $\alpha_t, \boldsymbol{\beta}_t, \delta_t$ are I_{t-1} − measurable processes. Then

$$\begin{aligned}\gamma_{t+1}\left(H_{t+1}\mathbf{Z}_{t+1}\right) &:= E^*\left[\Lambda^*_{t+1}H_{t+1}\mathbf{Z}_{t+1}\left|I^Y_{t+1}\right.\right] \\ &= \sum_{i=1}^{N}\left[\mathbf{p}_i\left(Y_{t+1}\right)'\gamma_t\left(H_t\mathbf{Z}_t\right)\right]\mathbf{Me}_i \\ &+\gamma_t\left(\alpha_{t+1}\mathbf{p}_i\left(Y_{t+1}\right)'\mathbf{Z}_t\right)\mathbf{Me}_i \\ &+\gamma_t\left(\delta_{t+1}\mathbf{p}_i\left(Y_{t+1}\right)'\mathbf{Z}_t\right)f\left(Y_{t+1}\right)\mathbf{Me}_i \\ &+\gamma_t\left(\boldsymbol{\beta}_{t+1}\mathbf{p}_i\left(Y_{t+1}\right)'\mathbf{Z}_t\right)\left[diag\left(\mathbf{Me}_i\right)-\mathbf{Me}_i\left(\mathbf{Me}_i\right)'\right]\end{aligned}$$

Proof. See Elliott et al. (1995) on pages 64-65.

Note that, a recursive filter for $\gamma_{t+1}\left(H_{t+1}\mathbf{Z}_{t+1}\right)$ is introduced because, unlike $\gamma_{t+1}\left(H_{t+1}\right)$, closed-form recursive estimates are obtained. However, once $\gamma_{t+1}\left(H_{t+1}\mathbf{Z}_{t+1}\right)$ is known, one computes the unnormalized estimator for $\gamma_{t+1}\left(H_{t+1}\right)$ by summing the components of $\gamma_{t+1}\left(H_{t+1}\mathbf{Z}_{t+1}\right)$.

Now, suppose J^{sr}_t to be the number of jumps of the chain from $\mathbf{e}_r$ to $\mathbf{e}_s$ at time t, i.e.

$$J^{sr}_t = \sum_{k=1}^{t}\mathbf{e}_r'\mathbf{Z}_{k-1}\mathbf{e}_s'\mathbf{Z}_k.$$

Let us focus on $\widehat{J}^{sr}_\tau \equiv E\left[J^{sr}_\tau\left|I_\tau{}^Y\right.\right]$. Working under P^* and applying the Conditional Bayes theorem

$$\widehat{J}^{sr}_\tau = \frac{E^*\left[\Lambda^*_\tau J^{sr}_\tau\left|I_\tau{}^Y\right.\right]}{E^*\left[\Lambda^*_\tau\left|I_\tau{}^Y\right.\right]}$$

where Λ^*_τ is defined in (17). The numerator, can now be easily computed by applying the previous theorem and setting $H_{t+1} = \widehat{J}^{sr}_{t+1}, H_0 = 0, \alpha_{t+1} = \left(\mathbf{e}_r'\mathbf{Z}_t\right)m_{sr}, \boldsymbol{\beta}_{t+1} = \left(\mathbf{e}_r'\mathbf{Z}_t\right)\mathbf{e}_s'$ and $\delta_{t+1} = 0$. Then,

$$\gamma_{t+1}\left(J^{sr}_{t+1}\mathbf{Z}_{t+1}\right) = \sum_{i=1}^{N}\left[\mathbf{p}_i\left(Y_{t+1}\right)'\gamma_t\left(J^{sr}_t\mathbf{Z}_t\right)\right]\mathbf{Me}_i + \left[\mathbf{p}_r\left(Y_{t+1}\right)'\widehat{\mathbf{Z}}^*_t\right]m_{sr}\mathbf{e}_s$$

The unnormalized recursive estimator of number of jumps, $\gamma_{t+1}\left(J_{t+1}^{sr}\right)$, follows, using the fact[10] $\gamma_{t+1}\left(J_{t+1}^{sr}\right) = \mathbf{1}'\gamma_{t+1}\left(J_{t+1}^{sr}\mathbf{Z}_{t+1}\right)$. Thus, in order to obtain the normalized recursive estimator under the original measure P, it is necessary to normalize this last expression by dividing it by $E^*\left[\Lambda_\tau^* \left| I_\tau{}^Y\right.\right] = \mathbf{1}'\widehat{\mathbf{Z}}_{t+1}^*$. We get

$$\widehat{J}_{t+1}^{sr} = \frac{\gamma_{t+1}\left(J_{t+1}^{sr}\right)}{\mathbf{1}'\widehat{\mathbf{Z}}_{t+1}^*}. \tag{22}$$

The same arguments lead to filtered estimates of

- the occupation time in state $\mathbf{e}_r$, $\widehat{O}_{t+1}^r \equiv E\left[O_{t+1}^r \left| I_{t+1}^Y\right.\right]$, where O_t^r is given by
$$O_t^r = \sum_{k=1}^{t} \mathbf{e}_r'\mathbf{Z}_{k-1}$$
- the observation process $\widehat{G}_{t+1}^r(f) \equiv E\left[G_{t+1}^r(f) \left| I_{t+1}^Y\right.\right]$, where $G_t^r(f)$ is defined by
$$G_t^r(f) = \sum_{k=1}^{t} \left[\mathbf{e}_r'\mathbf{Z}_{k-1}\right] f(Y_k)$$
and $f(Y)$ is some function of the observations.

A.3 Expectation Maximization approach

We are now in a condition to apply the EM approach to this context. As well known, the EM algorithm provides an approximate solution to compute the MLE when the maximization of the sample likelihood function is not feasible or difficult. Its main steps are as follows.

1. Choose an initial value, $\boldsymbol{\theta}_0$, for $\boldsymbol{\theta}$ and and a threshold $\varepsilon > 0$.
2. (E-step) At each step k (k=0, 1,. . .), set $\overline{\boldsymbol{\theta}} = \boldsymbol{\theta}_k$. Compute $E_{\overline{\boldsymbol{\theta}}}\left[\ln\left(\frac{dP_{\boldsymbol{\theta}}}{dP_{\overline{\boldsymbol{\theta}}}}\right) \left| I_\tau{}^Y\right.\right]$

[10] Since $\mathbf{1}'\mathbf{Z}_t = 1$, for each scalar sequence $\mathbf{H}$, I_t-adapted, we have

$$\begin{aligned}\mathbf{1}'\gamma_t(H_t\mathbf{Z}_t) &= \gamma_t\left(\mathbf{1}'(H_t\mathbf{Z}_t)\right)\\ &= \gamma_t\left(H_t\left(\mathbf{1}'\mathbf{Z}_t\right)\right) = \gamma_t(H_t)\end{aligned}$$

3. (M-step) Find $\boldsymbol{\theta}_{k+1} = \arg\max_{\boldsymbol{\theta}\in\Theta} E_{\overline{\boldsymbol{\theta}}}\left[\ln\left(\frac{dP_{\boldsymbol{\theta}}}{dP_{\overline{\boldsymbol{\theta}}}}\right)\left|I_\tau^{\,Y}\right.\right]$

4. Replace k with $k+1$, and come back to step 2 until $L(\boldsymbol{\theta}_{k+1}) - L(\boldsymbol{\theta}_k) < \varepsilon$.

The algorithm main feature is to generate a non-decreasing sequence of likelihood values $L(\boldsymbol{\theta}_{k+1}) \geq L(\boldsymbol{\theta}_k)$, ensuring the convergence towards a local maximum. We begin by updating the value of μ, keeping the rest of the parameters fixed. We consider only the case where μ_t is constant $\mu_t = \mathbf{c}'\mathbf{Z}_t = \mu$, since the case $\mathbf{c}_1 \neq \mathbf{c}_2 \neq ... \neq \mathbf{c}_N$, is simpler and is dealt with in Elliott et al. (1995).

Let us write

$$\overline{\Lambda}_t = \prod_{k=1}^{t} \exp\left\{\frac{(Y_k - \overline{\mu})^2 - (Y_k - \mu)^2}{2\left(\sigma'\mathbf{Z}_{k-1}\right)^2}\right\}.$$

Let us define a new measure P_μ so the restriction of its Radon-Nikodym derivative $dP_\mu/dP_{\overline{\mu}}$ conditional on I_t is given by $dP_\mu/dP_{\overline{\mu}}\,|I_t = \overline{\Lambda}_t$. Now

$$\begin{aligned}
\log\overline{\Lambda}_t &= \sum_{k=1}^{t}\left(\frac{2Y_k\mu - \mu^2}{2\left(\sigma'\mathbf{Z}_{k-1}\right)^2}\right) + f(\overline{\mu}) \\
&= \sum_{k=1}^{t}\sum_{i=1}^{N}\frac{2Y_k\mu - \mu^2}{2\sigma_i^2}\mathbf{e}_i'\mathbf{Z}_{k-1} + f(\overline{\mu}) \\
&= \sum_{i=1}^{N}\frac{2\ \mu G_k^i(Y) - \mu^2 O_k^i}{2\sigma_i^2}\ + f(\overline{\mu})
\end{aligned}$$

where $f(\overline{\mu})$ does not depend on μ. The E-step follows

$$E\left[\log\overline{\Lambda}_\tau\left|I_\tau^{\,Y}\right.\right] = \sum_{i=1}^{N}\frac{2\ \mu\widehat{G}_\tau^i(Y) - \mu^2\widehat{O}_\tau^i}{2\sigma_i^2}\ + \widehat{f(\overline{\mu})} \tag{23}$$

Upon differentiating $E\left[\log\overline{\Lambda}_\tau\left|I_\tau^{\,Y}\right.\right]$ with respect to μ, and equating the derivative to zero, the M-step gives the updated value

$$\widehat{\mu} = \frac{\sum_{i=1}^{N}\frac{\widehat{G}_\tau^i(Y)}{\sigma_i^2}}{\sum_{i=1}^{N}\frac{\widehat{O}_\tau^i}{\sigma_i^2}}. \tag{24}$$

Using similar arguments, one can derive new estimates of the parameters m_{ij} and σ_i, given the available information ${I_\tau}^Y$

$$\widehat{m}_{ij} = \frac{\widehat{J}_\tau^{ij}}{\widehat{O}_\tau^j} \quad i = 1, 2, .., N, \quad j = 1, 2, .., N \ - 1 \tag{25}$$

and

$$\widehat{\sigma}_i = \left(\frac{\widehat{G}_\tau^i \left(Y^2\right) \ - 2\mu \widehat{G}_\tau^i \left(Y\right) + \mu^2 \widehat{O}_\tau^i}{\widehat{O}_\tau^i} \right)^{\frac{1}{2}} . \quad i = 1, 2, .., N \tag{26}$$

The quantities $\widehat{\mathbf{Z}}_t, \widehat{J}_\tau^{ij}, \widehat{O}_\tau^i$ and $\widehat{G}_\tau^i$ are then reevaluated on the basis of the updated values $\widehat{\mu}, \widehat{\sigma}_i$ and $\widehat{m}_{ij}$. This iterative scheme proceeds until convergence is achieved.

B Recursive filter for the states with leverage effect

Let us assume $\boldsymbol{\theta}$ as known. Again, we use reference probability method to find the optimal recursive filter of the states. For this, let us consider a probability measure P^*, as defined in appendix A (equation (16)), under which $\mathbf{Y}$ is an $IIN\left(0,1\right)$ sequence of random variables. First, we see that

$$P^*\left(\mathbf{Z}_{t+1} = \mathbf{e}_i \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right) = E^*\left[\mathbf{e}_i' \mathbf{Z}_{t+1} \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right]$$

Using theorem A.1

$$\begin{aligned} E^*\left[\mathbf{e}_i' \mathbf{Z}_{t+1} \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right] &= \frac{E\left[\mathbf{e}_i' \mathbf{Z}_{t+1} \Lambda_{t+1} \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right]}{E\left[\Lambda_{t+1} \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right]} \\ &= E\left[\mathbf{e}_i' \mathbf{Z}_{t+1} \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right] \\ &= P\left(\mathbf{Z}_{t+1} = \mathbf{e}_i \left| I_t^{\mathbf{Z}}, I_{t+1}^Y \right.\right) \end{aligned}$$

Hence, under P^*, the process $\mathbf{Z}$ is consistent with (9). Second, write

$$\widehat{\mathbf{Z}}_{t+1} \equiv E\left[\mathbf{Z}_{t+1} \left| I_{t+1}^Y \right.\right]$$

Using similar arguments as those in appendix A, we prove the following

Lemma 5 *Under the assumption of model (9), a recursive filter for the state is given by*

$$\widehat{\mathbf{Z}}_{t+1} = \frac{\widehat{\mathbf{Z}}^*_{t+1}}{1'\widehat{\mathbf{Z}}^*_{t+1}}$$

being

$$\widehat{\mathbf{Z}}^*_{t+1} = \sum_{k=1}^{2}\sum_{i=1}^{N}\left[\mathbf{p}_i\left(Y_{t+1}\right)'\widehat{\mathbf{Z}}^*_t\right]\left[\mathbf{f}_k'\mathbf{U}_{t+1}\right]\left(\widetilde{\mathbf{M}}\mathbf{f}_k'\right)\mathbf{e}_i$$

where $\mathbf{p}_i\left(Y_{t+1}\right)$ *is defined in (19).*

Proof: from the Bayes theorem

$$E\left[\mathbf{Z}_{t+1}\left|I^Y_{t+1}\right.\right] = \frac{E^*\left[\mathbf{Z}_{t+1}\Lambda^*_{t+1}\left|I^Y_{t+1}\right.\right]}{E^*\left[\Lambda^*_{t+1}\left|I^Y_{t+1}\right.\right]}$$

The numerator of the above equation is equal to

$$\begin{aligned}
E^*\left[\mathbf{Z}_{t+1}\Lambda^*_{t+1}\left|I^Y_{t+1}\right.\right] &= E^*\left[\Lambda^*_t\frac{\phi\left[\left(Y_{t+1}-\mu_t\right)/\sigma\left(\mathbf{Z}_t\right)\right]}{\sigma\left(\mathbf{Z}_t\right)\phi\left(Y_{t+1}\right)}E^*\left[\mathbf{Z}_{t+1}\left|I^{\mathbf{Z}}_t, I^Y_{t+1}\right.\right]\left|I^Y_{t+1}\right.\right] \\
&= E^*\left[\Lambda^*_t\frac{\phi\left[\left(Y_{t+1}-\mu_t\right)/\sigma\left(\mathbf{Z}_t\right)\right]}{\sigma\left(\mathbf{Z}_t\right)\phi\left(Y_{t+1}\right)}\left(\widetilde{\mathbf{M}}\mathbf{U}_{t+1}\right)\mathbf{Z}_t\ \left|I^Y_{t+1}\right.\right] \\
&= \sum_{k=1}^{2}\sum_{i=1}^{N}\left[\mathbf{p}_i\left(Y_{t+1}\right)'\widetilde{\mathbf{Z}}^*_t\right]\left[\mathbf{f}_k'\mathbf{U}_{t+1}\right]\left(\widetilde{\mathbf{M}}\mathbf{f}_k'\right)\mathbf{e}_i
\end{aligned}$$

The result follows noting that $E^*\left[\Lambda^*_{t+1}\left|I^Y_{t+1}\right.\right] = \mathbf{1}'\widetilde{\mathbf{Z}}^*_{t+1}$.

C Monte Carlo analysis

We invoked large-sample theory both to compute parameter standard errors and to derive the asymptotic distribution of likelihood ratio test. However, there are no guidelines to verify the adequacy of the asymptotic approximation in our framework. In other words, it is not clear for which order of magnitude of the sample size we should trust standard errors and critical values of tests based on limit considerations. For this reason, we investigated the finite-sample properties of the MLE when the HM model is assumed to be the true DGP. We set $N = 5$, $\phi = 0.01, \alpha = -5$, $\delta = 1$ and $\mu = 7 \times 10^{-4}$. Setting $\alpha = -5$ and $\delta = 1$ implies (annualized) volatility levels ranging from 4% to 29%, which are typical values in applications. The value of ϕ was chosen near its lower boundary for two reasons: first, analysis of real (daily) time series returns displays a similar value; second, we expect the asymptotic properties to the MLE be violated, in "small" sample, especially when there are corner conditions. One thousand of Monte Carlo replications from samples of size $\tau = 500, \tau = 1000$ and $\tau = 2000$ were generated. The presence of multiple local maxima of the sample likelihood function may cause same problems in detecting the global one, especially for small values of τ, when SIMP is used[11]. Conversely, we are better off using the SA algorithm. Table 7 exhibits the results of this exercise.

[11] We stress that this problem is not dramatic as in the case of the EM algorithm applied to the unrestricted model, starting the maximization procedure with the true parameter vector. When τ is chosen to be greater than 300, the procedure is quite robust with respect to the choice of the starting values.

Table 7: MLE's finite-sample distributional properties based on 500, 1000 and 2000 replications from the HM model

DGP	$\boldsymbol{\phi = 0.01}$	$\boldsymbol{\alpha = -5}$	$\boldsymbol{\delta = 1}$	$\boldsymbol{\mu = 7 \times 10^{-4}}$
$\tau = 500$				
Mean	0.0112	-4.9988	0.9893	9.66×10^{-4}
Median	0.0108	-5.0009	0.9960	8.81×10^{-4}
Std.Err.	0.0065	0.0610	0.1102	5.09×10^{-4}
Skewness	4.4704	0.2186	-1.8484	1.615
Kurtosis	52.9731	13.8710	21.0299	9.273
$\tau = 1000$				
Mean	0.0109	-5.0012	0.9978	8.60×10^{-4}
Median	0.0107	-5.0020	0.9991	8.15×10^{-4}
Std.Err.	0.0035	0.0296	0.0433	2.85×10^{-4}
Skewness	0.3025	0.0058	-0.3242	1.089
Kurtosis	3.1099	4.0332	4.0980	5.776
$\tau = 2000$				
Mean	0.0105	-5.0033	1.0004	7.97×10^{-4}
Median	0.0104	-5.0032	1.0009	7.90×10^{-4}
Std.Err.	0.0025	0.0177	0.0272	1.50×10^{-4}
Skewness	0.1285	-0.1137	-0.0357	0.494
Kurtosis	3.0254	3.2821	2.9165	4.144

It is clear from the table above that estimation of the mean-reverting parameter ϕ must rely on a sufficiently large number of observations if volatility is highly persistent. With $\tau = 500$ the distribution of the MLE cannot be thought as drown from a normal variate. As a matter of fact, kurtosis and skewness are definitely different from respective normal typical values. Better when $\tau = 1000$. The MLE asymptotic normal distribution hypothesis is almost satisfied in the last panel of Table 7 in correspondence of $\tau = 2000$, albeit a relatively large bias is still present in the estimation of the (constant) drift. However, for purposes of valuating derivative assets, the volatility parameters (ϕ, α, δ) are those that really matter.

In the previous exercise, estimation of volatility parameters relied on a correct specification the drift term μ_t. In principle, a misspecification of the drift, could introduce a bias in the estimator of the parameters of interest (ϕ, α, δ). To investigate this possibility, we performed another Monte Carlo simulation based on one thousand samples of daily returns generated from

Table 8: MLE finite-sample distributional properties of the HM model when a misspecified constant drift is estimated ($\tau = 1000$)

DGP	$\boldsymbol{\phi = 0.01}$	$\boldsymbol{\alpha = -5}$	$\boldsymbol{\delta = 1}$	$\boldsymbol{\mu_t = 7 \times 10^{-4} + 20 \times \sigma^2(\mathbf{Z_t})}$
Mean	0.0106	-4.9897	1.0257	1.34×10^{-3}
Median	0.0105	-4.9889	1.0253	1.20×10^{-3}
Std. Err.	0.0038	0.0298	0.0468	4.97×10^{-4}
Skewness	0.2846	-0.1420	-0.1702	2.192
Kurtosis	3.1723	3.9246	4.2929	10.704

the restricted model (1) where

$$\begin{aligned} \mu_t &= \mu + \lambda\sigma^2(\mathbf{Z}_t) \\ &= \mu + \lambda \exp\{2\alpha + 2\delta g(\mathbf{Z}_t)\} \end{aligned}$$

$\tau = 1000$, $N = 5$, $\phi = 0.01, \alpha = -5.0, \delta = 1.0$ and $\mu = 7 \times 10^{-4}$. The risk premium λ was set such that μ_t was, on average 34% per year. Using these replications we estimated, for each sample, a constant drift. Table 8 sums up the results.

Monte Carlo analysis confirms that there are no biases in volatility parameter estimates, even when a misspecified model of the drift term is assumed, although, some higher moments are slightly worse (in the sense that estimators display a greater departure from normality) if compared with the second panel ($\tau = 1000$) of Table 7. The only marked difference can be found comparing the last column of the two tables. A higher standard deviation, skewness and kurtosis are reported in Table 8, exhibiting a stronger departure from normality. These results were in same way expected since, on small intervals, the magnitude of the drift term is of the same order of the conditional variance, hence, effects of the conditional mean on squared returns should be negligible.

Copyright © 2002
Alessandro Rossi,
Giampiero M. Gallo