An Article Submitted to

Studies in Nonlinear Dynamics & Econometrics

Manuscript 1512

A Component GARCH Model with Time Varying Weights

Luc Bauwens* Giuseppe Storti†

*CORE and Université Catholique de Louvain, luc.bauwens@uclouvain.be
†University of Salerno, storti@unisa.it

Copyright ©2008 The Berkeley Electronic Press. All rights reserved.
A Component GARCH Model with Time Varying Weights*

Luc Bauwens and Giuseppe Storti

Abstract

We present a novel GARCH model that accounts for time varying, state dependent, persistence in the volatility dynamics. The proposed model generalizes the component GARCH model of Ding and Granger (1996). The volatility is modelled as a convex combination of unobserved GARCH components where the combination weights are time varying as a function of appropriately chosen state variables. In order to make inference on the model parameters, we develop a Gibbs sampling algorithm. Adopting a fully Bayesian approach allows to easily obtain medium and long term predictions of relevant risk measures such as value at risk. Finally we discuss the results of an application to a series of daily returns on the S&P500.

*Bauwens's work was supported in part by the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, MICFINMA and by a FSR grant from UCL. This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister’s Office, Science Policy Programming. The scientific responsibility is assumed by the authors.
1 Introduction

In the past two decades the empirical evidence from financial markets has shown that the pattern of response of market volatility to shocks is highly dependent on the magnitude of these shocks. In particular, in several papers (among others, Hamilton and Susmel, 1994; Lamoureux and Lastrapes, 1990; 1993), it has been shown that the persistence of the volatility process tends to decrease after extreme events such as those observed in October 1987 and September 2001. Nelson (1992) has provided a theoretical investigation of this phenomenon. Formalizing the empirical intuition, the effects on the conditional variance process of shocks occurring during turbulent periods such as market crashes are much less persistent than the effects of shocks occurring during normal periods. Since GARCH models equally weight all the shocks, this feature is expected to have relevant effects on their ability to produce accurate long term predictions of volatility and related risk measures. Furthermore, it has been documented how structural breaks in the volatility process can give rise to spurious volatility persistence if a GARCH model is fitted to the data without accounting for the breaks (Lamoureux and Lastrapes, 1990; Mikosch and Starica, 2004).

These findings suggest that there are relevant settings in which the dynamic structure of volatility cannot be adequately captured by constant parameter GARCH models. Consequently, we have assisted to a growing interest in adaptive volatility models, characterized by time varying parameters, allowing to account for both structural breaks as well as state dependence of the volatility response. One class of such models is that of smooth transition GARCH models (ST-GARCH) developed by Luukkonen et al. 1988 (see also Lubrano, 2001). These models have smoothly changing parameters. Another class is that of regime-switching GARCH (RS-GARCH) models, first proposed by Hamilton and Susmel (1994) for a purely ARCH specification, and extended by Gray (1994). These models are a valuable tool for including state dependence in the dynamics of the volatility process. However, the diffusion of these models in practical financial modelling is still limited by the severe difficulties arising in their estimation. These mainly originate from the fact that (in the generalized ARCH case) the conditional variance at a given time point \( t \) and, hence, the conditional likelihood, cannot be explicitly calculated unless the full set of regimes visited by the process at previous time points is known. A review of recent contributions on the estimation of RS-GARCH models can be found in the papers by Marcucci (2005) and Bauwens et al. (2007). In the latter, furthermore, a Bayesian approach to the estimation of a RS-GARCH model, based on Gibbs sampling, is also proposed and discussed.

In this paper we propose a modification of the standard GARCH model, which allows for time varying persistence in the volatility dynamics. Namely, a lower
degree of persistence is assigned to extreme returns taking place in highly volatile periods rather than to shocks of lower magnitude occurring in tranquil periods. However, the model structure could be easily modified to account for more general situations in which variations in the volatility persistence originate from different sources such as, for example, leverage effects and intraday or intraweek seasonal effects in volatility. It is important to note that, on an observational ground, our model is able to reproduce most of the stylized features for which RS-GARCH model have been designed but, at the same time, it is still characterized by tractable inference procedures.

We name the dynamic specification proposed in this paper the weighted-GARCH (WGARCH) model. It is a generalization of the component GARCH model of Ding and Granger (1996), henceforth DG, further studied and modified by Engle and Lee (1999), in which the weights associated to the model components are time varying and can depend on adequately chosen state variables, such as lagged values of the conditional standard deviation or squared past returns. Maheu (2005) has recently shown, by means of a simulation study, that a modification of the basic two component DG model is potentially able to reproduce long-memory properties in the autocorrelation of squared returns and, consequently, to account for long range dependence in the volatility dynamics.

Moreover, likelihood based inference for the WGARCH is readily available since the conditional log-likelihood function can be obtained in a straightforward manner by means of a standard prediction error decomposition and maximized using routine optimization algorithms. In the paper we report estimation results for the WGARCH model under the assumption of Gaussian or Student’s $t$ errors.\footnote{In the first case, in settings in which the conditional normality assumption is likely to be violated, the estimates can be interpreted as quasi maximum likelihood estimates (QMLE).} Obtaining maximum likelihood (ML) estimates of the model parameters under alternative distributional assumptions, such as the skew-$t$ distribution proposed by Bauwens and Laurent (2005), is a relatively straightforward extension.

Furthermore, despite of the computational simplicity of the ML approach, we show that resorting to Bayesian inference can offer some relevant advantages if the modeler is interested in generating long term forecasts of volatility and associated risk measures such as value at risk (VaR) (for an introductory reading see Jorion, 1997).

Accounting for time varying persistence in the volatility dynamics, WGARCH models can potentially lead to more accurate medium and long term VaR forecasts compared to standard GARCH models. If our main interest lies only in computing one step ahead predictions, these are easily obtained as a by-product of a ML estimation algorithm. Nevertheless, some complications arise when we move to
the general case in which we are interested in computing multi-step ahead VaR predictions. In this case, it becomes necessary to compute the conditional expectation of the volatility at time $T + j$ given past and present information available at time $T$ ($I^T$), which is the optimal predictor of the conditional variance given $I^T$ for a quadratic loss function. For standard GARCH models it is possible to derive an analytical expression of this conditional expectation exploiting the associated ARMA representation for squared returns (Baillie and Bollerslev, 1992). Differently, in WGARCH models this representation is non-linear and, consequently, the analytical derivation of a closed form expression for multi-step volatility predictors becomes unfeasible.

At first glance, the use of Monte Carlo simulations from the estimated model appears to be the most natural and immediate solution to this kind of problem. However, a naive Monte Carlo procedure is not able to incorporate any information on parameter uncertainty and the value of the generated predictions is highly dependent on the estimated model parameters. Differently said, parameter uncertainty is naturally dealt with if we resort to a Bayesian approach based on MCMC techniques.

From a conceptual as well as an operational point of view, Bayesian inference offers a natural framework for dealing with the problem of estimating VaR for possibly long holding periods. This topic is recently receiving much attention in the statistical and econometric literature. Among the papers which have been concerned with the generation of reliable medium and long term predictions of financial risk, one notable example is the paper by Guidolin and Timmermann (2006) and, more recently, the paper by Colacito and Engle (2007). Both these contributions focus on the analysis of the term structure relating risk measures computed for different prediction horizons. In the approach we follow in this paper, multi-step ahead prediction of VaR is accomplished by a two step procedure. First, we generate a sample from the posterior distribution of the model parameters by means of a Gibbs sampling algorithm nesting a Metropolis step, for the conditional mean parameters, and a griddy Gibbs sampler (Bauwens and Lubrano, 1998), for the conditional variance parameters. Second, we simulate from the predictive distribution of returns conditional on the sample of parameter values drawn at the first step. VaR predictions can then be computed from the simulated predictive distribution of returns.

The structure of the paper is as follows. In Section 2 the WGARCH model is proposed and discussed. Problems related to likelihood inference and volatility prediction are discussed in Section 3. In Section 4 we illustrate a Bayesian inference procedure for estimating the model parameters while the algorithm for generating VaR predictions is discussed in Section 5. In Section 6 we present the results of an application of the proposed modeling approach to daily stock returns. Namely, WGARCH models are applied to VaR prediction for a series of daily S&P500 re-
turns and the results are compared with those obtained by using Threshold GARCH (TGARCH) models (Glosten et al., 1993). Our findings suggest that the two models perform almost equally well in predicting VaR for the S&P500 series although WGARCH models allow for a clearer interpretation of volatility dynamics. Special attention is then paid to the ex-post analysis of the volatility dynamics in a period immediately following the October 1987 stock market crash usually known as the ”Black Monday”. In Section 7 we conclude.

2 The model

Let \( r_t \) be a time series of returns on a given asset and denote by \( I_t \) the set of information available at time \( t \), consisting of \( X_t \) and the returns observed up to time \( t \), \( R_t = (r_0, r_1, \ldots, r_t) \). The following equations define a conditionally heteroskedastic model for \( r_t \) allowing regime switching in the conditional variance of the process:

\[
\begin{align*}
r_t &= \beta'X_t + u_t \\
u_t &= z_t\sqrt{s_{t-d}h_{1,t} + (1 - s_{t-d})h_{2,t}} \quad t = 1, \ldots, T.
\end{align*}
\]

In the previous equation, \( s_{t-d} \) is a Bernoulli random variable, with \( d \) a positive integer being the delay needed for \( s_t \) to affect the conditional variance dynamics, \( z_t \) is an iid \( \sim (0, 1) \) sequence of random variables, and \( h_{kt} (k = 1, 2) \) is assumed to be given by the following GARCH(p,q) equations:

\[
\begin{align*}
h_{k,t} &= a_{0k} + \sum_{i=1}^{p} a_{ik}u_{t-i}^2 + \sum_{j=1}^{q} b_{jk}h_{k,t-j} \quad (k = 1, 2)
\end{align*}
\]

where \( (a_{ik}, b_{jk}) \) are constant coefficients satisfying the constraints \( a_{0k} > 0, a_{ik} \geq 0 \) and \( b_{jk} \geq 0 \), for \( i = 1, \ldots, p, j = 1, \ldots, q \) and \( k = 1, 2 \). The conditional mean of \( r_t \) is modeled as a linear function of a \( (r \times 1) \) vector of observable explanatory variables \( X_t \), with \( \beta \) being a \( (r \times 1) \) vector of unknown coefficients. This specification is general enough to cover the case of a linear autoregressive scheme with exogenous explanatory variables (ARX). Furthermore, it can be easily extended to cover an ARMA dependence structure by simply including past values of \( u_t \) in \( X_t \). From equation (2) it follows that the conditional variance of \( r_t \) given past information \( I^{t-1} \) is given by

\[
h_t = w_{t-d}h_{1,t} + (1 - w_{t-d})h_{2,t}
\]

with \( w_{t-d} = E(s_{t-d}|I^{t-1}) \). For ease of exposition, the orders of the two GARCH components \( h_{kt} (k = 1, 2) \) have been assumed to be the same. However it must
be observed that, in theory, not only different values of $p$ and $q$ could be used but the two components could be even assumed to follow different models. This for example could be useful in order to reflect the different memory properties of the market in turbulent and tranquil periods. A related model is analyzed by Bauwens et al. (2007) although the specification considered in their paper differs from ours since we allow each volatility component to depend on its own past values ($h_{k,t}$) and not on lagged values of $h_t$. The motivation for this choice is twofold. First, it allows a clear-cut economic interpretation of the volatility components and their parameters (see the discussion in Haas et al., 2004, p. 498). Second, in this way we prevent the volatility dynamics from being path dependent. The model specification in (2) admits a second order equivalent representation which can be obtained by simply replacing the definition of $u_t$ given in (2) by the following

$$u_t = z_t \sqrt{w_{t-d} h_{1,t} + (1 - w_{t-d}) h_{2,t}} = z_t h_1^{1/2}. \quad (5)$$

The property of second order equivalence means that the model (5) admits the same first two conditional moments as model (2). Nevertheless, working with equation (5) leads to a substantial simplification of the associated inference procedures. The estimation of the parameters in (5), which henceforth will be denoted as a Weighted GARCH model of order($p,q$), abbreviated WGARCH($p,q$), can be performed by ML or Gaussian QML. Furthermore, we explain in Section 4 that working with (5) implies relevant practical advantages if one is interested in generating long term forecasts of VaR within a fully Bayesian setting. Equation (4) allows to account for the state dependent features of the volatility process by modelling it as a weighted average of two components whose weights change as a function of observable state variables. A suitable choice for the weight function is the logistic function

$$w_{t-d} = \frac{1}{1 + \exp(\gamma (\delta - v_{t-d})}, \quad \gamma > 0 \quad (6)$$

where $(\gamma, \delta)$ are unknown coefficients, and $v_t$ is an appropriately chosen state variable. The positivity restriction on $\gamma$ is explained below.

As state variable we consider the conditional standard deviation $h_{1/2 t-d}$ to reflect the tendency of volatility to be less persistent in turbulent periods than in tranquil ones. Alternatively, the absolute value of the past shock $u_{t-d}$ could be used. In this respect, it is worth stressing that using the past shock ($u_{t-d}$) as state variable allows the weight $w_t$ to depend on information up to time $t - d$ while only information up to time $t - d - 1$ can be considered if we choose the conditional standard deviation as a state variable. Nevertheless we prefer to model the weights as a function of $h_{1/2 t-d}$ rather than $|u_{t-d}|$ since the conditional standard deviation gives a smoother measure of market volatility. Of course our choice does not rule out the possibility
of considering alternative state variables such as the absolute returns or, if intraday
returns are available, realized volatility measures. However the problem of deter-
mining which is the best volatility proxy to be used for computing the weights series
is not formally addressed in this paper but it is left for future research.

The proposed specification can be extended to accommodate a variety of sit-
uations in which the volatility dynamics are characterized by one or several state
dependent features. For example, by appropriately selecting the state variables in
(6), many other situations such as leverage and seasonal effects could be dealt with.
Also, in the set-up we consider here, the weight function \( w_t \) is allowed to vary at
each time point in order to reflect short-term variations in market conditions. Al-
ternatively, the weights could be assumed to be dependent on variables observed
at lower frequencies, in order to incorporate, for example, the effects of relevant
macroeconomic variables on the dynamics of financial markets volatility. On an op-
erational ground this amounts to defining a WGARCH model in which the weights
are constant over a fixed span of time. The interactions between volatility dynamics
and the relevant macroeconomic environment have been recently addressed in many
papers such as Engle et al. (2006) and Engle and Rangel (2005). Also, allowing the
weights to depend on variables observed at different frequencies creates a potential
connection between WGARCH models and mixed-data approaches in the spirit of
Ghysels et al. (2007).

The logistic function has been extensively used in the already mentioned litera-
ture on ST-GARCH models. The value of \( \gamma \) can be interpreted as determining the
speed of transition from one component to the other one: the higher \( \gamma \) (in modu-
lus), the faster the transition. The positivity constraint \( \gamma > 0 \) is an identification
restriction and has the effect of associating the first volatility component \( h_{1,t} \) with
the high volatility regime. This is evident since when \( h_{1,t}^{1/2} \) tends to \( \infty \), \( w_{t-d} \) tends to
1 leading to virtually exclude the other component \( h_{2,t} \) whose weight tends to zero.
Similarly, the weight of \( h_{1,t} \) reaches its minimum value of \( (1 + \exp(\gamma \delta))^{-1} \) when
\( h_{1,t}^{1/2} = 0 \). Also, for \( \gamma \) tending to 0, \( w_{t-d} \) tends to \( (1 + \exp(0))^{-1} \) and the WGARCH
model tends to a constant weight component GARCH model. This creates a local
identification problem involving the constants \( (a_0) \) and the ARCH coefficients \( (a_j) \)
of the volatility components, \( i = 1, 2, \), \( j = 1, \ldots, p \).

Strictly speaking the model is always identified unless \( \gamma \) is exactly equal to
zero. However, even when the value of \( \gamma \) is positive but very close to zero, the
resulting likelihood surface can be quite flat giving rise to numerical instabilities
in the estimation of parameters. If we focus on the case of a model of order (1,1),
as is the case in most financial applications, this local identifiability problem can
be easily solved by forcing \( h_{1,t} \) to follow an integrated GARCH model with no
constant term in the conditional variance equation:

\[
\begin{align*}
  h_{1,t} &= a_{11}u_{t-1}^2 + (1 - a_{11})h_{1,t-1} \\
  h_{2,t} &= a_{02} + a_{12}u_{t-1}^2 + b_{12}h_{2,t-1}.
\end{align*}
\]

(7) (8)

Differently, the value of \( \delta \) can be interpreted as the \textit{threshold} around which the switch from one regime to the other one takes place at a speed which is controlled by \( \gamma \).

3 Likelihood inference and volatility prediction

Inference for the model (1)-(5) does not suffer from the difficulties affecting other alternatives such as RS-GARCH models. In particular, the prediction error decomposition form of the log-likelihood function is easily found to be given by

\[
\ell(u; \theta) = \sum_{t=1}^{T} \log f(u_t; \eta) - \sum_{t=1}^{T} \log h_t^{1/2},
\]

(9)

where \( u = (u_1 u_2 \ldots u_T) \), \( f(\cdot; \eta) \) denotes the probability density function of the standardized error \( z_t \), which may be indexed by a scalar parameter \( \eta \), and \( \theta = (\beta'; \nu'; \omega'; \eta)' \) where \( \nu = (a_{01}, a_{11}, b_{11}, a_{02}, a_{12}, b_{12})' \) and \( \omega = (\gamma, \delta)' \) are vectors collecting the unknown parameters in the volatility components and in the weight function \( w(\cdot) \), respectively. The log-likelihood function in (9) can be maximized by resorting to standard numerical procedures.

If we are interested in volatility prediction different considerations apply to one-step and multi-step ahead volatility predictions. In particular, if the forecast horizon \( j \) is such that \( j \leq d + 1 \), \( j \)-steps ahead volatility predictions can be computed analytically and the associated predictor is given by

\[
E(h_{T+j}|I_T) = \hat{h}_{T,j} = w_{T+j-d}\hat{h}_{T,j;1} + (1 - w_{T+j-d})\hat{h}_{T,j;2}
\]

(10)

where \( \hat{h}_{T,j;i} = E(h_{i,T+j}|I_T) \), for \( i = 1, 2 \). If the volatility components are assumed to follow a GARCH process, analytical expressions for the predictors \( \hat{h}_{T,j;i} \) can be obtained using the formulas given in Baillie and Bollerslev (1992). Differently, if \( j > d + 1 \), deriving an analytical predictor becomes a more difficult task since the weights \( w_{T+j-d} \) depend on future information, leading to the following expression:

\[
\hat{h}_{T,j} = E[w_{T+j-d}h_{T+j,1}|I_T] + E[(1 - w_{T+j-d})h_{T+j,1}|I_T].
\]

(11)

So, if \( j \leq d + 1 \), the volatility predictor is a linear combination of the predictors associated with each volatility component while, for \( j > d + 1 \), this relationship becomes non-linear.
The predictor in (11) can be evaluated by Monte Carlo simulation from the estimated model. However, this approach does not account for parameter uncertainty. Alternatively, it is possible to resort to the Bayesian approach in order to obtain an estimate of the predictive density of returns which is integrated over the admissible parameter space. This issue is dealt with in the next section.

In order to assess the practical relevance of this issue, it is worth discussing the value typically assumed for the delay $d$. The value of $d$ is expected to depend on the data collection frequency. However, if the model is fitted to daily data, it is reasonable to expect relatively low values of $d$ ($d \leq 5$). Hence, the generation of medium (e.g. weekly) and long-term (e.g. monthly) predictions of volatility will in general require to compute the expectation in (11). This is a relevant problem for risk managers, since long-term volatility predictions are required for the computation of some widely used risk measures such as VaR. For example, the Basle Committee (1996) specifies a multiple of three times the 99% confidence 10-day VaR as minimum regulatory market risk capital. Also, the RiskMetrics Group (1999) suggests that the forecast horizon should reflect an institution typical holding period: "banks, brokers, and hedge funds tend to look at a 1-day to 1-week worst-case forecast, while longer-term investors, like mutual and pension funds, may consider a 1-month to 3-month time frame. Corporations may use up to an annual horizon for strategic scenario analysis".

4 Bayesian inference for WGARCH models

In this section we illustrate a procedure for drawing inferences from WGARCH models by means of the Gibbs sampler. Our motivation for Bayesian inference is twofold. First, as later discussed, if the model is used to generate predictions of volatility and other related risk measures, such as VaR, using a Bayesian approach allows to naturally deal with parameter uncertainty by incorporating it in the estimated risk measures. Second, long-term predictions of volatility and associated risk measures are generated by simulation as a by-product of the inference on model parameters independently from the complexity of the model.

The complexity of GARCH models renders the derivation of analytical Bayesian inference results an insurmountable challenge. Hence the application of simulation based techniques is required. In order to draw a sample from the posterior distribution of the full parameter vector $\theta$, we implement a Gibbs sampling algorithm with blocks given by $\theta_1 = \beta$ and $\theta_2 = (v' ; w' ; \eta)'$. The elements of $\theta_1$ coincide with the conditional mean parameters while $\theta_2$ can be partitioned into two subvectors. The first $(v' ; w')'$ includes the conditional variance parameters and can be itself partitioned into $v = (a_{01}, a_{11}, b_{11}, a_{02}, a_{12}, b_{12})'$, the parameters of the volatility
components, and \( w = (\gamma, \delta)' \), the parameters of the weight function. The second part of \( \theta_2 \) coincides with \( \eta \) which is with the shape parameter of the innovations distribution. In the framework we consider in this paper \( \eta \) is a scalar.\(^2\) In particular, we focus on two different settings: Gaussian and Student’s \( t \) innovations. In the first case the shape parameter \( \eta \) is missing while, in the second, it coincides with the degrees of freedom parameter (\( \nu \)) of the \( t \) distribution. Considering the case of \( t \)-distributed innovations is particularly interesting for our analysis since the previous literature has documented several difficulties in inference for Markov-switching GARCH models under the assumption of Student’s \( t \) errors (see e.g. Gray, 1996).

For a given prior \( \pi(\beta) \), the conditional posterior of \( \beta \) is given by

\[
\varphi(\beta|\theta_2, I_T) \propto \pi(\beta) \prod_{t=1}^{T} f(r_t|\theta, I_{t-1})
\]  

(12)

where \( f(.) \) indicates the density of \( r_t \). Since direct sampling from \( \varphi(\beta|\theta_2, Y^T) \) is not feasible, we use the Metropolis-Hastings algorithm (Hastings, 1970), choosing a \( k \)-dimensional multivariate Gaussian distribution as a proposal. The mean and variance of the proposal density, denoted by \( \iota(.) \), are set equal to the ML estimate and the inverse of the associated observed information matrix, respectively. Assuming prior independence, \( \pi(\beta) \) is factorized as the product of \( k \) uniform marginal densities. At the \((i + 1)\)th iteration, we first generate a \((r \times 1)\) pseudo-random vector from the proposal, \( \zeta^{(i+1)} \sim \iota(\zeta) \), and compute

\[
p = \min \left\{ \frac{\varphi(\zeta^{(i+1)})}{\varphi(\beta^{(i)})} \frac{\iota(\zeta^{(i)})}{\iota(\beta^{(i+1)})}, 1 \right\}.
\]

(13)

Then, we accept \( \beta^{(i+1)} = \zeta^{(i+1)} \), with probability \( p \), and take \( \beta^{(i+1)} = \beta^{(i)} \), with probability \( 1 - p \).

Similarly, given a prior density \( \pi(\theta_2) \), the conditional posterior density of the parameter vector \( \theta_2 \) is given by:

\[
\varphi(\theta_2|\beta, I_T) \propto \pi(\theta_2) \prod_{t=1}^{T} f(r_t|\theta, I_{t-1}).
\]

(14)

To simulate from the density in (14) we use the griddy Gibbs sampler. This algorithm has been originally proposed by Ritter and Tanner (1992) to solve a bivariate problem while Bauwens and Lubrano (1998) have successively applied it to the estimation of an asymmetric GARCH type model including seven parameters. Let us

\(^2\)It could be a vector if different distributional assumptions were considered. This would be the case, for example, if a skew-t distribution (Bauwens and Laurent, 2005) was considered as error distribution.
denote by $\theta^{(i)}_2$ the value of the parameter vector sampled at the $i$-th iteration. At iteration $i + 1$, a draw from the conditional posterior of $\theta_2$ is generated through the following steps:

1. Let $\theta^{(i)}_{2,-1}$ be the vector of volatility parameters excluding its first element $a_{01}$. Then, use (14) to evaluate $\kappa(a_{01}|\beta^i, \theta^{(i)}_{2,-1}, I^T)$, which is the kernel of the conditional posterior density of $a_{01}$ given $\beta$, $\theta_2$ sampled at iteration $i$, over a grid $(a_{01}, a_{01}^2, \ldots, a_{01}^G)$, where $G$ is the number of points over which the interval of $a_{01}$ is discretized. The calculated values form the vector $G_\kappa = (\kappa_1, \ldots, \kappa_G)$ with $\kappa_g$ ($g = 1, \ldots, G$) being the kernel evaluated at the $g$-th grid point.

2. By a deterministic integration rule using $M$ points, compute

$$G_\Phi = (0, \Phi_2, \ldots, \Phi_G)$$

with

$$\Phi_g = \int_{a_{01}}^{a_{01}^G} \kappa(a_{01}|\beta^i, \theta^{(i)}_{2,-1}, I^T)da_{01} \quad g = 2, \ldots, G.$$

3. Simulate $u \sim U(0, \Phi_G)$ and invert $\Phi(a_{01}|\beta^i, \theta^{(i)}_{2,-1}, I^T)$ by numerical interpolation to obtain a draw $a_{01}^{(i+1)} \sim \varphi(a_{01}|\beta^i, \theta^{(i)}_{2,-1}, I^T)$.

4. Repeat steps 1-3 for all the other parameters in $\theta_2$.

The implementation of the griddy Gibbs sampler requires the definition of suitable integration intervals for each of the elements in $\theta_2$. The definition of the integration intervals plays a critical role in the procedure and so their determination must be made carefully. The values of the integration bounds for each parameter are usually tentatively specified and adjusted in order to cover the region of the parameter space over which the posterior is relevant. The prior distributions are assumed to be uniform over the chosen intervals for all the parameters except for the coefficients of the weight function $\gamma$ and $\delta$.

Following Lubrano (2001), choosing a uniform prior for this parameter leads to a non-integrable posterior distribution. A similar problem arises for the degrees of freedom parameter $\nu$ if the errors are assumed to follow a $t$ distribution, as shown in Bauwens and Lubrano (1998). In both cases, one possible solution is to assume an uniform prior restricted to a finite interval, with the advantage that the restriction on the prior domain is automatically implied by the use of the griddy Gibbs sampler.
Bauwens and Lubrano (1998) document a very long tail for $\nu$ when a flat prior is used. So, for this parameter we adopt their suggestion to use the half-Cauchy prior

$$\pi(\nu) \propto (1 + \nu^2)^{-1}1_{\{\nu > 0\}},$$

(15)
since this choice limits the impact of the integration interval on the posterior results.

A different solution is adopted for the prior on $\gamma$ and $\delta$. For these parameters we use truncated Gaussian distributions defined as

$$\pi(\gamma) \propto \exp\left(-0.5 \times (\gamma - \mu_\gamma)^2/\sigma_\gamma^2\right) 1_{\{\gamma_{\min} \leq \gamma \leq \gamma_{\max}\}}$$

and

$$\pi(\delta) \propto \exp\left(-0.5 \times (\delta - \mu_\delta)^2/\sigma_\delta^2\right) 1_{\{\delta_{\min} \leq \delta \leq \delta_{\max}\}},$$

where

$$\mu_\gamma = (\gamma_{\min} + \gamma_{\max})/2 \quad \mu_\delta = (\delta_{\min} + \delta_{\max})/2$$

$$\sigma_\gamma = (\gamma_{\max} - \gamma_{\min})/7 \quad \sigma_\delta = (\delta_{\max} - \delta_{\min})/7.$$

The prior mean is automatically set at the center of the chosen integration interval, and the untruncated standard deviation ensures that the interval covers 3.5 standard errors on each side of the mean. Thus, the only prior hyperparameters to be chosen are the limits of the integration intervals used by the gridy Gibbs sampler. We prefer to use a truncated Gaussian prior than a uniform prior over the integration interval, because a uniform prior results in a long right tail for the parameter if its prior interval is wide, or in a strongly truncated posterior if the interval is too short.\(^3\)

5 Bayesian prediction of VaR

VaR is probably the most popular measure of market risk used by researchers and risk managers. If $R_{T+j} = \sum_{i=1}^{j} r_{T+i}$ denotes the $j$-period ($j > 0$) aggregated return, at time $T$ the $j$-period VaR at the 100$(1 - \alpha)$% confidence level, indicated as $\text{VaR}_{T,j}^{(\alpha)}$, can be defined as the order $\alpha$ quantile of the conditional distribution of $R_{T+j}$ conditional on information at the time of prediction $T$:\(^4\)

$$P(R_{T+j} < \text{VaR}_{T,j}^{(\alpha)} | I_T) = \alpha.$$
Accordingly, the one-period VaR is just the order $\alpha$ quantile of the conditional distribution of $r_{T+1}$ given $I^T$.

In the last decade we assisted to the flourishing of a rich literature concerning the estimation of VaR. Nevertheless, little attention has been paid to the development of approaches able to account for the impact of parameter uncertainty on estimated risk measures. This problem is addressed by Christoffersen and Goncalves (2005) who suggest using residual bootstrap to evaluate confidence intervals for VaR. Their approach is however limited to GARCH(1,1) processes and one-period VaR. A closely related algorithm based on residual bootstrap is applied by Pas- 
cual et al. (2006) in order to estimate predictive intervals for the volatility of a 
GARCH(1,1) process. The Bayesian approach offers a convenient framework for 
the estimation of VaR since $\text{VaR}^{(\alpha)}_{T,j}$ can be estimated as the order $\alpha$ 
quantile of the predictive density $f(R_{T+j}|I^T)$. In general, assuming that an analytical expression 
for $f(R_{T+j}|I^T, \theta)$ is available, this density can be calculated by solving the integral 

$$
\int f(R_{T+j}|I^T, \theta) \varphi(\theta|I^T) d\theta.
$$

(17)

Alternatively, assuming we can sample from the posterior $\varphi(\theta|I^T)$, e.g. using the 
Gibbs sampler described in the previous section, a sample from the density in (17) 
can be drawn by repeating $N$ times the following steps:

1. Simulate $\theta^{(i)}$ from $\varphi(\theta|I^T)$.

2. Simulate $R_{T+j}^{(i)}$ from $f(R_{T+j}|I^T, \theta^{(i)})$ by simulating successively, for $k = 1$ 
to $j$, $r_{T+k}^{(i)}$ from $f(r_{T+k}|I^T, r_{T+k-1}^{(i)}, \ldots, r_{T+1}^{(i)}, \theta^{(i)})$ and computing $R_{T+j}^{(i)}$ as 
$\sum_{k=1}^{j} r_{T+k}^{(i)}$.

Step 2 amounts to generate a trajectory of $j$ one-period future returns and aggregat-
ing them to obtain the $j$-period return.

An estimate of $VaR_{T,j}^{(\alpha)}$ is then obtained by computing the order $\alpha$ quantile of 
the sample of size $N$ drawn from $f(R_{T+j}|I^T, \theta)$. So estimates of VaR for relatively 
long holding periods are obtained in a straightforward manner as a by-product of the 
Gibbs sampling algorithm. The same method could be used to estimate the expected 
shortfall. This is a risk measure closely related to VaR, but, differently from VaR, 
it satisfies the subadditivity property. Hence it can be considered a coherent 
risk measure (in the sense of Artzner et al., 1999). This way of computing VaR has two 
important advantages. First, it naturally accounts for parameter uncertainty since 
the predictive densities of returns are integrated over the parameter space. Second, 
long term predictions of volatility and related risk measures can be obtained by 
simulation independently of the model complexity.
6 An application to daily S&P500 returns

6.1 Model estimation

In this section we present an application to a time series of daily (percentage) log-returns on the S&P500 stock market index. The data used for the analysis can be freely downloaded from the website http://finance.yahoo.com. The observation period goes from January 5, 1971 to January 24, 2008 for a total of 9353 observations (figure 1). The data are characterized by a strong kurtosis and a remarkable negative skewness (table 1). The value of the Ljung-Box Q-statistics provides evidence in favour of the presence of a significant autocorrelation structure in the returns as well as in the squared returns series. Moreover, visual inspection of the sample correlogram of the squared returns reveals a highly persistent autocorrelation pattern (figure 2). In order to account for these features we model the returns \( r_t \) by a first order autoregressive model whose error terms follow a WGARCH model of order (1,1). The complete model is written here for ease of reference:

\[
\begin{align*}
    r_t &= \phi_0 + \phi_1 r_{t-1} + u_t, \\
    u_t &= h_t^{1/2} z_t, \\
    h_t &= w_{t-d} h_{1,t} + (1 - w_{1-d}) h_{2,t}, \\
    h_{k,t} &= a_{0k} + a_{1k} u_{t-1}^2 + b_{1k} h_{k,t-1}, & k = 1, 2, \\
    w_{t-d} &= 1 / \left[ 1 + \exp(\gamma(\delta - h_{t-d}^{1/2})) \right].
\end{align*}
\]

We assume that \( z_t \) belongs to an iid sequence of random variables with zero mean and unitary variance, the distribution being either Gaussian or \( t \) with \( \nu \) degrees of freedom (with \( \nu > 2 \) to ensure the existence of the variance). The corresponding models are called the Gaussian and the \( t \) model in the sequel (in brief, \( WG_n \) and \( WG_t \)). Before estimating these models, we need to fix the value of the delay parameter \( d \). We choose the value that maximizes the log-likelihood value over the range \( 1 \leq d \leq 5 \). The optimal value turned out to be \( d = 1 \) for the Gaussian model, and \( d = 2 \) for the \( t \) model.

Table 1: Descriptive and Ljung-Box Q-statistics for the S&P 500 returns (\( Q_r(.) \)) and squared returns (\( Q_{r^2}(.) \))

<table>
<thead>
<tr>
<th>mean</th>
<th>s.d.</th>
<th>min.</th>
<th>max.</th>
<th>skew.</th>
<th>kur.</th>
<th>( Q_r(10) )</th>
<th>( Q_{r^2}(10) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0288</td>
<td>0.9924</td>
<td>-22.8997</td>
<td>8.7089</td>
<td>-1.4108</td>
<td>37.4163</td>
<td>44.928</td>
<td>675.38</td>
</tr>
</tbody>
</table>
Figure 1: Daily returns of the S&P 500 index from January 5, 1971 to January 24, 2008.

The series has been divided into two subseries respectively including observations from January 5, 1971 to September 7, 2000, for a total of 7500 data points, and from September 8, 2000 to the end of the observation period. The first sub-sample has been used for model estimation while the second one has been kept for out of sample forecast evaluation. In order to estimate the posterior density of the parameters of the Gaussian and $t$ models defined above, we have implemented the Gibbs sampler described in section 4.

As a proposal for the Metropolis-Hastings step of the algorithm we have chosen a bivariate Gaussian distribution. The covariance matrix estimated by the observed information matrix has been multiplied by an inflation factor equal to 1.3 to allow for tails heavier than those implied by the asymptotic distribution of the ML estimator. This is a standard practice in the type of MH algorithm we use. The rationale behind this is to fine tune the proposal density by monitoring the acceptance rate of the sampler for different values of the inflation factor (see Bauwens et al., 1999, page 90). The mean and covariance matrix of the proposal density for $\beta = (\phi_0, \phi_1)'$
The posterior means and standard deviations are reported in table 3 (see the columns with $WG_n$ and $WG_t$ headers). In both models, the estimated intercept of the first component ($a_{01}$), associated with turbulent periods, is much higher than that of the second component ($a_{02}$), associated with tranquil periods. This finding
Table 2: Summary of prior distributions with the associated integration intervals for the Gaussian WGARCH ($WG_n$) and $t$-WGARCH ($WG_t$) models.

<table>
<thead>
<tr>
<th>coeff.</th>
<th>prior</th>
<th>$WG_n$ lower</th>
<th>$WG_n$ upper</th>
<th>$WG_t$ lower</th>
<th>$WG_t$ upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{01}$</td>
<td>uniform</td>
<td>0.01</td>
<td>1.50</td>
<td>uniform</td>
<td>0.01</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>uniform</td>
<td>0.10</td>
<td>0.95</td>
<td>uniform</td>
<td>0.01</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>uniform</td>
<td>0.01</td>
<td>0.92</td>
<td>uniform</td>
<td>0.30</td>
</tr>
<tr>
<td>$a_{02}$</td>
<td>uniform</td>
<td>0.0001</td>
<td>0.014</td>
<td>uniform</td>
<td>0.0005</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>uniform</td>
<td>0.003</td>
<td>0.089</td>
<td>uniform</td>
<td>0.018</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>uniform</td>
<td>0.91</td>
<td>0.99</td>
<td>uniform</td>
<td>0.91</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>trunc. norm.</td>
<td>0.90</td>
<td>1.67</td>
<td>trunc. norm.</td>
<td>0.65</td>
</tr>
<tr>
<td>$\delta$</td>
<td>trunc. norm.</td>
<td>0.60</td>
<td>19</td>
<td>trunc. norm.</td>
<td>1</td>
</tr>
<tr>
<td>$\nu$</td>
<td>half Cauchy</td>
<td>5</td>
<td>11.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

is in line with the previous literature since several papers (see e.g. Lamoureux and Lastrapes, 1990) have documented how the value of the unconditional volatility of the returns tends to switch to a higher level in turbulent periods. The positive sign of the estimated difference of $(a_{01} - a_{02})$ is obtained without imposing this restriction at the estimation stage. However, it is indirectly a consequence of the identification restriction $\gamma > 0$, imposed in order to avoid label switching. We remind that this restriction determines the association of the first component to turbulent periods.

Moreover, the estimated persistence of the first volatility component (shown in figures 3a and 4a as conditional standard deviations), measured in terms of the sum $(a_{11} + b_{11})$, is substantially lower than the corresponding estimate obtained for the second component (shown in figures 3b and 4b). Notice also the much larger sensitivity of the volatility component to the lagged squared shock in the first component than in the second one ($a_{11} > a_{12}$), and the smaller sensitivity to the lagged volatility component ($b_{11} < b_{12}$). So it follows that in turbulent periods the volatility tends to be less persistent and more sensitive to recent shocks, that is to say, to be mainly driven by short run factors, than in tranquil periods. Actually, the configuration of the parameter estimates of the second component is very similar to what one finds with a simple GARCH model for daily data.

Figures 3c (Gaussian model) and 4c ($t$ model) report the time plots of the weights $w_t$ evaluated at the posterior means. In both cases, during highly volatile periods the first component virtually excludes the other while the opposite happens as the market moves back to a tranquil state. The switch between high and low volatility periods, however, is differently characterized within the Gaussian and $t$ model.
Table 3: Posterior means and standard deviations (in brackets) for the Gaussian WGARCH, $t$-WGARCH, Gaussian TGARCH and $t$-TGARCH models.

<table>
<thead>
<tr>
<th></th>
<th>$WG_n$</th>
<th>$WG_t$</th>
<th>$TG_n$</th>
<th>$TG_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_0$</td>
<td>0.0401</td>
<td>0.0411</td>
<td>0.0260</td>
<td>0.0335</td>
</tr>
<tr>
<td></td>
<td>(0.0088)</td>
<td>(0.0084)</td>
<td>(0.0095)</td>
<td>(0.0086)</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.1134</td>
<td>0.1011</td>
<td>0.1191</td>
<td>0.1029</td>
</tr>
<tr>
<td></td>
<td>(0.01260)</td>
<td>(0.0117)</td>
<td>(0.0127)</td>
<td>(0.0118)</td>
</tr>
<tr>
<td>$a_{01}$</td>
<td>0.5637</td>
<td>0.2363</td>
<td>0.0114</td>
<td>0.0063</td>
</tr>
<tr>
<td></td>
<td>(0.2284)</td>
<td>(0.1741)</td>
<td>(0.0035)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.4627</td>
<td>0.0690</td>
<td>0.0274</td>
<td>0.0216</td>
</tr>
<tr>
<td></td>
<td>(0.1148)</td>
<td>(0.0324)</td>
<td>(0.0080)</td>
<td>(0.0055)</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.3613</td>
<td>0.7252</td>
<td>0.9238</td>
<td>0.9347</td>
</tr>
<tr>
<td></td>
<td>(0.1011)</td>
<td>(0.1145)</td>
<td>(0.0126)</td>
<td>(0.0097)</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.0758</td>
<td>0.0404</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0140)</td>
<td>(0.0093)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{02}$</td>
<td>0.0055</td>
<td>0.0036</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.0432</td>
<td>0.0357</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0092)</td>
<td>(0.0061)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>0.9499</td>
<td>0.9447</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0101)</td>
<td>(0.0087)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>9.369</td>
<td>7.413</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.044)</td>
<td>(2.514)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>1.392</td>
<td>1.252</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0674)</td>
<td>(0.1270)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>7.797</td>
<td>8.057</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.7718)</td>
<td>(0.8809)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WGARCH: see equations (18)-(22).
TGARCH: equations (18)-(19)-(23).

models. In particular, in the $t$ model the estimate of the slope parameter $\gamma$ turns out to be lower than in the Gaussian model, implying a substantial reduction of the speed of transition from the low to the high volatility regime. This also explains why, for the $t$ model, the time series of estimated weights appears smoother than its counterpart for the Gaussian model.

The estimated posterior marginal densities of the volatility model parameters computed by the griddy Gibbs sampler are reproduced in figures 5 and 6. It is worth noting how for both models the mode of the marginal posterior of the slope coefficient of the logistic function $\gamma$ is located far away from zero, providing further clear evidence in favour of the time varying weights model.
Figure 3: (a) First volatility component ($\sqrt{h_{1t}}$), (b) Second volatility component ($\sqrt{h_{2t}}$), and (c) Weight series ($w_t$) from the Gaussian WGARCH model.

### 6.2 VaR prediction

The estimated WGARCH models have been applied for predicting VaR at different horizons ($j=1,5,10,15,20$ days) and for different confidence levels ($\alpha=0.10, 0.05, 0.025, 0.01$). Following Giot and Laurent (2003), long as well as short trading positions have been considered in order to assess the ability of the estimated models to properly characterize both tails of the predictive distribution of returns.

VaR predictions have been generated using the simulation based procedure described in section 5. For each time point in the out of sample validation period and for each valid draw from the posterior of $\theta$, a trajectory of future returns, up to 20 days in the future, is drawn from the predictive density of returns. So, given that
Figure 4: (a) First volatility component ($\sqrt{h_1t}$), (b) Second volatility component ($\sqrt{h_2t}$), and (c) Weight series ($w_t$) from the $t$-WGARCH model.

the chosen validation period includes 1853 observations, our predictions of VaR are based on 1853 samples each of which is made up of 7500 trajectories from the predictive distribution of returns. Each trajectory includes a sequence of 20 data points. Nevertheless, the required computing time is still reasonably low. For the Gaussian model, for example, the generation of the full set of predictive densities required for VaR computation, over the whole validation period, took about 66 minutes on a laptop computer. This means that, conditional on the estimated model, only a few seconds are required for computing a single trajectory of VaR forecasts.

All the computations necessary for producing the results in this paper were performed in GAUSS. This means that, if needed, computing times could be further reduced by the use of C.
Figure 5: Estimated posterior densities for the Gaussian WGARCH model. From left to right and from top to bottom: $a_{01}$, $a_{11}$, $b_{11}$, $a_{02}$, $a_{12}$, $b_{12}$, $\delta$, $\gamma$. 
Figure 6: Estimated posterior densities for the $t$ WGARCH model. From left to right and from top to bottom: $a_{01}$, $a_{11}$, $b_{11}$, $a_{02}$, $a_{12}$, $b_{12}$, $\delta$, $\gamma$, $\nu$. 

Published by The Berkeley Electronic Press, 2008
The performance of the estimated Gaussian and $t$ AR(1)-WGARCH(1,1) models has been compared to those of AR(1) models with TGARCH(1,1) errors, again assuming either a Gaussian distribution or a $t$ one. The conditional variance equation of a TGARCH(1,1) model is given by

$$h_t = a_0 + (a_1 + c_1 1_{u_{t-1} < 0}) u_{t-1}^2 + b_1 h_{t-1}. \quad (23)$$

This specification was introduced by Glosten et al. (1993). It allows for a state-dependent response of volatility to past shocks although this aim is pursued by a different strategy. First, in a WGARCH model, the overall volatility is given by a convex linear combination of two separate GARCH models while, in a TGARCH model, the volatility process can only jump from one regime to the other. Moreover the two regimes are not completely separated since they only differ for the value of the ARCH coefficient and both depend on the same $h_{t-1}$ with the same impact coefficient ($b_1$). Second, in a TGARCH model and, in general, in any asymmetric GARCH model, the switch from one regime to the other is driven by the sign of past innovations rather than by the level of past volatility measured in terms of some proxy. The erratic nature of the threshold variable leads to frequent regime switches evenly distributed over the period of observation. Finally, it is widely recognized and confirmed by empirical observation that volatility tends to be negatively correlated with the sign of returns. In TGARCH models this still leads to loosely associate one regime to high volatility periods and the other one to low volatility periods as in WGARCH models. However, it is interesting to note that in asymmetric GARCH models such as the TGARCH (with $c_1$ positive), a higher persistence is associated with periods following negative returns rather than positive ones, that is to high volatility periods than to low volatility periods. This important difference with respect to WGARCH models is due to the fact that the main aim of asymmetric GARCH models is to capture the leverage effect in the volatility dynamics rather than to provide a general approach to modelling state dependent features in its persistence.

Inference on the AR(1)-TGARCH(1,1) model parameters has been conducted by the same Bayesian approach as described in section 4. The estimated posterior means and standard deviations, reported in table 3 (see columns headed by $TG_n$ and $TG_t$), show the usual pattern of a higher sensitivity of volatility to the lagged squared shock when it is negative than when it is positive.

In order to assess the quality of one-day ahead VaR predictions we have used the likelihood ratio statistic proposed by Kupiec (1995) for testing the hypothesis of correct unconditional coverage. Also in order to detect any serial dependence in

\textit{C ++} or any other more efficient programming language.
the series of VaR violations we have tested their independence using the likelihood ratio test statistic proposed by Christoffersen (1998).

By means of a binomial likelihood, the number of observed VaR violations is compared to the expected number of VaR violations under the hypothesis of correct coverage. The likelihood ratio test statistic is given by

\[
LR_U = -2 \log \left( \frac{p_0^{n_e} (1 - p_0)^{(T-n_e)}}{\hat{p}^{n_e} (1 - \hat{p})^{(T-n_e)}} \right)
\]

where \(T\) is the number of observations used for backtesting (1853 in our application), \(n_e = \sum_{t=1}^{T} L_t\), with \(L_t = 1_{\{VaR_t < r_t\}}\) and \(VaR_t = VaR_{t-1,1}\), \(\hat{p} = n_e/T\) is the observed proportion of VaR violations, \(p_0\) is the desired probability of VaR violations. Under the null, \(p = p_0\), which is the expected proportion of VaR violations, \(p = E(\hat{p}) = P(1_{\{VaR_t < r_t\}} = 1)\),

is equal to its target value \(p_0\), while, under the alternative \(p \neq p_0\). It can be shown that, as \(T \to \infty\), \(LR_U \xrightarrow{d} \chi^2_1\) under the null.

The test for independence of VaR violations (Christoffersen, 1998) is based on the statistic

\[
LR_I = -2 \log \left( \frac{\hat{p}^{n_{01}+n_{11}} (1 - \hat{p})^{n_{00}+n_{10}}}{\hat{p}_{01} (1 - \hat{p}_{01})^{n_{00}} \hat{p}_{11} (1 - \hat{p}_{11})^{n_{10}}} \right)
\]

where \(n_{ij}\) denotes the number of time points in which \(L_t = j\) \((j = 0, 1)\) has been observed following \(L_{t-1} = i\) \((i = 0, 1)\), \(\hat{p}_{ij} = \frac{n_{ij}}{n_{i+j}}, \) with \(t = 1, \ldots, T\). Under the null, \(L_t\) is an independent binary process with

\(p_{01} = p_{11} = p\)

while, under the alternative \(p_{01} \neq p_{11}\) which implies that \(L_t\) follows a first order Markovian process. Under the null, as \(T \to \infty\), \(LR_I \xrightarrow{d} \chi^2_1\). A test for correct conditional coverage can be then defined as \(LR_C = LR_U + LR_I\) with \(LR_C \xrightarrow{d} \chi^2_2\), as \(T \to \infty\). We use \(LR_C\) to perform a joint test for correct unconditional coverage and independence.

A different strategy is adopted for backtesting multi-step VaR forecasts. Since rolling VaR forecasts are considered, the resulting series of VaR violations are naturally affected by serial dependence. In particular, optimal \(j\)-steps ahead forecasts are expected to be at most \((j - 1)\)-dependent. To overcome this problem we adopt a slightly modified version of the Dynamic Quantile (DQ) test (Engle and Manganelli, 2004) in which the set of instruments used to assess the quality of forecasts

Published by The Berkeley Electronic Press, 2008
of \( VaR_{T,j}^{(\alpha)} \) is composed of past returns from \( T \) to \( (T - \lambda + 1) \) \( (\lambda > 0) \) plus a constant term and, of course, the forecast \( VaR_{T,j}^{(\alpha)} \). Under the null of a correctly specified \( VaR \) model, the out-of-sample DQ test statistic is asymptotically distributed as a \( \chi^2_{\lambda+2} \) random variable.

The test results are presented in Table 4, for one-day ahead \( VaR \) prediction, and in Table 5, for longer term predictions. For all the tests the null is rejected at the significance level \( \alpha \) for \( S > \chi^2_{n,1-\alpha} \), where \( \chi^2_{n,1-\alpha} \) is the \( (1 - \alpha) \) quantile of a \( \chi^2_n \) random variable and \( S \) is the value of the relevant test statistic. The \( p \)-value is defined as the probability \( P(\chi^2_n > S) \) which means that, in the following tables, a \( p \)-value smaller than \( \alpha \) implies a rejection of the associated null hypothesis.

For long trading positions, considering a one-day holding period, the models with Gaussian errors tend to perform better than their counterparts with \( t \) errors. A practical way of summarizing the performance of competing models is to count

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( WG_n )</th>
<th>( WG_t )</th>
<th>( TG_n )</th>
<th>( TG_t )</th>
<th>( WG_n )</th>
<th>( WG_t )</th>
<th>( TG_n )</th>
<th>( TG_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.1096</td>
<td>0.1306</td>
<td>0.1090</td>
<td>0.1257</td>
<td>0.0885</td>
<td>0.1025</td>
<td>0.0901</td>
<td>0.1020</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0566</td>
<td>0.0626</td>
<td>0.0551</td>
<td>0.0631</td>
<td>0.0443</td>
<td>0.0491</td>
<td>0.0421</td>
<td>0.0437</td>
</tr>
<tr>
<td>0.025</td>
<td>0.0324</td>
<td>0.0308</td>
<td>0.0297</td>
<td>0.0302</td>
<td>0.0264</td>
<td>0.0259</td>
<td>0.0238</td>
<td>0.0227</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0146</td>
<td>0.0092</td>
<td>0.0146</td>
<td>0.0092</td>
<td>0.0103</td>
<td>0.0070</td>
<td>0.0081</td>
<td>0.0053</td>
</tr>
</tbody>
</table>

For a given \( \alpha \) and model, the values are, from top to down: the out-of-sample empirical coverage \( \hat{p} \), the \( p \)-value of the corresponding \( LR_U \) (Kupiec) test (in brackets), the \( LR_T \) test statistic for independence (and its \( p \)-value), the \( LR_C \) test statistic for correct conditional coverage (and its \( p \)-value). See Table 3 for model estimates and definitions.
the overall number of test failures at the usual 0.05 significance level, that is, for each model and confidence level, we count the number of $p$-values lower than 0.05. For one-day VaR prediction (table 4) the Gaussian WGARCH and TGARCH models perform almost equally (0 failure for WGARCH, 1 for TGARCH). For longer holding periods (table 5), the Gaussian WGARCH (1 failure) and the Student’s $t$ TGARCH model (0 failures) result to be the best models. Again no clear winner emerges between WGARCH and TGARCH models.

The picture changes if we move to consider the performance in estimating VaR for short positions. For the one-day ahead case, the WGARCH model with $t$ errors is performing slightly better than the other competitors with 3 failures against 4 for the Gaussian TGARCH and WGARCH models, and 5 for the TGARCH model with $t$ errors.

For holding periods greater than one day, all the model specifications considered are performing poorly. Again, none of the models considered is performing strikingly better than the others.

6.3 Focus on the 1987 crash

In this section we compare the WGARCH and TGARCH model estimates of the volatility of the $S&P500$ daily returns focusing on a sub-sample including the October 1987 stock market crash. The period we consider includes 146 trading days, from September 16, 1987 till April 13, 1988. In figure 7 we graphically compare the volatilities estimated by WGARCH and TGARCH models over this period considering the Gaussian (panel a) as well as the $t$ case (panel b). In both cases it is evident that TGARCH models associate much more persistence to the October 19 shock than WGARCH models do. It is also interesting to notice that this is not a temporary effect, since the plots show that the gap in volatility level between the two models lasts for approximately four months.

A look at the series of component weights estimated by WGARCH models (figure 8) gives further insight on the way this persistence gap is generated. In both models, Gaussian and $t$, the weight of the high volatility component ($h_{1,t}$) suddenly increases and remains close to one for approximately four months. This means that, over this period, the second volatility component ($h_{2,t}$), which is the one associated to tranquil periods, is virtually excluded from the estimation of the overall volatility ($h_t$), which is dominated by the turbulent component. The situation is almost fully inverted at the end of March. The weight of the first component decreases toward zero which means that, after this date, the effect of $h_{1,t}$ becomes negligible and the market volatility dynamics are almost completely determined by $h_{2,t}$. These findings are in line with the previous literature, see e.g. Hamilton and Susmel (1994).
Table 5: VaR evaluation tests for holding periods of several days. From top to bottom: $j = 5, 10, 15, 20$ days.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$WG_5$</th>
<th>$WG_{10}$</th>
<th>$TG_5$</th>
<th>$TG_{10}$</th>
<th>$TG_5$</th>
<th>$TG_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.0876</td>
<td>0.0990</td>
<td>0.0887</td>
<td>0.0941</td>
<td>0.0368</td>
<td>0.0433</td>
</tr>
<tr>
<td></td>
<td>(0.3171)</td>
<td>(0.8488)</td>
<td>(0.3804)</td>
<td>(0.4875)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0433</td>
<td>0.0487</td>
<td>0.0400</td>
<td>0.0476</td>
<td>0.0081</td>
<td>0.0114</td>
</tr>
<tr>
<td></td>
<td>(0.7753)</td>
<td>(0.8192)</td>
<td>(0.2938)</td>
<td>(0.6706)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>0.025</td>
<td>0.0211</td>
<td>0.0238</td>
<td>0.0211</td>
<td>0.0254</td>
<td>0.0022</td>
<td>0.0049</td>
</tr>
<tr>
<td></td>
<td>(0.7906)</td>
<td>(0.9552)</td>
<td>(0.5415)</td>
<td>(0.8027)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0114</td>
<td>0.0114</td>
<td>0.0114</td>
<td>0.0130</td>
<td>0.0005</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>(0.8808)</td>
<td>(0.8893)</td>
<td>(0.9998)</td>
<td>(0.9984)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$WG_5$</th>
<th>$WG_{10}$</th>
<th>$TG_5$</th>
<th>$TG_{10}$</th>
<th>$WG_5$</th>
<th>$WG_{10}$</th>
<th>$TG_5$</th>
<th>$TG_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.0895</td>
<td>0.0992</td>
<td>0.0803</td>
<td>0.0927</td>
<td>0.0190</td>
<td>0.0266</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0150)</td>
<td>(0.0134)</td>
<td>(0.0045)</td>
<td>(0.0055)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0.0385</td>
<td>0.0456</td>
<td>0.0342</td>
<td>0.0423</td>
<td>0.0043</td>
<td>0.0087</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2710)</td>
<td>(0.8182)</td>
<td>(0.0037)</td>
<td>(0.3569)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>0.0217</td>
<td>0.0233</td>
<td>0.0190</td>
<td>0.0233</td>
<td>0.0011</td>
<td>0.0038</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.7430)</td>
<td>(0.8418)</td>
<td>(0.3543)</td>
<td>(0.3841)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.0098</td>
<td>0.0114</td>
<td>0.0109</td>
<td>0.0130</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.8634)</td>
<td>(0.9959)</td>
<td>(0.9999)</td>
<td>(0.2743)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For a given $\alpha$ and model, the values are the out-of-sample empirical coverage $\hat{p}$ and (between brackets) the $p$-value of the corresponding modified $DQ$ test statistic with $\lambda = 5$. See Table 3 for model estimates and definitions.
Figure 7: Period from 16/9/1987 to 13/4/1988: (a) Volatilities estimated from Gaussian WGARCH and TGARCH models; (b) Volatilities estimated from $t$ WGARCH and TGARCH models; (c) S&P500 log-returns.
Figure 8: Component weights ($w_t$) estimated by Gaussian (blue) and $t$ (red) WGARCH models over the period 16/9/1987-13/4/1988.

7 Concluding remarks

The WGARCH model discussed in this paper allows to capture state dependent volatility dynamics without suffering from the limitations which typically complicate inference from other alternatives such as RS-GARCH models.

To investigate the effectiveness of WGARCH models in standard risk management applications, we have presented an application of the proposed modelling strategy to the prediction of VaR for a time series of daily returns on the S&P 500 index. The empirical results show that, in general, WGARCH models are as accurate as Threshold GARCH models in predicting VaR for different holding periods. The unsatisfactory results in predicting VaR for short positions suggest that it could be of interest to consider skewed error distributions such as the skew-$t$ distribution investigated by Bauwens and Laurent (2005). At the same time, it has been shown that WGARCH models allow to better characterize the dynamical properties of stock market volatility in highly turbulent market periods, such as that occurr-
ring after the October 1987 crash. In particular, the volatility forecasts generated by WGARCH models appear to be much more consistent with the well consolidated stylized fact stating that market volatility is expected to be less persistent in turbulent periods rather than in tranquil ones.

Finally, our research has given rise to some open issues that we leave for future work. First, it would be useful to investigate the statistical properties of WGARCH models, such as its moments and volatility autocorrelation structure. Also, we believe that it could be of interest exploring the possibility of generalizing the WGARCH model structure to allow for different error distributions within the two components. Last but not least, WGARCH models could be used to gain further insight on the relationship between stock market volatility and economic factors, allowing for example the weights to incorporate information on the macro-economic environment.

References


