Estimation and Forecasting of Continuous Time Stochastic Volatility Model

Faouzi Jilani, Hanen Ould Ali and Salim Ben Sassi

Faculty of Economics and Management of Tunis, Tunisia

_____

Abstract

This paper improves the estimation of continuous time stochastic model that treats volatility as a latent variable and compares the forecasting performance of the Kalman filter procedure with Exponential model of Autoregressive Conditional Heteroscedastisity. Our empirical study examines the stock indice TUNINDEX by using the daily close price data over the period December 31, 1997, its creation date, to December 31, 2009. The results suggest the significant existence of leverage effect between TUNINDEX returns and its volatility. Indeed, an unanticipated increase in Tunindex return leads to increased uncertainty that is greater than that induced by an unanticipated drop in return. Thus, the volatility forecasts based on Kalman filter model may outperform those of EGARCH model.

Keywords: EGARCH,-Kalman Filter,-Stochastic Volatility.

_____

Introduction

One of the main phenomena observed in the process of stock returns is the presence of strong variations over times; these are periods of turbulence in financial markets. Statistically, the existence of these movements results in fat tails in the variations distribution function, calling into question the very strong assumption of Gaussian distribution returns. This empirical evidence has led to the invalidation of the famous formula of Black and Scholes (1973) that assumes that returns are generated from a normal distribution whose mean and variance are constant over time. Large efforts were then undertaken to consider the statistical properties of stock market fluctuations. The financial literature mainly concentrates on two approaches: the autoregressive models ARCH (Autoregressive Conditional Heteroscedasticity) introduced by Engel (1982) and the models with non deterministic volatility based on the works of Bachelier (1900). However, the big critic of ARCH model is a deterministic approach, while the other works consider making some volatility a random (unpredictable) variable. Several directions of research are developed for this approach: the stochastic volatility models introduced by Hull and White (1987), it is about a geometric Brownian process. It presents the major inconvenience not to be stationary. This type of specification was quickly abandoned for the benefit of the models generating a stationary volatility process, of type mean-reverting. Over time, the process tends to drift towards its long-term mean: such a process is named Ornstein-Uhlenbeck. Reference articles treating this type of model are the ones of Stein, E., and J.Stein (1991) for its simple version and of Harvey, Ruiz and Shephard (1994) for its logarithmic version. The econometric estimation of these models poses a challenge. Indeed, financial models are expressed in continuous time while observations can only be collected in a discrete time. On the other hand, the models studied are bivariate latent models where volatility is unobservable. First it is noticed that there is no analytical solution to this bivariate process. Therefore, we cannot consider exact discretization from which we could deduce a likelihood. The

only way to cope with this is discretization approximation. Several discretization schemes are considered, the most popular being the Euler-method and the discretization ARCH. Volatility forecasting is an essential task in financial markets, so there are many papers that study forecasting performance of different volatility models in the literature, but many studies have been written in the theme of deterministic volatility modeling. The aim of this paper is to estimate and predict stochastic volatility from these two discretizations approaches above. Our empirical study is based on historical daily data of TUNINDEX in the period between December 31, 1997 and December 31, 2009. Unfortunately, there is no single procedure available to calculate and predict volatility. In this paper, the researchers compare the Kalman filter procedure (Harvey & al., 1994) and an EGARCH estimation approach (Nelson, 1990). The remainder of the paper is organized as follows. In section 2, there is a description of the estimation methods of the stochastic volatility model. Section 3 presents the forecasting Tunindex volatility. Section 5 concludes.

Estimating Stochastic Volatility TUNINDEX

In this study, the logarithm of the instantaneous volatility follows the process of mean-reverting, also called the process "Ornstein-Uhlenbeck". These reflect the presence of a force reverting to a long-term drift of the volatility. In particular, if S represents the value of TUNINDEX and if $\ln \sigma_t^2$ symbolizes the logarithm of the instantaneous volatility, the dynamics of the bivariate diffusion process are governed by the following equations:

$$\begin{cases} dS_t = \alpha_c S_t dt + \sigma_t S_t dB_t \\ d \ln \sigma_t^2 = \kappa [\vartheta - \ln \sigma_t^2] dt + \gamma dW_t \end{cases}$$
(1)

Brownian movement dB_t and dW_t are possibly correlated, $E(dB_t \ dW_t) = \rho dt$. The coefficient ρ has a strong economic content because it represents the leverage

effect. The parameters to estimate, for this type of model is the drift α_c, the long-term average ϑ, the speed of mean reversion κ and volatility of volatility γ. These four parameters are considered constant, they will be determined from the database of the rate of return TUNINDEX between time t_0 and T, which are respectively the first and last date in the database

Volatility Estimation via the Kalman Filter

To estimate model (1), Harvey & al. (1994) conducted a linearization. This transformation helps to develop a model state measure. However, it should develop the log-OU by discretizing with a step equal to that of observations, denote by Δ the step of the discretization, for daily data, $\Delta = 1/252$ where approximately 252 is being the number of working days in a year, the discretization Euler gives:

$$\begin{cases} S_{t+\Delta} - S_t = \alpha_c S_t \Delta + \sigma_t \sqrt{\Delta} S_t u_t \\ \ln \sigma_{t+\Delta}^2 - \ln \sigma_t^2 = \kappa [\vartheta - \ln \sigma_t^2] \Delta + \gamma \sqrt{\Delta} v_t \end{cases}$$
(2)

u_t and v_t, two error terms are Gaussian with zero mean and variance 1. This model can be rewritten as follows:

$$\begin{cases} Y_t = \alpha + (\exp \dfrac{h_t}{2}) \sqrt{\Delta} u_t \\ h_{t+\Delta} = \mu + \beta h_t + \delta v_t \end{cases}$$
(3)

Y_t indicates the return of the support, $h_t = \ln \sigma_t^2$, $\alpha = \alpha_c \Delta$, $\mu = \kappa \vartheta \Delta$, $\beta = 1 - \kappa \Delta$ and $\delta = \gamma \sqrt{\Delta}$. Finally, the variations between two observations are modeled following the standard form:

$$\begin{cases} Y_t = \alpha + (\exp \dfrac{h_t}{2}) u_t \\ h_t = \mu + \beta h_{t-1} + \delta v_t \end{cases}$$
(4)

Formally adopting the notation $Y = (Y_1 ... Y_T)$ of the return vector of Support, $h = (h_1 ... h_T)$ unobservable volatility vector and denote by θ the set of parameters $(\alpha, \mu, \beta, \delta)$, the parameter α must be estimated prior to effecting the linearization of the model. The estimator of α is:

$$\hat{\alpha} = \frac{1}{T} \sum_{t=1}^{T} Y_t$$

(5)

To linearize the first equation of system (4), we square $Y_t - \hat{\alpha}$ and express it in logarithmic form. The following is obtained:

$$\begin{cases} \ln\left((Y_t - \hat{\alpha})^2\right) = h_t + \ln(u_t^2) \\ h_t = \mu + \beta h_{t-1} + \delta v_t \end{cases}$$

(6)

As further developing this equation as we know that $u_t \sim N(0.1)$, we can therefore deduce the distribution of $\ln(u_t^2)$. It corresponds to a logarithmic χ^2 distribution, whose expectation is -1.27 and the variance is $0.5\pi^2$, approximately 4.93. Note however that $\ln(u_t^2)$ cannot be correctly approximated by a normal law only if the sample is very large.

Following the approach of Racicot F.-E. and Theoret R. (2005), by adding and subtracting $E\ln(u_t^2)$ in the first equation of model (6), we obtain:

$$\ln\left((Y_t - \hat{\alpha})^2\right) = E\ln(u_t^2) + h_t + \left[\ln(u_t^2) - E\ln(u_t^2)\right]$$

Setting $\xi = \left[\ln(u_t^2) - E\ln(u_t^2)\right]$ and $\zeta = \delta v_t$, we get two white noise centered on variance 4.93 and δ^2 respectively.

We can therefore rewrite the model (6) as follows:

$$\begin{cases} Y_t^* = c - 1.27 + h_t + \xi_t \\ h_t = \mu + \beta h_{t-1} + \zeta \end{cases}$$

(7)

Where $Y_t^* = \ln\left((Y_t - \hat{\alpha})^2\right)$ and c is a constant introduced to take into account that it equals only -1.27 in very large samples. This model is as state-space linear. Equations 1 and 3 of the model (7) are in the appropriate form to use the Kalman filter. The first equation is called measurement equation as the variable Y is observed. The second equation is the equation of state or transition as h the state variable is latent. The method of estimating this model is then explained in two steps: first, the latent variables are estimated by Kalman filter approach, then the parameters are estimated by the method of Maximum Likelihood.

The researchers used EViews software for the implementation of this algorithm. The result of estimating the state-space model appears in the following table:

Table 1: Estimation of Stochastic Volatility of TUNINDEX Returns

Sspace: SS01

Method: Maximum likelihood (Marquardt)

Date: 01/30/10 Time: 16:34

Sample: 12/31/1997 12/31/2009

Included observations: 2993

Convergence achieved after 41 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	0.118033	17484.80	6.75E-06	1.0000
C(2)	0.917170	0.016806	54.57302	0.0000
C(3)	-1.906974	0.235490	-8.097885	0.0000
C(4)	-0.931217	1448.262	-0.000643	0.9995

	Final State	Root MSE	z-Statistic	Prob.
HTT	-11.47980	0.771733	-14.87535	0.0000

Log likelihood	-6777.460	Akaike info criterion	4.531547
Parameters	4	Schwarz criterion	4.539571
Diffuse priors	0	Hannan-Quinn criter.	4.534434

As obvious in Table (1), the coefficients C(1) and C(4) are not significant at the 95%. The coefficients C(1) to C(4) are the parameters of the model (7) with C(1)= c, C(2)= β and exp(C(3))= δ^2 and C(4)= μ. Using the system of equations linking the parameters of discrete-time model with the parameters of the model in continuous time, table (2) is drawn to summarize the results of estimation for the log-OU after the Euler discretization. Note that the parameters obtained from this discretization are annualized.

Table 2 : Estimation Results for Log-OU Model

|	α_c	ϑ	κ	γ	ρ
Log-OU	0.12	-11.24	20.87	6.11	0.60

It is noted that the instantaneous volatility moves around a long-term trend equal to 0.45%. To achieve this, use the following equation $\exp\left(\vartheta + \frac{\gamma^2}{4\kappa}\right)$ and take root. κ relatively high value is also found, which suggests a pronounced effect of mean reversion.

It is also noted that the coefficient of correlation ρ between movements in returns and movements in volatility is different to zero, which may justify the assumptions considered by some authors, such as Heston (1993) for example, who proved that the correlation between volatility and returns of assets is essential to generate the asymmetric distribution. The evolution of the observed variables and filtered of Y^* illustrated in Figure (1)
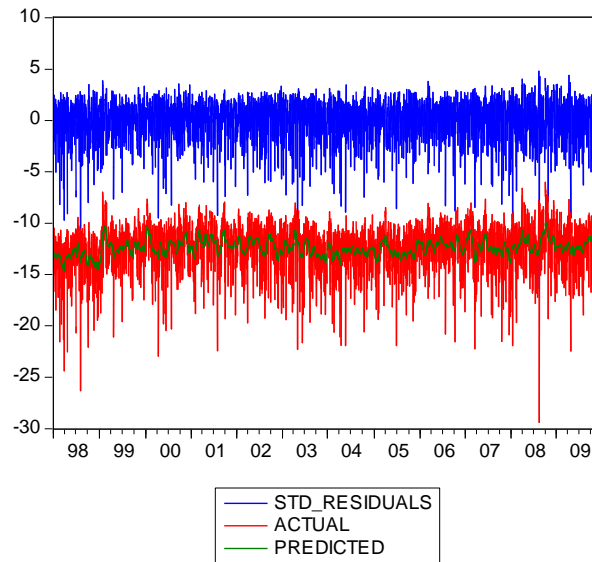

Figure 1: Values Observed and Filtred of Y^*

In option pricing models, volatilities are expressed yearly. In portfolio selection model in continuous time and for the calculation of VaR, the reference period is the day, volatility is expressed as daily. Moreover, according to the equation $h_t = \ln \sigma_t^2$, the stochastic volatility of TUNINDEX return is equal to:

$$\sigma_t = \exp(\frac{h_t}{2})$$

To annualize this volatility, multiply σ_t by $\sqrt{252}$. In other words, the daily volatility is about 6% of annual volatility. The evolution of the volatility of TUNINDEX return is shown in figure (2).
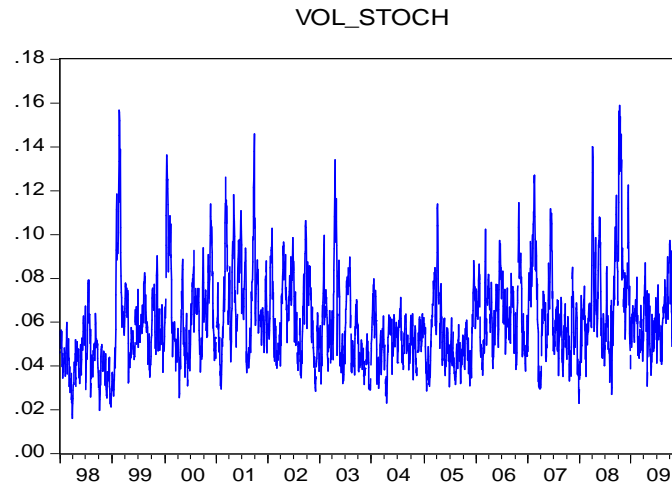
Figure 2: Stochastic Volatility of TUNINDEX Return

It is found that volatility is experiencing strong fluctuations around its average, which is 0.45% on daily basis, corresponding to 7.1% in annual frequency.

The first half of the decade was characterized by very high levels of volatility, the movement began in 1999, as shown in Figure (2), as a historical volatility. This period is also characterized by a multiplication of volatility peaks. A peak in volatility specifies a phase when the volatility settles at a level significantly above its long-term average. After a sharp drop between 2003 and 2005, the volatility has temporarily stabilized at a low level before increasing again in 2006. The increase in volatility during a period of time, results from a conjunction of proper phenomena to this period, such as the events of 11 September 2001, which significantly changed the aspect of risk taking behavior of investors just like the aspects of growth.

Exponential Model of Autoregressive Conditional Heteroscedastisity

Models of the ARCH family (Autoregressive Conditional Heteroscedasticity) have enjoyed considerable success since their first version put forward by Engel (1982). Nelson (1990) showed that ARCH models have the same stationary distributions that some stochastic volatility models expressed in continuous time. Specifically, the E-GARCH model has the same stationary distribution as a version of log-OU given by model (1). Indeed, this model is very useful to model not only excess kurtosis but also asymmetric effects that have returns on volatility. The simplest and most used EGARCH model is the EGARCH (1,1) defined as follows:

$$\begin{cases} Y_t = \mu + \varepsilon_t \\ \varepsilon_t = \sigma_t z_t \\ \ln \sigma_{t+1}^2 = \omega + \beta \ln \sigma_t^2 + \alpha z_t + \alpha' \left(|z_t| - \sqrt{\frac{2}{\pi}} \right) \end{cases}$$

Where Y_t is a random variable representing the arithmetic return of the support S_t, $z_t \rightarrow N(0,1)$ and $\mu, \omega, \beta, \alpha, \alpha'$ are real constants.

The process $\ln \sigma_t^2$ is AR (1) and admits a stationary solution if and only if $|\beta| \langle 1$.

Resume model (1):

$$\begin{cases} dS_t = \alpha_c S_t dt + \sigma_t S_t dB_t \\ d \ln \sigma_t^2 = \kappa \left[\vartheta - \ln \sigma_t^2 \right] dt + \gamma dW_t \end{cases}$$

Where E(dB$_t$ dW$_t$)= ρdt .

Nelson (1990) proposes a suite of models indexed by h which converges in distribution to model (1), when the amplitude of the elementary time interval approaches zero. Consider the family of models:

$$\begin{cases} \dfrac{S_{(k+1)h} - S_{kh}}{S_{kh}} = \alpha_c h + \sigma_{kh}\sqrt{h}\, z_k \\[2ex] \ln \sigma^2_{(k+1)h} = \ln \sigma^2_{kh} + \kappa\left[\vartheta - \ln \sigma^2_{kh}\right]h + \gamma\sqrt{h}\, z_k^* \end{cases}$$

Where $z_k^* = \theta\, z_k + \varphi\left(|z_k| - \sqrt{\dfrac{2}{\pi}}\right)$. Indeed

the two random variables: z_k and $|z_k| - \sqrt{\dfrac{2}{\pi}}$ are centered, reduced and uncorrelated. The parameter θ allows to

model the asymmetric effect related to the sign of innovation z_k and the parameter φ takes into account the asymmetry related to the amplitude of innovation z_k as measured by the difference $|z_k| - \sqrt{\dfrac{2}{\pi}}$.

Therefore, the EGARCH model (1.1) describes a mean-reverting process. So the logarithmic conditional variance returns in long-term ϑ with speed κ. The researchers will apply this model to Tunindex return in the period between December 31, 1997 and December 31, 2009. Table (3) represents the results of estimating EGARCH (1,1) model.

Table 3- Estimation of Conditional Volatility of Tunindex Returns

Dependent Variable: Rdt Tunindex
Method: ML - ARCH
Date: 11/29/10 Time: 18:37
Sample: 12/31/1997 12/31/2009
Included observations: 2993
Convergence achieved after 19 iterations
Bollerslev-Wooldrige robust standard errors & covariance
Presample variance: unconditional
LOG(GARCH) = C(2) + C(3)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(4)
    *RESID(-1)/@SQRT(GARCH(-1)) + C(5)*LOG(GARCH(-1))

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000239	6.78E-05	3.532979	0.0004
Variance Equation				
C(2)	-1.389300	0.252892	-5.493651	0.0000
C(3)	0.442591	0.060366	7.331807	0.0000
C(4)	0.013632	0.024462	0.557279	0.5773
C(5)	0.904112	0.019923	45.38145	0.0000
R-squared	0.002761	Mean dependent var		0.000499
Adjusted R-squared	0.004104	S.D. dependent var		.004941
S.E. of regression	0.004951	Akaike info criterion		-8.116412
Sum squared resid	0.073257	Schwarz criterion		-8.106382
Log likelihood	12151.21	Hannan-Quinn criter.		-8.112804
Durbin-Watson stat	1.355499			

As can be seen from the table above, the coefficients C, C(2), C(3) and C(5) are significant at the 95%. The coefficients C,

C(2), C(3), C(4) and C(5) are respectively the parameters $\mu, \omega, \alpha', \alpha\ and\ \beta$ of the EGARCH (1.1) model.
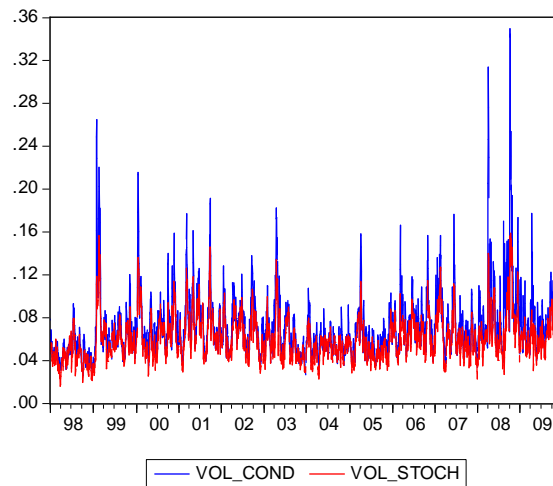
After estimating the EGARCH (1.1) model with maximum likelihood method, the parameters of the model (1) can be determined by the following equivalence:

$$\sigma_t = \exp\left(\frac{\hat{\omega} + \hat{\beta}\ln\sigma_{t-1}^2 + \hat{\alpha}z_{t-1} + \hat{\alpha}'\left(|z_{t-1}| - \sqrt{\frac{2}{\pi}}\right)}{2}\right)$$

$$\begin{cases} \alpha_c = \hat{\mu} \\ \kappa = 1 - \hat{\beta} \\ \vartheta = \dfrac{\hat{\omega}}{1 - \hat{\beta}} \\ \gamma = \sqrt{\hat{\alpha}^2 + \hat{\alpha}'^2} \end{cases}$$

$\hat{\omega}, \hat{\alpha}', \hat{\alpha} \text{ and } \hat{\beta}$ indicate the maximum likelihood estimating.

From EGARCH (1,1) model, an estimate of the conditional volatility can be proposed, according to the equation:

Figure 3: Stochastic Volatility and Conditional Volatility of Tunindex Returns

In Figure (3), we compare the conditional volatility associated with the EGARCH (1.1) model to stochastic volatility calculated previously. It is noted that the profiles of evolution of the two volatility curves are very close. We also observe that the stochastic volatility fluctuates less than the conditional volatility. In this case, the conditional volatility associated with the EGARCH (1,1) model had increased further during the market crash of October 2008.

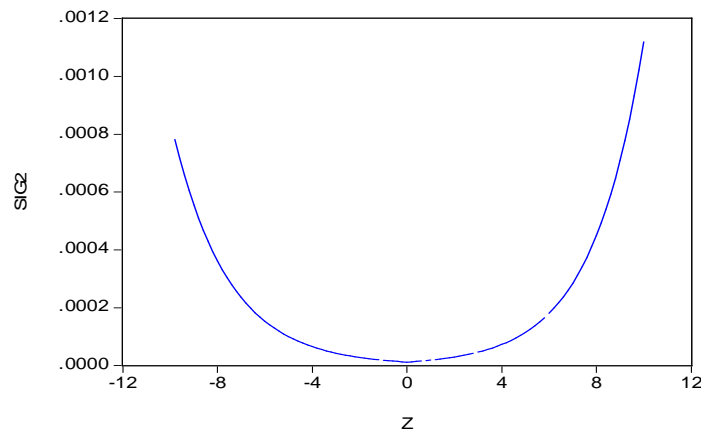In this study, the leverage effect is represented by θ; it is positive and statistically different from zero. In figure (4), the new impact for Tunindex return by NIC (News Impact Curve) is determined; the curve representing the variance σ^2 against impact $z = \varepsilon/\sigma$,

Which

$$\ln\sigma_t^2 = \hat{\omega} + \hat{\beta}\ln\sigma_{t-1}^2 + \hat{\alpha}z_{t-1} + \hat{\alpha}'\left(|z_{t-1}| - \sqrt{\frac{2}{\pi}}\right)$$

Figure 4: News Impact Curve of Tunindex Return

An asymmetric leverage effect is clearly observed in Figure (4). Thus, the conditional variance of Tunindex return reacted more to positive shocks than negative shocks of equal magnitude. The economic consequence of this result is that an unanticipated increase in Tunindex return leads to increased uncertainty greater than that induced by an unanticipated drop in return.

Forecasting TUNINDEX Volatility

It is now possible to produce forecasts of volatility from the models established,

which are strictly recursive. Take the case of TUNINDEX. The forecast starts January 02, 2010, the sample ending on the date of December 31, 2009, and we continue to forecast until June 30, 2010. The result of forecast is shown in Figure (5). As can be seen, the stochastic volatility model provides an increase in volatility like the EGARCH (1,1) model. But it should be noted that the volatility arising from the EGARCH (1.1) model was initially slightly higher than the stochastic volatility. Thus, the two volatilities tend to be closer to their equilibrium value in the long term.
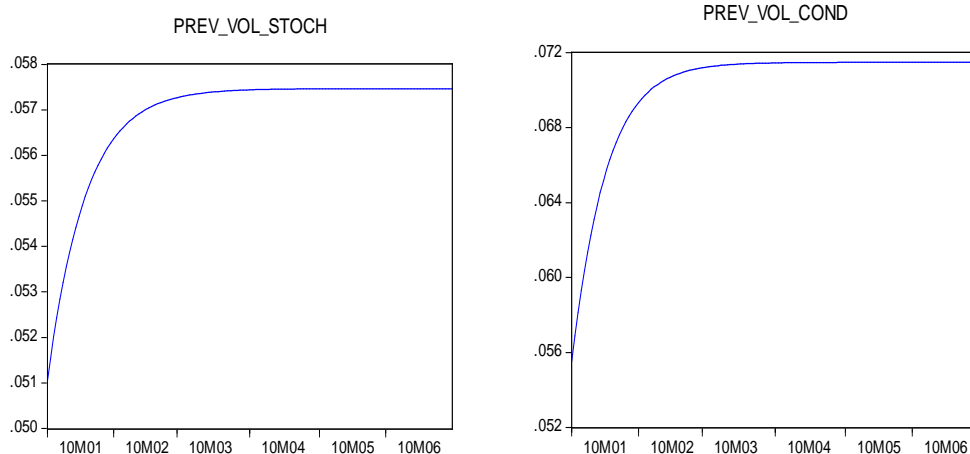
Figure 5: Forecasting Tunindex Volatility: Kalman Filter Model and EGARCH (1,1) Model

To explore the robustness of forecast accuracy, the root mean squared forecast error or RMSFE criterion was employed.

Applying this formula:

$$RMSFE = \sqrt{\frac{\sum_{t=1}^{N}(Y_t - F_t)^2}{N}}$$

Where Y_t is the actual value at period t, and F_t is the forecast for period t. Hence, RMSFE based on Kalman filter forecast model is lower than the one based on EGARCH forecast model. That is why we strongly recommend the Kalman filter as a powerful tool to forecast stochastic volatility.

Conclusion

In this paper, the researchers demonstrate their interest in stochastic volatility model where volatility is governed by logarithmic Ornstein-Uhlenbeck process. This specification has been frequently used in the literature to describe the volatility process, mainly because of its convenience and simplicity.

The estimation procedures associated are the state-space model and the EGARCH(1,1) model. Our empirical study is based on the historical daily data of TUNINDEX in the period between 31 December 31, 1997 and December 31, 2009.

This wide research reflects the importance of volatility in investment. So the empirical results have confirmed the hypothesis of correlation between movements in Tunindex returns and movements in volatility. In fact, an unanticipated increase in Tunindex return leads to increased uncertainty greater than that induced by an unanticipated drop in return. In addition, estimates obtained from the two discretization schemes (Euler and ARCH) are quite similar.

Forecasting the returns volatility, regardless of the model used, is an exercise fraught with risk. The model of stochastic volatility predicted an increase in volatility like the EGARCH (1, 1) model. But it should be noted that the volatility arising from the EGARCH (1, 1) model was initially slightly higher than the stochastic volatility. Thus, the two volatilities tend to be closer to their equilibrium value in the long term.

This survey has concentrated on the question: which method will provide the best forecasts? Forecasting can be a tedious task for financial analysts because of discrepancies in the results provided by different tools. The researchers strongly recommend the Kalman filter as a powerful tool to forecast stochastic volatility. However, both models remain easily implementable alternatives to more complex and computer-intensive techniques such as MCMC.

A good forecast of the volatility of asset prices over the investment holding period is a good starting point for assessing investment risk. Thus, volatility forecasting will continue to remain a specialist subject and will continue to be studied vigorously.

References

Aihara, S. I. & Bagchi, A. (2006). "Filtering and Identification of Heston's Stochastic Volatility Model and its Market Risk," *Journal of Economic Dynamics & Control* ,30, 2363–2388.

Ait-Sahalia, Y. & Kimmel, R. (2007). "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413–452.

Andersen, T. G. & Lund, J. (1997). "Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77, 343-377.

Bakshi, G., Cao, C. & Chen, Z. (1997). "Empirical Performance of Alternative Pricing Models," *Journal of Finance*, 52, 499-547.

Broto, C. & Ruiz, E. (2002). "Estimation Methods for Stochastic Volatility Models: A Survey," *Universidad Carlos III de Madrid*, working Paper, (14), 02-54.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). "Maximum Likelihood for Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society* B(39), 1-38.

Ferland, R. & Lalancette, S., (2006). "Dynamics of Realized Volatilities and Correlations: An Empirical Study," *Journal of Banking & Finance* ,30, 2109–2130.

Figlewski, S. (1997). "Forecasting Volatility," *Financial Markets, Institutions and Instruments* , 6(2).

Fornari, F. & Mele, A. (2006). "Approximating Volatility Diffusions with CEV-ARCH Models," *Journal of Economic Dynamics & Control* , 30, 931–966.

Gelfand, A. E. & Smith, A. F. M. (1990). "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398-409.

Mills, T. C. (1995). "Modelling Skewness and Kurtosis in the London Stock Exchange FT-SE Index Return Distributions," *The Statistician*, 44(3), 323–332.

Peiro, A. (1999). "Skewness in Financial Returns," *Journal of Banking and Finance*, 23, 847–862.

Saatcioglu, C., Korap, H. L. & Volkan, A. G. (2007). "Information Content of Exchange Rate Volatility: Turkish Experience," *International Business & Economics Research Journal*, (6), 9-14.