Volatility Estimation via Hidden Markov Models *

Alessandro ROSSI[†] Giampiero M. GALLO [‡]

This revision: July 2005

Abstract

We propose a stochastic volatility model where the conditional variance of asset returns switches across a potentially large number of discrete levels, and the dynamics of the switches are driven by a latent Markov chain. A simple parameterization overcomes the commonly encountered problem in Markov-switching models that the number of parameters becomes unmanageable when the number of states in the Markov chain increases. This framework presents some interesting features in modelling the persistence of volatility, and that, far from being constraining in data fitting, it performs comparably well as other popular approaches in forecasting short–term volatility.
Keywords: Stochastic volatility, Markov chain, GARCH, SWARCH, Forecasting.

JEL: C22, C53, G13

*We thank Christophe Planas, Stewart Hodges, Richard Baillie and an anonymous referee for valuable comments on the paper. Suggestions by Frank Diebold, Peter Christoffersen, René Garcia and other participants in the conference *New Directions in Risk Management* held in Frankfurt on Nov 4–5, 2003 are gratefully acknowledged. 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 modelling financial asset returns it is now customary to assume that volatility is time–varying and can take values over a continuous positive range: GARCH (see e.g. Bollerslev *et al.* 1992, 1994) and stochastic volatility models (see e.g. Ghysels *et al.*, 1996) fall in this category. Maintaining that volatility is time–varying, other contributions in the literature suggest models where volatility is assumed to take distinct values over a finite number of states or regimes, driven by a hidden Markov process (see e.g. Elliot *et al.*, 1998). There are several conceptual advantages in adopting such a framework: the presence of regimes is consistent with the stylized facts of persistence in the behavior of the time–varying variance of returns at the basis of other volatility models. GARCH–type models are linear in the squares of innovations while some sort of nonlinearity may need to be accommodated: this occurs when volatility dynamics are allowed to be state–dependent and differences in the persistence of innovations is made dependent on the size of the innovation (Friedman and Laibson, 1989). Think of "exceptional" events which send the markets into a short–lived turmoil which is absorbed relatively quickly: GARCH models are reckoned to have too high persistence to be consistent with the observed behavior following these occurrences, while regimes allow for varying degrees of persistence across regimes. GARCH effects and regime-switching volatility can be combined to give rise to the models of Hamilton and Susmel (1994), Cai (1994) and Gray (1996). Option pricing has found volatility varying over a finite number of regimes an appealing feature. To price and hedge interest rate derivatives, Naik and Lee (1994) extend the Vasicek model allowing the variance of the short interest rate to switch between low and high regimes. Duan *et al.* (1999) develop an option pricing model where the underlying stock price dynamic is driven by a regime switching process. Britten-Jones and Neuberger (2000) show how the classic Black and Scholes (1973) option pricing model can be extended to make it compatible with stochastic volatility and observed option prices with the aim to evaluate and hedge path-dependent derivatives. Specifically, given the prices of European options, and allowing for a wide range of stochastic volatility dynamics, they derive a class of price processes for the underlying asset which is consistent with the observed smile surface (see also Rossi, 2002). Evans (2003) extends affine models of the term structure including regime switching in the mean and variance of the nominal and real short rates.

It is well known that the main disadvantage with this approach is that the number of parameters increases with the number of states of the Markov chain complicating model estimation. For these reasons, empirical applications of Markov switching volatility models to financial series usually consider a small number of states (between 2 and 4); an important exception is the Markov-Switching multifractal volatility model (Calvet and Fisher 2001, 2004) which considers switching across a very large number of states, yet retaining a parsimonious specification. In what follows we propose a time-varying volatility model where conditional variance switches across a (potentially large) number of discrete levels, and the dynamics of the switches are driven by a latent Markov chain. Far from being constraining in data fitting, this framework performs well in forecasting when compared to other approaches: we can handle a reasonable number of states on daily data (most Markov switching applications are on lower frequency data) while catching most of the stylized facts exhibited by asset returns. To make this feasible, we suggest an appropriate parameterization which makes the number of parameters independent of the number of regimes and can be estimated by maximum likelihood. Given the unobservability of the volatility process, we adopt the recursive filtering algorithm proposed by Hamilton (1994) to draw inferences about the unobserved state variable.

In Section 2 we define the main assumptions behind our framework, showing in Section 3 how the imposition of some restrictions on the parameter space allows us to reduce the number of parameters for model estimation while handling a reasonable number of states for the Markov chain. Moreover, we show how we can easily accommodate asymmetric responses of volatility to negative innovations, and the implications of our specification for volatility persistence. We then discuss filtering and smoothing of the unobserved states, model estimation and the computation of volatility forecasts in our framework. In Section 4 we estimate volatility parameters for the S&P500 stock index, showing that a good performance in forecasting is obtained with several indicators used as targets. Section 5 concludes.

2 A Markovian Framework for Volatility

Let s_t be the price of a certain asset at time t. We consider the return on the asset, observable at time t, as a random variable $r_t = \ln(s_t/s_{t-1})$, for a given sample $t = 1, \ldots, T$. Its volatility is driven by a discrete-time, N-state Markov chain z_t. We denote with I_t the P-augmented increasing sigma-field generated by $\{z_s, r_s : s \leq t\}$, whereas restrictions to I_t generated by the specific random variables are denoted by superscripts (e.g. I_t^z is the filtration generated by $\{z_s : s \leq t\}$). We suppose that N and the distribution of z_0 are known. For convenience, the realizations of the Markov chain are assumed to be the N-dimensional unit vectors, e_i ($i = 1, 2, ..., N$), with a unit element in the i-th position and zeros elsewhere. The stochastic volatility model of interest here can be written in the state-space format:

$$\begin{aligned} r_t &= \mu_t + \sigma(z_t)^{1/2} u_t \\ z_t &= M_{t-1} z_{t-1} + v_t, \end{aligned} \tag{1}$$

where $\mu_t = E[r_t | I_{t-1}^r]$ is the conditional mean of the observable process, and $\sigma(z_t)^{1/2}$ is the value of the volatility prevailing at time t, with $\sigma(\cdot)$ a positive-valued scaling function that takes values σ_1 when $z_t = e_1$, σ_2 when $z_t = e_2$, and so on.

Since several empirical studies with financial time series data indicate that the distribution of asset returns is usually rather leptokurtic, even after controlling for volatility clustering (see e.g. Bollerslev *et al.*, 1992, for a survey), we allow asset returns innovations, u_t, to be distributed as a Student's–t r.v.'s[1] with unit variance and ν degrees of freedom. The case of Gaussian innovations can then be retrieved as a limiting case ($\nu \to \infty$). The transition equation for z_t highlights how the short–term dynamics of the first-order Markov chain is fully described by the $N \times N$ one-step transition matrix M_t, the generic element of which

$$m_t^{ij} \equiv \Pr(z_{t+1} = e_i | I_t) = \Pr(z_{t+1} = e_i | z_t = e_j, r_t) \tag{2}$$

describes the transition probability of the chain (cf. Hamilton, 1994, p. 679). The entries of M_t satisfy $m_t^{ij} \geq 0$, and $\sum_i m_t^{ij} = 1$, for each $1 \leq i, j \leq N$ and t. Using (2), we have $E[z_t | I_{t-1}] = E[z_t | z_{t-1}, r_{t-1}] = M_{t-1} z_{t-1}$. Hence, defining $v_t \equiv z_t - M_{t-1} z_{t-1}$, we have $E[v_t | I_{t-1}] = 0$, which provides a semi-martingale representation for the transition equation. Innovations u_t and v_t are assumed to be independent. This is similar to the model used by Elliot *et al.* (1998) to model monthly IBM stock prices with $\mu_t = c' z_t$, and $\sigma(z_t) = \sigma' z_t$, (with c and σ two vectors of constants), and $M_t = M$, without any restriction on transition probabilities. Also, this model can be seen as a restricted version of the SWARCH model (Hamilton and Susmel, 1994) which is described later on. Although similar in spirit, our model is quite different from those of Calvet and Fisher (2001, 2004), which can be considered the first attempt to make accessible stochastic volatility models based on high-dimensional regime-switching. A comparison between their approach and

[1]In a GARCH setting, Student's–t innovations for asset returns have been introduced by Bollerslev (1987), Baillie and Bollerslev (1989), and Harvey *et al.* (1992), just to quote a few.

ours is not attempted here, though.

3 The Model

The volatility model described above presents some unattractive features. First, the number of parameters to be estimated increases quadratically with the number of states of the Markov chain, and hypotheses on the number of states itself run into the problem of nuisance parameters being unidentified under the null (Andrews and Ploberger, 1994). There are no guidelines to establish this size in practical applications; however, to make an example, a value of N equal to 7, (which might 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, over–parameterized models lead to non–efficient estimators even in large samples (cf. Harvey, 1990), whereas from a computational point of view, it might not be straightforward to find the global maximum of the likelihood function. In practice, this would require several hundred initial values to start the numerical maximization procedure even when the size of the Markov chain is small Hamilton and Susmel (1994) report that in their SWARCH specification the global maximum of the likelihood function might still go undetected when the number of states of the Markov chain is above three. With these characteristics, it is difficult to expect that the extreme parameter uncertainty could translate into good forecasting performance.

We propose some possible alternatives here, exploring how the model could be parsimoniously parameterized in a way which removes the dependence upon N. We introduce a dependence structure between the Markov process and the observations by selecting two different transition matrices depending on the sign of lagged-one return. This allows us to take into account the "leverage" effect (Black, 1976), that is, a negative correlation between returns and volatility innovations, a feature often encountered in financial data. By the same token, following Gray (1996), we allow the entries of the transition matrix of the Markov chain to be dependent on past returns. This is a major departure from the model by Elliott *et al.* (1998) in that it implies a time-varying persistence of the variance of asset returns. In particular, we will be able to assess whether large returns are associated with a lower persistence of volatility than small returns. This feature should also improve the short-term forecasting ability of our model.

3.1 A Simple Model Parameterization

We start by specifying the conditional mean in a simple autoregressive form:

$$\mu_t = \mu + \sum_{k=1}^{p} \gamma_k r_{t-k}. \tag{3}$$

The major departure in our model (cf. also Britten-Jones and Neuberger, 2000) from the assumption $\sigma(z_t) = \sigma' z_t$ is to constrain volatility regimes to follow

$$\sigma_i = \exp\{\alpha + \delta g(e_i)\}, \quad i = 1, 2, \ldots, N \tag{4}$$

which specifies volatility regimes as a function of only two coefficients, α and δ, no matter the number of states. The function $g(\cdot)$, defined by

$$g(e_i) = \frac{2i - (N+1)}{N-1}, \tag{5}$$

has the effect of associating distinct values between -1 and 1 to each regime linearly increasing with i. When δ is positive, we identify σ_1 with the variance in the lowest volatility regime, and σ_N with the highest one.

Finally, to complete the model specification we need to specify the transition probabilities of the Markov chain which translate into the main dynamics of the model. We proceed from some desired characteristics of the model: a higher volatility should be generated by the model when past returns innovations are negative; the persistence of volatility should be dependent on the magnitude of asset returns. Finally, the number of unknown parameters of the transition matrix M_t should not depend on N.

In order to achieve these features, we allow for two different transition matrices, M_t^+, and M_t^-, according to the sign assumed by r_t:

$$M_t = \begin{cases} M_t^+ & \text{if } r_t > 0 \\ M_t^- & \text{if } r_t \leq 0 \end{cases} \tag{6}$$

Let us suppose that the changes in volatility can occur just one step at a time: this is equivalent to assuming that M_t^+ and M_t^- are tridiagonal matrices with the elements on the main diagonal representing the probability of staying in the state and the off–diagonal elements the probability of moving to a higher or a lower level, respectively. Let us now introduce a negative correlation between returns and volatility innovations. If at time t the asset price decreases (bad news), then the probability that the state vector moves toward a higher volatility level should be higher than in the case of good news[2]. Hence the requirement for $i = j + 1$

$$m_t^{ij-} = \Pr(z_{t+1} = e_i | z_t = e_j, r_t \leq 0) > Pr(z_{t+1} = e_i | z_t = e_j, r_t > 0) = m_t^{ij+}$$

while the opposite should occur when $i = j - 1$. Incorporating all constraints on transition probabilities, let us thus consider the following specification for the generic elements

$$m_t^{ij-} = \begin{cases} 1 - \phi_t & i = j \\ \frac{1}{2}\phi_t[1 + g(e_j)] & i = j - 1 \\ \frac{1}{2}\phi_t[1 - g(e_j)] & i = j + 1 \\ 0 & |i - j| > 1 \end{cases} \qquad m_t^{ij+} = \begin{cases} 1 - \sum_{i, i \neq j} m_t^{ij+} & i = j \\ \frac{1}{2}\phi_t \psi[1 + g(e_j)] & i = j - 1 \\ \frac{1}{2}\frac{\phi_t}{\psi}[1 - g(e_j)] & i = j + 1 \\ 0 & |i - j| > 1 \end{cases} \tag{7}$$

where $g(\cdot)$ is as before, and $\psi > 0$. The parameter ψ allows returns and volatility innovations to be correlated. Values of $\psi > 1$ entail a negative correlation between returns and volatility:

$$\begin{aligned} m_t^{ij-} &= \frac{1}{2}\phi_t[1 - g(e_j)] > \frac{1}{2}\frac{\phi_t}{\psi}[1 - g(e_j)] = m_t^{ij+} \quad \text{when } i = j + 1 \\ m_t^{ij-} &= \frac{1}{2}\phi_t[1 + g(e_j)] < \frac{1}{2}\phi_t \psi[1 + g(e_j)] = m_t^{ij+} \quad \text{when } i = j - 1. \end{aligned}$$

The presence of asymmetric effects may be tested by constraining $\psi = 1$, (which implies $M_t^+ = M_t^- = M_t$). A necessary condition for $0 \leq m_t^{ij} \leq 1$, for each i, j, and t, is that the time-varying parameter ϕ_t takes on values in the interval $(0, 1)$[3]. This can be achieved by making it dependent on r_t through

$$\phi_t = \Phi(a + b|r_t|), \tag{8}$$

for some coefficients a and b, where $\Phi(\cdot)$ denotes the (standard) Normal cumulative distribution function. Note that when $b > 0$, ϕ_t is a monotonically increasing function of $|r_t|$, so that returns having greater magnitude imply a lower persistence for the volatility evolution as will be argued below. Setting $b = 0$ would imply a constant volatility persistence.

We refer to the structure (1)-(8) with N states as our Hidden Markov Model (HMM). The cases with and without asymmetric effect will be denoted T(hreshold)-HMM(N), and HMM(N) respectively.

[2] We could easily extend the definition to a finer grid of r_t values.

[3] This is not sufficient to guarantee $m_t^{ij} \in [0, 1]$ since for $\psi \to \infty$ transition probabilities fall outside the unit interval. However, leaving ψ unbounded in the empirical application the natural constraints on transition probabilities are rarely violated. In those cases we set transition probabilities to their boundaries 0 and 1.

3.2 Volatility Persistence

Some of the characteristics of the process just described are very important for the dynamics of stochastic volatility. A key aspect is related to the volatility persistence implied by asset returns, a point widely debated in the financial econometric literature: Hamilton and Susmel (1994), and Cai (1994) argue in favor of regime-switching with the motivation of too large a persistence following large shocks implied by GARCH models. Lamoureux and Lastrapes (1990a) attribute the huge persistence implied by GARCH as the incapability to catch structural changes in the unconditional variance of asset returns. The same authors (1990b) show that with other variables inserted in the information set (current volume) the measure of persistence decreases.

Our non-linear framework departs from the standard GARCH constant persistence and is rather flexible in that it allows volatility to decay at different rates of speed according to the magnitude of asset returns. Taking the maximum eigenvalue (smaller than one) of the transition matrices M_t as the measure of volatility persistence, in view of (8), the persistence of the variance forecasted at time t for future horizons, depends both on the sign and the magnitude of r_t. For a particular choice of parameters ($N = 7, a = -2.48,\ b = .85,\ \psi = 2.39$ – cf. Table 2d below), Figure 1 shows a profile of persistence implied by our model as a function of asset returns. It can be noted that the rate of speed of volatility decay towards the long-run level is a non-decreasing function of absolute returns: so long as $b > 0$, bigger shocks decay faster than smaller shocks. By the same token, negative returns are associated with a higher volatility persistence than positive returns. This would put our model in line with the evidence provided by Friedman and Laibson (1989) who claim that large and small returns innovations have a different impact on volatility persistence, with the former being much less persistent than the latter.

Figure 1 about here

3.3 Filtering and Maximum Likelihood Estimation

In the T-HMM(N) model as defined by equations (1)-(8) we can denote the vector of model parameters by $\theta = (\mu,\ \gamma_1,\ \dots\ \gamma_p,\ \alpha,\ \delta,\ a,\ b,\ \psi,\ \nu)$. Given the latent structure of the Markov process, estimation of θ by maximum likelihood involves filtered estimates for the states. This can be easily seen by writing the conditional distribution of r_t given past observations over N states of the Markov chain

$$f(r_t|I^r_{t-1}) = \sum_{i=1}^{N} f(r_t|z_t = e_i, I^r_{t-1}) \Pr(z_t = e_i|I^r_{t-1}),$$

and substituting to the last term its expression

$$\Pr(z_t = e_i|I^r_{t-1}) = \sum_{j=1}^{N} \Pr(z_t = e_i|z_{t-1} = e_j, r_{t-1}) \Pr(z_{t-1} = e_j|I^r_{t-1}).$$

Therefore, we have

$$f(r_t|I^r_{t-1}) = \sum_{i=1}^{N}\sum_{j=1}^{N} f(r_t|z_t = e_i, I^r_{t-1}) \Pr(z_t = e_i|z_{t-1} = e_j, r_{t-1}) \Pr(z_{t-1} = e_j|I^r_{t-1}). \quad (9)$$

In view of the assumption about the Student's–t distribution of asset returns innovations, the first term in (9) is

$$f(r_t|z_t = e_i, I^r_{t-1}) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)}[\pi(\nu-2)\sigma(e_i)]^{-1/2}\left[1 + \frac{(r_t - \mu_t)^2}{\sigma(e_i)(\nu-2)}\right]^{-\frac{\nu+1}{2}}$$

which for $\nu \to \infty$ tends to

$$[(2\pi\sigma(e_i)]^{-1/2} \exp\left[-\frac{(r_t - \mu_t)^2}{2\sigma(e_i)}\right].$$

The second term in (9) is m^{ij}_{t-1}, and the last one is known as the filtered estimate of the state obtained by processing past and present observations. Recursive filters for the unobserved states were proposed in this context by Hamilton and Susmel (1994), and Elliot *et al.* (1995). The former, however, is computationally superior[4] and will be adopted here (cf. Hamilton, 1994, p.692–3).

Let $z_{t|t}$ denote the $E[z_t|I^r_t]$, F_t be an $N \times 1$ vector the j-th element of which is $F^j_t \equiv f(r_t|z_t = e_j, I^r_{t-1})$, and $\mathbb{1}$ be the $N \times 1$ vector of ones. Then a recursive filter for the unobservable variables has the form

$$z_{t|t} = \frac{(M_{t-1}z_{t-1|t-1}) \odot F_t}{\mathbb{1}'[(M_{t-1}z_{t-1|t-1}) \odot F_t]} \quad (10)$$

where $\odot$ denotes the Hadamard, element–by–element, product. Using (9) and observing that

$$\Pr(z_t = e_i|I^r_t) = E[e'_i z_t|I^r_t] = e'_i z_{t|t},$$

[4] We thank an anonymous referee for pointing this out to us.

the generic term in the likelihood function can be expressed as

$$f(r_t|I^r_{t-1}) = \sum_{i=1}^{N}\sum_{j=1}^{N} f(r_t|z_t = e_i, I^r_{t-1}) m^{ij}_{t-1} e'_j z_{t-1|t-1},$$

where $m^{ij}_{t-1} = m^{ij-}_{t-1}$, if $r_{t-1} \leq 0$, or $m^{ij}_{t-1} = m^{ij+}_{t-1}$, if $r_{t-1} > 0$. As in any linear state-space models where the Kalman filter is used to estimate unobserved states, the filtering algorithm described by (10) provides the log-likelihood as a byproduct. The log-likelihood function can then be maximized numerically with respect to the unknown parameter vector θ. Asymptotic standard errors of model parameters can be computed by inverting the negative of the Hessian of the log-likelihood function evaluated at the maximum likelihood estimates. Note also that the conditional variance of asset returns at time t,

$$E[(r_t - \mu_t)^2|I^r_{t-1}] = E[\sigma(z_t)|I^r_{t-1}] = \sigma' z_{t|t-1},$$

is given by the scalar product between the vector of estimated (discrete) variance values across regimes and the one–step–ahead prediction of the state made at time $t-1$.

3.4 Forecasting

The computation of variance forecasts from our model deserves careful attention. In order to compute k–step–ahead variance forecasts,

$$\begin{aligned} E[(r_{t+k} - \mu_{t+k})^2|I^r_t] &= E[\sigma(z_{t+k})|I^r_t] \\ &= \sigma' E[z_{t+k}|I^r_t] \\ &= \sum_{i=1}^{N} \sigma_i \Pr(z_{t+k} = e_i|I^r_t) \end{aligned}$$

we need to find conditional probabilities $\Pr(z_{t+k} = e_i|I^r_t)$, $k = 1, 2, \cdots$, and $i = 1, 2, \cdots, N$. For $k = 1$, and $i = 1, 2, \cdots, N$, the task is straightforward: since

$$\Pr(z_{t+1} = e_i|I^r_t) = \sum_{j=1}^{N} \Pr(z_{t+1} = e_i|z_t = e_j, I^r_t) \Pr(z_t = e_i|I^r_t)$$

one–step–ahead variance forecasts are given by $\hat{h}^2_{t+1|t} = \sum_{i=1}^{N} \sigma_i \Pr(z_{t+1} = e_i|I^r_t) = \sigma' M_t z_{t|t}$.

Moving to k–step–ahead variance forecasts, a closed-form recursive formula is more complicated to obtain. To proceed in steps, let us first consider a constant conditional mean, and let us suppose that $E[z_{t+l}|I_t^r]$ be known for $l = 1, 2, \cdots, k$. Then, for each $k \geq 1$, and $i = 1, 2, \cdots, N$

$$\Pr(z_{t+k+1} = e_i | I_t^r) = \sum_{j=1}^{N} \Pr(z_{t+k+1} = e_i | z_{t+k} = e_j, I_t^r) \Pr(z_{t+k} = e_j | I_t^r).$$

For any j, the second term in the previous expression, $\Pr(z_{t+k} = e_j|I_t^r)$, is already available from previous steps, while the first term can be written as

$$\begin{aligned}
\Pr(z_{t+k+1} = e_i | z_{t+k} = e_j, I_t^r) &= \int_{\mathbf{R}} \Pr(z_{t+k+1} = e_i | z_{t+k} = e_j, r_{t+k}) f(r_{t+k} | z_{t+k} = e_j, I_t^r) dr_{t+k} \\
&= \int_{\mathbf{R}} m_{t+k}^{ij} f(r_{t+k} | z_{t+k} = e_j)\ dr_{t+k}. \qquad (11)
\end{aligned}$$

Given that m_t^{ij} depends on r_t via $\phi_t = \Phi(a + b|r_t|)$, through (7), expression (11) is known as long as $\int_{\mathbf{R}} \phi_{t+k} f(r_{t+k}|z_{t+k} = e_j)\ dr_{t+k}$ is known. Under the assumption of constant mean for the process of returns, the latter integral does not depend either on k or t, but only on j. This means that it has to be computed numerically just N times, which is not a daunting task.

Let us now consider an expression for the conditional mean which involves time dependence, as in (3). As previously, we want to compute $E[z_{t+k+1}|I_t^r]$ assuming $E[z_{t+l}|I_t^r]$ as known for $l = 1, 2, \cdots, k$. Standard probability laws yield:

$$\begin{aligned}
\Pr(z_{t+k+1} = e_i | I_t^r) &= \sum_{j_1=1}^{N} \cdots \sum_{j_k=1}^{N} \Pr(z_{t+k+1} = e_i | z_{t+k} = e_{j_k}, \cdots, z_{t+1} = e_{j_1}, I_t^r) \\
&\times \Pr(z_{t+k} = e_{j_k}, \cdots, z_{t+1} = e_{j_1} | I_t^r).
\end{aligned}$$

The second conditional probability of the sum above is available from previous steps, whereas the first, using Bayes's rule, can be written as

$$\begin{aligned}
&\Pr(z_{t+k+1} = e_i | z_{t+k} = e_{j_k}, \cdots, z_{t+1} = e_{j_1}, I_t^r) \\
&= \int_{\mathbf{R}} \Pr(z_{t+k+1} = e_i | z_{t+k} = e_{j_k}, r_{t+k}) f(r_{t+k} | z_{t+k} = e_{j_k}, \cdots, z_{t+1} = e_{j_1}, I_t^r)\ dr_{t+k} \\
&= \int_{\mathbf{R}} m_{t+k}^{ij_k} f(r_{t+k} | z_{t+k} = e_{j_k}, \cdots, z_{t+1} = e_{j_1}, I_t^r)\ dr_{t+k} \qquad (12)
\end{aligned}$$

which involves an integral to be computed. Denoting by $\epsilon_t \equiv (\sigma' z_t)^{\frac{1}{2}} u_t$, and given that $z_{t+k}, \cdots, z_{t+1}, I_t^r$ are known, the observational equation can be written

$$\begin{aligned}
r_{t+k} &= \mu + \sum_{i=1}^{p-k+1} \gamma_{k-1+i} r_{t+1-i} + \sum_{i=1}^{k-1} \gamma_i r_{t+k-i} + \epsilon_{t+k} \\
&= \gamma_k(1)^{-1}[\mu + \sum_{i=1}^{p-k+1} \gamma_{k-1+i} r_{t+1-i}] + \gamma_k(L)^{-1} \epsilon_{t+k},
\end{aligned}$$

where $\gamma_k(L) = 1 - \gamma_1 L - \cdots - \gamma_{k-1} L^{k-1}$, and $\gamma_k(L)^{-1} = 1 + \theta_1 L + \theta_2 L^2 + \cdots$. Thus the process of returns admits the moving average representation:

$$r_{t+k} \quad = \quad c + \sum_{i=k}^{\infty} \theta_i \epsilon_{t+k-i} + \sum_{i=0}^{k-1} \theta_i \epsilon_{t+k-i}, \quad \theta_0 = 1 \tag{13}$$

where $c = \gamma_k(1)^{-1}[\mu + \sum_{i=1}^{p-k+1} \gamma_{k-1+i} r_{t+1-i}]$. Equation (13) may be used to simulate random variables $r_{t+k}^{(g)}$, $g = 1, 2, \cdots, G$, from the conditional distribution of returns. If we assume that the innovations u_t are normally distributed,

$$r_{t+k} | z_{t+k} = e_{j_k}, \cdots, z_{t+1} = e_{j_1}, I_t^r \sim N(c + \sum_{i=k}^{t+k-1} \theta_i \hat{\epsilon}_{t+k-i}, \sum_{i=0}^{k-1} \theta_i^2 \sigma_{j_{i+1}}).$$

For Student's-t innovations, instead, $r_{t+k}^{(g)}$ is drawn simulating first $\theta_i \epsilon_{t+k-i}$, $i = 0, 1, k-1$ from a $t_\nu(0, \theta_i^2 \sigma_{j_{i+1}})$, then summing up over i, and adding $c + \sum_{i=k}^{t+k-1} \theta_i \hat{\epsilon}_{t+k-i}$. In either case, the sum in (12) is approximated by

$$\frac{1}{G} \sum_{g=1}^{G} m_{t+k}^{i,j_k}(r_{t+k}^{(g)}).$$

It is worth noting that this approach is viable for small values of k since, in general, the numbers of terms to be computed in (12) grows with k at a rate of $(N-1)^k$. However, the magnitude of the autoregressive coefficients $\gamma_1, \cdots, \gamma_p$, turns out to be quite small in empirical applications. This implies that the coefficients of the moving average representation (13), $\theta_1, \theta_2, \cdots, \theta_j, \cdots$, die out after small values of j. Hence computation of $\Pr(z_{t+k+1} = e_i | I_t^r)$ remains feasible by conditioning on a small set of lagged terms for z_{t+k+1}.

An alternative strategy, which has been followed in our application, is to compute multi–step–ahead variance forecasts by means of Monte Carlo simulation. For any desired degree of accuracy, the algorithm below provides an estimate of k-step ahead variance forecasts by simulating future observations and latent variables recursively:

1. set $i = 1$;
2. set $j = 1$;
3. draw $r_{t+j}^{(i)}$ from $t_\nu(\mu_{t+j}^{(i)}, \sigma' z_{t+j}^{(i)})$;
4. draw $z_{t+j+1}^{(i)}$ from $E[z_{t+j+1} | I_{t+j}^r] = M_{t+j}^{(i)} z_{t+j}^{(i)}$;

5. iterate steps 3-4 to have $z_{t+k}^{(i)}$;

6. iterate steps 2-5 for $i = 1, 2, \cdots, G$;

7. estimate $\hat{E}[z_{t+k}|I_t^r] = \frac{1}{G}\sum_{i=1}^{G} z_{t+k}^{(i)}$;

8. compute $\hat{h}_{t+k|t}^2 = \sigma' \hat{E}[z_{t+k}|I_t^r]$.

To sample z_{t+j} from $E[z_{t+j}|I_t^r]$ let X be the $N \times 1$ vector such that $X_i = \sum_{k=1}^{i} \Pr(z_{t+j} = e_k|I_t^r)$. By construction $X_N = 1$. Draw $u \sim U[0,1]$, and let n be $\min_i : X_i \geq u$. Then set the n-th element of z_{t+j} equal to one and the other elements to zero. In the application we set $G = 10^5$. Note that for $G = 10^7$, the relative change of variance forecasts is roughly 10^{-3}.

3.5 Smoothing

Remarkably, the procedure allows us to reconstruct the probabilities that the Markov process is in a given state, say e_j, at time t after having observed $r_1, r_2, \ldots, r_T$, for $t \leq T$, i.e. $\Pr(z_t = e_j|I_T^r)$. Smoothed inference about the regimes has been proposed by Kim (1994) for a general class of dynamic linear models with Markov switching effects, which carries over to the problem at hand as well (cf. Hamilton, 1994, p.694). Denoting $z_{t|T} \equiv E[z_t|I_T^r]$, the probability of interest is given by

$$\Pr(z_t = e_j|I_T^r) = E[e_j' z_t|I_T^r] = e_j' z_{t|T}.$$

Given filtered estimates $z_{t|t}$, $t = 1, 2, \ldots, T$, and parameter vector θ, Kim's recursive smoother takes the form

$$z_{t|T} = z_{t|t} \odot \{M_t[z_{t+1|T} \div (M_t z_{t|t})]\}$$

where $\odot$ and $\div$ represent the element–by–element multiplication and division operators respectively. The recursion starts at $t = T$ and proceeds backward with $t = T-1, T-2, ..., 1$.

4 Forecasting S&P 500 Volatility

This section presents an empirical application of our stochastic volatility model to high-frequency (daily) returns of the S&P 500 composite stock index. The time series exhibits volatility clustering, but the daily frequency has somewhat hindered the application of traditional Markov switching models (which, as a matter of fact, have mainly used weekly or monthly returns). We will compare the forecasts from our approach to those produced by two other classes of models, i.e. GARCH-type (where asymmetric effects are taken into account) and SWARCH (where switching is inserted but the number of parameters increases rapidly with the number of states).

4.1 The Models Used as Comparisons

The attribution of the 2003 Nobel Prize to Rob Engle's work is also a recognition of the fact that the class of GARCH models is by far the most successful attempt to describe the dynamics of the volatility of asset returns. One may specify the model in general terms as an asymmetric Threshold GARCH (Glosten *et al.*, 1994) with Student's–t innovations:

$$\begin{aligned} r_t|I_{t-1} &= \mu_t + h_t u_t, \quad u_t|I^r_{t-1} \sim t_\nu \\ h_t^2 &= \omega + \alpha u_{t-1}^2 + \psi 1\!\!1_{(u_{t-1}\leq 0)} u_{t-1}^2 + \beta h_{t-1}^2, \end{aligned} \tag{14}$$

where $1\!\!1_{(A)}$ is the indicator function of the set A. The conditional mean μ_t is given by (3). For the process (14) to be well defined we need ω, α, and β to be positive and $\alpha + \beta + \frac{\psi}{2} < 1$ for stationarity. The parameter ψ accounts for the leverage effect, while Student's-t innovations are introduced for a better fit of the leptokurtosis in the unconditional distribution of asset returns. The Gaussian Threshold GARCH(1,1) is obtained by taking $\nu = \infty$ and the Gaussian GARCH (1,1) constrains ψ to be zero as well.

Given model parameters, the multi–step–ahead variance forecasts made at time t, say $\hat{h}^2_{t+k|t}$, are given by the recursions

$$\begin{aligned} \hat{h}^2_{t+1|t} &= \omega + \alpha \hat{u}_t^2 + \psi 1\!\!1_{(\hat{u}_t\leq 0)} \hat{u}_t^2 + \beta \hat{h}^2_{t|t} \\ \hat{h}^2_{t+k|t} &= \omega + (\alpha + \beta + \psi/2)\hat{h}^2_{t+k-1|t} \qquad k > 1. \end{aligned} \tag{15}$$

Alternatively, one can use the Markov switching framework to model conditional variance (Hamilton and Susmel, 1994, as an extension of Hamilton, 1989). The idea is to allow the parameters of an ARCH(p) to change according to an N-states Markov chain $\{z_t : t = 0, 1, 2, \ldots\}$ with a constant transition matrix M, by means of a scaling function g_{z_t} that takes values $(1, g_1, g_2, \ldots, g_N)$ according to the state assumed by z_t

$$\begin{aligned} & r_t|I_{t-1} = \mu_t + g_{z_t}^{1/2} v_t, \\ & v_t|I_{t-1} = h_t u_t, \qquad u_t|I_{t-1} \sim t_\nu, \\ & h_t^2 = \alpha_0 + \alpha_1 v_{t-1}^2 + \psi 1\!\!1_{(v_{t-1}\leq 0)} v_{t-1}^2 + \alpha_2 v_{t-2}^2 + \cdots + \alpha_p v_{t-p}^2. \end{aligned} \tag{16}$$

The errors u_t are conditionally independent and identically distributed as Student's–t with ν degrees of freedom. Student's–t innovations and ψ play the same role as in the GARCH specification. Denoting $\tilde{u}_t = \frac{u_t}{\sqrt{g_{z_t}}}$, and $\tilde{h}^2_{t+k|t} = E[\tilde{u}^2_{t+k}|\tilde{u}^2_t, ..., \tilde{u}^2_{t-p}, z_t, ..., z_{t-p}]$, then the k–step–ahead variance forecasts, say $\hat{h}^2_{t+k|t}$, can be computed by the recursive formula

$$\hat{h}^2_{t+k|t} = \sum_{i=1}^{N}\sum_{j=1}^{N}\cdots\sum_{m=1}^{N} \Pr(z_t = e_i, z_{t-1} = e_j, \cdots, z_{t-p} = e_m|I^r_t)g'M^k e_i \tilde{h}^2_{t+k|t} \tag{17}$$

where $\tilde{h}^2_{t+k|t}$ is

$$\begin{aligned} &\alpha_0 + \alpha_1(\tilde{u}^2_t + \psi I_{(\tilde{u}_t \le 0)}) + ... + \alpha_p \tilde{u}^2_{t-p} \qquad k = 1 \quad \text{and} \\ &\alpha_0 + (\alpha_1 + \tfrac{\psi}{2})\tilde{h}^2_{t+k-1|t} + ... + \alpha_p \tilde{h}^2_{t+k-p|t} \quad k > 1. \end{aligned}$$

4.2 The Data

We estimate our models on daily returns (close–to–close) of the S&P 500 composite index. We have 1303 observations available, ranging from January 3, 1995 to December 31, 1999. The first 870 observations (about two–thirds of the total) are used for model estimation, while the remaining 433 are left for the out-of-sample analysis.

Some descriptive statistics are reported in Table 1: as usual, we notice negative skewness and the presence of fat tails for the empirical unconditional distribution of returns. Some evidence of autocorrelation in the returns and volatility clustering (represented by correlation in the squared returns) are detected by the Ljung-Box statistic (Ljung and Box, 1978).

Insert **Table 1** approximately here

It is worth noting that our sample period includes the large drop in the S&P 500 index on Monday, October 27, 1997 (-7.1%), the value of which is responsible for a large portion of the excess kurtosis. Regressing returns against a constant and a single dummy variable for this day causes a decrease from 8.63 to 2.87 in the excess kurtosis. Interestingly, when the time–varying volatility was explicitly modelled, the dummy variable was no longer significant and hence it was dropped from the mean equation. We did find relevant day–of–the–week effects. For this we model the conditional mean of returns by a fifth–order autoregressive process: $\mu_t = \mu + \gamma r_{t-5}$.

4.3 Estimation Issues

Tables 2a-2d present maximum likelihood estimates, standard errors, log-likelihoods, and some goodness-of-fit statistics for GARCH, SWARCH, and our HMM proposal.

In estimating GARCH models a single maximum is found when starting the maximization procedure from several points of the parameter space. Looking at Table 2a some remarks are in order: according to the GARCH models, volatility is quite persistent. This persistence (measured by $\alpha + \beta + \psi/2$) seems to be robust to changes in model specifications. It ranges from .979 (Gaussian Threshold GARCH) to 0.995 (Gaussian GARCH). Both the inclusion of asymmetric effects and fat tails innovations are strongly supported by the data, as shown by likelihood ratio tests and standard errors on ψ, and ν. Yet, the Jarque-Bera normality test strongly rejects the normality of standardized residuals even after controlling for volatility clustering in the case of Gaussian innovations. The positive sign of ψ implies a negative correlation between asset returns and conditional variance, as expected. Judging by the Akaike Information Criterion (AIC, Akaike, 1974) the Student's–t Threshold-GARCH would be confirmed to be the best specification in terms of goodness-of-fit. No serial correlation among the standardized squared residuals (up to lag 10) is exhibited by any of these GARCH models, as shown by the Ljung-Box statistics.

Table 2b reports the results for SWARCH with 3 states and 2 lags which is the one specification which has the best out-of-sample performance relative to other specifications (models with 2 and 4 states and 1 lag). As in Hamilton and Susmel (1994) we started the maximization procedure from several points of the parameter space and we identified only one local maxima for the Gaussian T-SWARCH(3,2).[5] The ARCH parameters α_1 and α_2 are small and never statistically significant, implying marginal ARCH effects when switching regimes are involved. This contrasts with results reported by the authors based on weekly returns of the S&P 500 from July 1962 to December 1987. Volatility persistence implied by the ARCH component of the SWARCH process is very low, ranging from 0.12 (Gaussian SWARCH) to 0.20 (Student's-t T-SWARCH) [6]. The leverage effect is detected only when Student's–t innovations are involved. This is consistent with the likelihood ratio

[5]The code used to estimate SWARCH models is the one kindly provided by Jim Hamilton at his web site http://weber.ucsd.edu /˜ jhamilto/software.htm.

[6]Following Hamilton and Susmel (1994), the measure of persistence in volatility is computed by selecting the greatest eigenvalue of the matrix having as first row $[\alpha_1 + \psi/2 \quad \alpha_2]$, and as second row $[1 \quad 0]$.

test, where the increment of the likelihood from the Gaussian SWARCH to the Gaussian Threshold-SWARCH is very moderate. As in the case of GARCH models, Student's–t innovations are strongly supported by the data. The degrees-of-freedom in the Student's–t distribution is estimated at $\hat{\nu} = 6.33$, slightly higher than the corresponding estimate in the GARCH case ($\hat{\nu} = 5.60$). The estimated entries of the transition matrix of the Student's–t SWARCH(3,2), $m_{ij} = \Pr(z_{t+1} = i | z_t = j)$, are as follows

$$\hat{M} = \begin{bmatrix} .997 & .002 & 0 \\ .003 & .991 & .008 \\ 0 & .007 & .992 \end{bmatrix}$$

where the zeros are set as constraints to increase comparability with our proposal. Note that, in spite of the low persistence implied by the ARCH component, the latent regimes are strongly persistent, with a probability of less than 0.01 of moving from a given state to a lower or higher state. Diagnostics on standardized squared residuals (measured by the Ljung-Box statistic) does not signal major misspecification problems in the dynamics of the conditional variance. Once again the AIC rewards the case where asymmetric effect and Student's–t innovations are inserted into the model.

Tables 2c and 2d report estimation and diagnostics results for our Switching Markov model in correspondence of $N = 3$, and $N = 7$. Values of $N > 7$ improve neither the fit to the data nor the forecasting ability, while the case $N = 5$ exhibits intermediate results between the cases with 3 and 7 states, and hence they are not reported. When $N = 3$, we restrict $b = 0$ throughout, given the lack of statistical significance of its estimates: we do so in the expectation that a more parsimonious model should have a better ability in forecasting volatilities at longer time-horizons. We observe δ and, to a lesser extent, ν to increase with N reflecting the fact that a higher number of states allows one to capture a wider range of volatility. The parameter responsible for time-varying persistence, b, is statistically insignificant even when $N = 7$ and Student's-t innovations are involved. The estimated degrees of freedom of the Student's-t distribution are larger than the corresponding ones for the GARCH and SWARCH models, and always statistically significant. This is consistent with the fact that a higher number of states better captures the behavior of fatter tails in the unconditional distribution of returns. The parameter describing asymmetric effects in volatility, ψ, is always greater that 1, as expected. With Gaussian innovations and $N = 7$, both the likelihood ratio test and the standard error on ψ confirm that this asymmetric effect cannot be dropped out. This evidence is less clear with 3 states. This is in agreement with estimates of the Gaussian SWARCH. As with the GARCH and SWARCH models, the Ljung-Box statistic does not detect any autocorrelation in the standardized squared residuals. Based on the AIC one would select the case of Student's–t innovations and asymmetric effects both with 3 and 7 states. Likewise for the GARCH and SWARCH models the normality test on standardized residuals rejects the null hypothesis.

Insert **Tables 2a-2d** approximately here

Insert Figure 2 approximately here

Figure 2 shows in-sample and one–step–ahead out-of-sample variance estimates for the best models within each class, chosen according to the AIC. The top panel exhibits S&P 500 absolute returns, while the second, third and fourth panels show variance estimates for the GARCH, SWARCH and our proposal, respectively. As it can be seen from the figure, estimated variance levels displayed by the three panels are quite similar. However, a closer inspection reveals a striking difference in the way each reacts to sudden large shifts in volatility. In the GARCH model, large isolated shocks are slowly absorbed through time. The opposite happens for the SWARCH model, where large shocks decay very quickly. This is a reflection of the big difference in persistence implied by these models. The profile of the conditional variance estimated according to our proposal falls in between the other two approaches.

The smoothed probabilities that the Markov process could lie in a given state, estimated on the basis of the entire sample period, i.e. $\Pr(z_t = e_j | I_T^r)$, for $j = 1, 2, \cdots, N$ are also of interest. This evidence is reported in Figure 3, where the Gaussian Threshold–HMM(3) and the Gaussian Threshold–SWARCH(3,2) are compared. It is apparent that they came very close to one another during the whole in-sample period, as far as the estimated probabilities are concerned.

Figure 3 approximately here

From the evidence produced here, it is apparent that low volatility has characterized the periods January 95 - January 96, August 96 - December 96. Most of the second half of the sample (December 96 - May 98) is spent in a period of medium volatility with the highest peaks (third regime) occurring as a consequence of a short–lived turmoil. Notice that the series corresponding to our proposal has more of an "anticipatory" behavior, that is a less sharp movements in and out of a regime.

4.4 Forecasting Performance

Variance forecasts are obtained using the estimation results in Tables 2a-2d. Parameter estimates are held fixed during the forecasting exercise. Out-of-sample accuracy of variance forecasts is assessed using three forecasting horizons: 1 day, one week (5 days), and one month (20 days). We consider three proxies of volatility: squared excess returns, realized volatility, and the model-based estimator of volatility suggested by Barndorff–Nielsen and Shephard (2002). The first two measures are model-free consistent estimators of the conditional variance of returns, with the latter being a more efficient estimator (see Anderson and Bollerslev, 1998, and Barndorff–Nielsen and Shephard, 2002). Squared excess returns are computed as $(r_t - \hat{\mu})^2$, where $\hat{\mu}$ is the constant expected return over the full sample period (T=1303) estimated at 8.93%. Realized variances come from high-frequency (5-minutes) returns as in Blair *et al.* (2001). The method is briefly detailed in Appendix B. The model-based estimator of actual variance is described in Appendix C. Figure 4 shows the dynamics of the three benchmark measures during the out-of-sample period. Squared excess returns turn out to be very noisy estimates of actual variance, whereas model-based variance estimates show very smooth behavior.

Figure 4 approximately here

The forecasts $\hat{h}^2_{\tau+k|\tau}$ produced by each of the 12 models analyzed here were obtained with horizons $k = 1, 5, 20$ steps ahead. For each of the volatility proxies σ^2_τ as targets, we have run the Mincer-Zarnowitz regression for forecast unbiasedness

$$\sigma^2_{\tau+k} = \gamma_0 + \gamma_1 \hat{h}^2_{\tau+k|\tau} + u_{\tau+k},\ \ \tau = 870, \ldots, 1303 - k,\ \ k = 1, 5, 20$$

and tested the null hypothesis $H_0 : \gamma_0 = 0,\ \gamma_1 = 1$. The results are reported in Tables 3a–3c where for each model and for each target we report the estimated $\hat{\gamma}_0$ and $\hat{\gamma}_1$ with the Newey and West (1987) heteroskedastic and autocorrelation consistent standard errors in parentheses. We summarize results about the unbiasedness of the forecasts by reporting the p-value for a Wald-type test of the null joint hypothesis $H_0 = \gamma_0 = 0, \gamma_1 = 1$, and the R^2 of the regression to judge on the overall regression fit.

As one would expect, the performance of the forecasts (measured by the corresponding R^2) decreases with the forecast horizon. Results on forecast unbiasedness are somewhat mixed: for one-step forecasts and squared returns, most models (with the notable exception of the SWARCH class) are unbiased with a relatively poor fit. The fit increases when realized variance is adopted as a target but the test statistics reject unbiasedness across models. The columns related to the Barndorff-Nielsen and Shephard estimator show a much higher fit of the forecasts. Unbiasedness is exhibited by models belonging to our class. When the forecast horizon increases to $k = 5$, results for squared returns and the B–N/S estimates by and large remain the same while there is a dramatic improvement of the performance of all models when realized variance is used as a target. Finally, unbiasedness is generally lost when $k = 20$ and squared returns are considered with the notable exception of the Gaussian HMM(3) and GARCH. When the benchmarks are the B–N/S estimates, only the Gaussian HMM(3) produces unbiased variance forecasts, whereas unbiasedness is not rejected by most of the models for the realized variance. Whichever the volatility target, some further evidence on performance is shown through the behavior of the Mean Squared Error (MSE) loss function:

$$MSE = \frac{1}{433 - k + 1} \sum_{\tau=870}^{1303-k} (\sigma^2_{\tau+k} - \hat{h}^2_{\tau+k|\tau})^2.$$

Table 4 reports the percent improvement of accuracy of variance forecasts with respect to a naïve model with variance constant over time.[7] Comparisons are given for all the specifications in Tables 2a-2d. The percent improvement is computed by $100 \times (MSE_n - MSE_i)/MSE_n$, where the subscript n denotes the naïve model and while i refers to a generic competitor.

Insert **Table 4** approximately here

The first row of Table 4 reports values of the MSE for the naïve model for any given forecasting horizon and volatility target measure. The magnitude of these figures are indicators of the variability of the three target measures around a constant forecast over time. Bold numbers highlight the best model in each category. Almost all numbers are positive, showing that all specifications outperform the naïve one.

Dealing with short-term forecasts, when $k = 1$, the inclusion of asymmetric effects improves results of any model. The Gaussian GARCH with asymmetric effect turns out to be the best when the squared excess residuals (SqRet in the table) are taken as the benchmark measure. When comparisons are made in terms of the realized variance (RealVar) and the model–based estimates of variance (B–N/S), the Gaussian T-HMM(7),

[7] The naïve model has the form: $r_t = \alpha + \gamma r_{t-5} + u_t$, with $u_t \sim iiN(0, \sigma^2)$.

and the Student-t T-HMM(7) display a better forecasting ability.

The situation changes slightly when the horizon moves forward. For the squared excess returns, the improvements are fairly modest and pinpoint, if anything, a poor performance of the SWARCH specifications. For the other two volatility benchmark measures, when $k = 20$, we notice that there is a general preference for less parameterized models: with a model-based benchmark the 3–states version of our proposal comes ahead, whereas the GARCH with asymmetric effect and Student's-t innovations prevails when realized variance is used as a target.

To complete the picture, we have performed the Diebold and Mariano (1995) test for predictive accuracy comparison, keeping the parameter estimates fixed at their in-sample values and choosing the Gaussian GARCH as a benchmark model. The test statistic has an expected value of 0 under the null hypothesis of no difference between the benchmark model and the competitors. The results are presented in the Tables 5. One can note that the SWARCH models perform consistently worse than the benchmark (and fairly significantly so). The inclusion of asymmetric effect and Student's-t innovations significatively (at a 5% level) improves the predictive ability of the GARCH and HMM(7) models just for the shortest forecasting horizon. For longer horizons none of the models in the table performs significatively better than the Gaussian GARCH.

5 Conclusions

In this paper we have suggested a stochastic model for modelling financial time series volatility based on the idea that conditional variance can take on a finite number of discrete values and that its dynamics is ruled by a inhomogeneous Markov chain with time-varying transition matrix. We showed that the main advantage of such an approach is to provide a non-linear framework which accommodates a different treatment of innovations according not only to their sign (as in the asymmetric GARCH models) but also according to their size, translating the idea that extraordinary movements in returns are shortly lived and less persistent than small sized changes. This agenda is made feasible by a parameterization which guards against the procedure being prone to a shortcoming of traditional Markov Switching models where the number of parameters quadratically increases when the number of states increases (an exception being Calvet and Fisher, 2001, 2004).

From an empirical point of view, the in-sample performance of our proposal is encouraging, in that the stylized facts of the financial time series analyzed are well captured by the features of the model. The model estimation is quite simple, even in the presence of a large number of states (though for the data at hand on the S&P 500 index, the performance of the model does not improve substantially if the number of states increases beyond $N = 7$). The model diagnostics are reassuring in that the standardized residuals do not show departures from the assumptions on the innovations process.

We chose to perform a forecasting comparison by using three different target variables as proxies for volatility, namely, squared returns, a model–based measure of volatility suggested by Barndorff–Nielsen and Shephard and realized volatility. Not surprisingly, since the former measure is very noisy, the tracking record is not very good, whereas for the other two the results show a slight superiority of our approach when compared to standard GARCH models; Switching ARCH models exhibit a relatively poorer prediction ability overall. A fuller version of our model (with 7 states and time-varying transition probabilities) performs well at very short horizons (one step-ahead), while at longer horizons (five and twenty step–ahead) a more parsimonious model (with three states and constant transition probabilities) appears to be preferable.

A Building Realized Variance Series

The daily S&P 500 realized variances are built from 5-minutes high frequency returns supplied to us by Olsen Ltd. The latest index level available before the 5-minutes mark is used to compute 5-minutes returns. As in Blair *et al.* (2001), realized variance for day t is constructed by summing the squared overnight return of the previous day and the squares of the 78 intra-daily 5-minutes returns occurring between 08:35 CST and 15:05 CST of the current day. The overnight return is defined as the log-difference between the last index level at day $t-1$ (15:05 CST) and the first index level at day t (08:35 CST), multiplied by 100. In days where 5-minutes returns were not available (such as public holidays) or where more than 41 intra-daily observations were missing, a measure of realized variance has been built by interpolating across realized variances of the previous and subsequent days. This occurred 16 times within the out-of-sample period for which we computed the series of realized variances. The time series of realized variances is reported in the bottom panel of Figure 4.

B The Barndorff-Nielsen and Shephard Model-based Estimator of Actual Variance

Suppose the dynamics of log-prices of an asset, say p_t, is driven by the stochastic differential equation

$$dp_t = (\alpha + \beta\sigma_t^2)dt + \sigma_t dW_t, \tag{18}$$

where W_t is the standard Wiener process, and σ_t^2 is the instantaneous variance prevailing at time t. Over a small interval of time δ, returns are given by $r_n = p_{n\delta} - p_{(n-1)\delta}$. From (18) it follows that

$$r_n|\sigma_n^2 \sim N[\alpha\delta + \beta\sigma_n^2, \sigma_n^2]$$

where the actual variance σ_n^2, equals $\int_{(n-1)\delta}^{n\delta} \sigma_t^2 dt$. Defining

$$\{p\}_n^G = \sum_{i=1}^{G} [p_{(n-1)\delta+\delta i \backslash G} - p_{(n-1)\delta+\delta(i-1)\backslash G}]^2$$

as the realized variance at time n using G intra–daily observations, the authors show that $\{p\}_n^G$ is a consistent ($G \to \infty$) and unbiased (when $\alpha = \beta = 0$) estimator of the actual variance, i.e.

$$\{p\}_n^G = \sigma_n^2 + e_n, \qquad E[e_n|\sigma_n^2] = 0 \tag{19}$$

Moreover, if the variance process follows a non-Gaussian Ornstein-Uhlenbeck process, or a Constant Elasticity of Variance process, then the actual variance admits an ARMA(1,1) representation. Thus, daily squared returns ($G = 1$) might be used to estimate a "cleaner" measure of the actual variance σ_t^2 by writing

$$\begin{aligned} r_t^2 &= \sigma_t^2 + e_t \\ (1 - \phi L)\sigma_t^2 &= \mu_\sigma + (1 + \theta L)u_t \end{aligned} \tag{20}$$

where e_t, and u_t are independent Gaussian white noises with variances V_e, and V_u.

In general, model (20) is not identified: part of the noise e_t can be assigned to the unobserved component σ_t^2, or, vice versa, part of the shocks on σ_t^2 can be assigned to e_t without altering the likelihood. We identify a decomposition using the canonical assumption, i.e. by considering the non–invertible ARMA(1,1) representation of σ_t^2 (see Maravall and Planas, 1999). This yields the smoothest estimate of the actual variance σ_t^2.

Using the full sample available, we estimated model parameters by maximum likelihood after casting equations (20) in a state-space format. Parameter estimates (standard errors), are as follows: $\phi = .928$ (.028), $\mu_\sigma = .069$ (.028), $V_e = 6.219$ (.210), and $V_u = .043$ (.021). The estimated inverse signal to noise ratio, $V_e/V_u \simeq 143$, implying that squared residuals represent a somewhat noisy estimate of actual variance. Inference for the unobserved component σ_t^2 is derived after smoothing with the fixed-point smoother algorithm (see Harvey, 1989, pp.151-154). Our measure of the actual variance at time t is then given by $E[\sigma_t^2 | r_1, r_2, \cdots, r_T]$, with $t = 1, 2, \cdots, T$. The out-of-sample period estimates of actual variance are shown in Figure 5 together with squared excess returns.

Figure 5 approximately here

References

Akaike H. 1974. A New Look at the Statistical Model Identification, IEEE Transactions on Automatic Control AC-19, 716-723.

Andersen T.G., and Bollerslev T. 1998. Answering the skeptics: yes standard volatility models do provide accurate forecasts. International Economic Review 39, 885-905.

Andrews D.W.K. and W. Ploberger 1994. Optimal Tests when a Nuisance Parameter is Present only under the Alternative, Econometrica, 62, 1383-1414.

Baillie R. T., and Bollerslev T. 1989. The message in daily exchange rates: a conditional-variance tale, Journal of Business & Economic Statistics 7, 297-305.

Barndorff-Nielsen O.E., and Shephard N. 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models, Journal of the Royal Statistical Society B 64, 253-280.

Black F. 1976, Studies of stock market volatility changes, Proceedings of the American Statistical Association, Business and Economic Statistics section, 177-181.

Black F. and Scholes M. 1973. The Pricing of Options and Corporate Liabilities, Journal of Political Economy 81, 637-654.

Blair J.B., Poon S.H., and Taylor J. 2001. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns, Journal of Econometrics 105, 5-26.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307-327.

Bollerslev T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return, Review of Economics & Statistics 69, 542-547.

Bollerslev T., Chou R.Y., and Kroner K.P. 1992. ARCH modelling in finance: a review of theory and empirical evidence, Journal of Econometrics 52, 5-59.

Bollerslev T., Engle R., and Nelson D. 1994. ARCH models, Ch. 49 in R.F. Engle, and D.L. McFadden (eds.), Handbook of Econometrics, 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, 839-866.

Cai J. 1994. A Markov model of Switching-regime ARCH, Journal of Business and Economic Statistics 12, No. 3, 309-316.

Calvet L., and Fisher A.J. 2001. Forecasting multifractal volatility, Journal of Econometrics, 105, 27-58.

Calvet L., and Fisher A.J. 2004. How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes, Journal of Financial Econometrics, 2, 49-83.

Duan J., Popova I., and Ritchken P. 1999. 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, 209-218.

Engle R. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation,Econometrica 50, 987-1008.

Evans M.D.D. 2003. Real Risk, Inflation Risk, and the Term Structure, The Economic Journal 113, 345-389.

Friedman B.M, and Laibson D.I. 1989. Economic Implications of Extraordinary Movements in Stock Prices, Brookings Papers on Economic Activity 2, 137-172 and 186-189.

Ghysels E., Harvey A., and Renault E. 1996. Stochastic Volatility, Handbook of Statistics 14, 119-191.

Glosten L.R., Jagannathan R., and Runkle D.E. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48, 1779-1801.

Gray S.F. 1996. Modelling the conditional distribution of interest rates as a regime-switching process, Journal of Financial Economics 42, 27-62.

Kim C-J. 1994. Dynamic linear models with Markov Switching, Journal of Econometrics 60, 1-22.

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

Hamilton J.D. 1990. Analysis of Time Series Subject to Changes in Regimes, Journal of Econometrics 45, 39-70.

Hamilton J.D. 1994. Time Series Analysis, Princeton University Press.

Hamilton J.D. 1996. Specification Testing in Markov-Switching Time-Series Models, Journal of Econometrics 70, 127-157.

Hansen B.E. 1992. The Likelihood-Ratio Test under Non-standard Conditions: Testing the Markov Switching Model of GNP, Journal of Applied Econometrics, 7, S61-S82.

Hamilton J.D., and Susmel R. 1994. Autoregressive conditional heteroskedasticity and changes in regimes, Journal of Econometrics 64, 307-333.

Harvey A.C. 1989. Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

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

Harvey A.C., Ruiz E., and Sentana E. 1992. Unobservable component time series models with ARCH disturbances, Journal of Econometrics 52, 129-158.

Lamoureux C.G., and Lastrapes W.D. 1990a. Persistence in variance, structural changes, and the GARCH model, Journal of Business and Economic Statistics 8, No. 2, 226-234.

Lamoureux C.G., and Lastrapes W.D. 1990b. Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects, Journal of Finance 45, 221–229.

Ljung G., and Box G.E.P. 1978. On a Measure of Lack of Fit in Time Series Models, Biometrika 65, 297-303.

Maravall A., and Planas C. 1999. Estimation error and the specification of unobserved component models, Journal of Econometrics 92, 325-353.

Naik V., and Lee M. 1994. The yield curve and bond option prices with discrete shifts in economic regimes, manuscript, University of British Columbia.

Nelson D.B. 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347-370.

Newey W., and West K. 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

Sentana E. 1995. Quadratic ARCH models, Review of Economic Studies 62, 639-661.

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, 1-31.

Figure 1: Variance persistence profiles for the Student-t Threshold HMM(7)

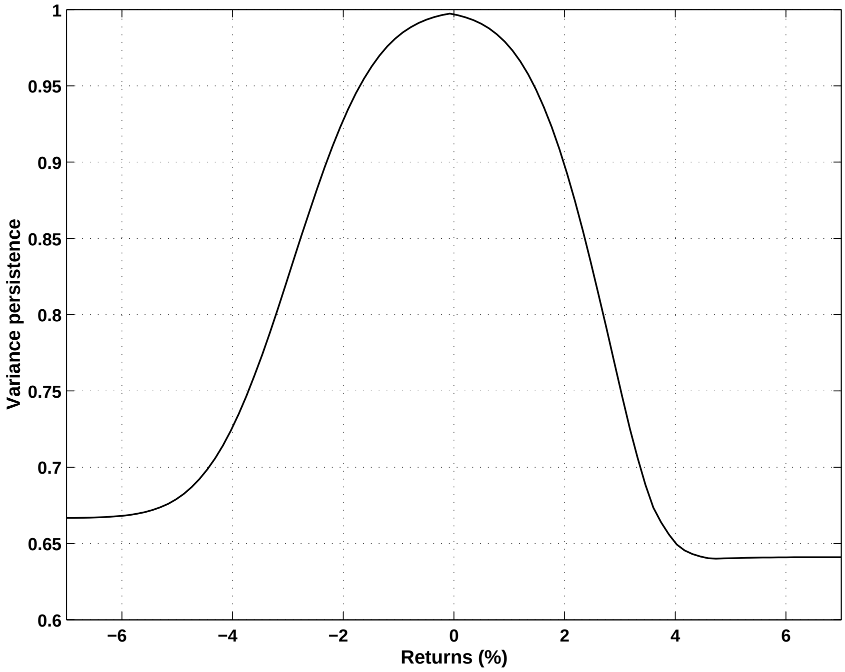

Figure 2: In and Out-of-sample (one-step-ahead) variance estimates

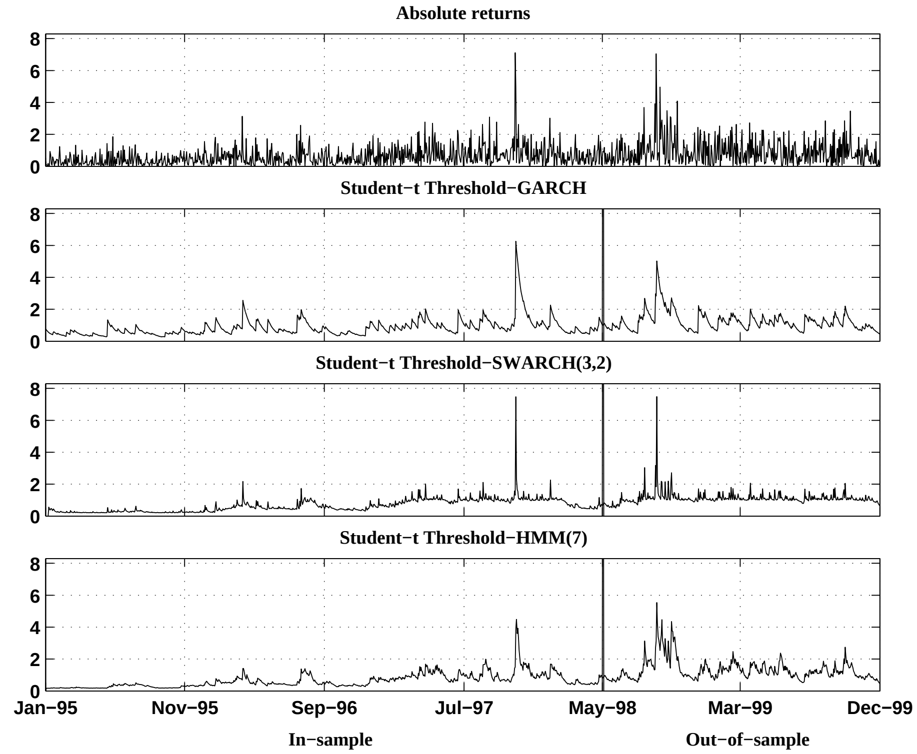

Figure 3: Smoothed probabilities for the Gaussian Threshold-HMM(3) (full line) and Gaussian Threshold-SWARCH(3,2) (dashed line) over time.

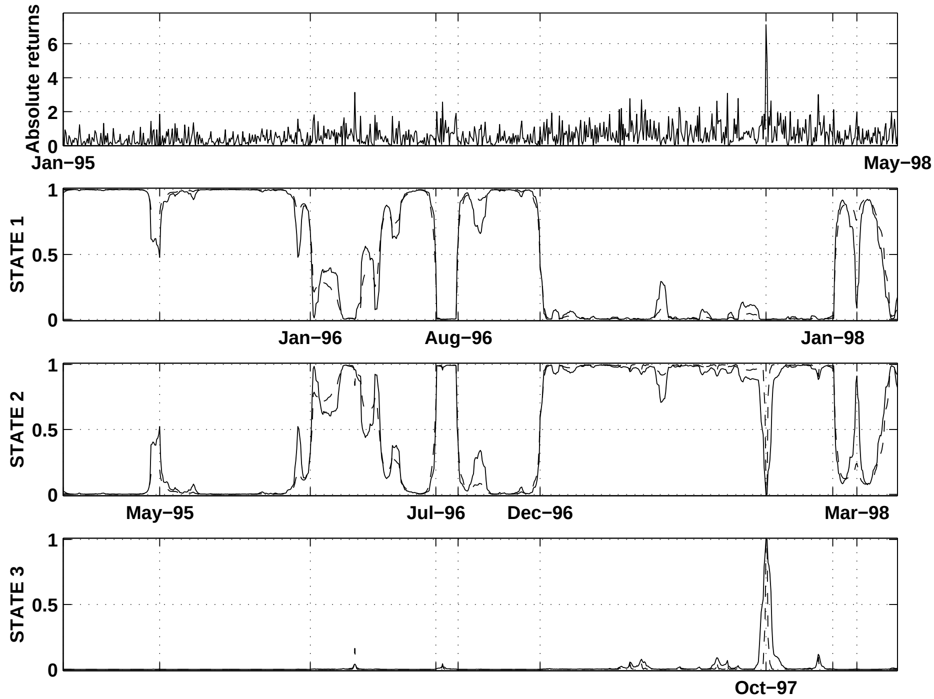

Figure 4: Volatility benchmarks (out-of-sample period)

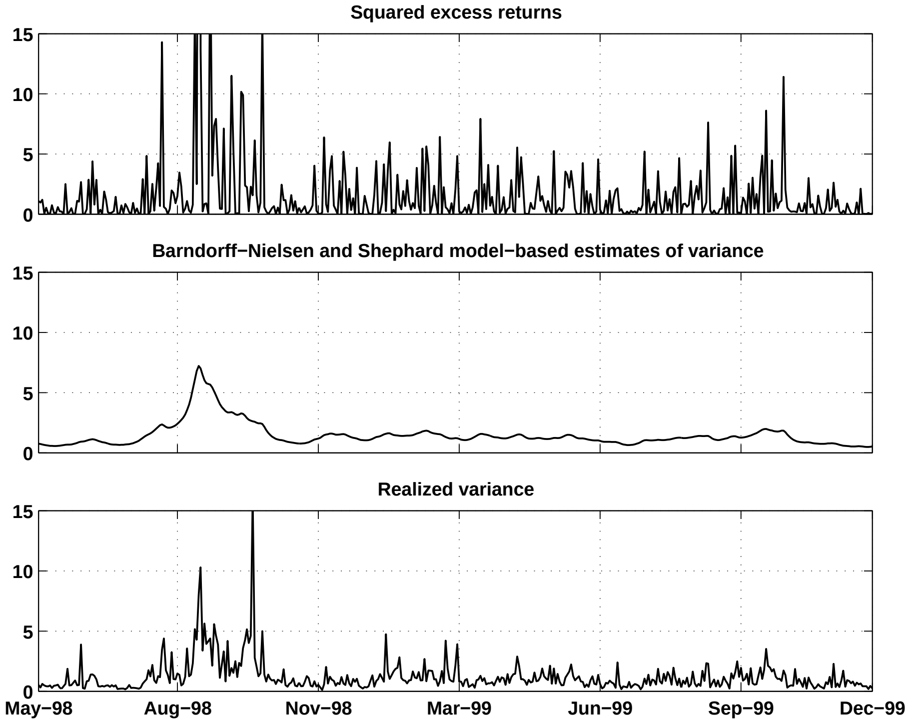

Figure 5: Squared excess returns and the smoothed estimates of actual variance

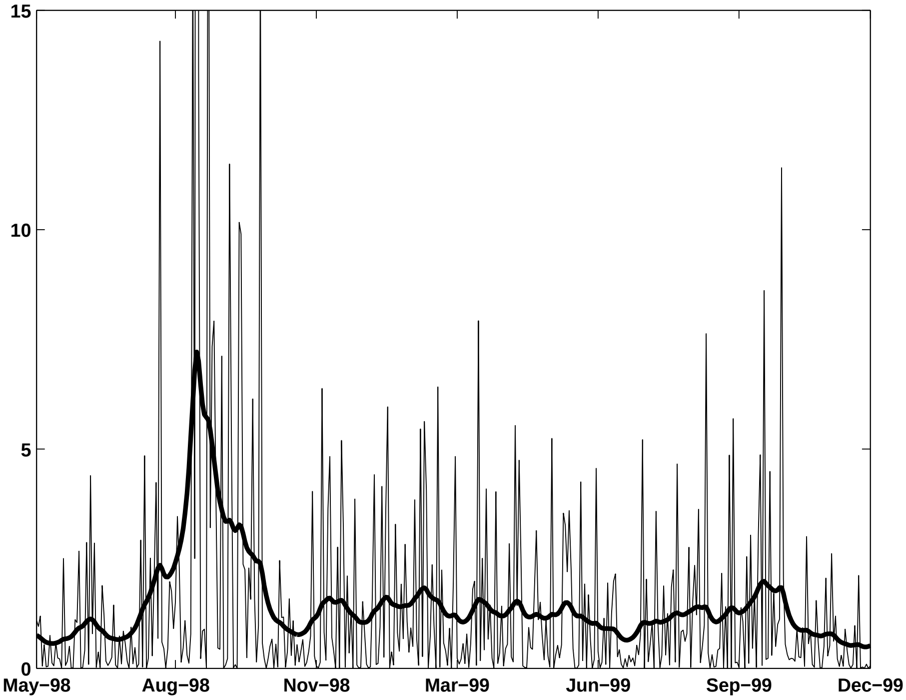

Table 1

Sample statistics for daily returns

Number of observations	870	
Mean	.102	
Standard deviation	.825	
Skewness	-0.70	
Excess kurtosis	8.63	
$Q(10)$ on returns	32.48	[0.000]
$Q(10)$ on squared returns	99.93	[0.000]

Note: $Q(10)$ denotes the Ljung-Box statistics for the first 10 lags. p-values within brackets.

Table 2a

Estimation results and diagnostics

Model: $r_t = \mu + \gamma r_{t-5} + h_t u_t, \quad u_t | I^r_{t-1} \sim t_\nu$

$$h^2_t = \omega + \alpha u^2_{t-1} + \psi 1_{(u_{t-1} \leq 0)} u^2_{t-1} + \beta h^2_{t-1}$$

Gaussian GARCH

μ	γ	ω	α	β	ψ^*	ν^*
.126	-.116	$.62\times10^{-3}$	.068	.927	0	∞
(.024)	(.035)	$(.24\times10^{-3})$	(.010)	(.012)	-	-

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$	J-B
	-979.83	2.2640	6.0 [.81]	348.1 [0.00]

Gaussian Threshold-GARCH

μ	γ	ω	α	β	ψ	ν^*
.108	-.104	.017	.010	.891	.155	∞
(.024)	(.033)	(.004)	(.013)	(.015)	(.022)	-

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$	J-B
	-968.04	2.2392	6.70 [.75]	181.6 [0.00]

Student's-t Threshold-GARCH

μ	γ	ω	α	β	ψ	ν
.110	-.081	.014	.008	.911	.126	5.603
(.028)	(.034)	(.007)	(.019)	(.026)	(.046)	(1.125)

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$
	-946.67	2.1924	7.25 [.70]

Notes: estimation is based on the first 870 daily observations. Standard errors are in parentheses and p-values in brackets. AIC denotes the Akaike Information Criterion, $Q_{r^2}(10)$ denotes the Ljung-Box statistics for the first 10 lags of squared residuals. J-B is the value of the Jarque and Bera Normality test statistics on standardized residuals. * denotes constrained model parameters.

Table 2b

Estimation results and diagnostics

Model: $r_t = \mu + \gamma r_{t-5} + g_{z_t}^{1/2} v_t, \quad v_t = h_t u_t, \quad u_t \sim iit_\upsilon$

$h_t^2 = \alpha_0 + \alpha_1 v_{t-1}^2 + \alpha_2 v_{t-2}^2 + \psi 1_{(v_{t-1} \leq 0)} v_{t-1}^2, \quad g_{z_t} \in (1 \;\; g_2 \;\; g_3)$

z_t is a first order homogeneous Markov chain with 3 states

Gaussian SWARCH(3,2)

μ	γ	g_2	g_3	α_0	α_1	α_2	ψ^*	υ^*
0.129	-.090	3.531	72.410	.252	1×10^{-7}	.015	0	∞
(0.023)	(.033)	(.414)	(63.014)	(.025)	(1×10^{-4})	(.033)		

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$	J-B
	-955.60	2.2221	4.72 [.91]	388.8 [.00]

Gaussian Threshold-SWARCH(3,2)

μ	γ	g_2	g_3	α_0	α_1	α_2	ψ	υ^*
.132	-.093	3.523	62.079	.252	1×10^{-7}	.019	.066	∞
(.022)	(.032)	(.400)	(61.308)	(.025)	(7×10^{-5})	(.031)	(.065)	

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$	J-B
	-955.08	2.2232	5.66 [.84]	292.3 [.00]

Student's-t Threshold-SWARCH(3,2)

μ	γ	g_2	g_3	α_0	α_1	α_2	ψ	υ
.127	-.098	2.075	5.049	.206	1×10^{-8}	.026	.127	6.334
(.023)	(.032)	(.449)	(1.028)	(.035)	(4×10^{-5})	(.037)	(.067)	(1.120)

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$
	-944.07	2.2002	7.13 [.71]

Note: see Table 2a

Table 2c

Estimation results and diagnostics

Model: $r_t = \mu + \gamma r_{t-5} + \sigma(z_t)^{1/2} u_t, \quad u_t \sim iit_v$

$\sigma_i = \exp[\alpha + \delta g(e_i)], \;\; g(e_i) = \frac{2i-(N+1)}{N-1}, \;\; i = 1, 2, \ldots, N$

$z_t = M_{t-1} z_{t-1} + v_t, \quad$ where M_t is given by (6)-(8)

Gaussian HMM(3)

μ	γ	α	δ	a	b^*	ψ^*	v^*
.127	-.087	-.057	1.358	-2.188	0	1	∞
(.022)	(.033)	(.085)	(.102)	(.178)	-	-	-

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$	J-B
	-963.32	2.2260	5.92 [.82]	381.2 [.00]

Gaussian Threshold-HMM(3)

μ	γ	α	δ	a	b^*	ψ	v^*
.129	-.085	-.023	1.413	-1.955	0	3.144	∞
(.022)	(.033)	(.091)	(.116)	(.235)	-	(1.226)	-

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$	J-B
	-960.04	2.2208	6.25 [.79]	244.9 [.00]

Student's-t Threshold-HMM(3)

μ	γ	α	δ	a	b^*	ψ	v
.121	-.080	-.076	1.263	-2.243	0	2.836	7.115
(.021)	(.033)	(.116)	(.128)	(.187)	-	(1.450)	(1.830)

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$
	-950.68	2.2016	6.11 [.81]

Note: see Table 2a

Table 2d

Estimation results and diagnostics

Model: $r_t = \mu + \gamma r_{t-5} + \sigma(z_t)^{1/2} u_t, \quad u_t \sim iit_v$

$\sigma_i = \exp[\alpha + \delta g(e_i)], \;\; g(e_i) = \frac{2i-(N+1)}{N-1}, \;\; i = 1, 2, \ldots, N$

$z_t = M_{t-1} z_{t-1} + v_t, \quad$ where M_t is given by (6)-(8)

Gaussian HMM(7)

μ	γ	α	δ	a	b	ψ^*	v^*
.128	-.095	.068	1.920	-2.566	1.273	1	∞
(.022)	(.034)	(.211)	(.218)	(.464)	(.557)	-	-

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$	J-B
	-953.81	2.2065	5.99 [.82]	180.6 [.00]

Gaussian Threshold-HMM(7)

μ	γ	α	δ	a	b	ψ	v^*
.125	-.090	.538	2.365	-2.185	.912	2.291	∞
(.021)	(.034)	(.264)	(.310)	(.242)	(.094)	(0.259)	-

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$	J-B
	-949.93	2.1999	8.49 [.58]	135.1 [.00]

Student's-t Threshold-HMM(7)

μ	γ	α	δ	a	b	ψ	v
.120	-.085	.598	2.378	-2.477	.848	2.389	8.114
(.021)	(.033)	(.368)	(.510)	(.446)	(.543)	(0.877)	(2.567)

Diagnostics:	Log-Like	AIC	$Q_{r^2}(10)$
	-944.67	2.1901	9.23 [.51]

Note: see Table 2a

Table 3a

Mincer-Zarnowitz regressions - One-step-ahead forecasts

	Squared Returns				Realized Variance				Barndorff-Nielsen/Shephard			
	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2
Gaussian GARCH	-0.72	1.84	0.12	0.06	-0.86	1.67	0.00	0.34	-0.81	1.84	0.00	0.68
	(0.49)	(0.48)			(0.27)	(0.26)			(0.17)	(0.16)		
Gaussian T-GARCH	-0.45	1.62	0.04	0.09	-0.39	1.28	0.00	0.40	-0.09	1.25	0.00	0.62
	(0.46)	(0.46)			(0.14)	(0.15)			(0.14)	(0.14)		
Student's t T-GARCH	-0.59	1.75	0.06	0.09	-0.53	1.40	0.00	0.40	-0.26	1.40	0.00	0.66
	(0.46)	(0.46)			(0.16)	(0.17)			(0.14)	(0.13)		
Gaussian SWARCH	0.89	0.52	0.00	0.05	0.67	0.42	0.00	0.21	0.89	0.45	0.00	0.41
	(0.27)	(0.26)			(0.11)	(0.08)			(0.10)	(0.06)		
Gaussian T-SWARCH	0.76	0.63	0.00	0.05	0.55	0.51	0.00	0.22	0.77	0.55	0.00	0.43
	(0.29)	(0.28)			(0.12)	(0.09)			(0.10)	(0.06)		
Student's-t T-SWARCH	-0.45	1.80	0.02	0.05	-0.62	1.63	0.00	0.28	0.05	1.27	0.00	0.28
	(0.41)	(0.44)			(0.13)	(0.13)			(0.16)	(0.13)		
Gaussian HMM(3)	0.13	0.99	0.41	0.05	-0.07	0.88	0.00	0.29	0.12	0.94	0.49	0.55
	(0.49)	(0.43)			(0.22)	(0.19)			(0.19)	(0.17)		
Gaussian T-HMM(3)	0.13	1.07	0.05	0.06	-0.09	0.97	0.01	0.35	0.17	0.98	0.01	0.58
	(0.50)	(0.47)			(0.20)	(0.19)			(0.18)	(0.17)		
Student's-t T-HMM(3)	-0.11	1.22	0.30	0.05	-0.32	1.11	0.00	0.31	-0.10	1.15	0.33	0.54
	(0.55)	(0.50)			(0.25)	(0.23)			(0.23)	(0.21)		
Gaussian HMM(7)	-0.05	1.23	0.10	0.04	-0.44	1.26	0.00	0.30	-0.18	1.26	0.11	0.50
	(0.53)	(0.51)			(0.26)	(0.25)			(0.21)	(0.20)		
Gaussian T-HMM(7)	-0.08	1.32	0.01	0.08	-0.28	1.19	0.03	0.42	0.11	1.09	0.00	0.58
	(0.47)	(0.47)			(0.15)	(0.15)			(0.14)	(0.14)		
Student's-t T-HMM(7)	-0.22	1.33	0.20	0.07	-0.47	1.26	0.00	0.43	-0.08	1.16	0.10	0.60
	(0.51)	(0.47)			(0.19)	(0.17)			(0.18)	(0.16)		

Notes: the first two columns report estimated values of γ_0, and γ_1 for the Mincer-Zarnowitz regressions $\sigma^2_{\tau+k} = \gamma_0 + \gamma_1 \hat{h}^2_{\tau+k|\tau} + u_{\tau+k}$, $\tau = 870, \ldots, 1303 - k, k = 1$, where $\sigma^2_{\tau+k}$ is one of the volatility proxy (squared returns, realized variance, and the Barndorff-Nielsen and Shephard model–based estimates of variance) and $\hat{h}^2_{\tau+k|\tau}$ are volatility forecasts. p-val indicates the p-value of the Wald-type statistic for testing the null $H_0 : \gamma_0 = 0,\ \gamma_1 = 1$.

Table 3b

Mincer-Zarnowitz regressions - Five-step-ahead forecasts

	Squared Returns				Realized Variance				Barndorff-Nielsen/Shephard			
	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2
Gaussian GARCH	-0.33	1.53	0.16	0.04	-0.29	1.21	0.41	0.18	-0.28	1.42	0.01	0.41
	(0.55)	(0.46)			(0.27)	(0.25)			(0.18)	(0.16)		
Gaussian T-GARCH	0.20	1.13	0.13	0.04	0.07	0.94	0.91	0.18	0.17	1.09	0.01	0.40
	(0.43)	(0.33)			(0.17)	(0.18)			(0.13)	(0.12)		
Student's t T-GARCH	0.04	1.27	0.11	0.04	-0.02	1.02	0.99	0.19	0.03	1.20	0.01	0.43
	(0.43)	(0.34)			(0.18)	(0.18)			(0.13)	(0.12)		
Gaussian SWARCH	-1.07	2.62	0.01	0.02	-1.35	2.54	0.08	0.15	-0.93	2.38	0.00	0.22
	(1.13)	(1.14)			(0.82)	(0.86)			(0.54)	(0.57)		
Gaussian T-SWARCH	-1.47	2.98	0.01	0.02	-1.30	2.46	0.11	0.10	-1.19	2.61	0.00	0.19
	(1.34)	(1.34)			(0.80)	(0.82)			(0.78)	(0.79)		
Student's-t T-SWARCH	-0.48	1.97	0.03	0.01	-0.26	1.41	0.34	0.03	-0.29	1.70	0.00	0.06
	(0.68)	(0.76)			(0.40)	(0.45)			(0.46)	(0.50)		
Gaussian HMM(3)	0.39	0.81	0.33	0.03	0.12	0.74	0.02	0.19	0.26	0.83	0.21	0.38
	(0.31)	(0.27)			(0.22)	(0.19)			(0.19)	(0.17)		
Gaussian T-HMM(3)	0.46	0.87	0.21	0.03	0.07	0.90	0.80	0.21	0.29	0.93	0.00	0.37
	(0.36)	(0.26)			(0.22)	(0.20)			(0.17)	(0.16)		
Student's-t T-HMM(3)	0.17	1.05	0.33	0.03	-0.08	0.97	0.22	0.20	0.13	1.01	0.06	0.36
	(0.34)	(0.31)			(0.25)	(0.23)			(0.21)	(0.21)		
Gaussian HMM(7)	0.18	1.10	0.24	0.02	-0.02	0.98	0.81	0.11	0.07	1.11	0.05	0.23
	(0.30)	(0.29)			(0.21)	(0.21)			(0.20)	(0.20)		
Gaussian T-HMM(7)	0.70	0.73	0.07	0.02	0.21	0.84	0.50	0.16	0.52	0.80	0.00	0.24
	(0.32)	(0.21)			(0.21)	(0.22)			(0.14)	(0.15)		
Student's-t T-HMM(7)	0.47	0.86	0.17	0.02	0.01	0.94	0.70	0.19	0.33	0.89	0.00	0.28
	(0.29)	(0.21)			(0.23)	(0.21)			(0.15)	(0.16)		

Note: as in table 3a with k=5.

Table 3c

Mincer-Zarnowitz regressions - Twenty–step–ahead forecasts

	Squared Returns				**Realized Variance**				**Barndorff-Nielsen/Shephard**			
	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2	$\hat{\gamma}_0$	$\hat{\gamma}_1$	p-val	R^2
Gaussian GARCH	0.57	0.82	0.04	0.01	0.05	0.95	0.99	0.12	0.62	0.70	0.00	0.11
	(0.28)	(0.27)			(0.34)	(0.33)			(0.16)	(0.17)		
Gaussian T-GARCH	0.41	1.08	0.00	0.02	0.18	0.97	0.11	0.13	0.53	0.88	0.00	0.17
	(0.29)	(0.37)			(0.30)	(0.34)			(0.18)	(0.23)		
Student's t T-GARCH	0.39	1.10	0.00	0.02	0.12	1.01	0.18	0.13	0.51	0.90	0.00	0.17
	(0.28)	(0.36)			(0.31)	(0.35)			(0.18)	(0.23)		
Gaussian SWARCH	0.38	1.38	0.00	0.01	0.39	0.95	0.00	0.02	0.57	1.05	0.00	0.04
	(0.18)	(0.39)			(0.13)	(0.24)			(0.09)	(0.21)		
Gaussian T-SWARCH	0.38	1.35	0.00	0.01	0.39	0.93	0.00	0.02	0.57	1.03	0.00	0.04
	(0.18)	(0.39)			(0.13)	(0.24)			(0.09)	(0.21)		
Student's-t T-SWARCH	0.26	1.37	0.00	0.01	0.31	0.94	0.00	0.03	0.52	1.00	0.00	0.05
	(0.22)	(0.38)			(0.16)	(0.24)			(0.12)	(0.21)		
Gaussian HMM(3)	0.76	0.55	0.10	0.01	0.26	0.65	0.05	0.10	0.64	0.57	0.02	0.13
	(0.36)	(0.28)			(0.30)	(0.26)			(0.26)	(0.22)		
Gaussian T-HMM(3)	0.41	1.07	0.00	0.03	0.38	0.76	0.02	0.10	0.84	0.58	0.00	0.10
	(0.33)	(0.41)			(0.27)	(0.32)			(0.15)	(0.20)		
Student's-t T-HMM(3)	0.59	0.82	0.02	0.02	0.33	0.74	0.40	0.09	0.81	0.56	0.00	0.08
	(0.33)	(0.34)			(0.32)	(0.33)			(0.19)	(0.20)		
Gaussian HMM(7)	0.57	0.84	0.01	0.01	0.52	0.58	0.01	0.03	0.73	0.62	0.00	0.06
	(0.37)	(0.41)			(0.18)	(0.20)			(0.20)	(0.22)		
Gaussian T-HMM(7)	0.68	0.75	0.00	0.03	0.72	0.41	0.00	0.05	1.07	0.34	0.00	0.06
	(0.26)	(0.33)			(0.13)	(0.17)			(0.09)	(0.12)		
Student's-t T-HMM(7)	0.67	0.72	0.00	0.02	0.63	0.46	0.00	0.06	1.02	0.36	0.00	0.07
	(0.26)	(0.30)			(0.16)	(0.19)			(0.11)	(0.13)		

Note: as in table 3a with k=20.

Table 4

Comparison of variance forecasts relative to the constant variance model

		k=1			k=5			k=20		
		SqRet	**RealVar**	**B-N/S**	**SqRet**	**RealVar**	**B-N/S**	**SqRet**	**RealVar**	**B-N/S**
Constant variance	**MSE**	12.69	2.02	1.68	12.81	2.03	1.69	13.25	2.10	1.76
					Percent improvement in MSE					
Gaussian GARCH		9.76	37.93	67.54	8.81	28.71	57.43	6.37	24.63	39.66
Gaussian T-GARCH		**12.72**	46.06	71.41	8.74	29.35	58.12	6.33	23.89	38.64
Student's-t T-GARCH		12.25	45.37	71.86	**8.97**	30.00	59.11	6.37	**24.69**	38.88
Gaussian SWARCH		6.28	-5.07	21.87	5.06	19.79	33.43	2.70	9.55	17.67
Gaussian T-SWARCH		8.50	14.84	43.18	4.97	17.56	32.44	2.91	10.21	19.06
Student's-t T-SWARCH		8.26	33.42	46.22	4.50	14.21	29.26	3.89	13.02	24.78
Gaussian HMM(3)		11.12	35.57	71.27	8.88	24.61	**60.21**	**6.39**	17.39	**41.40**
Gaussian T-HMM(3)		11.42	42.53	72.01	8.11	30.02	57.35	5.04	21.22	33.30
Student's-t T-HMM(3)		10.76	38.61	69.91	8.90	29.41	58.12	5.24	19.06	34.26
Gaussian HMM(7)		9.42	36.61	64.77	7.39	24.58	49.63	4.98	16.97	31.48
Gaussian T-HMM(7)		11.97	**48.80**	69.26	6.77	27.45	46.44	2.30	9.31	14.43
Student's-t T-HMM(7)		11.95	48.20	**72.29**	8.48	**30.14**	54.79	4.40	16.84	27.53

Notes: The first row contains MSE's values. The remaining entries report percent improvement of accuracy of forecasts w.r.t. the constant variance model. The accuracy is measured by $MSE = \frac{1}{T-k}\sum_{t=1}^{T-k}\left(\sigma^2_{t+k} - \hat{h}^2_{t+k|t}\right)^2$, where k is the forecasting horizon, σ^2_{t+k} is the "true" variance prevailing at time $t+k$ measured by SqRet: squared excess returns, RealVar: realized variance, and B-N/S: the Barndorff-Nielsen and Shephard model-based estimator of variance, $\hat{h}^2_{t+k|t}$ denotes the k–step–ahead variance forecast, and $T = 433$. Percent improvement is computed by $100 \times (MSE_n - MSE_i)/MSE_n$, where subscripts n and i refer to the constant variance model and a generic volatility model. Bold indicates the best in each category.

Table 5

Diebold-Mariano tests: model benchmark Gaussian GARCH - MS comparisons

	k=1			k=5			k=20		
	SqRet	**RealVar**	**B-N/S**	**SqRet**	**RealVar**	**B-N/S**	**SqRet**	**RealVar**	**B-N/S**
Gaussian T-GARCH	1.26	**2.44**	1.31	-0.06	0.30	0.17	-0.08	-0.18	-0.37
	[0.21]	[0.01]	[0.19]	[0.94]	[0.76]	[0.87]	[0.93]	[0.85]	[0.71]
Student's-t T-GARCH	1.52	**2.92**	**1.94**	0.22	0.74	0.51	0.00	0.02	-0.31
	[0.13]	[0.00]	[0.05]	[0.83]	[0.46]	[0.61]	[0.99]	[0.98]	[0.75]
Gaussian SWARCH	-0.41	-2.57	-3.43	-1.77	-1.39	-2.15	-1.21	-0.93	-1.17
	[0.68]	[0.01]	[0.00]	[0.07]	[0.16]	[0.03]	[0.22]	[0.35]	[0.24]
Gaussian T-SWARCH	-0.19	-2.04	-2.98	-1.69	-1.44	-2.06	-1.18	-0.91	-1.13
	[0.84]	[0.04]	[0.00]	[0.09]	[0.15]	[0.04]	[0.24]	[0.36]	[0.25]
Student's-t T-SWARCH	-0.76	-0.82	-4.02	-1.55	-1.54	-1.88	-1.01	-0.84	-0.96
	[0.44]	[0.40]	[0.00]	[0.12]	[0.12]	[0.06]	[0.31]	[0.40]	[0.33]
Gaussian HMM(3)	0.45	-0.39	1.15	0.04	-0.47	0.28	0.01	-0.59	0.10
	[0.66]	[0.69]	[0.25]	[0.97]	[0.63]	[0.78]	[0.99]	[0.55]	[0.92]
Gaussian T-HMM(3)	0.51	0.79	1.44	-0.86	0.28	-0.02	-1.68	-0.73	-1.60
	[0.61]	[0.43]	[0.15]	[0.39]	[0.78]	[0.98]	[0.09]	[0.46]	[0.11]
Student's-t T-HMM(3)	0.52	0.16	1.12	0.07	0.15	0.12	-2.80	-2.28	-3.50
	[0.60]	[0.88]	[0.26]	[0.95]	[0.88]	[0.90]	[0.00]	[0.02]	[0.00]
Gaussian HMM(7)	-0.37	-0.74	-1.80	-1.01	-1.74	-1.42	-0.89	-0.81	-0.83
	[0.70]	[0.45]	[0.07]	[0.31]	[0.08]	[0.15]	[0.37]	[0.41]	[0.40]
Gaussian T-HMM(7)	0.86	**2.23**	0.60	-1.45	-0.54	-2.09	-1.39	-1.00	-1.38
	[0.39]	[0.03]	[0.55]	[0.14]	[0.59]	[0.04]	[0.16]	[0.31]	[0.17]
Student's-t T-HMM(7)	1.02	**1.95**	**2.06**	-0.34	0.82	-0.82	-1.20	-0.82	-1.13
	[0.31]	[0.05]	[0.04]	[0.72]	[0.41]	[0.40]	[0.23]	[0.41]	[0.25]

Notes: The Diebold–Mariano statistic (1995) is used to test the hypothesis of no difference between the forecast errors of a model relative to a reference model (in this case the Gaussian GARCH) against a two-sided alternative. A negative sign means that the reference model performs better than the model in the row, the opposite being true for a positive sign of the Diebold Mariano statistic. p-values in brackets. Bold figures highlight models performing better than the benchmark at a 5% level.