Threshold volatility models: forecasting performance

Márquez M.D.[1], Muñoz M.P.[2], Martí-Recober M.[2], and Acosta L.M.[2]

[1] Department of Business Economics, Univ. Autònoma de Barcelona, Barcelona, Spain, mariadolores.marquez@uab.es
[2] Department of Statistics and Operational Research, Univ. Politècnica de Catalunya, Barcelona, Spain,
{pilar.munyoz,manuel.marti-recober,lesly.acosta}@upc.edu

Summary. The aim of this paper is to compare the forecasting performance of competing volatility models, in order to capture the asymmetric effect in the volatility. We focus on examining the relative out-of-sample forecasting ability of the models (SETAR-TGARCH and SETAR-THSV), which contain the introduction of regimes based on thresholds in the mean equation and volatility equation, compared to the GARCH model and SV model. For each model, we consider two cases: Gaussian and *t-Student* measurement noise distribution. An important problem when evaluating the predictive ability of volatility models is that the "true" underlying process is not observable and thus a proxy must be defined for the unobservable volatility. To attain our proposal, the proxy volatility measure and the loss function must also be decided to ensure a correct ranking of models.

Our empirical application suggests the following results: when time series include leverage effects on the mean, the introduction of threshold in the mean and variance equations produces more accurate predictions. If the leverage in the mean is not important, then the SVt is flexible enough to beat the threshold models.

Key words: Volatility, Threshold Models, Bayesian estimation

1 Aims and Background

Volatility is understood as a measure of the random component of a return time series. Modelling and forecasting volatility is crucial in financial markets (option pricing, risk management and portfolio management). Models that try to explain the volatility of return series have become exceedingly complex in recent years. However, volatility models are not merely theoretical constructs; the success of these models depend on their capability to capture and predict the features of empirical data. Many authors now focus their interest on the study of volatility forecasting (see Patton [Pat05] and the references therein).

Threshold models [Ton90] are an interesting alternative for modelling both returns and volatilities. The fundamental idea behind these models is the introduction of regimes based on thresholds, thus allowing the analysis of complex stochastic systems from simple subsystems. In this group, the self-exciting threshold autoregressive model (SETAR) is defined on the basis of an autoregressive (AR) process in each regimen, and the variable that governs the switching regime is the delayed process variable. Cao and Tsay [CT92] showed that this new kind of models provided better forecasts than the previous family of conditional heteroscedastic models.

Li and Li [LL96] and Chan [Cha05] propose to combine SETAR models with Threshold GARCH (SETAR-TGARCH) to explain nonlinearity in terms of the mean and variance. So et al. [SLL02] combine SETAR models with Threshold SV model[3]. These authors also study the forecasting performance of those models.

The aim of this paper is to compare the forecasting performance of competing models. We focus on examining the relative forecasting ability of the SETAR-TGARCH model and the SETAR-THSV model compared to the GARCH model and SV model, with the main problem consisting of evaluating the predictive ability of volatility models being that the "true" underlying volatility process is not observable and a proxy must be defined for the unobservable volatility. To attain our proposal, the proxy volatility measure and the loss function must also be decided to ensure a correct ranking of models.

This paper is organized as follows: Section 2 presents the SETAR-TGARCH and SETAR-THSV models. In Section 3, we discuss the choice of the volatility proxy and loss function. Section 4 focuses on empirical findings, and finally in Section 5 we present the concluding remarks.

2 Models

The conditional heteroscedastic models (GARCH (1,1), SV) are the non-linear time series models most commonly used in the finance literature. Generally, these two models cannot capture the asymmetry of volatility [ANV99]. This characteristic, known as the leverage effect, is related to the asymmetric behavior of the market, in the sense that it is more volatile after a continuous decrease in prices than after a rise (both of the same magnitude). The non-linear SETAR models, however, allow the asymmetries in the mean (non-linearity in the mean of the returns) to be captured. The former models belong to the class of "first-generation models".

To capture the asymmetric effect in the volatility, the introduction of thresholds in the volatility equation has been proposed, thus obtaining the Threshold GARCH (TGARCH) model or Threshold Stochastic Volatility (THSV) model. If we try to explain both kinds of non-linearity, in terms of the mean and in variance, the SETAR model is combined with the previous ones, obtaining the SETAR-TGARCH model and the SETAR-THSV model, which are known as "second-generation models" (see Table 3).

These second-generation models capture the main features of volatility. The difference is that the SETAR-THSV is more flexible than the SETAR-TGARCH to capture kurtosis though the estimation procedure is computationally more costly; see Appendix A for a complete model formulation.

[3]SV: Stochastic Volatility; THSV: Threshold Stochastic Volatility; TGARCH: Threshold GARCH.

2.1 Parameters estimation

In this work, the identification and estimation of the SETAR model was done using the methodology proposed by Tsay [Tsa89] and the algorithm designed by Márquez (2002) [Mar02] that improves the identification and estimation processes. This algorithm was developed in Fortran 77 and the estimation is based on the conditional least-squares whereas the model order selection is done automatically minimizing the AIC (Akaike Information Criteria). The parameters estimation for the GARCH, TGARCH and SETAR-TGARCH models was based on maximizing the conditional likelihood function with BFGS algorithm.

The SV model and other models based on it were formulated into a state-space representation. To estimate the unknown parameters we choose to resort to Bayesian estimation because the likelihood function cannot be expressed in closed form. An usual approach is to introduce the unknown parameters as part of the state vector $(x_t, \theta)'$ and then to obtain the a posteriori PDF $p(x_t, \theta|D_t)$[4] as a new observation y_t arrives. Herein we estimate this a posteriori PDF using our modified version of the SIR (Sampling Importance Resampling) filter described in Muñoz et al. [MMM04]; this procedure avoids the use of optimization algorithms. We developed a specific code in the R language (`http://cran.r-project.org`) for the models exposed in Appendix A.

For all models, a simulation study was carried out using as true parameters values those already published in the literature. This is done in order to assess the accuracy and stability of our implemented algorithms and thus guarantying the quality of our results.

3 Comparison of different volatility models

In this section, we focus on the procedure for comparing conditional volatility forecasts, based on three steps:

Step 1. First of all, as volatility is an unobservable characteristic of the return time series, we need to choose a proxy for it. Several proxies are used in the financial literature as a proxy of σ_t^2, for instance: absolute value of returns $|r_t|$, realized volatility (which is the sum of squared intradaily returns), and squared returns r_t^2. All of them exhibit advantages and disadvantages [AB98, HL06, Pat05, Poo05]. We choose to use one of the volatility proxies proposed in the Risk Metrics Manual 1995, the so called *rolling daily volatility* (RDV). The RDV is defined as a moving window obtained from the standard deviation of the returns:

$$\hat{\sigma}_t^{RDV} = \sqrt{\frac{1}{k}\left[\sum_{i=t-(k-1)/2}^{i=t+(k-1)/2} (r_i - \overline{r}_k)^2\right]}, \tag{1}$$

where k is the lag length of the rolling window, in days, and $\overline{r}_k$ is the sample mean of the k returns used.

[4] D_t means the set of all the observations until time t, $D_t = \{y_1, ..., y_t\}$

Step 2. Second, a loss function must be chosen among those used by several authors, in order to compare the goodness-of-fit of the one-step-ahead forecast obtained from each one of the volatility models used in this paper to the *rolling daily volatility* defined by Eq. (1).

We propose to use the following quadratic function:

$$L\left(\hat{\sigma}_{t+1}^{RDV},\hat{\sigma}_{t+1/t}\right)=\left((\hat{\sigma}_{t+1}^{RDV})^2-(\hat{\sigma}_{t+1/t})^2\right)^2, \tag{2}$$

where $\hat{\sigma}_{t+1/t}$ is the one-step-ahead forecast obtained from the volatility model.

Step 3. Finally, the out-of-sample predictive accuracy of the different models is statistically tested using the Diebold and Mariano's sign test [DM95], used for comparing forecast errors of different models. Following Poon [Poo05], let $\{\hat{\sigma}_{t+1/t}^{i}\}_{t=1}^{N}$ and $\{\hat{\sigma}_{t+1/t}^{j}\}_{t=1}^{N}$ be two sets of forecasts for the volatility $\{\sigma_{t+1}\}_{t=1}^{N}$ from models i and j respectively. The loss function defined in Eq. (2) is used to calculate the *loss differential* as:

$$d_{t+1}=\left((\hat{\sigma}_{t+1}^{RDV})^2-(\hat{\sigma}_{t+1/t}^{i})^2\right)^2-\left((\hat{\sigma}_{t+1}^{RDV})^2-(\hat{\sigma}_{t+1/t}^{j})^2\right)^2 \tag{3}$$

This loss differential is simply the difference between the two forecasts errors obtained from models i and j, respectively. In this case, we test the null hypothesis:

$$H_0: E(d_t)=0 \quad \text{vs.} \quad E(d_t)\neq 0 \tag{4}$$

Assuming that $d_t \sim iid$, then the test statistic is:

$$S=\sum_{t=1}^{N} I_+(d_t) \quad \text{where} \quad I_+(d_t)=\begin{cases}1 & if\ d_t>0\\ 0 & otherwise\end{cases} \tag{5}$$

Following Diebold and Mariano [DM95], the large sample studentized version of an exact finite sample test, the sign test, is asymptotically normal.

$$S_a=\frac{S-0.5N}{\sqrt{0.25N}}\sim N(0,1). \tag{6}$$

4 Empirical results

We considered two time series:

- The IBEX 35 is the official index returns of the Spanish Stock Exchange continuous market. This data set consists of 3932 observations of IBEX and spans from January 2, 1990 through May 10, 2005. The first 3539 observations are used in the estimation process and 393 in the one-step-ahead forecasting.
- The S&P 500 composite index returns spans from October 10, 1990 through December 3, 2003 (3085 observations). The first 2785 observations are used for the parameters estimation and the last 300 for the one-step-ahead forecasting.

4.1 Main features of the series

The two data sets exhibit the main features of financial returns series: heavy tails, heteroscedasticity, volatility clusters and the leverage effect.

In order to capture the behavior of the return time series as well as the main features of the volatility, we analyzed as in Muñoz et al. [MMM05] the following models: GARCH, SV, SETAR-TGARCH $(1, 1)$ and SETAR-THSV. For each model we consider two cases, when $\{u_t\}$ is Gaussian or follows a standardized *t-Student* distribution.

The validation of the different models is based on the standardized observations (called *residuals*). These are defined as $\hat{u}_t = r_t/\hat{\sigma}_t$, where $\hat{\sigma}_t$ is the estimated volatility obtained by replacing the estimated parameters considered in the equation that defines volatility. In particular, we use the Box-Ljung statistics of the residuals to check the goodness-of-fit of the equation relating to the mean; then we applied the test to the squared residuals, to validate the equation that models volatility. The skewness and kurtosis coefficients and the residuals normal probability plot will allow validating the assumptions about the distribution of the random noise variables. The models that pass the validation procedure are: GARCHt, SVt, SETAR-TGARCHt (all with t-Student innovations) and the SETAR-THSV.

4.2 Comparing forecasting performance

For each model which passes the validation procedure and for each data set, the one-step-ahead is calculated, then the rolling daily volatility (RDV) using Eq. (1). We consider $k = 11$, roughly half a month of activity. To choose the k value is a delicate matter, because a large value smoothes the volatility series too much whereas a small value introduces too much noise. In fact, we only decided on this value after several tests Next, the *loss differential* is calculated from Eq. (3) and finally, the *sign test* is done (see Eq. (6)). The results for this test are:

Table 1. Diebold's and Mariano's sign test (z-values)

IBEX35	SV_t	SETAR-$TGARCH_t$	SETAR-THSV	S&P500	SETAR-SV_t	SETAR-$TGARCH_t$	THSV
$GARCH_t$	7.31	6.80	5.77	$GARCH_t$	7.28	-3.76	2.82
SV_t		4.041	3.63	SV_t		-8.57	-2.11
SETAR-$TGARCH_t$			2.09	SETAR-$TGARCH_t$			6.34

Table 2 shows that all the values are significant at the 0.95 confidence level, indicating that all the models are not equivalent with regard to forecasting performance.

Once the estimated models are proven to be statistically different, we then select the model that best fits our series. In order to select the "best" model as the model which gives the most accurate forecasts, we use the mean square error between the rolling window volatility and the one-step-ahead forecast, for each model i:

$$MSE^i = \frac{1}{N}\left[\sum_{t=1}^{N}\left((\hat{\sigma}_{t+1}^{RDV})^2 - (\hat{\sigma}_{t+1/t}^{i})^2\right)^2\right] \quad (7)$$

Table 2 shows that the forecasting ability of the models for the IBEX35 is similar, when taking into account the MSE statistics, with the best model being the SETAR-THSV. For the S&P500 series, the SV_t model is superior to the others.

Table 2. Ranking of models

MSE	RDV/ GARCH_t	RDV/ SV_t	RDV/ SETAR-TGARCH_t	RDV/ SETAR-THSV
IBEX35	0.068	0.060	0.065	0.057
S&P500	3.364	2.216	3.308	3.470

We also evaluate the forecasting performance in sample, obtaining results similar to the out of sample. In Table 3, we show the estimates coefficients of the SETAR-THSV model for the IBEX35 return series. The leverage effect in the mean is clear in this time series. Returns display asymmetric behavior in response to negative lags shocks. This model captures the change in the level of stochastic volatility, volatility increases ($\hat{\alpha}^1 > 0$) after negative shocks (bad news) and decreases ($\hat{\alpha}^2 < 0$) after positive shocks (good news). The model displays a high persistence $\hat{\beta}$, which does not switch between the regimes.

The best model for S&P 500 is the simplest one, the stochastic volatility model with error term following a t distribution (SV_t). The leverage effect for this data is not relevant in the return and neither in the volatility. More important is the introduction of the heavy error distribution for the error term. The persistence parameter $\hat{\beta}$ is 0.99 (st.dev.: 0.003) and the estimated degrees of freedom $\hat{\nu}$ are 8.73 (0.82) signaling significant departures from the normal model. Moreover, $\hat{\alpha} = -0.01$ (0.11) and $\hat{\sigma}_\eta = 0.193$ (0.018).

Table 3. IBEX35: Estimated coefficients of the SETAR-THSV model

SETAR		THSV	
First regime	Second regime	First regime	Second regime
$\hat{\phi}_3^1 = -0.0507$ (0.0259)	$\hat{\phi}_4^2 = -0.0478$ (0.0214)	$\hat{\alpha}^1 = -0.079$ (0.009)	$\hat{\alpha}^2 = -0.037$ (0.002)
$\hat{\phi}_8^1 = 0.0602$ (0.0258)	$\hat{\phi}_6^2 = -0.0436$ (0.0219)	$\hat{\beta} = 0.962$ (0.002)	
$\hat{\phi}_{15}^1 = 0.0977$ (0.0263)	$\hat{\phi}_8^2 = 0.0454$ (0.0220)	$\hat{\sigma}_\eta = 0.172$ (0.014)	
	$\hat{\phi}_{11}^2 = 0.0507$ (0.0217)		
	$\hat{\phi}_{13}^2 = 0.0529$ (0.0217)		
$T = 0.000$			

() Std. dev.

5 Conclusions and future work

In this paper we have compared the forecasting performance of competing models for two return time series: IBEX 35 and SP&500[5]. We have obtained a different best model for each case. For the IBEX 35, the introduction of threshold in the mean and variance equations produces more accurate predictions. In the S&P500 return series, however, the leverage in the mean is not important and thus the SV_t is flexible enough to beat the threshold models. Empirical evidence applied to eight international financial market indices, including the G-7 countries supports the hypothesis of thresholds nonlinearity in both the mean and also in volatility [CSG05]. Our results go in this line though further research is needed.

References

[AB98] Andersen, T.G., Bollerslev, T.: Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Int Econ Rev, **39**, 885–905 (1998)

[ANV99] Anderson, H., Nam, K., Vahid, F. Asymmetric nonlinear smooth transition GARCH models. In: Rothman, P. (ed) Non Linear Time Series Analysis of Economic and Financial Data, Kluwer Academic Press, 191–207 (1999)

[Cha05] Chan, Y.H.: Bayesian comparison of Asymmetric Heterocedastic Models. MA Thesis, Feng Chia University, Taichung, Taiwan (2005)

[CSG05] Chen, W.S., So, M.K., Gerlach, R.H.: Assessing and testing for threshold nonlinearity in stock returns. Aust NZ J Stat **47(4)**, 473–488 (2005)

[CT92] Cao, C.Q., Tsay, R.S.: Nonlinear time-series analysis of stock volatilities. J Appl Econom, Supplement 1, S165–S185 (1992)

[DM95] Diebold, F.X., Mariano, R.: Comparing predictive accuracy. J Bus Econ Stat, **13**, 253–265 (1995)

[HL06] Hansen, P.R., Lunde, A.: Consistent ranking of volatility models. J Econometrics, **131**, 97–121 (2006)

[LL96] Li, C.W., Li, D.K.: On a double-threshold autoregressive heterocedastic time series models. J Appl Econom, **11**, 253–274 (1996)

[Mar02] Márquez, M.D.: Modelo SETAR aplicado a la volatilidad de la rentabilidad de las acciones: algoritmos para su identificación. PhD Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain (2002)

[MMM04] Muñoz, M.P., Márquez, M.D., Martí Recober, M., Villazón, C., Acosta, L.:. Stochastic volatility and TAR-GARCH models: Evaluation based on simulations and financial time series. In: Antoch, J. (Ed.) COMPSTAT 2004 - Proceedings in Computational Statistics 16th Symposium Prague, Czech Republic, Physica-Verlag, Springer, ISBN: 3-7908-1554-3 (2004)

[MMM05] Muñoz, M.P., Márquez, M.D., Martí Recober, M., Villazón, C.: A walk through non-linear models in financial series. 3rd World Conference on Computational Statistics and Data Analysis, Limassol, Cyprus (2005)

[5]The *Rolling daily volatility* was used as volatility proxy: an important decision since the quality of the comparison greatly depends on the choice of the proxies

[Pat05] Patton, A.J.: Volatility forecast evaluation and comparison using imperfect volatility proxies. Econometric Society. 2005 World Congress, London (2005)
[Poo05] Poon, S.H.: A Practical Guide to Forecasting Financial Market Volatility. John Wiley and Sons (2005)
[SLL02] So, M.P., Li, W.K., Lam, K.: A threshold stochastic volatility. J Forecasting **21**, 473–500 (2002)
[Ton90] Tong, H.: Non linear time series: A dynamical system approach. Oxford University Press (1990)
[Tsa89] Tsay, R.S.: Testing and modelling threshold autoregressive processes. JASA **84**, 231–240 (1989)

Model formulation

Table 4. Summary of first and second-generation models

First-generation	Return Equation	Volatility Equation
GARCH/GARCH$_t$	$r_t = \sigma_t u_t,\ u_t \sim NID(0,1)$ *or* $u_t \sim t_v$	$\sigma_t^2 = \gamma_0 + \sum_{i=1}^{c} \gamma_i \gamma_{t-i}^2 + \sum_{i=1}^{d} \lambda_i \sigma_{t-i}^2$
SV/SV$_t$	$r_t = \sigma_t u_t,\ u_t \sim NID(0,1)$ *or* $u_t \sim t_v$	$\ln(\sigma_t^2) = \alpha + \beta \ln(\sigma_t^2) + \sigma_\eta \eta_t, \eta_t \sim NID(0,1)$
Second-generation	**Return Equation**	**Volatility Equation**
SETAR-TGARCH/ SETAR-TGARCH$_t$	$r_t = \begin{cases} \phi_0^1 + \sum_{i=1}^{p} \phi_i^1 r_{t-i} + y_t^1,\ r_{t-1} < T \\ \phi_0^2 + \sum_{i=1}^{q} \phi_i^2 r_{t-i} + y_t^2,\ r_{t-1} \geq T \end{cases}$	$\sigma_t^2 = \gamma_0^1 + \sum_{i=1}^{c} \gamma_i^1 \gamma_{t-i}^2 + \sum_{i=1}^{d} \lambda_i^1 \sigma_{t-i}^2,\ r_{t-1} < T$ $\sigma_t^2 = \gamma_0^2 + \sum_{i=1}^{f} \gamma_i^2 \gamma_{t-i}^2 + \sum_{i=1}^{g} \lambda_i^2 \sigma_{t-i}^2,\ r_{t-1} \geq T$
	$y_t^i = \sigma_t^i u_t,\ u_t \sim NID(0,1)$ or $u_t \sim t_v$	
SETAR-THSV/ SETAR-THSV$_t$	$r_t = \begin{cases} \phi_0^1 + \sum_{i=1}^{p} \phi_i^1 r_{t-i} + y_t^1,\ r_{t-1} < T \\ \phi_0^2 + \sum_{i=1}^{q} \phi_i^2 r_{t-i} + y_t^2,\ r_{t-1} \geq T \end{cases}$	$\ln(\sigma_t^2) = \alpha^1 + \beta \ln(\sigma_t^2) + \sigma_\eta \eta_t,\ r_{t-1} < T$ $\ln(\sigma_t^2) = \alpha^2 + \beta \ln(\sigma_t^2) + \sigma_\eta \eta_t,\ r_{t-1} \geq T$
	$y_t^i = \sigma_t^i u_t,\ u_t \sim NID(0,1)$ or $u_t \sim t_v$	