Modelling and Forecasting Stochastic Volatility

Maria Helena Lopes Moreira da Veiga

Supervisor: Dr. Michael Creel

Universitat Autònoma de Barcelona

Departament d'Economia i d'Història Econòmica

International Doctorate in Economic Analysis

January 2004

Contents

Aos meus pais e Marc

Acknowledgements

I would like to thank those who have contributed with their help and advices to the conclusion of this thesis.

First, I would like to thank my advisor, Professor Michael Creel, for introducing me to the Efficient Method of Moments methodology and consequently to the subject of stochastic volatility and for his advice, encouragement and patience everytime I knocked on his door with (Fortran and econometric) doubts. I am also grateful to Professor Laura Mayoral, who followed my work, for her advises and Professor Jordi Caballé for helping me understand a financial relation underlying an estimated model in the second chapter.

The third chapter of this thesis benefited from the valuable help of Professor George Tauchen and from the advices of Professor Tim Bollerslev, during my stay at Duke University. I also thank seminar participants at the Simposio de Analisis Económico in Salamanca 2002 and at Universitat Autònoma de Barcelona for helpful remarks.

I really appreciate the hospitality of Duke University, where I spent some time as a visiting student.

The empirical part of my fourth chapter of this thesis started at Duke University and benefited from the help of Professor George Tauchen that gave me part of the data and help me programming the long memory stochastic volatility model.

I would like to thank the Economics Department of Universitat Autònoma de Barcelona for the excellent quality of IDEA program and for the good atmosphere that professors and students live there.

I also would like to thank my colleagues and friends: Coralio Ballester, Vitoria Ateca, Pilar Socorro, Ester Camiña and Olivia Estrella that were able to bear me and made my life in Barcelona very pleasant. Agradeco también a Ana Vilaró su amistad.

Financial support from the Faculdade de Economia da Universidade do Porto and from Programa PRAXIS XXI (JNICT) is also gratefully acknowledged.

Eu gostaria de agradecer aos meus pais todo o carinho, amor e apoio que me deram ao longo destes cinco anos sem os quais todo este processo seria ainda mais duro. Muito Obrigada. Finally, thanks to Marc for everything.

Chapter 1

Introduction

Volatility plays an important role for asset pricing theory. The "implied volatility" obtained from the Black-Scholes model for pricing an European option has been by far the most widely measure of future volatility, even when it was known that the assumptions for the model to be valid were violated. However, after the October 1987 stock market crash, the model has not adjusted reality so well and the "implied volatility" seems to have little predictive power relative to the historical volatility, Canina and Figlewski (1993). The trend since then has been to search for models that were able to fit the main empirical stylized facts and the model's ability to reproduce them has been a criterion to dismiss or not a particular specification. The empirical facts have been well documented in several works such as: Bollerslev et al. (1994), Ghysels et al. (1995), etc., and among them we consider: volatility clustering, periods of high volatility are usually followed by periods of high volatility and vice-versa, thick tails of the distribution of asset returns, persistence that it is characterized by the fact that shocks to volatility dissipate at a slow rate, small first order autocorrelation of the squared returns and leverage effects that suggest that stock price movements are negatively correlated with volatility. Therefore, several models have been proposed with the aim to capture these empirical facts. Two well known examples are the original GARCH(1,1) model proposed by Bollerslev (1986) and Taylor

(1986) and the ARSV(1) also proposed by Taylor (1986). Later on in the nineties, several other models emerged with the aim of capturing volatility persistence, including the long memory stochastic volatility model of Harvey (1993) and the model of Breidt et al. (1994) that incorporate in the original stochastic volatility specification a factor of volatility that is fractionally integrated. Other examples are also the IGARCH, that is an extension of the previous GARCH that allows the possibility of an unit root in the volatility specification and the FIEGARCH that allows for long memory by considering that volatility is a fractionally integrated process.

More recently, Ming Liu (2000) and Breidt et al. (2000) proposed regime switching as another explanation for the observed long memory. As Ming Liu (2000) says, the long memory pattern is present in the autocorrelation function of volatility whenever the regime switches in a heavy-tail manner.

All these specifications were designed in discrete time,[1] though a majority of theoretical works in Finance have modelled the logarithm of asset prices as a univariate diffusion, Andersen et al. (2000). In order to solve this dilemma at the end of the nineties, some financial econometric models started to be designed in continuous time. Examples are the papers by Gallant and Tauchen (2001) and by Chernov et al. (2003) that present several systems of stochastic differential equations for equity returns. In these works the equity returns' volatility is model as a function of variables denoted volatility factors, that try to capture some of the stylized facts reported above. Both papers provide empirical evidence that continuous stochastic volatility models with only one volatility factor are not able to capture simultaneously the extra kurtosis and the volatility persistence of the data. Their empirical results report that the introduction of a second factor of volatility allows that one factor might be slow mean reverting and consequently persistent while the other might accommodate the fat tails.

[1] This procedure is justified by the event that financial time series are observed at discrete time intervals.

The third chapter goes deep into the question: why stochastic volatility models with one factor of volatility tend to fail specification tests and in which conditions, if any, they are able to fit the main characteristics of the data? As a starting point, we use the logarithmic stochastic volatility (with one factor) model of Gallant and Tauchen (2001), as a benchmark and we change it slightly by introducing a feedback factor into the volatility factor specification. This feedback factor reveals itself of extreme importance because it captures volatility clustering by allowing that volatility is low when the factor itself is low. The chapter also reports the importance of this feature as a possible imperfect substitute of an increase in volatility caused by stock market crashes. In fact, if we observe the graph of volatility we see a possible change in the volatility pattern at the very end of the sample. This event, as it has been reported in several studies such as Beine and Laurent (2000), Granger and Hyung (1999) and Diebold and Inoue (1999) increases volatility persistence if we do not account for it. This extra persistence is captured by the feedback factor and the model as it is described, is able to pass the specification test.

In this chapter, as well as in all the thesis, we use Efficient Method of Moments of Gallant and Tauchen (1996) as an estimation method because of its testing advantages[2]. In fact, the minimized criterion function scaled by the number of observations follows asymptotically a chi-square distribution that allows us to test if the model is corrected specified.

These stochastic volatility models in continuous time have been tested and the results show that they are flexible and potent enough to describe the main features of the data

[2]Other estimation techniques are for instance: the procedure based on a spectral regression proposed by Geweke and Porter-Hudak (1983) and a frequency-domain estimator for the fractionally integrated stochastic volatility model (FISV) suggested by Breidt et al. (1998). They showed that it is consistent but the asymptotic distribution is not known. Wright (1998) also proposed a new estimator of the FISV model based on the minimum distance estimator (MDE) proposed by Tieslau et al.(1996). It consists in minimizing a quadratic distance function. Wright (1998) showed that the estimator is $T^{1/2}$-consistent and asymptotically normal, provided that $d < \frac{1}{4}$.

but there still remains a question to answer: are they able to forecast volatility accurately? In fact, this is a crucial point since a good volatility model should be able to describe and forecast volatility. In order to inquire if these models forecast as well as they describe volatility, we evaluate, in the fourth chapter of this thesis, the volatility forecasting performance of a continuous time stochastic volatility model. We use the two factors stochastic volatility model of Gallant and Tauchen (2001) for the returns of Microsoft and we filter the underlying volatility using the reprojection technique of Gallant and Tauchen (1998). Remember that under the assumption that the model is correctly specified, we obtain a consistent estimator of the integrated volatility. Then, we compare the forecasting performance of the stochastic volatility model in continuous time to the forecasting performance of other well known models: the GARCH and the ARFIMA. The evaluation procedure consists in using both the R^{*2} of the individual regressions of realized volatility[3] on a constant and on the volatility forecasts obtained from the estimated models and using the t-statistics of the coefficients of the auxiliary regressions. If they are not statistically different from zero and one, respectively, the volatility forecasts are unbiased estimators of realized volatility.

Moreover, the realized volatility is calculated using 15-minutes intraday data instead of tick-by-tick data in order to avoid using some forms of interpolation that could cause negative correlation in the returns series and consequently contaminate this measure. Finally, the empirical results indicate, without doubts, that the continuous time model in the out-of-samples periods does perform better in comparison to the traditional GARCH and ARFIMA that show difficulties in tracking the growth pattern of the realized volatility at the very end of the sample.

Finally, in the fifth chapter we choose the statistical methodology in discrete time justified by the fact that we still obtain an accurate approximation of the options' payoffs

[3] Since realized volatility is considered a good measure of volatility. R^{*2} is the corrected R^2 of Andersen and Bollerslev (2002).

and because it is also interesting to compare the performance of specifications in discrete time to the performance of models in continuous time. We also ask ourselves what will be the effect of introducing a second volatility factor in the long memory stochastic volatility model (LMSV) of Breidt et al. (1994); Will it allow the model to deal simultaneously with kurtosis and persistence, as in the continuous time environment? Having in mind these questions in this chapter, we start by extending the previous model by introducing a volatility factor that in its simplest form is autocorrelated of order one, as the original ARSV by Taylor (1986). Remember that we still model the volatility persistence by assuming that the volatility of the returns shows a long memory feature captured by a fractionally integrated process. The innovation is this extra short run volatility factor that increases kurtosis and helps the model to capture volatility persistence, an effect that can be seen in the autocorrelation function of the squared returns implied by the model. Furthermore, considering some restrictions of the parameters it is possible to fit the empirical fact of small first order autocorrelation of squared returns. All these results are proved theoretically and the model is implemented empirically using the S&P 500 composite index returns. Finally, the empirical results show us that: the short run volatility factor improves the EMM criterion as in Ming Liu (2000) and the long memory stochastic volatility model with two factors of volatility performs better than the two benchmark models. Anyway, according to our results it seems that stochastic volatility models in continuous time seems to fit better the main empirical facts. This occurs, certainly, because they are quite flexible, in the sense that we do not impose rigid specifications on the volatility factors as in discrete time.

Bibliography

[1] Andersen, T. G., Bollerslev, T.,Diebold, F.X. and Labys, P. (2000): Exchange Rate Returns Standardized by Realized Volatility are (nearly) Gaussian, *Multinational Finance Journal*, 4, 159 -179.

[2] Andersen, T. G.and Bollerslev, T. (2002): Correcting the Errors: A Note on Volatility Forecast Evaluation Based on High-Frequency Data and Realized Volatilities, *Discussion paper*, Duke University.

[3] Beine, M. and Laurent, S. (2000): Structural Change and Long Memory in Volatility: New Evidence from Daily Exchange Rates, *Discussion paper*, Université of Liège.

[4] Bollerslev, T. (1986): Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics,* 31, 307-327.

[5] Bollerslev, T., Engle, R.F. and Nelson, D.B (1994): ARCH Models, in *Handbook of Econometrics*, Vol IV. Amsterdam: Elsevier Science B.V.

[6] Breidt, F., N. Crato and De Lima P. (1994): Modeling Long-Memory Stochastic Volatility, *Manuscript*, Iowa State University, Ames, IA.

[7] Breidt, F., N. Crato and De Lima P. (1998): The Detection and Estimation of Long Memory in Stochastic Volatility, *Journal of Econometrics*, 83, 325-348.

[8] Breidt, F., N. Crato and DeLima P. (2000): Modeling the Persistent Volatility of Asset Returns, *Working paper,* New Jersey Institute of Technology.

[9] Canina and Figlewski (1993): The Informational Content of Implied Volatility, *Review of Financial Studies.*

[10] Chernov, M., Gallant, A. R., Ghysels, E. and G. Tauchen (2003): Alternative Models for Stock Price Dynamics, *Journal of Econometrics*, 116, 225-257.

[11] Diebold, F.X., Inoue, A. (1999): Long Memory and Structural Change, *Manuscript*, Departmant of Finance, Stern School, NYU, May 1999.

[12] Gallant, A. R. and G. Tauchen (1996): Which Moments to Match, *Econometric Theory*, 12, 657-681.

[13] Gallant, A. R. and G. Tauchen (1998): Reprojecting Partially Observed Systems with Application to Interest rate Diffusions, *Journal of American Statistical Association,* 93, 10-24.

[14] Gallant, A. R. and G. Tauchen (2001): Efficient Method of Moments, *Discussion paper*, University of North Carolina at Chapel Hill.

[15] Geweke, J.F.and Porter-Huddak, S. (1983): The Estimation and Application of Long Memory Models and Fractional Integration, *Journal of Time Series Analysis* 1, 15-29.

[16] Ghysels, E., Harvey, A. and Renault, E. (1995): Stochastic Volatility, in *Handbook of Statistics*,14, Statistical Methods in Finance, G.S. Maddala and C. Rao (eds), North Holland, Amsterdam.

[17] Granger, C.W.J., Hyung, N. (1999): Occasional Structural Breaks and Long Memory, UCSD, *Discussion paper 99-14,* June 1999.

[18] Harvey, A. (1993): Long memory and Stochastic Volatility, *Manuscript,* London School of Economics.

[19] Liu, M. (2000): Modeling Long Memory in Stock Market Volatility, *Journal of Econometrics* 99, 139-171.

[20] Taylor, S.J (1986): Modelling Financial Time Series, *John Wiley.*

[21] Tieslau, M.A., Schmidt P. and Baillie, R. T. (1996): A Minimum Distance Estimator for Long-Memory Processes, *Journal of Econometrics*, 71, 249-264.

[22] Wright, J. H. (1999): A New Estimator of the Fractionally Integrated Stochastic Volatility Model, *Economics Letters* 63, 295-303.

Chapter 2

Methodology

In this chapter we explain the Efficient Method of Moments (EMM) estimation procedure by Gallant and Tauchen (1996), since it is used in all the work that follows. It is based on two compulsory phases. The first is projection, that consists of projecting the observed data onto a transition density that is a good approximation of the distribution implicit in the true data generating process. The simulated density is denominated the auxiliary model and its score is called "the score generator for EMM". The advantage is that the score has an analytical expression. The second phase consists of estimating the parameters of the model with the help of the score generator. This score enters the moment conditions in which we replace the parameters of the auxiliary model by their quasi-MLEs obtained in the projection step and the estimates of the model proposed are obtained by minimizing the GMM criterion function. Since the minimized criterion function scaled by the number of observations asymptotically follows a chi-square distribution, it leads to diagnostic tests that help explaining the reasons for the failure of the fitted model. Finally, the last step, called reprojection, is a post-estimation simulation analysis that allows to filter volatility encompassed by the model to evaluate the proposed models, to obtain the density and to forecast.

We can not apply Maximum Likelihood estimation methods in our work because there are some unobserved variables in the proposed models. So, for this reason the likelihood for the entire state vector is frequently not feasible. Nevertheless, the simulation of the evolution of the state vector is quite possible and the EMM is based on this. Other advantages of the EMM are the availability of diagnostic tests to assess the specification of the model and graphs that suggest reasons for failure.

Aït-Sahalia (1996a, 1996b) also developed an alternative estimation strategy for estimation stochastic differential equations. The method of estimation proposed by this author differs from the EMM because the moment functions are computed directly from the data rather than simulated. Note that full observation of the state is necessary in order to estimate all the parameters.

Recently, new methods of simulation have been developed Brandt and Santa-Clara (1999) is one example. These authors apply the simulated likelihood estimation procedures to multivariate diffusion processes. Nevertheless, these procedures have difficulties to deal with latent variables and moreover, the simulations have to be performed for every conditioning variable and for every parameter value.

Let $\{y_t\}_{t=-\infty}^{\infty}$, $y_t \,\epsilon\, \Re^M$, be a multiple, discrete stationary time series and $x_t = (y_{t-L}, ..., y_t)$ a stretch from the previous process with density $p(y_{-L}, ..., y_0|\rho)$ defined over $\Re^l, l = M(L+1)$. ρ is a vector of unknown parameters and $\{\tilde{y}_t\}_{t=-L}^{n}$ the real data from which it is to be estimated. The main problem that makes traditional methods of estimation inviable is that this density is in general not available. However, expectations of the forms

$$E_\rho(g) = \int ... \int g(y_{-L}, ..., y_0)p(y_{-L}, ..., y_0)dy_{-L}....dy_0,$$

can be approximated quite well by averaging over a long simulation

$$E_\rho(g) = \frac{1}{N}\sum_{t=1}^{N} g(\hat{y}_{t-L}, ...\hat{y}_{t-1},\ \hat{y}_t).$$

Let $\{\hat{y_t}\}_{t=-L}^{N}$ denote the simulation from $p(y|\ x, \rho)$, where $x = x_{-1} = (y_{-L}, ..., y_{-1})$, $y = y_0$ and $p(y/\ x, \rho) = p(y_{-L}, ..., y_0|\rho)/p(y_{-L}, ..., y_{-1}|\rho)$. Notice that the length of simulation should be large enough for the Monte Carlo error to be negligible.

Gallant and Tauchen (1996) proposed an estimator for the vector of parameters ρ in the situation above. This method relies on a minimum chi-square estimator for the vector of parameters, which permits the optimized chi-square criterion to be used to test the specification adopted. The moment conditions entering the minimum chi-square criterion come from the score vector $\frac{\partial}{\partial\theta}\log f(y_t|x_{t-1}, \theta)$ of an auxiliary model $f(y_t|x_{t-1}, \theta)$ that closely approximates the true density. If this is true, the EMM estimator will be nearly as efficient as the ML estimator. One commonly used auxiliary model in applications is the SNP density $f_K(y|x, \theta)$ that was proposed by Gallant and Nychka (1987). It has been showed that the efficiency of the EMM estimator can be close to that of the ML estimator if K is made large enough, Gallant and Long (1997).

The first step is to obtain the auxiliary model. Therefore, we use the SNP density that is obtained by expanding in a Hermite expansion the square root of $h(z)$, an innovation density,

$$\sqrt{h(z)} = \sum_{i=0}^{\infty}\theta_i z^i\sqrt{\phi(z)}.$$

Here $\phi(z)$ is the standard normal density function[1]. The reshaped density is given by

$$h_K(z) = \frac{P_K^2(z)\phi(z)}{\int P_K^2(u)\phi(u)du},$$

where

$$P_K(z) = \sum_{i=0}^{K}\theta_i z^i,$$

and $h_K(z)$ integrates to one since it is normalized. The SNP density is, according to the following location-scale transformation $y = \sigma z + \mu$,

[1]This expansion exists because Hermite functions are dense in L_2 and $\sqrt{h(z)}$ is an L_2 function.

$$f_K(y|\theta) = \frac{1}{\sigma} h_K(\frac{y-\mu}{\sigma}).$$

Following our notation, $h(z) = p(x, y|\rho^0)$ is the transition density and ρ^0 is the true vector of parameters. Therefore, the location-scale transformation becomes

$$y = R_x z + \mu_x,$$

where z is an innovation and R_x is an upper triangular matrix. R_x, for a GARCH specification which is the one that model the data used in this paper, is given by

$$vech(R_{x_{t-1}}) = \rho_0 + \sum_{i=1}^{L_r} P_i |y_{t-1-L_r} - \mu_{x_{t-2-L_r+i}}| + \sum_{i=1}^{L_g} diag(G_i) R_{x_{t-2-L_g+i}},$$

where $vech(R)$ is a vector of dimension $M(M+1)/2$ which contains the unique elements of the matrix R, ρ_0 denotes a vector of dimension $M(M+1)/2$, P_1 through P_{L_r} are $M(M+1)/2$ by M matrices and G_1 through G_{L_g} are vectors of length $M(M+1)/2$.

The density function of this innovation is

$$h_K(z|x) = \frac{P_K^2(z,x)\phi(z)}{\int P_K^2(u,x)\phi(u)du},$$

where $P(z,x)$ is a polynomial in (z,x) of degree K and $\phi(z)$ is the multivariate density of M independent standard normal random variables. As before, the polynomial $P_K(z,x)$ equals

$$P_K(z,x) = \sum_{\alpha=0}^{K_z} (\sum_{\beta=0}^{K_x} a_{\beta\alpha} x^\beta) z^\alpha,$$

where α and β are multi-indexes with degrees K_z and K_x, respectively. Since $h_K(z|x)$ is a homogeneous function of the coefficients of $P_K(z,x)$, it is necessary to impose a restriction $(a_{00} = 1)$ to have a unique representation.

The location function is linear

$$\mu_x = b_0 + Bx_{t-1},$$

with b_o a vector and B a matrix, both formed of parameters to be estimated.

Taking in account the location-scale transformation the SNP density becomes at last

$$f_{_K}(y|x,\theta) = \frac{h_K[R_x^{-1}(y-\mu_x)|x]}{\det(R_x)}.$$

The maximal number of lags is $L = max(L_u, L_u + L_r, L_p)$. L_u denotes the number of lags in μ_x, $L_u + L_r$ is the number of lags in R_x and finally L_p denotes the number of lags that go into the x part of the polynomial $P_K(z,x)$.

The following step is estimation. In this phase the main aims are: first of all estimate the vector of parameters ρ, test if the specification proposed for modeling the data is adequate by using the minimum chi-square criterion, and finally analyze the reasons of the system failure and shed light on the possible modifications that can better fit the data.

The EMM estimator $\hat{\rho}_n$ is determined as follows. First, we use the score generator determined in the projection step

$$f(y_t|x_{t-1},\theta) \qquad \theta \in \Re^{p_\theta}$$

and the data $\{\tilde{y}_t\}_{t=-L}^{n}$ in order to obtain the quasi-maximum likelihood estimate

$$\tilde{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{t=0}^{n} \log[f(\tilde{y}_t|\tilde{x}_{t-1},\theta)].$$

The information matrix is

$$\tilde{I}_n = \frac{1}{n}\sum_{t=0}^{n}[\frac{\partial}{\partial\theta}\log f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta}_n)][\frac{\partial}{\partial\theta}\log f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta}_n)]'.$$

In the literature it is assumed that $f(y|x, \tilde{\theta}_n)$ is a good approximation to the true density of the data. Otherwise, more complicated expressions for the weighting matrix should be used[2].

Defining the moment conditions by

$$m(\rho, \tilde{\theta}) = E_\rho\{\frac{\partial}{\partial\theta} \log f(y_t|x_{t-1}, \tilde{\theta})\},$$

which are obtained by averaging over a long simulation

$$m(\rho, \tilde{\theta}_n) = \frac{1}{N}\sum_{t=0}^{N}[\frac{\partial}{\partial\theta} \log f(\hat{y}_t|\hat{x}_{t-1}, \tilde{\theta}_n)],$$

the EMM estimator is obtained by

$$\hat{\rho}_n = \arg\min m'(\rho, \tilde{\theta}_n)(\tilde{I}_n)^{-1} m(\rho, \tilde{\theta}_n). \tag{2.1}$$

The asymptotic properties of the estimator are derived in Gallant and Tauchen (1996) and presented below. Define ρ^0 as the true value of the parameter ρ and θ^0 as an isolated solution of the moment conditions $m(\rho^0, \theta) = 0$. Then under regularity conditions it can be shown that

$$\lim_{n\to\infty} \hat{\rho}_n = \rho^0 \ a.s.,$$

$$\sqrt{n}(\hat{\rho}_n - \rho^0) \xrightarrow{D} N\{0, [(M^0)'(I^0)^{-1}(M^0)]^{-1}\},$$

[2]See Gallant and Tauchen (1996) and Gallant and Tauchen (2001). However, Gallant and Long (1997), Gallant and Tauchen (1999) and Coppejans and Gallant (2002), proved if the auxiliary model corresponds to the SNP density the information matrix above will be the adequate.

$$\lim_{n\to\infty} \hat{M}_n = M^0 \ a.s. \text{ and}$$

$$\lim_{n\to\infty} \tilde{I}_n = I^0 \ a.s.,$$

where $\hat{M}_n = M(\hat{\rho}_n, \tilde{\theta}_n)$, $M^0 = M(\rho^0, \theta^0)$, $M(\rho, \theta) = (\frac{\partial}{\partial \rho'})m(\rho, \theta)$ and

$$I^0 = E_{\rho^0}[\frac{\partial}{\partial \theta} \log f(y_0|x_{-1}, \theta^0)][\frac{\partial}{\partial \theta} \log f(y_0|x_{-1}, \theta^0)]'.$$

These asymptotic results permit testing if the model is correctly specified. Under the H_0 that $p(y_{-L}, ..., y_0|\rho)$ is the correct model

$$L_0 = nm'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{I}_n)^{-1}m(\hat{\rho}_n, \tilde{\theta}_n)$$

follows asymptotically a chi-square with $p_\theta - p_\rho$ degrees of freedom. It is also possible to test restrictions on the parameters, i.e.,

$$H_0 : h(\rho^0) = 0$$

where h is a mapping from R into $\Re^q$ and the test statistic is given by

$$L_h = n[m'(\hat{\hat{\rho}}_n, \tilde{\theta}_n)(\tilde{I}_n)^{-1}m(\hat{\hat{\rho}}_n, \tilde{\theta}_n) - m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{I}_n)^{-1}m(\hat{\rho}_n, \tilde{\theta}_n)] \qquad \overset{a}{\sim} \chi^2(q)$$

and

$$\hat{\hat{\rho}}_n = \arg\min_{h(\rho)=0} m'(\rho, \tilde{\theta}_n)(\tilde{I}_n)^{-1}m(\rho, \tilde{\theta}_n).$$

Finally, it is also possible to obtain confidence intervals for the parameters by computing the standard deviations using numeric methods. These intervals present a drawback because sometimes a parameter approaches a value for which the model is explosive and this fact is not accompanied by an increase in the EMM objective function. Gallant and Tauchen (1996) came up with a solution that consists of inverting the difference test L_h[3]. These "inverted" intervals are not free of problems. In fact, it was shown that they do not present more accurate coverage probabilities, especially when the degrees of freedom are low.

Since

$$\sqrt{n}m(\hat{\rho}_n, \tilde{\theta}_n) \xrightarrow{d} N\{0, I^0 - (M^0)[(M^0)'(I^0)^{-1}(M^0)]^{-1}(M^0)'\},$$

the t-ratios are given by

$$T_n = S_n^{-1}\sqrt{n}m(\hat{\rho}_n, \tilde{\theta}_n),$$

where $S_n = (diag\{\tilde{I}_n - (\hat{M}_n)[(\hat{M}_n)'(\tilde{I}_n)^{-1}(\hat{M}_n)]^{-1}(\hat{M}_n)'\})$. The characteristics of the data are reflected in the different elements of score. If the model fails to fit these characteristics this fact comes out in the large values taken by the t-ratios (of the elements of the score). In this case, the failure can suggest alternative modelizations.

[3] In order to invert the test we select for the interval that values of ρ_i^* for which the H_0: $\rho_i^0 = \rho_i^*$ is not reject under the test L_h.

Bibliography

[1] Aït- Sahalia, Yacine (1996a): Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies 9*, 385-342.

[2] Aït- Sahalia, Yacine (1996b): Nonparametric Pricing of Interest Rate Derivative Securities, *Econometrica* 64, 527-560.

[3] Brandt, M. and Santa-Clara, P. (1999): Simulated Likelihood Estimation of multivariate Diffusions with an Application to Interest Rates and Exchange Rates with Stochastic Volatility, *Working paper*, UCLA.

[4] Coppegans, M. and Gallant, A.R. (2002): Cross-Validated SNP Density Estimates, *Journal of Econometrics*, 110, 27 - 65.

[5] Gallant, A.R. and J.R. Long (1997): Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square, *Biometrika*, 84, 125-141.

[6] Gallant, A. R. and D.W. Nychka (1987): Semi-Nonparametric Maximum Likelihood Estimation, *Econometrica* 55, 363-390.

[7] Gallant, A. R. and G. Tauchen (1996): Which Moments to Match?, *Econometric Theory*, 12, 657- 681.

[8] Gallant, A. R. and G. Tauchen (1999): The Relative Efficiency of Methods of Moments Estimators, *Journal of Econometrics* 92, 149-172.

[9] Gallant, A. R. and G. Tauchen (2001): Efficient Method of Moments, *Discussion paper*, University of North Carolina at Chapel Hill.

Chapter 3

Are One Factor Logarithmic Volatility Models Useful to Fit the Features of Financial Data? An Application to Microsoft Data

3.1 Introduction

A volatility model should be able to model the main characteristics of financial series of returns such as volatility persistence, volatility clustering, leverage effects, fat tails and small first order autocorrelation of squared returns. During the three last decades several models have been proposed with the aim of capturing these empirical facts. Stochastic volatility models, for instance, were designed to fit mainly volatility persistence but recent empirical work has found that these models fail in capturing the fat tails of the returns' distribution, Chernov and Ghysels (2000).

Gallant and Tauchen (2001) and Chernov et al. (2003) propose several models in continuous time and evaluate the importance of several volatility factors to the modelization of equity returns. Both papers provide empirical evidence that continuous-time stochastic volatility models with only one volatility factor are not able to capture simultaneously extra kurtosis and volatility persistence. The introduction of a second factor of volatility allows that one might be slow mean reverting while the other might accommodate the fat tails.

This chapter differs from previous works in that it provides empirical evidence that continuous time models with one factor of volatility, in some conditions, are able to fit the main characteristics of financial data. It also reports the importance of the feedback factor as a possible imperfect substitute of an increase in volatility caused by stock market crashes. The estimated models are direct extensions of Gallant and Tauchen's (2001) model, by including the feedback feature. The paper of Chernov et al. (2003) also presents logarithmic models in continuous time but our specifications differ from the previous because they do not allow for stochastic instantaneous expected returns (the drift of the return equation is not stochastic) and leverage effects. The advantage of these modelizations compared to an affine specification is that they allow that the volatility be dependent on state, although there are not closed-form solutions. However, pricing formulas may be computed by simulation. Chernov et al. (2003) consider this an advantage when compared to the risk-neutral measure transformations used by the affine models.

The empirical results report that the one factor logarithmic volatility model without feedback does not fit the Microsoft data which confirms prior findings in the literature. A new result comes out when we introduce the feedback factor. The model, now, does pass the specification test. This feature is of extreme importance because it allows capturing the low variability of the volatility factor when the factor is itself low (volatility clustering). The feedback factor also allows capturing the increase in volatility persistence that occurs when there is an apparent change in the pattern of volatility. The introduction of a second factor of volatility with feedback does not seem relevant for the Microsoft data.

This chapter is organized as follows. Section two presents and characterizes the models we study. Section three covers the projection, estimation and reprojection steps and reports the empirical results for the Microsoft data. Section four concludes the chapter.

3.2 Continuous Time Stochastic Volatility Logarithmic Models

Recently researchers tend to model volatility as stochastic. The literature is vast referring to the estimation of models with or without stochastic volatility or with or without jumps, see Bates (2000), Chernov et al. (2003), Gallant and Tauchen (2001), Ghysels et al. (1995), for example.

Our model is the following:

$$\frac{dP_t}{P_t} = \alpha_{10}dt + \exp(\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t})dW_{1t} \tag{3.1}$$

$$dU_{2t} = (\alpha_{20} + \alpha_{22}U_{2t})dt + (\beta_{20} + \beta_{22}U_{2t})dW_{2t} \tag{3.2}$$

$$dU_{3t} = (\alpha_{30} + \alpha_{33}U_{3t})dt + (\beta_{30} + \beta_{33}U_{3t})dW_{3t} \tag{3.3}$$

where P_t is the Microsoft price series evolving in continuous time and W_i with $i = 1, 2, 3$ are three independent wiener processes. This specification is an extension to Gallant and Tauchen (2001) model since it includes the feedback features $\beta_{22}U_2$ and $\beta_{33}U_3$ in the equations 3.2 and 3.3, respectively.

In this system the instantaneous standard deviation of the rate of return is an exponential function of the factors U_{2t} and U_{3t}[1]. This specification nests two groups of models: the first includes the logarithmic model with one volatility factor ($L1$), with $\beta_{13} = 0$ and

[1]As Chernov et al. (2003) refer, the logarithmic models with feedback violate the standard regularity

$\beta_{22} = 0$, and the logarithmic model with two volatility factors ($L2$), with $\beta_{13} \neq 0$, $\beta_{22} = 0$ and $\beta_{33} = 0$ and the second group contains the one factor logarithmic volatility model with feedback ($L1F$), where $\beta_{13} = 0$ and $\beta_{22} \neq 0$, and the logarithmic model with two factors of volatility and feedback ($L2F$), where $\beta_{13} \neq 0$, $\beta_{22} \neq 0$ and $\beta_{33} \neq 0$. One advantage of the feedback feature is to allow for volatility clustering. The empirical results, later on, reveal that this feature may be quite relevant especially when there seems to exist a change in the pattern of volatility. We could suspect that the increase in volatility caused by stock market crashes and the feedback feature could be imperfect substitutes, in the sense that the introduction of this feature could help in capturing the increase in volatility caused by stock market crashes[2]. We do not introduce jumps in the specifications because the possible change in the pattern of volatility occurs at the very end of the sample. Moreover, the volatility factors of equation 3.1 present drifts and volatilities that are linear functions of themselves, respectively and the drifts in equations 3.2 and 3.3 allow for mean reversion when $\alpha_{ii} \neq 0$ for $i = 2, 3$. A small value of α_{ii} for $i = 2, 3$ means that a shock to the volatility of the return takes time to dissipate. This is referred in the financial econometrics literature as persistence or long memory and a large percentage of the financial series seem to show this feature, Zaffaroni (2000). Finally, β_{10} is also an important parameter since it takes care of the long-run mean of the volatility of the price equation 3.1.

conditions. Hence, the solutions of the system of SDEs associated with these models and the stochastic integrals are not defined. In order to solve this problem they propose a method that consists on splicing the exp(.) function that models the volatility behavior of equation 3.1 with, as they say, "...the linear growth condition at the level of volatility is so high that it is unlike to be observed in the U.S. equity index returns." For more details please see the appendix A of their paper.

[2] We think to test this suspicion using a Monte Carlo experiment.

Identification restrictions

To achieve identification it is necessary to impose some restrictions. In this concrete case for the logarithmic specification we set

$$\alpha_{20} = 0, \alpha_{30} = 0, \beta_{20} = 1, \beta_{30} = 1.$$

These restrictions are the minimum necessary to achieve identification.

Therefore the previous specification becomes, for the first group:

$$\frac{dP_t}{P_t} = \alpha_{10}dt + \exp(\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t})dW_{1t} \tag{3.4}$$

$$dU_{2t} = \alpha_{22}U_{2t}dt + dW_{2t} \tag{3.5}$$

$$dU_{3t} = \alpha_{33}U_{3t}dt + dW_{3t} \tag{3.6}$$

with $\beta_{13} = 0$ or $\beta_{13} \neq 0$ if we refer to $L1$ or $L2$, respectively.

For the second group:

$$\frac{dP_t}{P_t} = \alpha_{10}dt + \exp(\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t})dW_{1t} \tag{3.7}$$

$$dU_{2t} = \alpha_{22}U_{2t}dt + (1 + \beta_{22}U_{2t})dW_{2t} \tag{3.8}$$

$$dU_{3t} = \alpha_{33}U_{3t}dt + (1 + \beta_{33}U_{3t})dW_{3t} \tag{3.9}$$

with $\beta_{13} = 0$ or $\beta_{13} \neq 0$ if we refer to $L1F$ or $L2F$, respectively. The first group of SDE was already estimated by Gallant and Tauchen (2001) for a small sample of Microsoft data.

We use these restrictions first because they are common in previous similar SDE and second because they provide flexibility and numerical stability in the estimation phase.

3.3 Empirical Results

3.3.1 SNP estimation results

In this subsection of the chapter we present the results of the projection step.

The auxiliary model that best fits the raw data is found using the SNP model described in the second chapter. The data is composed of 3,778 observations on a daily price of a share of Microsoft, adjusted for stock splits, from 13^{th} of March, 1986, till 23^{rd} of February, 2001 (see Figures 1, 2 and 3). The first 47 observations were reserved for forming lags. The values taken by L_u, L_g, L_r, L_p, K_z and K_x were determined by going along a expansion path and the selection criterion used was the BIC (Bayesian Information Criterion), Schwarz (1978).

As always, models that present a small value for the BIC criterion are preferred to the ones with higher values. The expansion path has a tree structure. As Gallant and Tauchen (1996) suggested, better than expanding the entire tree structure is to start expanding L_u keeping $L_r = L_p = K_z = K_x = 0$ until the BIC increases value. The following step is to expand in L_r with $L_p = K_z = K_x = 0$. Next, one expands K_z with $K_x = 0$ and finally L_p and K_x. Sometimes it can happen that the smallest value of the BIC is somewhere inside the tree. So, it is convenient for this reason to expand K_z, L_p and K_x at a few intermediate values of L_r.

The best model according to this procedure[3] has

$$L_u = 1, L_r = 1, L_g = 1, L_p = 1, K_z = 6 \text{ and } K_x = 0$$

and can be characterized as a Semiparametric GARCH.

[3] This strategy reveals itself reasonable in much applied work, Fenton and Gallant (1996b). Gallant and Tauchen (2001) also arrived at the same specification.

3.3.2 The estimation step

All the estimated results were obtained using the computer package EMM programmed by Gallant and Tauchen (1996) with Fortran 77 available at ftp.econ.duke.edu. The global minima of equations 3.4 and 3.9 were found through an exhaustive search grid of the starting values and the help of randomization.

Table 2 gives a summary of the specifications presented in section two and shows the value of the diagnostic test which follows an asymptotic chi-square distribution with $p_\theta - p_\rho$ degrees of freedom. From the table and in particular from the chi-square test, we can infer that the results for the one factor volatility model without feedback confirm prior findings in the literature. The model is sharply rejected at a 5% level of confidence. A new result comes out when we introduce the feedback factor. It turns out not only significant but also it is of vital importance for the good fit of the model that now passes the specification test without violating any of the moment conditions (see Tables 2 and 4). When we analyze the estimates for this latter model, we see that all coefficients are statistically significant. The feedback factor turns out to be very relevant, with a negative value. This implies that if now the volatility factor U_2 is high its instantaneous volatility decreases and in the future the volatility factor U_2 is expected to decrease. This combined with the negative value of β_{12} (the coefficient of the volatility factor in (4)) makes perfect sense and matches financial theory[4]. So this feature allows that the variability of the volatility factor is low when itself is low (volatility clustering). Another characteristic that comes out from the estimation is the value of the parameter that corresponds to the mean reversion feature, α_{22}. Its value is inferior to unity. Thus, shocks to volatility of returns take time to dissipate - the long memory property. If we also observe the graph of volatility we will see an increase in the volatility for the last period of the sample (see

[4]This is so because if the volatility factor is high now the instantaneous volatility of the return decreases, implying an expected decrease of the return in the future.

Figures 3 and 3.1) caused by stock market crashes. Recent studies, for instance: Beine and Laurent (2000), Granger and Hyung (1999) and Diebold and Inoue (1999) report that there is an increase in volatility persistence if we do not account for the possibility of pattern changes. In order to investigate this, we consider the sample used in Gallant and Tauchen (2001) that ranges from March 13, 1986 till November 16, 2000 and our sample. We compute the autocorrelation functions (ACF's) of the squared returns of the absolute values of returns and we observe specially for the latter that the ACF decays slower towards zero (see Figures 5 and 6). We also compare their $L1$ model results with ours and we observe that the parameter of mean reversion, α_{22}, is abruptly greater than one in absolute value, which means fast mean reversion and consequently low persistence in volatility. In contrast, the same specification estimated considering the sample used in this chapter reports an empirical result for that parameter of -0.902, which is much smaller in absolute value than the previous one (see Table 3). Both pieces of evidence are signals of an increase in persistence in the presence of structural changes in volatility. This extra persistence leads to volatility clustering with periods of low volatility being followed by periods of low volatility and vice-versa. Therefore, the estimation results may suggest that change in the pattern of volatility and feedback factor may be imperfect substitutes (the latter can capture the former by allowing for volatility clustering that results from an increase in persistence).

Although the frequency of data is daily, it is scaled so that the coefficients are on an annual basis. That is, a value of 0.4102 for α_{10} represents an annual average rate of return equal to 41.02%. The step size is $\Delta = 1/6048$, which corresponds to 24 steps per day and 252 trading days per year.

Since the feedback factor reveals itself of extreme importance, we estimate a two factors logarithmic volatility model incorporating this feature. Analyzing the results we can say that for all the lengths the parameter β_{13} is not significant, which means that for this data and for this sample, the second factor of volatility is unimportant. We report its results for $N = 100\ 000$ in table 3.

Finally, we estimate the $L2$ specification as in Gallant and Tauchen (2001) and we infer from the results that this model is another possible candidate to model the data. We observe that one factor of volatility is extremely slow mean reverting while the other is very fast mean reverting[5].

Finally, the Table 4 summarizes the EMM quasi-t-ratios diagnostics for $L1$, $L1F$ and $L2$. There is evidence that these statistics are asymptotically downward biased. Gallant and Tauchen (1996) presented some corrections to this t-ratios but recent evidence showed that they might not be especially reliable when there are few degrees of freedom. Consequently, in this chapter we present only the unadjusted t-ratios, without forgetting the downward bias. Relatively to the one factor logarithmic volatility model without feedback, it does not seem to fit the scores corresponding to the GARCH scale. This may be due to the strong persistent stochastic volatility in the data. When we introduce the feedback factor, none of the scores are violated, i.e., the model seems to fit the Hermite parameters as well as the GARCH parameters. The same for model $L2$.

From the estimation step, two models come out, $L1F$ and $L2$. It is not possible to choose between them based on the diagnostics computed at this step. The reprojection step will give us more tools that will help us to evaluate their performance.

3.3.3 The reprojection step

The reprojection step allows us to filter the volatility factors U_{2t} and U_{3t} for any desired sampling frequency. In fact, as a by-product of the estimation step we obtain a long simulation of the volatility factors $\{\hat{U}_{2t}\}_{t=1}^{N}$ and $\{\hat{U}_{3t}\}_{t=1}^{N}$. Having as the main aim to

[5] As in Gallant and Tauchen (2001). All the coefficients are statistically significant at a 5% significance level, except α_{22} that is significant at a 10% significance level. We consider it different from zero, otherwise the model would be similar to the L1 model, which has been sharply rejected by the specification test.

obtain

$$E(U_{2t}|\{y_\tau\}_{\tau=1}^t),$$

and

$$E(U_{3t}|\{y_\tau\}_{\tau=1}^t),$$

we start generating simulations of $\{\hat{U}_{2t}\}_{t=1}^N$, $\{\hat{U}_{3t}\}_{t=1}^N$ and $\{\hat{y}_t\}_{t=1}^N$ at the estimated vector of parameters $\hat{\rho}$ and with N equal 100 000. Then, we impose the same SNP-GARCH model founded in the projection step, on the simulated values $\hat{y}_t$. According to Gallant and Tauchen (2001), this provides a good representation of the one-step ahead conditional variance $\hat{\sigma}_t^2$ of $\hat{y}_{t+1}$ given $\{\hat{y}_\tau\}_{\tau=1}^t$. We follow by running regressions of $\hat{U}_{2t}$ and $\hat{U}_{3t}$ on $\hat{\sigma}_t^2$, $\hat{y}_t$ and $|\hat{y}_t|$ and lags of these series:

$$\begin{aligned}\hat{U}_{2t} &= \alpha_0 + \alpha_1\hat{\sigma}_t^2 + \alpha_2\hat{\sigma}_{t-1}^2 + ... + \alpha_p\hat{\sigma}_{t-p}^2 + \theta_1\hat{y}_t + \theta_2\hat{y}_{t-1} + ... \\ &\quad +\theta_q\hat{y}_{t-q} + \pi_1|\hat{y}_t| + \pi_2|\hat{y}_{t-1}| + ... + \pi_r|\hat{y}_{t-r}| + u_t,\end{aligned}$$

$$\begin{aligned}\hat{U}_{3t} &= \beta_0 + \beta_1\hat{\sigma}_t^2 + \beta_2\hat{\sigma}_{t-1}^2 + ... + \beta_p\hat{\sigma}_{t-p}^2 + \gamma_1\hat{y}_t + \gamma_2\hat{y}_{t-1} + \\ &\quad ... + \gamma_q\hat{y}_{t-q} + \lambda_1|\hat{y}_t| + \lambda_2|\hat{y}_{t-1}| + ... + \lambda_r|\hat{y}_{t-r}| + \mu_t.\end{aligned}$$

With this procedure we obtain calibrated functions inside the simulation that gives predicted us values of U_{2t} and U_{3t} given $\{y_\tau\}_{\tau=1}^t$. In fact, given the length of the simulation, these regressions are, as Gallant and Tauchen (2001) say, analytic projections. Finally, we evaluate these functions on the observed data series $\{\hat{y}_\tau\}_{\tau=1}^t$ to obtain reprojected values of the volatility factors, $\tilde{U}_{2t}$ and $\tilde{U}_{3t}$.

Figures 7.0, 7.1, 8 and 9 show the reprojected volatility factors of models $L2$ and $L1F$, respectively. As to be expected $\tilde{U}_{3t}$ for the $L2$ is quite choppy and $\tilde{U}_{2t}$ is slightly slower moving than $\tilde{U}_{3t}$ as we can verify by figures 7.0 and 7.1[6]. Curiously, the increase in volatility in the last part of the sample and the crash of 1987 are attributed in its majority to the fast mean reverting factor, U_{3t} . This suggests that both events were temporary. Finally, the reprojected volatility factor from the $L1F$ model is the most "alive" of the three. It tracks quite well the patterns of both factors in the previous model $L2$ and it captures some extra noise in the volatility. So, it seems, for the purpose of volatility modelling , that the $L1F$ specification works quite well.

3.4 Conclusion

This chapter studies four systems of SDE for modelling the daily return on the Microsoft shares, $L1$, $L1F$, $L2$ and $L2F$. From the diagnostics at the estimation step two models seem to fit the data well, $L1F$ and $L2$. One possible reason for the failure of the model with only one volatility factor could be its inadequacy to model the strong persistent stochastic volatility caused by stock market crashes. This drawback, however, is overcome by introducing the feedback factor. It allows for volatility clustering and it is able to capture the strong persistence. The model, now, seems to fit all the score moment conditions associated with the GARCH parameters as well as the score moment conditions corresponding to the Hermite parameters responsible for the tail behavior. The second valid model that comes out from estimation is the logarithmic with two volatility factors.

Reprojection assumes an important role in the model selection since it gives us more tools for comparing models. By computing the reprojected volatility factors implied by the previous specifications we see that there is no advantage in estimating the two factors

[6] $\tilde{U}_{2t}$ could be much more slowly moving as in Gallant and Tauchen (2001) . The fact that it is not, can be justified by the value of the t-statistic for α_{22}, 1.86667 (that is not statistically significant at a 5% significance level). We considered it significant due to the possibility of computational error in the wald standard deviation justified by the relatively big amplitude of the confidence interval and its asymmetry to the estimate.

stochastic volatility model for this sample. The $L1F$ model is able to reproject volatility quite well. It even does not miss the crash of 1987.

For the more complicated specification $L2F$, the empirical results show that the second factor is not significant.

Bibliography

[1] Bates, D. (2000): Post-'87 Crash Fears in the S&P 500 Futures Option market, *Journal of Econometrics*, 94, 181-238.

[2] Beine, M. and Laurent, S. (2000): Structural Change and Long Memory in Volatility: New Evidence from Daily Exchange Rates, *Discussion paper*, University of Liège.

[3] Chernov, M., Gallant, A. R., Ghysels, E. and G. Tauchen (2003): Alternative Models for Stock Price Dynamics, *Journal of Econometrics*, 116, 225-257.

[4] Chernov, M. and Ghysels, E. (2000): A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation, *Journal of Financial Economics* 56, 407-458.

[5] Diebold, F.X., Lee, Inoue, A. (1999): Long Memory and Structural Change, *Manuscript*, Department of Finance, Stern School, NYU, May 1999.

[6] Fenton, V.M. and Gallant, A. R. (1996): Qualitative and Asymptotic Performance of SNP Density Estimators, *Journal of Econometrics*, 74, 74 -118.

[7] Gallant, A. R. and G. Tauchen (1996): Which Moments to Match, *Econometric Theory*, 12, 657-681.

[8] Gallant, A. R. and G. Tauchen (2001): Efficient Method of Moments, *Discussion paper*, University of North Carolina at Chapel Hill.

[9] Granger, C.W.J. and Hyung, N. (1999): Occasional Structural Breaks and Long Memory, UCSD, *Discussion paper* 99-14, June 1999.

[10] Ghysels, E. Harvey, A. and Renault, E. (1995): Stochastic Volatility, in Handbook of Statistics, 14, Statistical Methods in Finance, G.S. Maddala and C. Rao (eds), North Holland, Amsterdam.

[11] Schwarz, G. (1978): Estimating the Dimension of a Model, *The Annals of Statistics*, vol 6, no. 2, pp 461-464.

[12] Zaffaroni, P. (2000): Time Series Models of Changing Volatility, *Lecture notes* 4, 1569-1622.

3.5 Figures and Tables

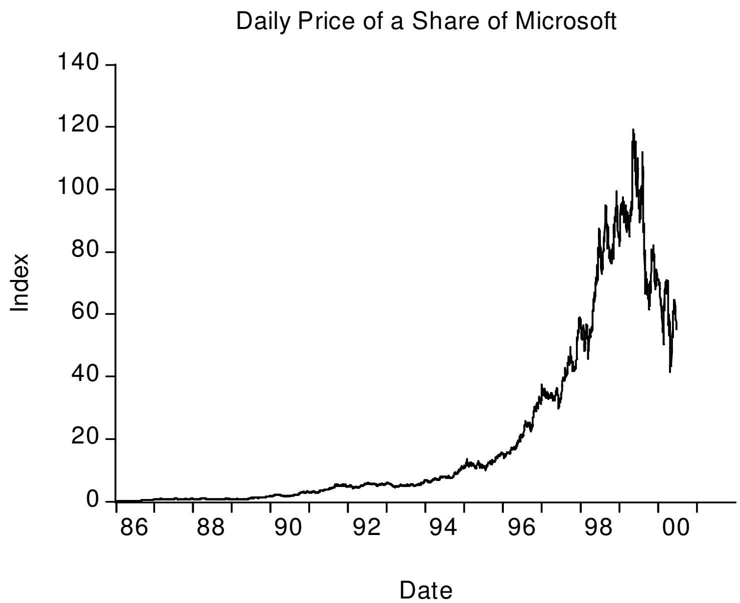

Figure 1

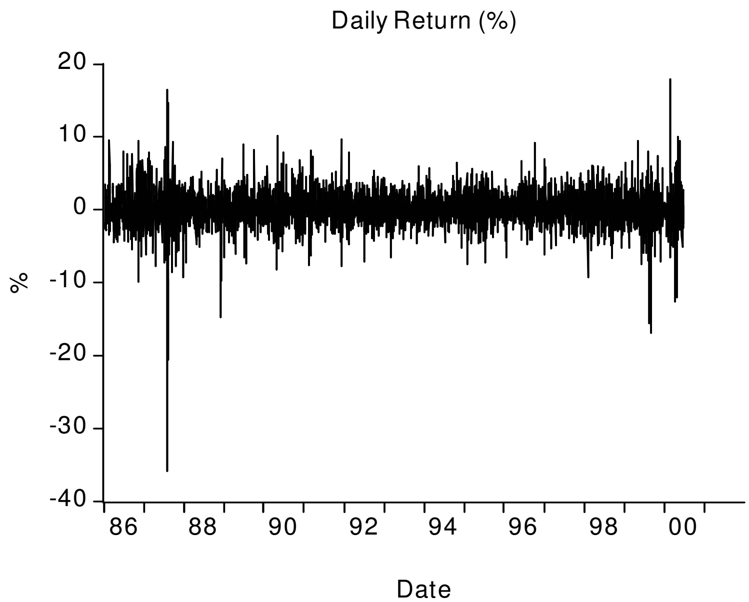

Figure 2

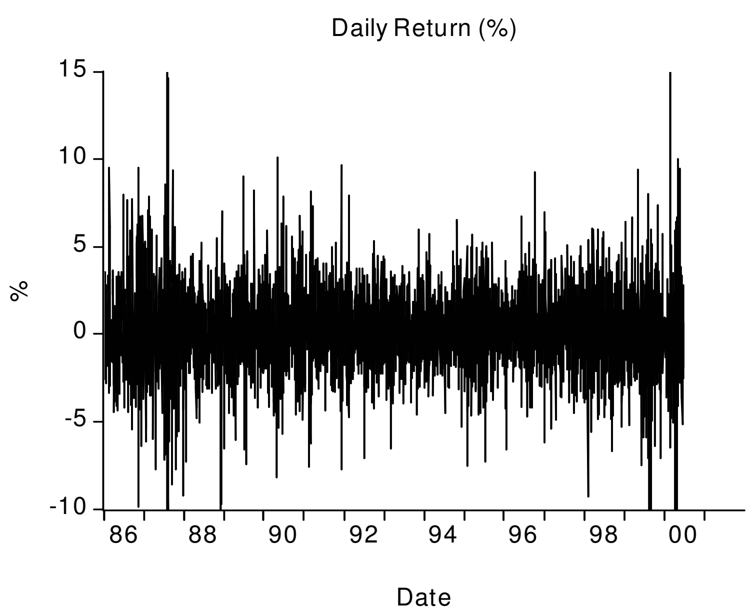

Figure 2.1: Different Scale

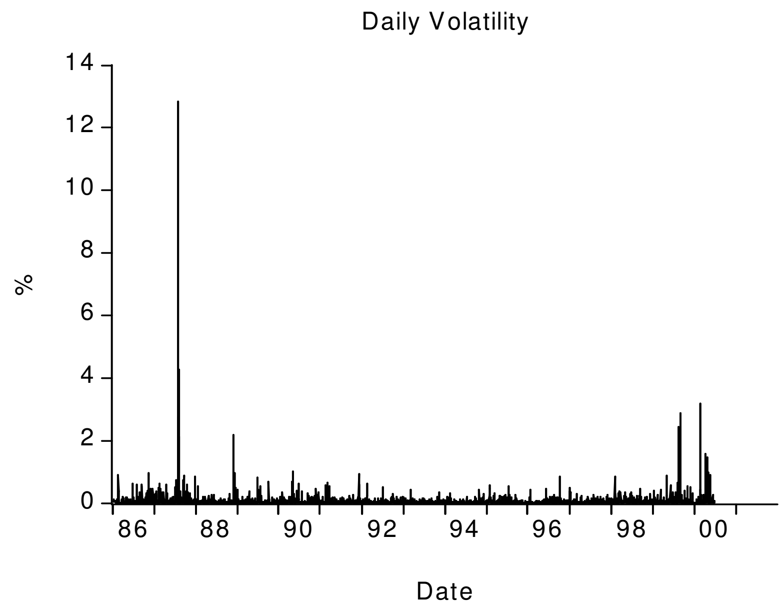

Figure 3

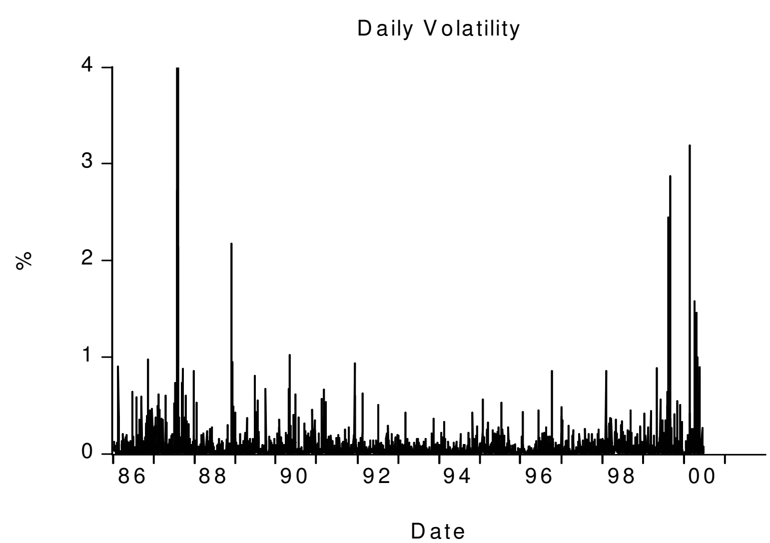

Figure 3.1: Different Scale

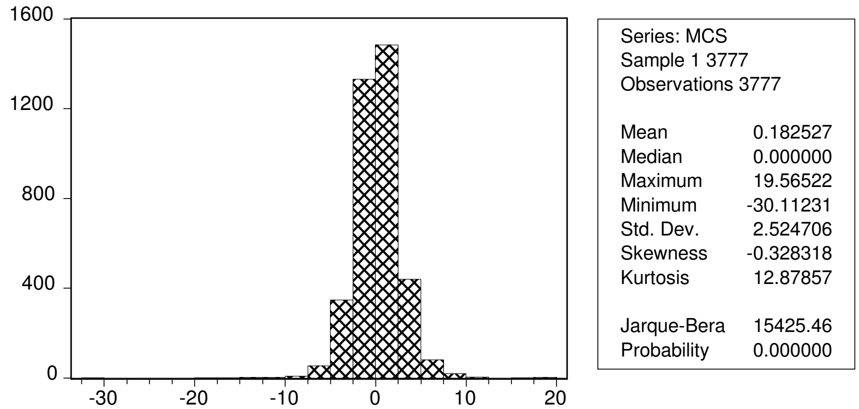

Figure 4: Histogram of the Returns on Microsoft, March 14, 1986 - February 23, 2001.

			Parameters			
$L_u = 0$	$L_g = 0$	$L_r = 0$	$L_p \geq 0$	$K_z = 0$	$K_x = 0$	iid Gaussian
$L_u > 0$	$L_g = 0$	$L_r = 0$	$L_p \geq 0$	$K_z = 0$	$K_x = 0$	Gaussian VAR
$L_u > 0$	$L_g = 0$	$L_r = 0$	$L_p \geq 0$	$K_z > 0$	$K_x = 0$	Semiparametric VAR
$L_u \geq 0$	$L_g = 0$	$L_r > 0$	$L_p \geq 0$	$K_z = 0$	$K_x = 0$	Gaussian ARCH
$L_u \geq 0$	$L_g = 0$	$L_r > 0$	$L_p \geq 0$	$K_z > 0$	$K_x = 0$	Semiparametric ARCH
$L_u \geq 0$	$L_g > 0$	$L_r > 0$	$L_p \geq 0$	$K_z = 0$	$K_x = 0$	Gaussian GARCH
$L_u \geq 0$	$L_g > 0$	$L_r > 0$	$L_p \geq 0$	$K_z > 0$	$K_x = 0$	Semiparmetric GARCH
$L_u \geq 0$	$L_g > 0$	$L_r > 0$	$L_p \geq 0$	$K_z > 0$	$K_x > 0$	Nonlinear Nonparametric

Table 1: note:" L_u is the lag length of the location function. L_g is the lag lenght of the GARCH part of the scale function. L_r is the lag lenght of the ARCH part of the scale function. L_p is the lag length of the polynomials in x. K_x is the degree of polynomials in x that determine the coefficients of the Hermite expansion of the inovation density", Gallant and Tauchen (2001). The table is similar to the table in SNP guide, ftp:econ.duke.edu/get.

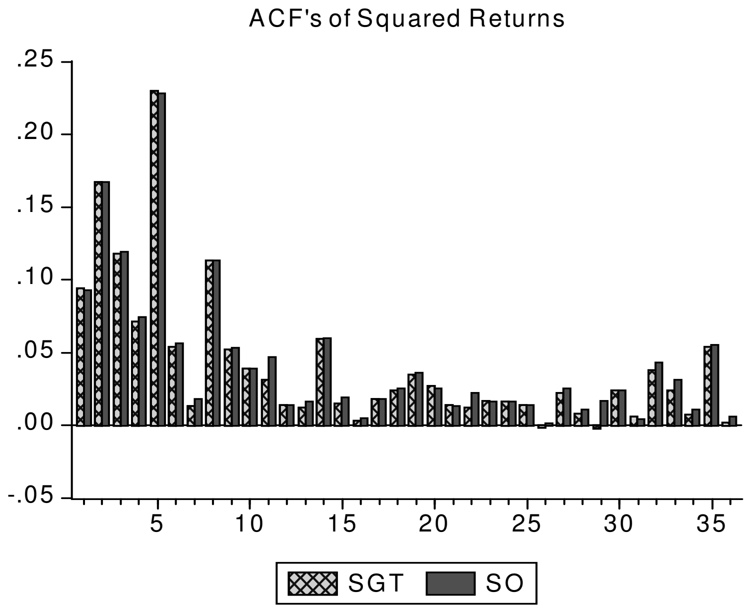

Figure 5: SGT - ACF of Gallant and Tauchen (2001). SO - ACF of this paper.

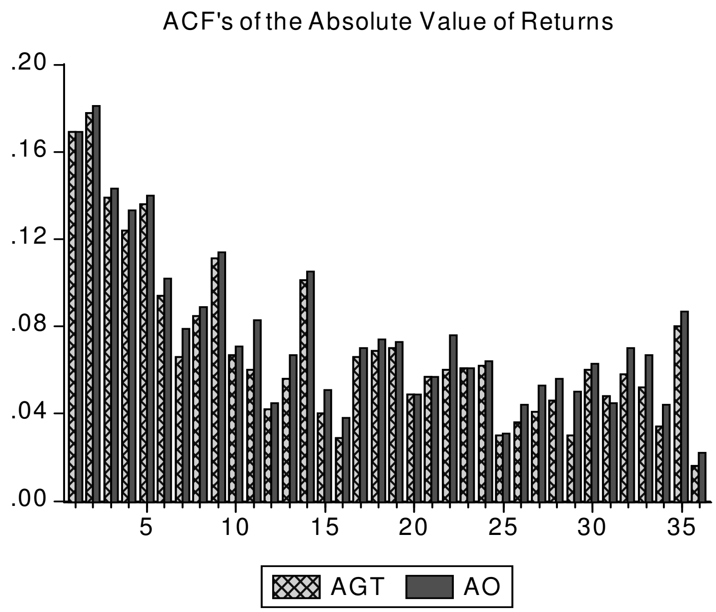

Figure 6: AGT - ACF of Gallant and Tauchen (2001). AO - ACF of this paper.

Model	α_{10}	α_{22}	α_{33}	β_{10}	β_{12}	β_{13}	β_{22}	β_{33}	N	χ^2	df	p-val
L1	*	*		*	*				50k	15.983	7	0.0253
L1	*	*		*	*				100k	16.650	7	0.0198
L1F	*	*		*	*		*		50k	9.417	6	0.1514
L1F	*	*		*	*		*		100k	10.635	6	0.1003
L1F	*	*		*	*		*		150k	8.078	6	0.2324
L1F	*	*		*	*		*		175k	8.463	6	0.2061
L2	*	*	*	*	*	*			100k	6.704	5	0.2436
L2F	*	*	*	*	*	*	*	*	50k	5.568	4	0.2338
L2F	*	*	*	*	*	*	*	*	100k	2.015	4	0.7330

Table 2: *is used for free parameters. 175k refers to a simulation of length 175,000 at step size $\Delta = 1/6048$, corresponding to 24 steps per day and 252 trading days per year.

Model	α_{10}	α_{22}	α_{33}	β_{10}	β_{12}	β_{13}	β_{22}	β_{33}
L1								
Estimate	0.42	-0.902		-0.124	0.429			
Std. Dev.	0.073	0.097		0.003	0.038			
95% Lower	0.270	-1.048		-0.132	0.373			
95% Upper	0.571	-0.757		-0.116	0.485			
L1F								
Estimate	0.410	-0.159		-0.112	-0.160		-0.224	
Std. Dev.	0.037	0.002		0.003	0.003		0.004	
95% Lower	0.326	-0.164		-0.119	-0.166		-0.234	
95% Upper	0.499	-0.155		-0.104	-0.153		-0.215	
L2								
Estimate	0.424	-0.00028	-89.21	-0.120	0.0063	-4.628		
Std. Dev.	0.074	0.00015	3.932	0.0087	0.0010	0.076		
95% Lower	0.269	-0.00049	-97.154	-0.123	0.0043	-4.778		
95% Lower	0.579	-0.000078	-81.432	-0.097	0.0083	-4.480		
L2F								
Estimate	0.415	-2.161	0.214	-0.126	0.723	-0.504	0.733	7.190
Std. Dev.	0.084	0.077	2.691	0.089	0.035	6.668	0.623	0.635
95% Lower	0.414	-2.162	0.172	-0.127	0.723	-0.607	0.727	7.180
95% Lower	0.415	-2.161	0.252	-0.124	0.727	-0.408	0.740	7.199

Table 3: Estimates, Standard Deviations and Confidence Intervals

Coefficient		L1	L1F	L2
Location Function:				
b_0	psi(1)	0.066	0.361	0.101
b_1	psi(2)	1.685	0.839	1.239
Scale Function:				
τ_0	tau(1)	1.824	0.377	0.764
τ_{gz}	tau(2)	2.540	0.132	1.032
τ_{gx}	tau(3)	2.315	0.282	0.921
Hermite Polynomial:				
$a_{0,1}$	A(2)	0.081	0.453	0.358
$a_{0,2}$	A(3)	1.868	1.212	-0.018
$a_{0,3}$	A(4)	-0.069	0.544	0.500
$a_{0,4}$	A(5)	1.809	1.981	0.146
$a_{0,5}$	A(6)	-0.418	0.499	0.418
$a_{0,6}$	A(7)	1.226	1.858	-0.032

Table 4: Scores Diagnostic

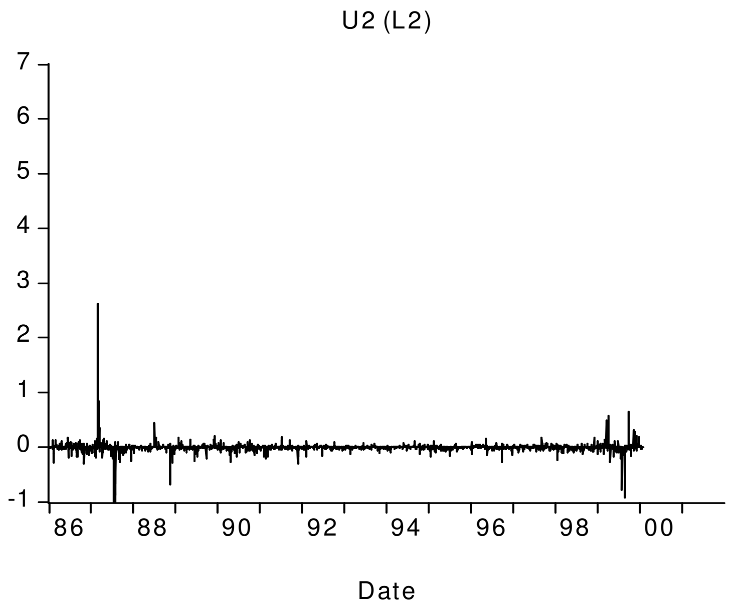

Figure 7.0

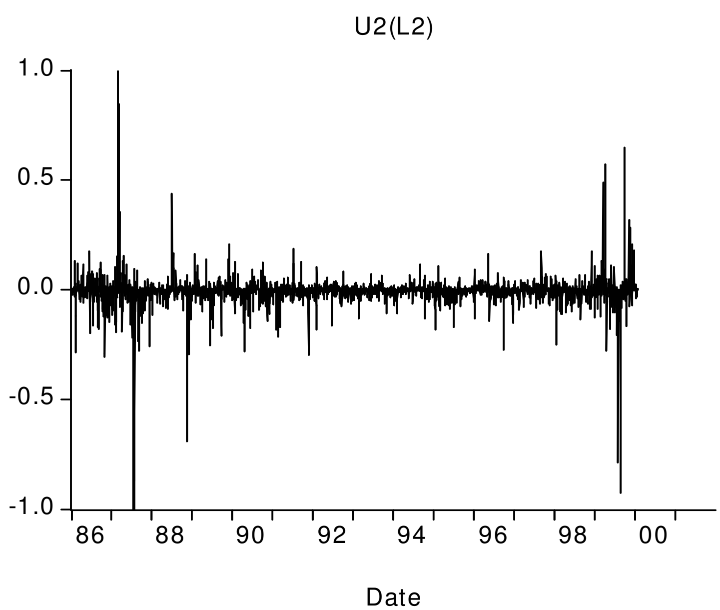

Figure 7.1 - Diferent Scale

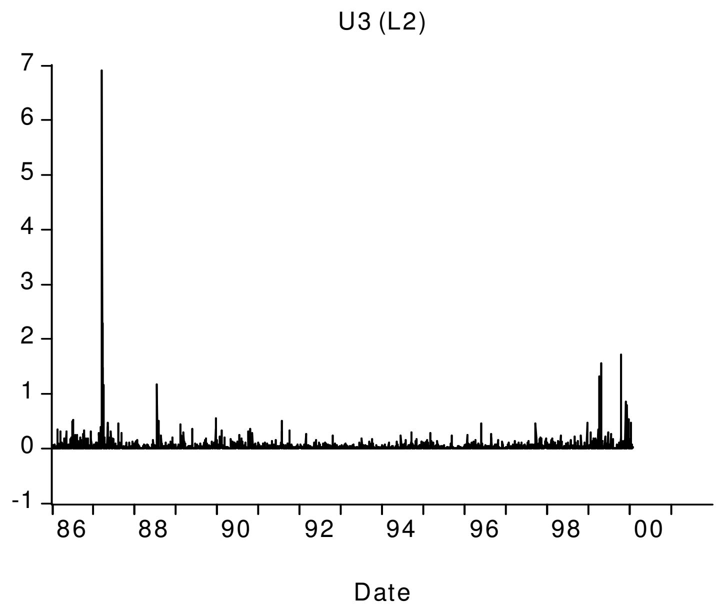

Figure 8

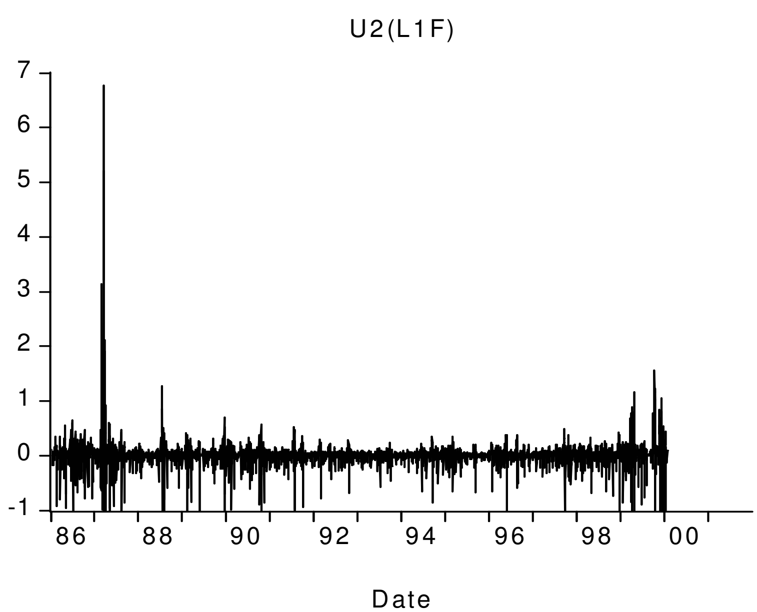

Figure 9

Chapter 4

Forecasting Volatility Using A Continuous Time Model

4.1 Introduction

Volatility plays an important role for asset pricing theory as it is directly linked to the risk-return relation. Good measures and forecasts of future volatility are of vital importance for finance theory[1]. One measure of volatility has been "the implied volatility" obtained from Black-Scholes model, but as empirical evidence has been showing, the performance of this model is not the same in all periods. In fact, after the October 1987 stock market crash, the model is not adjusting reality so well and the implied volatility seems to have little predictive power relative to the historical volatility, Canina and Figlewski (1993). On the other hand, the standard well known ARCH/GARCH models seem to perform poorly in forecasting volatility (see Day and Lewis (1992), Jorian (1996), Pagan and Schwert (1990), West and Cho (1995), etc).

Nowadays with the availability of high-frequency data and the advances of compu-

[1] Notice that volatility is used as a measure of risk. The higher is the volatility, the higher is the risk and consequently the higher is going to be the expected return.

tational tools it is possible to improve volatility forecasting. There are two main ways of forecasting volatility. The first way treats volatility as observed and uses intra-period data to get the so called "realized volatility" obtained basically by summing intra-period squared returns and fitting to them models that incorporate the main features of this data, for instance, the long-memory property. Since much theoretical work assumes that the logarithm of asset prices follows a continuous time model, like a diffusion, one advantage of this procedure is that the realized volatility can be made arbitrarily close to the integral of instantaneous volatility over the period by reducing the intra-period. However, if we choose tick-by-tick prices we might have to use some forms of interpolation since these prices are not generally available at regular time intervals. Empirical evidence shows that this can cause negative correlation in the returns series and consequently lead to poor forecasts of models using this data. Other problems are related with changes of patterns in volality due to market microstructures, for instance the existence of lunch periods, the close of the market, etc.. To attempt to avoid these drawbacks, in this chapter, we use 15-minutes intraday data. The second way of forecasting volatility treats volatility as latent in the sense that it can be filtered after estimation. Getting a correct specification is pivotal since volatility estimates are model dependent. In this chapter, we follow this second alternative and we fit the continuous time model with two factors of volatility of Gallant and Tauchen (2001) to the returns of Microsoft.

To sum up, the aim of this chapter is to evaluate the volatility forecasting performance of the continuous time stochastic volatility model comparatively to the ones obtained with the traditional GARCH and ARFIMA models. In order to inquire into this, we estimate using the Efficient Method of Moments (EMM) of Gallant and Tauchen (1996) a continuous time stochastic volatility model for the logarithm of asset price and we filter the underlying volatility using the reprojection technique of Gallant and Tauchen (1998). Under the assumption that the model is correctly specified, we obtain a consistent estimator of the integrated volatility by fitting a continuous time stochastic volatility model to the data. The forecasting evaluation for the three estimated models is done with the help of the R^{*2} of the individual regressions of realized volatility on a constant and on the volatility forecasts obtained from the estimated models.[2] The empirical results show evidence that the performance of the continuous time model in the out-of-sample periods is better compared to the ones of the traditional GARCH and ARFIMA models. Further, these two last models show difficulties in tracking the growth pattern of the realized volatility for the sample considered. This probably is due to the increase in volatility caused by stock market crashes in this last part of the sample.

The plan of the chapter is as follows. Section 2 introduces the concept of realized volatility and the way to calculate it. Section 3 presents the continuous time model and the estimation results. Section 4 evaluates the forecasting performance of the three estimated specifications and section 5 concludes the chapter.

4.2 Realized Volatility

4.2.1 Theoretical relation between realized volatility and integrated volatility

Let $r_{t,j}$, $0 \leq j \leq n$, represent a set of $n+1$ intraday returns for day t. $j = 0$ refers to the closed market period that ranges from day $t-1$ until the open on day t. $j = 1$ is the fifteen minutes commencing at the open and $j = n$ is the last fifteen minutes return before market closes.

Much theoretical work models the logarithm of asset prices (p_k) as a univariate diffusion, Andersen et al. (2000),

$$dp_k = \mu_k dt + \sigma_k dW$$

where W is a Wiener process. So the daily return of asset k is given by

[2] R^{*2} is the corrected R^2 of Andersen and Bollerslev (2002).

$$p_k(t) - p_k(t-1) \equiv r_k(t) = \int_{t-1}^{t} \mu_k(s)ds + \int_{t-1}^{t} \sigma_k(s)dW(s). \tag{4.1}$$

These authors proved that under innocuous regularity conditions, the realized volatility $\sum_{j=0}^{n} r_{t,j}^2$ converges to the integrated volatility $\int_{t-1}^{t} \sigma_k^2(s)ds$ as n converges to ∞. So, the performance of the realized volatility estimator depends only on the number of observations. For a given sample period the higher the frequency of the data and the larger the number of observations, the better the approximation of the realized volatility estimator to the integrated volatility.

4.2.2 Data

Realized volatility has been calculated from the intraday 15-minutes price of a share of Microsoft[3], from 10^{th} of April, 1997, till 23^{rd} of February, 2001, according to Nelson and Taylor (2000).

The models, GARCH, ARFIMA and the continuous time model with two factors of volatility described below use daily data on Microsoft, adjusted for stock splits, from 13^{th} of March, 1986, till 23^{rd} of February, 2001 for 3,778 observations (see Figure 1).

Calculating the realized volatility

In this chapter the realized variance for the trading day t (the period ranges from the close on day t-1 to the close on day t) is calculated as a weighted average of the intraday squared returns. Accordingly to Nelson and Taylor (2000) it is given by

$$\hat{\sigma}_t^2 = \sum_{j=0}^{n} w_j r_{t,j}^2, \tag{4.2}$$

[3] The data was obtain freely from www.Price-data.com

and we must impose the constraint $\sum_{j=0}^{n} \lambda_j w_j = 1$ in order to ensure conditionally unbiased estimates when intraday returns are uncorrelated. λ_j represents the proportion of a trading's day total return variance that is attributed to period j. They assume that the $\lambda's$ are equal for all days t. In order that $\hat{\sigma}_t^2$ is a consistent, unbiased and efficient (with the least variance) estimate of the integrated volatility, Nelson and Taylor (2000) deduced that

$$w_j = \frac{1}{(n+1)\lambda_j}. \tag{4.3}$$

In particular, because the weight w_0 for the closed market return is much less than for the other returns (because λ_0 is very big) they specify w_j as

$$\begin{aligned} w_j &= \frac{1}{(1-\lambda_0)nk_j}, & 1 \leq j \leq n \\ w_j &= 0 & j = 0, \end{aligned}$$

where k_j is the proportion of the open-market variance given by

$$k_j = \frac{\lambda_j}{1-\lambda_0} \quad \text{with} \quad \sum_{j=1}^{n} k_j = 1.$$

Natural estimates of these variance proportions are:

$$\hat{\lambda}_j = \frac{\sum_t r_{t,j}^2}{\sum_t \sum_{i=0}^{n} r_{t,i}^2} \quad \text{and} \quad \hat{k}_j = \frac{\sum_t r_{t,j}^2}{\sum_t \sum_{i=1}^{n} r_{t,i}^2},$$

where the sums over days t can be for all days or for particular days (see Figure 3).[4]

4.3 The Model

Recently, researchers tend to model volatility as stochastic. The literature is vast referring to the estimation of models with or without stochastic volatility or with or without jumps,

[4] For more about these estimates, see Taylor and Xu (1997).

see Bates (2000), Gallant and Tauchen (2001), Chernov et al. (2003), Ghysels et al. (1995), etc.

This chapter estimates the stochastic volatility model with two factors of volatility[5] given by:

$$\frac{dP_t}{P_t} = \alpha_{10}dt + \exp(\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t})dW_{1t} \tag{4.4}$$

$$dU_{2t} = (\alpha_{20} + \alpha_{22}U_{2t})dt + dW_{2t} \tag{4.5}$$

$$dU_{3t} = (\alpha_{30} + \alpha_{33}U_{3t})dt + dW_{3t}, \tag{4.6}$$

where P_t is the Microsoft price series evolving in continuous time and W_i with $i = 1, 2,$ are three independent Wiener processes.

In this system the instantaneous standard deviation of the rate of return is an exponential function of the factors U_{2t} and U_{3t}. The drifts in equations 4.5 and 4.6 allow for mean reversion when $\alpha_{22} \neq 0$ and $\alpha_{33} \neq 0$. Small values of α_{22} and α_{33} mean that a shock to the volatility of the return takes time to dissipate. This is referred as persistence in the financial econometrics literature, and a big percentage of the financial series seem to show this feature, Zaffaroni (2000). β_{10} is also an important parameter since it takes care of the long-run mean of the volatility of the price equation 4.4.

Identification restrictions

To achieve identification it is necessary to impose some restrictions. In this concrete case for the logarithmic specification we set

[5] Gallant and Tauchen (2001) already estimated this model for a subsample of the data used in this paper.

$$\alpha_{20} = 0, \ \alpha_{30} = 0.$$

Hence the previous specification becomes:

$$\frac{dP_t}{P_t} = \alpha_{10}dt + \exp(\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t})dW_{1t} \tag{4.7}$$

$$dU_{2t} = \alpha_{22}U_{2t}dt + dW_{2t} \tag{4.8}$$

$$dU_{3t} = \alpha_{33}U_{3t}dt + dW_{3t} \tag{4.9}$$

We use these restrictions, as do Gallant and Tauchen (2001), first because they are common in previous similar SDE and second because they provide flexibility and numerical stability in the estimation phase.

4.4 Forecasting

Forecasting using the continuous time stochastic volatility model requires the reprojection step[6]. It allows us to filter the volatility factors U_{2t} and U_{3t} and consequently to obtain a forecast of the underlying integrated volatility for any desired sampling frequency. In fact, as a by-product of the estimation step we obtain a long simulation of the volatility factors $\{\hat{U}_{2t}\}_{t=1}^{N}$ and $\{\hat{U}_{3t}\}_{t=1}^{N}$. Having as the main aim to obtain

$$E(U_{2t}|\{y_\tau\}_{\tau=1}^{t}),$$

$$E(U_{3t}|\{y_\tau\}_{\tau=1}^{t})$$

[6]Notice that we use the same data as chapter three. In this way, the projection and estimation part are similar (see Tables 1 and 2).

we start generating simulations of $\{\hat{U}_{2t}\}_{t=1}^{N}$, $\{\hat{U}_{3t}\}_{t=1}^{N}$ and $\{\hat{y}_t\}_{t=1}^{N}$ at the estimated parameter $\hat{\rho}$ and with N equal 100000. Then we impose the same SNP-GARCH model founded in the projection step, on the simulated values $\hat{y}_t$. According to Gallant and Tauchen (2001), this provides a good representation of the one-step ahead conditional variance $\hat{\sigma}_t^2$ of $\hat{y}_{t+1}$ given $\{\hat{y}_\tau\}_{\tau=1}^{t}$. Then, we run regressions of $\hat{U}_{2t}$ and $\hat{U}_{3t}$ on lags of $\hat{\sigma}_t^2$, $\hat{y}_t$, $|\hat{y}_t|$.

$$\hat{U}_{2t} = \alpha_0 + \alpha_1 \hat{\sigma}_{t-1}^2 + ... + \alpha_p \hat{\sigma}_{t-p}^2 + \theta_1 \hat{y}_{t-1} + ... + \theta_q \hat{y}_{t-q} + \pi_1 |\hat{y}_{t-1}| + ... + \pi_r |\hat{y}_{t-r}| + u_t$$

$$\hat{U}_{3t} = \beta_0 + \beta_1 \hat{\sigma}_{t-1}^2 + ... + \beta_p \hat{\sigma}_{t-p}^2 + \gamma_1 \hat{y}_{t-1} + ... + \gamma_q \hat{y}_{t-q} + \lambda_1 |\hat{y}_{t-1}| + ... + \lambda_r |\hat{y}_{t-r}| + \mu_t.$$

With this procedure we obtain calibrated functions inside the simulation that give predicted values of U_{2t} and U_{3t} given $\{y_\tau\}_{\tau=1}^{t-1}$. In fact, given the length of simulation, these regressions are as Gallant and Tauchen (2001) say analytic projections. Finally, we evaluate these functions on the observed data series to obtain forecasts of the volatility factors, $\tilde{U}_{2t}$ and $\tilde{U}_{3t}$. The volatility forecast, for day t +1 will be

$$\exp(\hat{\beta}_{10} + \hat{\beta}_{12} \tilde{U}_{2(t+1)} + \hat{\beta}_{13} \tilde{U}_{3(t+1)}).$$

We are going to split the sample in two subsamples, the first subsample is used to estimate the models and the second part (the out-of- sample period) is used to evaluate the models' forecasts. We use three out-of-sample periods. The first out-of-sample period ranges from the 11^{th} of January 2001 till the 23^{rd} February 2001. This out-of-sample is quite short because we compute one-day-ahead forecasts. The second ranges from the 3^{rd} of January 2000 till the 23^{rd} February 2001. Our intention in this case is to test the forecasting performance of the models at longer forecasting horizons (around three months). Finally, the third out-of-sample ranges from the 4^{th} of January 1999 till the 31^{st} of December 1999. In this case, we try to answer the question: is still there evidence that the SV2F model forecasts better than the benchmark models, in a horizon that does not seem to show an increase in volatility caused by stock market crashes?.

4.4.1 Alternative models

We also tried two different specifications. The first one is the traditional GARCH model. The parameters of this model are estimated with the historical daily data to build out-of-sample volatility forecasts (see Table 3).[7]

There is strong empirical evidence that the volatility has long memory, in the sense that the effect of a shock to volatility persists for a long number of periods. The second specification is the ARFIMA model and it tries to fit this feature by modelling volatility as a fractionally integrated process.

Definition 4.1. *A stationary process $\{y_t\}$ is said to be* ***long memory*** *if its ACF (auto-correlation function γ) decays toward zero so slowly that*

$$\sum_{u=-n}^{n} |\gamma_y(u)| \uparrow \infty \quad as \quad n \to \infty,$$

Zaffaroni (2001).

In order to make inferences about the long memory characteristic of the volatility series we will use a formal test.

Testing the existence of long memory

There are many tests that we can apply to check for long memory of volatility. In this chapter, we first use the traditional R/S method.

Consider $Y_1, Y_2, ..., Y_n$ the observations in n successive periods and $\bar{Y}$ the empirical average. The adjusted range R is defined as

$$R(n) = \max_{0 \le l \le n} \{\sum_{i=1}^{l} Y_i - l\bar{Y}\} - \min_{0 \le l \le n} \{\sum_{i=1}^{l} Y_i - l\bar{Y}\}$$

[7] We use the GARCH package for Ox, Doornik and Oons (2001), to estimate the model available at Jurgen A. Doornik's web page.

and an estimate of the variance of the process underlying the data is

$$S^2(n,q) = \sum_{j=-q}^{q} \omega_q(j)\hat{\gamma}(j),$$

where $\hat{\gamma}(j)$ is an estimate of the autocovariance function at lag j and $\omega_q(j)$ are weights. Finally the R/S statistic is then defined as

$$Q(n,q) = \frac{R(n)}{S(n,q)}.$$

Helms et al.(1984) set $q = 0$ and $\omega_0(0) = 1$. The R/S statistic with these restrictions suffers from two disavantages: first its distribution is not known and secondly it can be affected by short-memory components. Lo (1991) modified this statistic by putting $q \neq 0$ in order to deal with these problems. His weights were given by

$$\omega_q(j) = 1 - \frac{j}{q+1}, \qquad q < n$$

and q was chosen as the greatest integer less than or equal to

$$(\frac{3n}{2})^{\frac{1}{3}}(\frac{2\hat{\rho}(1)}{1-\hat{\rho}^2(1)})^{\frac{2}{3}},$$

with $\hat{\rho}(1)$ as an estimate of the first order autocorrelation of the process.

For short memory processes the values of $Q(n,q)$ converge to n^J. d is the long memory parameter and J is related to it by $J = d+1/2$. Mandelbrot and Taqqu (1979) also proved that the process has long memory when $J > 1/2$ and their estimator for J was

$$\hat{J} = \frac{\log(R(n)/S(n))}{\log n}.$$

According to the correlogram and the long memory test it seems that the series of squared returns (see Figure 4 and Table 4) is fractionally integrated, that is

Definition 4.2. *A stationary process $\{y_t\}$ is said to be fractionally integrated with long memory if it can be written as*

$$(1-L)^d \phi(L) y_t = \theta(L)\epsilon_t,$$

where L is the lag operator, $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator with roots inside the unit circle, ϵ_t are independently and identically distributed as $N(0,\sigma^2)$ and

$$0 < d < \frac{1}{2}.$$

Therefore, we can think of an ARFIMA model as a quite good description of the volatility dynamics.

Estimating the ARFIMA model

We use the ARFIMA package for Ox[8] in order to estimate the parameters of the ARFIMA model. The best model according to the BIC criterion does not have a moving average part and the autoregressive part is of order one (see Table 5).

4.4.2 Evaluating and comparing alternative volatility forecasts

In this subsection we assess the performance of the volatility forecasts generated from the continuous time stochastic volatility model and compare it with the performance of the GARCH and ARFIMA forecasts for the three out-of-sample periods[9].

For the first out-of-sample period we are going to use one-day-ahead volatility forecasts and then we compare them to the estimate of realized volatility determined before. For this, we proceed following the analysis in Andersen and Bollerslev (1998) by regressing the realized volatilities on a constant and on the various model forecasts. In this case, the

[8] Available at Jurgen A. Doornik web page: www.nuff.ox.ac.uk/Users/Doornik.

[9] The first out-of-sample period ranges from the 11^{th} of January 2001 till the 23^{rd} February 2001, the second ranges from the 3^{rd} of January 2000 till the 23^{rd} February 2001 and the third out-of-sample ranges from the 4^{th} of January 1999 till the 31^{st} of December 1999.

models and the filters have been estimated and computed 27 times. Tables 6 to 8 report the estimated regressions for the one-day-ahead out-of-sample forecasts that assumes the following form:

$$\widehat{rvolatility}_{t+1} = \beta_0 + \beta_1 \widehat{\sigma^2_{t+1/t,ARFIMA}} + u_{t+1} \tag{4.10}$$

$$\widehat{rvolatility}_{t+1} = \beta_0 + \beta_1 \widehat{\sigma^2_{t+1/t,GARCH}} + u_{t+1} \tag{4.11}$$

$$\widehat{rvolatility}_{t+1} = \beta_0 + \beta_1 \widehat{\sigma^2_{t+1/t,SV2F}} + u_{t+1}$$

The analysis of the results is based on the R^{*2} of the regressions above[10] and on the t-statistics for the hypothesis of $\beta_0 = 0$ and/or $\beta_1 = 1$. We use both OLS and instrumental variables (IV) methods of estimation. The use of IV can be justified by the existence of a possible error in the forecast of future volatility that would lead the OLS estimates to be inconsistent. The instruments that were used are the past volatility forecasts for the two first equations and the squared return for the last equation because it seemed more correlated to the volatility forecast than to its past value.

For the considered out-of-sample period we find that the hypothesis of $\beta_0 = 0$ and $\beta_1 = 1$ are both rejected at a 5% significance level for the GARCH and ARFIMA models.

[10] Andersen and Bollerslev (2002) show that there is a bias in empirical realized volatility measures built directly from high- frequency data due to the existence of market microstructure frictions. This leads to a downward bias in the R^2 obtained from the above regressions. In fact, they show that these R^2 will under-estimate the true R^{*2} by the multiplicative factor, that is $R^{*2}=\beta * R^2$:

$\beta = Var[RV_t(h)]/Var[IV_t] \approx Var[RV_t(h)]/\{Var[RV_t(h)] - hE[RQ_t(h)]\}$ where RV_t is the realized volatility, IV_t is the integrated volatility and

$$RQ_t(h) \equiv \frac{1}{h}\frac{2}{3}\sum_{i=1}^{1/h} r^{(h)4}_{t-1+ih},$$

with $1/h = 96$ corresponding to the use of "15-minute" returns. For more details please check Andersen and Bollerslev (2002).

Moreover the coefficients of volatility forecasts of both regression models 4.10 and 4.11 show negative signs, which could lead us think that both models are inappropriate to forecast volatility (however both variables are statistically insignificant). These strange results may be explained by a an increase in volatility (caused by stock market crashes) observed in the out-of-sample period and not taken in to account in both specifications.

Contrarily, the SV2F model seems to forecast much better in the out-of-sample period of 28 days. The empirical results report that the variable volatility forecast is probably an unbiased estimator of future volatility since the hypothesis of $\beta_0 = 0$ and(or) $\beta_1 = 1$ are not reject at a 5% significance level (see Tables 6 and 7). The R^{*2} is equal to 0.235883, which is larger than the ones observed for the GARCH and the ARFIMA.

The better performance of the continuous stochastic volatility model is due probably to its flexibility and ability to capture volatility persistence. As it has been reported in several papers, for instance, Diebold and Inoue (1999), Granger and Hyung (1999), Kim and Kon (1999) and Beine and Laurent (2000), changes in volatility and persistence are imperfect substitutes. By this we mean that the persistence captured in a model is strongly reduced when we include jumps. Since we do not allow for jumps, since the increase in volatility occurs over a short period at the very end of sample, and thus cannot be explicitly modeled, the SV2F tries to accommodate the "missing" shifts by allowing that one factor of volatility be extremely slow mean reverting (a sign of strong persistence in volatility). The GARCH and ARFIMA models are not able to account for the apparent switch in volatility.

Next, we use the second out-of-sample period to investigate the forecasting performance of the SV2F model at longer horizons. We denote the whole out-of-sample period as $[t, T]$, where t corresponds to the 3^{rd} of January 2000 and T to the 23^{rd} February 2001. We split the out-of sample period $[t, T]$ into the subsets $[t, t_1]$, $[t_{1+1}, t_2]$ and $[t_{2+1}, T]$ which are used for volatility forecasting. Notice that t_1 corresponds to the 18^{th} of May 2000 and t_2 to the 4^{th} of October 2000. In other words, we estimate the SV2F model and calculate the volatility factors U_{2t} and U_{3t} three times, at $t-1$ (12^{th} of December 1999), t_1 and t_2. Therefore, the volatility forecast for day $t+7$ is for example

$$\exp(\hat{\beta}_{10} + \hat{\beta}_{12}\tilde{U}_{2(t+7)} + \hat{\beta}_{13}\tilde{U}_{3(t+7)}).$$

Remember that the $\widehat{\beta}'s$ and the calibrated coefficients of $\tilde{U}_2$ and $\tilde{U}_3$ remain the same till the next estimation date.[11]

Table 8 reports the forecasting results for the SV2F, GARCH and ARFIMA models. Once more the GARCH and the ARFIMA models perform poorly. The hypothesis of $\beta_0 = 0$ and/or $\beta_1 = 1$ are both rejected at a 5% significance level for these two models and the coefficients of volatility forecasts of both regression models 4.10 and 4.11 show negative signs as before.

The volatility forecasting performance based on the stochastic volatility model seems to improve over the other two although we can no longer claim that the volatility forecast could be an unbiased estimator of future volatility. Observing once more Figure 3, we see that the out-of-sample period of 289 periods corresponds exactly to the part of sample where volatility pattern seems to change. This probably explains the poor performance of both GARCH and ARFIMA. In fact, if we observe the graphs of the series of residuals of equations 4.10 and 4.11, respectively, we see evidence that the forecasts from GARCH and ARFIMA are not able to assimilate this increase in the volatility (see Figures 5 and 6). The correlogram of the residuals of GARCH also shows evidence that they are not white noise (see Figures 7 and 8). The SV2F model performs better due, as it was explained before, to its ability of capturing volatility persistence.

Finally, we also evaluate the forecasting performance of the continuous time model in the out-of-sample period that ranges from the 4^{th} of January 1999 till the 31^{st} of December 1999. We choose this period because it precedes the period during which

[11] Note that it would be possible to update coefficients at each day or at 10 days, but this would be very demanding computationally. In this way, our forecasts are not optimal.

volatility's pattern changes. As before, we denote the whole out-of-sample period as $[t, T]$, where t corresponds to the 4^{th} of January, 1999 and T to the 31^{st} December, 1999. We split the out-of sample period $[t, T]$ into the subsets $[t, t_1]$, $[t_{1+1}, t_2]$ and $[t_{2+1}, T]$ which are used for volatility forecasting. We use, this time, 10-day-ahead forecasts. Notice that t_1 corresponds to the 4^{th} of May 1999 and t_2 to the 2^{nd} of September 1999. In other words, we estimate the SV2F model and calculate the volatility factors U_{2t} and U_{3t} three times, at $t-1$ (31^{st} of December 1998), t_1 and t_2. Considering this out-of-sample period we observe, for the continuous time model, that the hypothesis of $\beta_0 = 0$ is not rejected at a 1% significance level and the hypothesis of $\beta_1 = 1$ is not rejected at any conventional significance levels (see Table 9). The GARCH and ARFIMA models perform worse accordingly to R^{*2} and the previous hypothesis of $\beta_1 = 1$ is rejected at all conventional significance levels.

Moreover, following the analysis in Andersen et al. (2001), we also focus our forecasting evaluation on regressions of the realized volatility on a constant, on the SV2F model forecasts and on the other benchmark model's forecasts:

$$\widehat{rvolatility}_{t+1} = \beta_0 + \beta_1 \widehat{\sigma^2_{t+1/t,SV2F}} + \beta_2 \widehat{\sigma^2_{t+1/t,GARCH}} + u_{t+1} \tag{4.12}$$

$$\widehat{rvolatility}_{t+1} = \beta_0 + \beta_1 \widehat{\sigma^2_{t+1/t,SV2F}} + \beta_2 \widehat{\sigma^2_{t+1/t,ARFIMA}} + u_{t+1} \tag{4.13}$$

Table 10 reports the empirical results. When including both the SV2F and the GARCH or the ARFIMA forecasts in the same regression, the estimates of the coefficients β_2 in equations 4.12 and 4.13 are not different from zero statistically and the hypothesis of $\beta_0 = 0$ and/or $\beta_1 = 1$ in both regressions are not rejected at any conventional significance levels. Futhermore, the inclusion of the GARCH or ARFIMA forecasts does not improve significantly the R^{*2} relatively to the one based only on the SV2F forecasts. So, according to these results, when there is not an increase in volatility's pattern, it seems that any bias in the SV2F volatility forecast is of future volatility, is minor and not statistically significant.

4.5 Conclusion

In this chapter we evaluate the predictive ability of the continuous time stochastic volatility model with two factors of volatility (SV2F) and compare its volatility forecasts to the forecasts obtained from the traditional GARCH model and ARFIMA models. We choose as a proxy of ex-post volatility the realized volatility obtained from the intraday returns. We argue that this is a good measure of ex-post volatility because much theoretical work models the logarithm of asset prices as a univariate diffusion and it has been shown that under innocuous regularity conditions, the realized volatility converges to the integrated volatility. We have been careful to avoid microstructures problems by considering only observations at 15 minute intervals.

The main contributions of this chapter include: First, the computation of the realized volatility accordingly to Nelson and Taylor (2000) in order to ensure conditionally unbiased estimates when intraday returns are uncorrelated. Secondly, we apply the reprojection technique proposed by Gallant and Tauchen (1998) to obtain volatility forecasts from the SV2F model and finally, we compare the forecasting performance of this last model with two others. The empirical results show evidence that the volatility forecasting performance of the stochastic volatility model is significantly better than that of the other two models, which both perform poorly in short and mid-ranges forecast horizons.

Bibliography

[1] Andersen, T. G. and Bollerslev, T. (1998): Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, *International Economic Review*, Vol. 39, no. 4, pp. 885-905.

[2] Andersen, T. G.and Bollerslev, T. (2002): Correcting the Errors: A Note on Volatility Forecast Evaluation Based on High-Frequency Data and Realized Volatilities, *Discussion paper*, Duke University.

[3] Andersen, T. G., Bollerslev, T.,Diebold, F.X. and Labys, P. (2000): Exchange Rate Returns Standardized by Realized Volatility are (nearly) Gaussian, *Multinational Finance Journal*, 4, 159 -179.

[4] Andersen, T. G., Bollerslev, T.,Diebold, F.X. and Labys, P. (2001): Modeling and Forecasting Realized Volatility, *Working paper 8160*, NBER.

[5] Bates, D. (2000): Post-'87 Crash Fears in the S&P 500 Futures Option market, *Journal of Econometrics*, 94, 181-238.

[6] Beine, M. and Laurent, S. (2000): Structural Change and Long Memory in Volatility: New Evidence from Daily Exchange Rates, *Discussion Paper*, Université of Liège.

[7] Canina and Figlewski (1993): The Informational Content of Implied Volatility, *Review of Financial Studies*.

[8] Chernov, M., Gallant, A. R., Ghysels, E. and G. Tauchen (2003): Alternative Models for Stock Price Dynamics, *Journal of Econometrics*, 116, 225-257.

[9] Day, T. E. & Lewis, C. M. (1992): Stock Market Volatility and the Information Content of Stock Index Options, *Journal of Empirical Finance* 5, 131-154.

[10] Diebold, F.X., Inoue, A. (1999): Long Memory and Structural Change, *Manuscript*, Departmant of Finance, Stern School, NYU, May 1999.

[11] Doornik, J.A. and Ooms, M. (2001): A Package for Estimating, Forecasting and Simulating ARFIMA Models: ARFIMA Package 1.01 for Ox, *Guide*, Erasmus University.

[12] Fenton, V.M. and Gallant, A. R. (1996): Qualitative and Asymptotic Performance of SNP Density Estimators, *Journal of Econometrics*, 74, 74 -118.

[13] Gallant, A. R. and G. Tauchen (1996): Which Moments to Match?, *Econometric Theory*, 12, 657- 681.

[14] Gallant, A. R. and G. Tauchen (1998): Reprojecting Partially Observed Systems with Application to Interest rate Diffusions, *Journal of American Statistical Association,* 93, 10-24.

[15] Gallant, A. R. and G. Tauchen (2001): Efficient Method of Moments, *Discussion paper*, University of North Carolina at Chapel Hill.

[16] Granger, C.W.J., Hyung, N. (1999): Occasional Structural Breaks and Long Memory, UCSD, *Discussion paper* 99-14, June 1999.

[17] Ghysels, E., Harvey, A. and Renault, E. (1995): Stochastic Volatility, in *Handbook of Statistics*,14, Statistical Methods in Finance, G.S. Maddala and C. Rao (eds), North Holland, Amsterdam.

[18] Helms, B., Kaen, F., and Rosenman, R. (1984): Memory in Commodity Futures Contracts, *Journal of Futures Markets*, 4, 559-567.

[19] Jorion, P. (1996): Risk and Turnover in the Foreign Exchange Market, *in* J.A. Frankel, G. Galli & A. Giovannini, eds, "The Microstruture of Foreign Exchange Markets", The University of Chicago Press, Chicago.

[20] Kim, D., Kon, S.J (1999): Structural Change and Time Dependence in Models of Stock Returns, *Journal of Empirical Finance*, 6, 283-308.

[21] Laurent, S. and Peters, J.P (2002): A Tutorial for G@rch 2.3, a Complete Ox Package for Estimating and Forecasting ARCH Models,

www.crest.fr/pageperso/lfa/sebastien.laurent/sebatien.laurent.htm.

[22] Lo, A.W. (1991): Long Term Memory in Stock Market Prices, *Econometrica* 59, 1279-1313.

[23] Mandelbrot, B.B. and Taqqu, M (1979): Robust R/S Analysis of Long-Run Serial Correlation, *Proceedings of the 42^{nd} Session of the International Statistical Institute*, International Statistical Institute.

[24] Nelson, A. and Taylor, S. J.(2000): The Realized Volatility of FTSE -100 Future Prices, *Discussing paper*, Department of Accounting and Finance, Lancaster University.

[25] Pagan, A. R. and Schwert, G.W. (1990): Alternative Models for Conditional Stock Volatility, *Journal of Economics and Business Statistitcs* 9, 63-71.

[26] Taylor, S. J. and Xu, X. (1997): The Incremental Volatility Information in One Million Foreign Exchange Quotations, *Journal of Empirical Finance* 4, 317-340.

[27] West, K. D. and Cho, D. (1995): The Preditive Ability of Several Models of Exchange Rate Volatility, *Journal of Econometrics,* 69, 367-391.

[28] Zaffaroni, P. (2000): Time Series Models of Changing Volatility, *Lecture notes.*

4.6 Figures and Tables

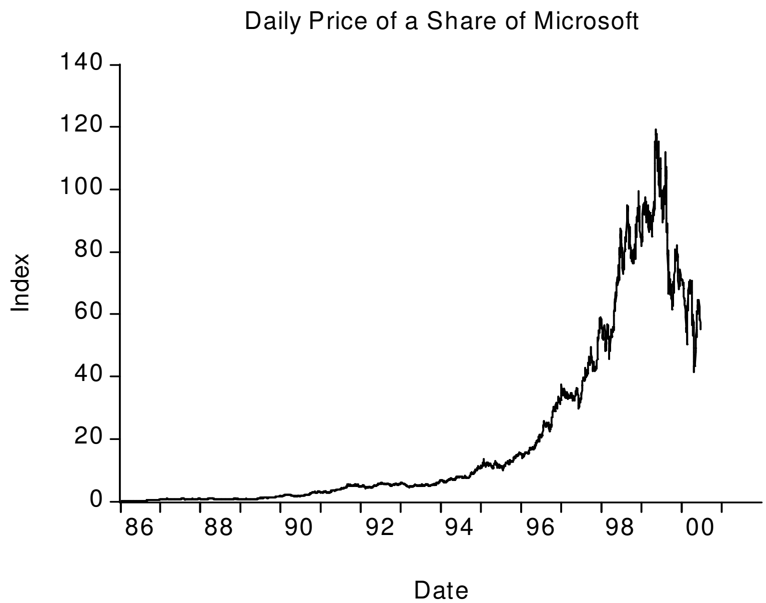

Figure 1

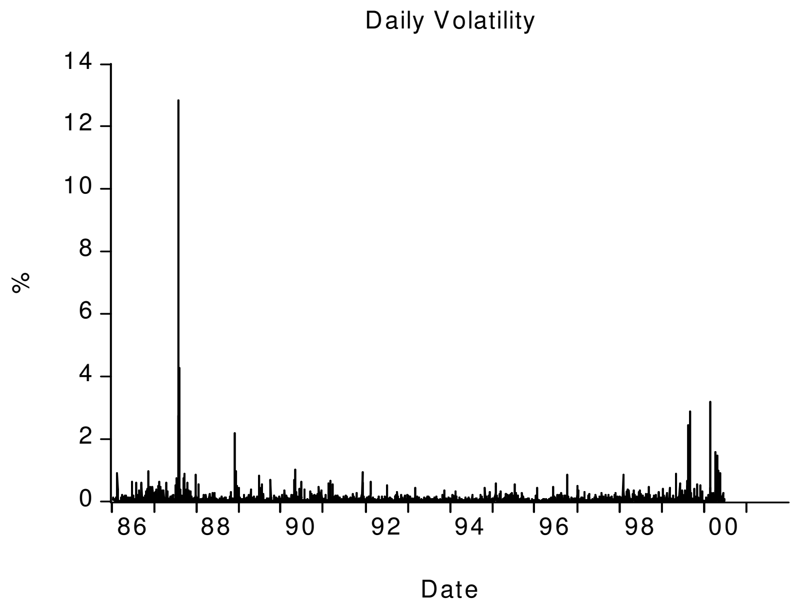

Figure 2

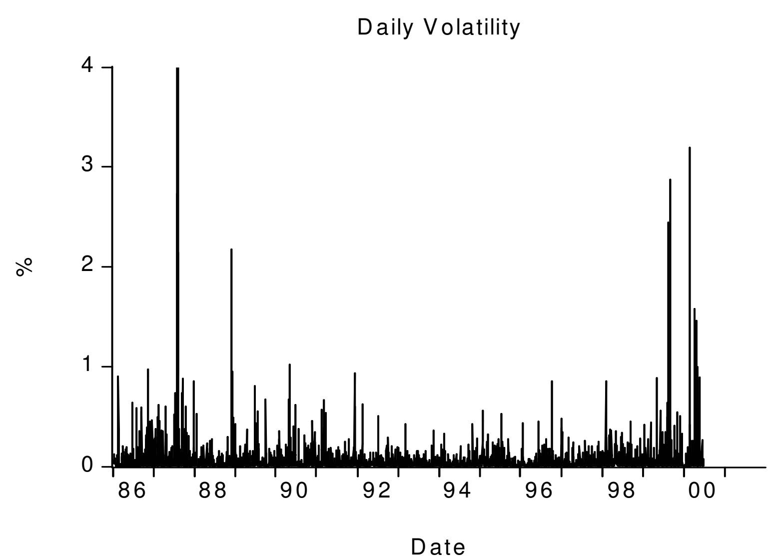

Figure 2.1: Different Scale

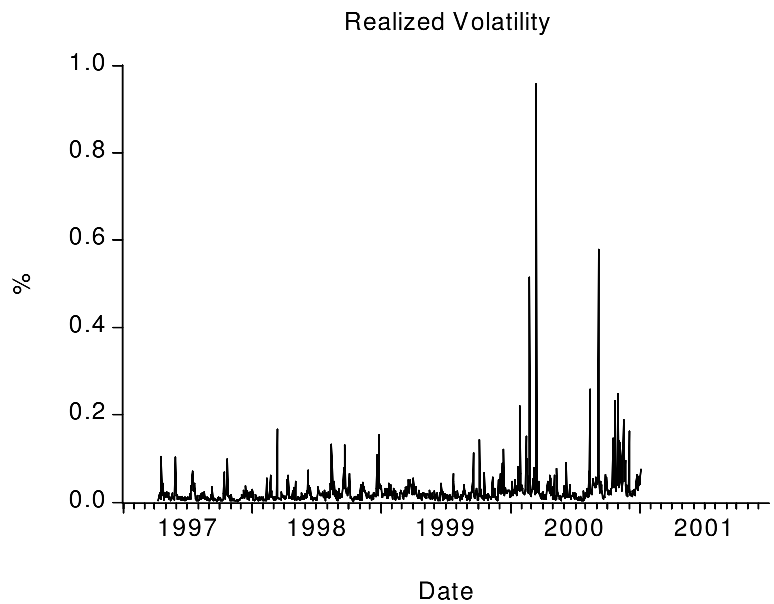

Figure 3

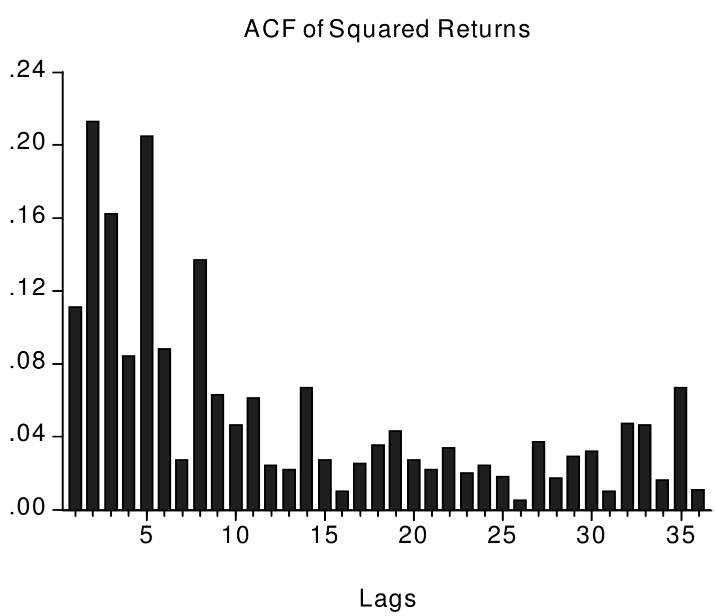

Figure 4

Model	α_{10}	α_{22}	α_{33}	β_{10}	β_{12}	β_{13}	N	χ^2	df	p-val
SV2F	*	*	*	*	*	*	100k	6.704	5	0.2436

Table 1:*is used for free parameters. 100k refers to a simultion of length 100 000 at step size $\Delta = 1/6048$,corresponding to 24 steps per day and 252 trading days per year.

Speficication	α_{10}	α_{22}	α_{33}	β_{10}	β_{12}	β_{13}
SV2F						
Estimate	0.42419	-0.00027568	-89.2101	-0.1099	0.006289	-4.6277
Std. Dev.	0.0736671	0.00014924	3.932837	0.008741	0.001034	0.0755
95%Lower	0.2693015	-0.00049470	-97.15399	-0.123322	0.004320	-4.77841
95% Upper	0.5793685	-0.00007876	-81.432454	-0.097090	0.008299	-4.480246

Table 2: Estimates, Standard Deviations and Confidence Intervals

	Estimates	Std. Error	t-value	Prob
Cst(M)	0.186230	0.035534	5.241	0.000
Cst(V)	0.267139	0.062602	4.267	0.000
GARCH(beta1)	0.864093	0.020700	41.74	0.000
ARCH(alpha1)	0.096255	0.013620	7.067	0.000

Table 3

Long Memory Test	q=0	q=q*
Q	1116.861	680.330
$\widehat{J}$	0.852075	0.79189
$\widehat{d}$	0.352075	0.291894

Table 4

	Estimates	Std. Error	t-value	Prob	BIC
ARFIMA(0,d,0)					
d parameter	0.126917	0.01142	11.1	0.000	3.178017
ARFIMA(1,d,0)					
d parameter	0.233008	0.01821	12.8	0.000	3.163838
AR-1	-0.186391	0.02271	-8.21	0.000	
ARFIMA(2,d,0)					
d parameter	0.240679	0.02533	9.50	0.000	3.165289
AR-1	-0.195044	0.03005	-6.49	0.000	
AR-2	-0.009916	0.02259	-0.439	0.661	
ARFIMA(0,d,1)					
d parameter	0.265111	0.02659	9.97	0.000	3.164190
MA-1	-0.216276	0.03214	-6.73	0.000	
ARFIMA(0,d,2)					
d parameter	0.243477	0.03137	7.76	0.000	3.165364
MA-1	-0.197413	0.03562	-5.54	0.000	
MA-2	0.0227693	0.01933	1.18	0.239	
ARFIMA(1,d,1)					
d parameter	0.242072	0.02843	8.51	0.000	3.165280
AR-1	-0.138536	0.1095	-1.27	0.206	
MA-1	-0.057881	0.1298	-0.446	0.656	

Table 5

Dependent variable RV n = 28	Estimates	Std. Error	T-value	Prob	R^{*2}
ARFIMA					0.054
β_0	6.018	2.135	2.819	0.0091	
β_1	-0.219	0.180	-1.212	0.2362	
GARCH					0.028
β_0	5.129	1.944	2.6378	0.0139	
β_1	-0.148	0.171	-0.866	0.3946	
SV2F					0.24
β_0	0.634	1.127	0.562	0.5786	
β_1	2.668	0.942	2.833	0.0088	

Table 6: OLS estimation

Dependent variable RV n = 28	Estimates	Std. Error	T-value	Prob
ARFIMA				
β_0	7.710	2.944	2.619	0.0148
$\beta 1$	-0.370	0.254	-1.458	0.1574
GARCH				
β_0	6.928	2.933	2.944	0.0069
β_1	-0.318	0.212	-1.503	0.1455
SV2F				
β_0	-5.225	2.951	-1.770	0.0884
$\beta 1$	8.097	2.648	3.057	0.0051

Table 7: IV Estimation

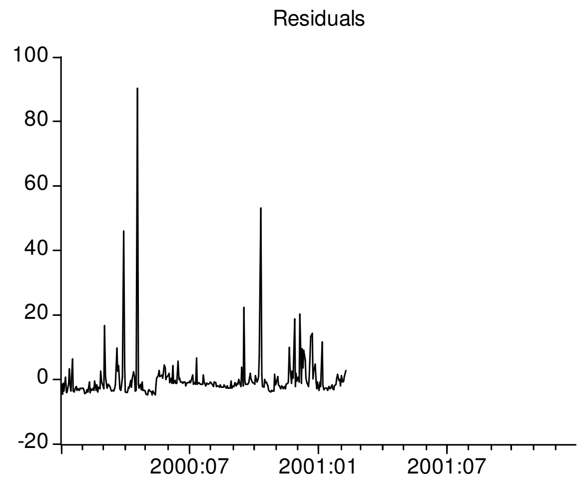

Figure 5: These are the residuals of the regression of the realized volatility on a constant and on the ARFIMA forecasts for the period that ranges from the 3^{rd} of January 2000 to the 23^{rd} of February 2001.

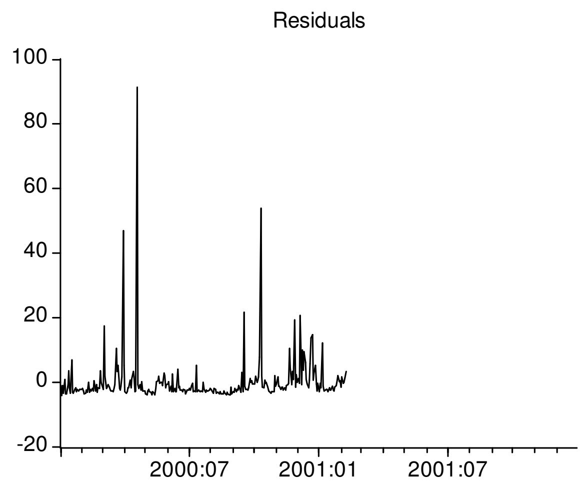

Figure 6: These are the residuals of the regression of the realized volatility on a constant and on the GARCH forecasts for the period that ranges from the 3^{rd} of January 2000 to the 23^{rd} of February 2001.

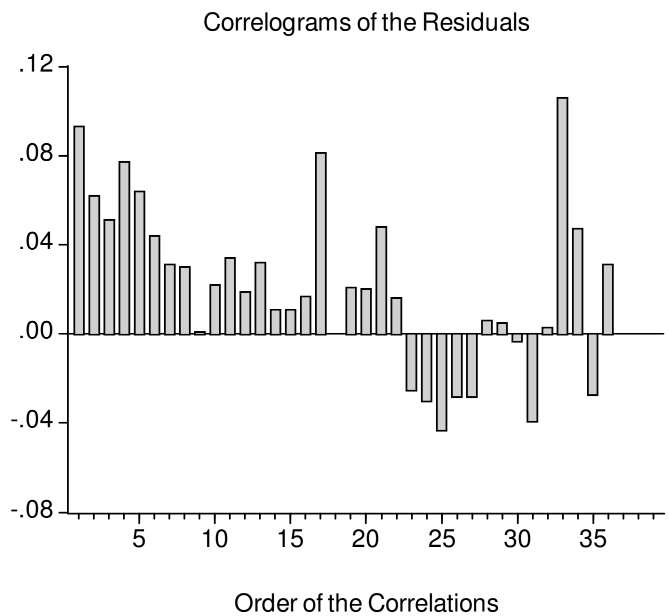

Figure 7: This is the correlogram of the residuals of the regression of the realized volatility on a constant and on the ARFIMA forecasts for the period that ranges from the 3^{rd} of January 2000 to the 23^{rd} of February 2001. * means that the autocorrelation is significative.

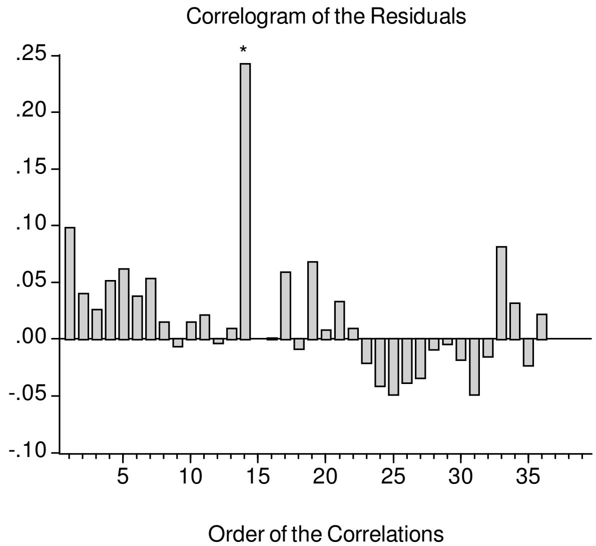

Figure 8: This is the correlogram of the residuals of the regression of the realized volatility on a constant and on the GARCH forecasts for the period that ranges from the 3^{rd} of January 2000 to the 23^{rd} of February 2001. * means that the autocorrelation is significative.

Dependent variable RV n = 289	Estimates	Std. Error	T-value	Prob	R^{*2}
GARCH					0.004
β_0	10.601	4.726	2.243	0.0257	
β_1	-1.027	0.733	-1.401	0.1622	
ARFIMA					0.03
β_0	10.66	2.396	4.450	0.0000	
β_1	-0.933	0.292	-3.194	0.0016	
SV2F					0.12
β_0	-14.79	0.731	-20.23	0.0000	
β_1	18.86	0.234	80.63	0.0000	

Table 8: OLS estimation with Newey-West HAC Standard Errors

Dependent variable RV n = 252	Estimates	Std. Error	T-value	Prob	R^{*2}
GARCH (10 days ahead)					0.01
β_0	1.016	0.833	1.220	0.2237	
β_1	0.176	0.154	1.144	0.2539	
ARFIMA (10 days ahead)					0.041
β_0	-0.179	1.267	-0.142	0.8876	
β_1	0.385	0.231	1.667	0.0968	
SV2F (10 days ahead)					0.10
β_0	0.947	0.460	2.059	0.0405	
β_1	0.940	0.484	1.942	0.0533	

Table 9: OLS estimation with Newey-West HAC Standard Errors

Dependent variable RV n = 252	Estimates	Std. Error	T-value	Prob	Adj.R^{*2}
regression 4.12					0.103
β_0	0.406	0.926	0.438	0.6618	
β_1	0.948	0.485	1.957	0.0516	
β_2	0.098	0.134	0.727	0.4679	
regression 4.13					0.114
β_0	-0.685	1.281	-0.535	0.5929	
β_1	0.870	0.453	1.920	0.0560	
β_2	0.306	0.190	1.608	0.1091	

Table 10: OLS estimation with Newey-West HAC Standard Errors

Chapter 5

A Two Factor Long Memory Stochastic Volatility Model

5.1 Introduction

A volatility model should be able to model the main characteristics of financial series of returns such as volatility persistence, volatility clustering, leverage effects, fat tails and small first order autocorrelation of squared returns. Many models have been proposed with the aim of capturing these empirical facts. Two examples are the original GARCH(1,1) model proposed by Bollerslev (1986) and Taylor (1986) and the ARSV(1) also proposed by Taylor (1986). Later on in the nineties, several other models emerged in the field of stochastic volatility like the Long Memory Stochastic Volatility model of Harvey (1993) and the model of Breidt et al. (1994) that incorporate in the original stochastic volatility specification a factor of volatility that is fractionally integrated. Both papers try to model the fact that the volatility of the series of returns decays slowly towards zero (the long memory property).

More recently, Ming Liu (2000) and Breidt et al. (2000) proposed regime switching as another explanation for observed long memory. As Ming Liu (2000) says, the long memory pattern is present in the autocorrelation function of volatility whenever the regime switches in a heavy-tail manner.

In this chapter we model volatility persistence by assuming that the volatility of the returns has a long memory feature captured by a fractionally integrated model as in Breidt et al. (1994). The innovation is that we introduce a short run volatility factor that allows the model to generate extra kurtosis and simultaneously to accommodate better the volatility persistence. Moreover, for some restrictions of the parameters it is possible to fit the empirical fact of small first order autocorrelation of squared returns.

A motivation for our second factor derives from the empirical results on estimation of continuous time models. Models with only one factor of volatility fail to fit the main features of the data. Chernov et al. (2003) show that the problem is overcome by introducing an extra volatility factor. So, it is expected that the introduction of this second factor contributes to a better explanation of data. Moreover, Ming Liu (2000) also estimates a Regime Switching Model with an AR(1) volatility factor and finds that the introduction of this short memory dynamics seems to be helpful.

We test the performance of our model empirically by fitting it to the returns of the S&P 500 Composite Index. The empirical results show us evidence (for the S&P 500), that the long memory stochastic volatility model with two factors of volatility performs better than the two benchmark models, the ARSV (the autoregressive stochastic volatility model) of Taylor (1986) and the LMSV (the long memory stochastic volatility model) of Breidt et al. (1994) in capturing volatility persistence and fat tails of the unconditional distribution of returns.

The chapter is organized as follows: Section two covers the empirical facts of the financial series of returns. Section three describes the Long Memory Stochastic Volatility model (LMSV) and presents the two factor LMSV model (2FLMSV) and its main properties. Section 4 reports the estimation results and Section 5 concludes the chapter.

5.2 Empirical Facts of The Financial Series of Returns

Many empirical analyses of financial time series have been realized in the last three decades and in particular for the financial time series of returns, it has been observed that they exhibit:

- Volatility persistence - The sample autocorrelation function (ACF) of the squared returns is large and statistically significantly different from zero, which implies that the effect of a shock to volatility persists for a long number of periods, Ghysels et al. (1995).
- Leverage effect - The effect of shocks to volatility is not symmetric. A negative shock has a larger effect than a positive shock on the future volatility of the asset. This can be explained by the fall of the price of a stock leading to a rise of the debt-to-equity, ratio which increases the returns volatility to the equity holders. On the other hand, the expected increase in volatility reduces the demand of the stock due to risk aversion and the consequent decrease in the stock value gives rise to an increase of volatility, Zaffaroni (2000).
- Mean reversion in volatility - we understand that there exists a normal value of volatility to which it will return.
- Thick tails - The unconditional distribution of assets returns presents high kurtosis ranging from 4 to 50 as was reported by Engle and Patton (2001). If the model is to capture the main characteristics of data this feature should be introduced in the model.
- Small first order autocorrelation of squared returns, Carnero et al. (2001).

5.3 The LM Models

5.3.1 The LMSV model

Breidt et al. (1994) extended the traditional stochastic volatility model by assuming that the volatility component is a stationary long-memory process, such as:

$$y_t = \sigma_t \zeta_t \quad \text{with} \quad \sigma_t = \sigma \exp(h_t/2),$$

where $\{h_t\}$ is independent of $\{\zeta_t\}$, ζ_t is $i.i.d(0,1)$ and $\{h_t\}$ is the fractionally integrated Gaussian noise process

$$(1-L)^d h_t = \epsilon_t, \qquad \eta_t \sim iidN(0, \sigma_\epsilon^2), \tag{5.1}$$

that is weakly stationary in the range $d \in (0, 0.5)$. L stands for the lag operator.

The long memory property

According to Baillie (1996), a discrete time series process y_t with autocorrelation function (ACF), ρ_k at lag k, possesses long memory if

$$\lim_{n\to\infty} \sum_{k=-\infty}^{n} |\rho_k| = \infty.$$

This means that the ACF takes far longer to decay towards zero than the ACF of a strict stationary ARMA process, which decays at an exponential rate. In a more formal way, we can say that:

Definition 5.1. *A weakly stationary process has long memory if its ACF $\rho(.)$ has a hyperbolic decay, i.e.,*

$$\rho(k) \sim f(k)k^{2d-1} \quad as \quad k \to \infty,$$

$$0 < d < \frac{1}{2} \text{ and } f(k) \text{ is slow varying at infinity.}^1$$

One example of a long memory process is the process represented by equation 5.1.

In 1981, Hosking showed for $d < \frac{1}{2}$ that h_t is not only stationary but also it has a $MA(\infty)$ representation given by:

$$h_t = (1-L)^{-d}\epsilon_t = \sum_{k=0}^{\infty}\psi_k\epsilon_{t-k} \quad \text{with} \quad \psi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)},$$

where $\Gamma()$ is the Gamma function[2]. Moreover, the coefficients ψ_k converge hyperbolically to zero. Another interesting point is the behavior of the ACF of h_t. After some computations Baillie (1996) obtains that the variance and the autocovariance functions are given respectively by:

$$\gamma(0) = \sigma_\epsilon^2 \frac{\Gamma(1-2d)}{\{\Gamma(1-d)\}^2} \tag{5.2}$$

and

$$\gamma(k) = \sigma_\epsilon^2 \frac{\Gamma(1-2d)\Gamma(k+d)}{\Gamma(d)\Gamma(1-d)\Gamma(k+1-d)}, \quad \text{if} \quad k \geq 1.$$

Consequently, the autocorrelation function is:

$$\rho(k) = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} = \prod_{1\leq i\leq k} \frac{i+d-1}{i-d}, \quad \text{for} \quad k \geq 1$$

[1] A function $f(x)$ is defined as being regularly varying at infinity with index α if $\lim_{t\longrightarrow\infty} \frac{f(tx)}{f(t)} = x^\alpha, \forall x > 0$. It is slow varying at infinity if $\alpha = 0$. Therefore, asymptotically it becames a constant. An example of a slowly varying function at infinity is $f(x) = \log(x)$.

[2] For $d > -1$ the binomial expansion of $(1-L)^d$ is given by:

$$(1-L)^d = \sum_{k=0}^{\infty} \begin{pmatrix} d \\ k \end{pmatrix} (-L)^k = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - ...$$

Remember that $[\frac{\Gamma(a+x)}{\Gamma(b+x)}]x^{a-b}$. So when $k \to \infty$, $\psi_k \frac{k^{d-1}}{\Gamma(d)}$.

and we can write it as

$$\rho(k) \approx \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1},$$

when $k \to \infty$. Although a shock to volatility takes a quite long time to dissipate this last expression shows that the volatility converges slowly to a normal value (mean reversion).

5.3.2 The 2FLMSV

In this Subsection, we propose a long memory stochastic volatility model with two factors of volatility (2FLMSV). The first factor accounts for the persistence in the stochastic volatility since it is assumed to be a stationary fractionally integrated process, while the second factor accommodates the short run dynamics and generates extra kurtosis. Let y_t denote, for instance, the return in percentage of a financial asset traded on a financial market. Then the 2FLMSV model for y_t is:

$$y_t = \zeta_t \sigma \exp\left(\frac{\alpha_1 h_{1t} + \alpha_2 h_{2t}}{2}\right), \tag{5.3}$$

with ζ_t $i.i.d$ $(0,1)$ and α_1 and α_2 constants. Futhermore, the first volatility factor follows the process:

$$(1-L)^d (h_{1t} - \mu) = \epsilon_t, \tag{5.4}$$

where ϵ_t is $i.i.d$ $(0, \sigma_\epsilon^2)$. The second factor of volatility follows an AR(1):

$$h_{2t} = \phi h_{2t-1} + \eta_t, \text{ with } \eta_t \sim iid(0, \sigma_\eta^2) \text{ and } |\phi| < 1. \tag{5.5}$$

This last condition guarantees the stationarity of h_2. Moreover, η_t, ϵ_t and ζ_t are mutually independent for all t, and h_1 and h_2 are unobservable latent variables. The use of a second factor of volatility that is autorregressive of order one is in agreement with the first modelizations in this area, such as the ARSV(1) by Taylor (1986).

Statistical properties

Given that ζ_t has a standard normal distribution, y_t is a martingale difference, the stationarity of which depends on the stationarity of its volatility factors, h_1 and h_2. After some computations, we can obtain that the variance[3] of y_t is equal to:

$$var(y_t) = \sigma^2 \exp(\frac{\alpha_1^2\sigma_{h_1}^2 + \alpha_2^2\sigma_{h_2}^2}{2})$$

and the excess kurtosis displayed by y_t is

$$\frac{E(y_t^4)}{E(y_t^2)^2} - 3 = 3[\exp(\alpha_1^2\sigma_{h_1}^2 + \alpha_2^2\sigma_{h_2}^2) - 1],$$

which is larger than 3 for $\alpha_1 \neq 0$ and $\alpha_2 \neq 0$. Moreover, for $\alpha_1 = \alpha_2 = 1$, the two factor long memory stochastic volatility model is able to generate higher kurtosis than the LMSV of Breidt et al. (1994).

The autocorrelation function of the returns is:

$$corr(y_t, y_{t+k}) = 0 \quad \forall k \neq 0.$$

Proceeding the same way for the series of the squared returns, we obtain that:

$$E(y_t^2) = \sigma^2 \exp(\frac{\alpha_1^2\sigma_{h_1}^2 + \alpha_2^2\sigma_{h_2}^2}{2}),$$

$$var(y_t^2) = \sigma^4 \exp(\frac{3}{2}(\alpha_1^2\sigma_{h_1}^2 + \alpha_2^2\sigma_{h_2}^2))[E(\zeta_t^4)\exp(\frac{\alpha_1^2\sigma_{h_1}^2 + \alpha_2^2\sigma_{h_2}^2}{2}) - 2]$$

and

[3] Given the properties of the lognormal distribution and the fact that the normal distribution is reproducible with respect to its arguments; $E(\exp(bh_t)) = \exp(b^2\frac{\sigma_h^2}{2})$, where b is a constant and σ_h^2 is the variance of h_t.

$$
\begin{aligned}
cov(y_t^2, y_{t+k}^2) &= \sigma^4 E(\zeta_t^2 \zeta_{t+k}^2) E[\exp(\alpha_1 h_{1t} + \alpha_2 h_{2t}) \exp(\alpha_1 h_{1t+k} + \alpha_2 h_{2t+k})] \\
&\quad -\sigma^4 \exp(\alpha_1^2 \sigma_{h_1}^2 + \alpha_2^2 \sigma_{h_2}^2),
\end{aligned}
$$

for $k \neq 0$. These properties fit the empirical fact that the correlation between the squared returns is different from zero.

It is more straight forward to analyze these properties if we rewrite model 5.3 in the following way:

$$
\begin{aligned}
x_t &= \ln y_t^2 = \ln \sigma + \alpha_1 h_{1t} + \alpha_2 h_{2t} + \ln \zeta_t^2 \\
&= \ln \sigma + E \ln \zeta_t^2 + \alpha_1 h_{1t} + \alpha_2 h_{2t} + \ln \zeta_t^2 - E \ln \zeta_t^2 \\
&= w + \alpha_1 h_{1t} + \alpha_2 h_{2t} + z_t,
\end{aligned} \tag{5.6}
$$

where w and z_t are respectively equal to $\ln \sigma + E \ln \zeta_t^2$ and $\ln \zeta_t^2 - E \ln \zeta_t^2$. Since ζ_t is standard normal, $E \ln \zeta_t^2 = -1.27$ and $\sigma_z^2 = \pi^2/2$. So, the process $\{x_t\}$ is a sum of a long - memory process, an AR(1) process and a non- Normal error. Furthermore,

$$E(x_t) = w$$

and

$$cov(x_t, x_{t+k}) = \alpha_1^2 \gamma_{h_1}(k) + \alpha_2^2 \gamma_{h_2}(k) + \sigma_z^2 I(k = 0), \tag{5.7}$$

where $\gamma(.)$ are the autocovariance functions of h_{1t} and h_{2t} and $I(k = 0)$ is 1 if $k = 0$ and zero otherwise. Consequently, the autocorrelation function of the squared returns is going to depend on two autocorrelation functions: one that decays slowly towards zero and other that decays faster towards zero.

Notice that for the LMSV model of the previous Section the ACF is given by:

$$cov(x_t, x_{t+k}) = \gamma_h(k) + \sigma_z^2 I(k=0),$$

which implies that the autocorrelations of order 1 of the logarithm of the squared returns, $\rho_{LMSV}(1)$ and $\rho_{2FLMSV}(1)$, are respectively:

$$\rho_{LMSV}(1) = \frac{\gamma_h(1)}{\sigma_h^2 + \sigma_z^2}$$

and

$$\rho_{2FLMSV}(1) = \frac{\alpha_1^2 \gamma_{h_1}(1) + \alpha_2^2 \gamma_{h_2}(1)}{\alpha_1^2 \sigma_{h_1}^2 + \alpha_2^2 \sigma_{h_2}^2 + \sigma_z^2}.$$

As $\gamma_h(1) = \gamma_{h_1}(1)$ because $h_t = h_{1t}$ $\forall t$, the $\rho_{2FLMSV}(1)$ can be lower or bigger than $\rho_{LMSV}(1)$, depending on the values of α_1 and α_2. Hence, for the restriction $\alpha_1 = \alpha_2 = 1$, the $\rho_{2FLMSV}(1)$ is smaller than $\rho_{LMSV}(1)$, considering that h_2 is a stationary AR(1) process. Therefore, for this case our model is able to fit the main empirical facts reported above and to generate higher kurtosis.

5.4 An Empirical Example

In this Section we evaluate the performance of our model in capturing the empirical features of financial data. For this purpose, we use daily close price data on the S&P 500 Composite Index over the period January 3, 1928 to February 19, 2002, representing 18 609 observations.

Figures 1 and 2 plot the price level and the returns on the index (adjusted for dividends and splits) over the sample period.

5.4.1 Summary of data

Figure 3 shows some summary statistics of the data. The average return is about one-thirtieth of a percent per day and the daily variance is 1.3631. Moreover, the distribution of returns is negatively skewed and the kurtosis is also quite high.

Finally, we have also computed the correlograms of the returns and the squared returns series (Figures 4 and 5 respectively). We find that the autocorrelation function of the first converges quickly towards zero while the autocorrelation function of the squared returns tends more slowly towards zero (a symptom of long memory) and the $\rho(1)$ is smaller than the $\rho(2)$.

Detecting the existence of long memory

Once observed the ACF's of volatility, we may suspect that the data exhibits the long memory property.

There are many tests we can apply to check this suspicion. In this chapter, we first use the traditional R/S method.

Consider $Y_1, Y_2, ..., Y_n$ the observations in n successive periods and $\bar{Y}$ the empirical average. The adjusted range R is defined as

$$R(n) = \max_{0 \leq l \leq n} \{\sum_{i=1}^{l} Y_i - l\bar{Y}\} - \min_{0 \leq l \leq n} \{\sum_{i=1}^{l} Y_i - l\bar{Y}\}$$

and an estimate of the variance of the process underlying the data is

$$S^2(n, q) = \sum_{j=-q}^{q} \omega_q(j)\hat{\gamma}(j),$$

where $\hat{\gamma}(j)$ is an estimate of the autocovariance function at lag j and $\omega_q(j)$ are weights. Finally the R/S statistic is then defined as

$$Q(n, q) = \frac{R(n)}{S(n, q)}.$$

Helms et al. (1984) set $q = 0$ and $\omega_0(0) = 1$. The R/S statistic with these restrictions suffers from two disavantages: first its distribution is not known and secondly it can be affected by short-memory components. Lo (1991) modified this statistic by putting $q \neq 0$ in order to deal with these problems. His weights were given by

$$\omega_q(j) = 1 - \frac{j}{q+1}, \qquad q < n$$

and q was chosen as the greatest integer less than or equal to

$$(\frac{3n}{2})^{\frac{1}{3}}(\frac{2\hat{\rho}(1)}{1-\hat{\rho}^2(1)})^{\frac{2}{3}},$$

with $\hat{\rho}(1)$ as an estimate of the first order autocorrelation of the process.

For short memory processes the values of $Q(n, q)$ converge to n^J. d is the long memory parameter and J is related to it by $J = d+1/2$. Mandelbrot and Taqqu (1979) also proved that the process has long memory when $J > 1/2$ and their estimator for J was

$$\hat{J} = \frac{\log(R(n)/S(n))}{\log n}.$$

Table 1 reports the results of the test that suggest a possible fractionally integrated process for the volatility since $\hat{J} > \frac{1}{2}$.

We also compute a Wald type test in time domain similar to the Dickey-Fuller approach, Dolado et al. (2002). In their paper, they test the null hypothesis of a fractional integrated process of order d_0, $FI(d_0)$ versus a fractional integrated process of order d_1, $FI(d_1)$ with $d_1 < d_0$. The test is the t-statistic associated to the coefficient of $\Delta^{d_1} y_{t-1}$ in a regression of $\Delta^{d_0} y_t$ on $\Delta^{d_1} y_{t-1}$ and some lags of $\Delta^{d_0} y_t$. In our case we consider two different null hypotheses: $d_0 = 0.3$ and $d_0 = 0.4$. The t-statistics are normally distributed because under the null the process is stationary and d_1 is estimated by fitting an ARFIMA(1,d,0) to the squared returns series. Fractional integration is not rejected at 5% significance level for the squared returns (see Table 2).

5.4.2 Empirical results

Projection step

Here we present the results of the projection step.

The auxiliary model that best fits the raw data is found using the SNP model described in the chapter 2. The first 50 observations were reserved for forming lags. The values taken by L_u, L_g, L_r, L_p, K_z and K_x were determined by going along a expansion path and the selection criterion used was the BIC (Bayesian Information Criterion), Schwarz (1978).

As always, models that present a small value for the BIC criterion are preferred to the ones with higher values. The expansion path has a tree structure. As Gallant and Tauchen (1996) suggested, better than expanding the entire tree structure is to start expanding L_u keeping $L_r = L_p = K_z = K_x = 0$ till the BIC increases value. The following step is to expand in L_r with $L_p = K_z = K_x = 0$. Next, we expand K_z with $K_x = 0$ and finally L_p and K_x. Sometimes it can happen that the smallest value of the BIC is somewhere inside the tree. So, it is convenient for this reason to expand K_z, L_p and K_x at a few intermediate values of L_r.

The best model according to this procedure has

$$L_u = 2, L_r = 28, L_g = 0, L_p = 1, K_z = 8 \text{ and } K_x = 0$$

and can be characterized as a semiparametric ARCH.

Estimation step

We start by estimating the first benchmark model: the ARSV model of Taylor (1986) supposing that the errors are Gaussian. Table 3 reports the results of the specification test. The ARSV model fails to approximate the distribution of data. The hypothesis null of correct specification is sharply rejected. Next, we analyze the EMM quasi-t-ratios

$\widehat{T_n}$ since they provide suggestive diagnostics for the failure of the model. Figure 6 shows these EMM quasi-t-ratios as a bar graph for the ARSV model. As we can see from the graph, the possible reason for the failure of the model is its difficulty in matching the features of the polynomial part of the SNP score, such as: a_{20} till a_{90}. This means that either the specification $exp(h_t)$ is incorrect or ζ_t is not Gaussian.

We consider the second possibility since the exponential transformation does not seem to be a problem in Gallant et al. (1997). We choose a spline error transformation to the Gaussian innovation. Therefore, the model becomes:

$$y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(h_{2t}/2)\sigma T_{\zeta_t}(\zeta_t),$$

$$T_{\zeta_t}(\zeta_t) = b_0 + b_1\zeta_t + b_2\zeta_t^2 + b_3 I_+(\zeta_t)\zeta_t^2$$

and

$$h_{2t} = \phi h_{2t-1} + \eta_t,$$

where ζ_t is $i.i.d$ $(0,1)$, η_t is $i.i.d(0,\sigma_\eta^2)$, $|\phi| < 1$ and η_t and ζ_t are mutually independent for all t.

In this way, we are allowing a deviation from the Gaussian specification by permitting ζ_t to be a spline error and consequently we are generating extra kurtosis and introducing an asymmetry. We have to impose some restrictions on the b parameters for identification issues: the expected value of $T_\zeta(\zeta)$ and its variance are restricted to be respectively, 0 and 1.

The EMM objective function value is overwhelming reduced, as it may be seen from Table 5, and the moments of the polynomial part of the SNP score are better fitted (see Figure 7). In spite of this, the ARSV model still fails the scores of the ARCH specification (r_{26} and r_{28}) and r_6 is also quite close to the critical value 2, which may indicate that there is a possible misspecification in the conditional variance process due to the statistical significance of r_{26} and r_{29}. Notice that the lags order of the ARCH part decreases with the order of r, we mean that r_{29} corresponds to lag 1 of the ARCH part and r_3 to lag 27.

Next, we investigate the long memory stochastic volatility model by Breidt et al. (1994), that models volatility as a fractional integrated process. The estimated model is a sightly modified version of the model presented in Section 5.3. We introduce a constant and two lags of y_t due to the fact that financial series usually show some autocorrelation. Consider

$$y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \sigma\zeta_t \exp(h_{1t}/2),$$

$$(1-L)^d h_{1t} = \epsilon_t, \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2).$$

The same properties for the errors described in Section 5.3 apply here and we use the same estimation procedure of Gallant et al. (1997). Since the fractionally integrated process can be written as a moving average of infinite order for $\mid d \mid < \frac{1}{2}$, that is

$$h_t = (1-L)^{-d}\epsilon_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} \;\text{ with }\; \psi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)},$$

and the Cholesky factorization of the covariance matrix of h_t is impossible to compute, we truncate the infinite moving average at $k = 1000$ and we trim off the first 10 000 realizations. As is noted by Gallant et al. (1997), some people would say that this procedure does not generate realizations from a long-memory process. Remember that the generated process is going to be stationary for $|d| < 1$ due to the truncation procedure. Nevertheless, Bollerslev and Mikkelson (1996) prove that this procedure still generates a process with high volatility persistence.

From Table 3 we can see that the value of the specification test decreased substantially but the model still does not accommodate all the features of the data. When we use the spline transformation (see Table 5), we can observe that the model does not fit the scores of the ARCH lags of higher order, which means that volatility persistence is still not well approximated with this stochastic specification. Also, the moments associated with the terms in the polynomial approximation are non-zero. Certainly, we need to transform the model in order to capture the extra kurtosis and the strong volatility persistence of data. As it has been showed in Subsection 5.3.2, the introduction of an extra factor of volatility might allow the model to generate extra kurtosis and simultaneously might help in accommodating the volatility persistence. Having in mind this purpose, we estimate the following specification

$$y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \sigma T_{\zeta_t}(\zeta_t)\exp(\frac{h_{1t} + h_{2t}}{2}),$$

$$T_{\zeta_t}(\zeta_t) = b_0 + b_1\zeta_t + b_2\zeta_t^2 + b_3 I_+(\zeta_t)\zeta_t^2,$$

$$(1-L)^d h_{1t} = \epsilon_t, \quad \epsilon_t \ = \sum_{i=1}^{L\epsilon} a_i \epsilon_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) \text{ and } L\epsilon = 0 \quad \text{or} \quad L\epsilon = 1$$

and

$$h_{2t} = \phi_1 h_{2t-1} + \eta_t, \text{with} \ \ \eta_t \sim N(0, \sigma_\eta^2) \ \text{ and } \ |\phi| < 1.$$

We impose to $T_{\zeta_t}(\zeta_t)$ the same restrictions as before, for identification purposes.

The 2FLMSV is a combination of two models: the ARSV of Taylor (1986) and the LMSV of Breidt et al. (1994). The empirical results reported in Table 5 and Figure 9 suggest us that the fit has improved and the fatness of the tails appears to be somewhat better accommodated. Furthermore, the volatility persistence generated by the model seems to approximate better the persistence in the data. In fact, if we look at the parameter estimates in Table 6 for the second variant of the 2FLMSV, we observe that the estimate of d, the fractional integrated coefficient, is 0.561. This means that the first volatility factor is quite persistent and consequently shocks to volatility are going to take time to dissipate. Notice that the h_1 process is still stationary due to the truncation procedure. The second volatility factor is, as expected, fast mean reverting. The estimate for ϕ_1 is -0.208, smaller than 1 in absolute value. Moreover, since h_1 is very slowly mean reverting, if volatility is high today, tomorrow it is going to be high as well, given the value of h_2. So, the 2FLMSV model also allows for cycles of high volatility and vice-versa - volatility clustering.

Finally, the estimates of the variances of the volatility factors, σ_η^2 and σ_ϵ^2, allow us to obtain an estimate of kurtosis. Considering the Gaussian case as an indicator of the spline case, the kurtosis estimated with the 2FLMSV model is around 30% higher than the kurtosis estimated with the LMSV model. So, empirically the 2FLMSV model generates higher kurtosis than the LMSV.

Although the 2FLMSV model seems to capture the main empirical facts (if we observe Figure 9 once more, we observe that the quasi t-ratios of the mean scores are in majority smaller than 2) the hypothesis null of correct specification is still rejected at all relevant significance levels.

Chumacero (1997) studies the small sample properties of EMM estimators of the ARSV model and he confirms the previous findings for the χ^2 specification test, using a Monte Carlo experiment. Inference based on the over identifying restrictions test as well as other χ^2 statistics shows important over rejections. So, if this is the case for the ARSV model, there is a high probability that the same happens with the 2FLMSV model.

A doubt remains: Is the model corrected specified, or shall we disbelieve on the ability of stochastic volatility models (in discrete time) to model the stock returns as the specification test suggests? Gallant et al. (1997) arrived to pessimistic results. In particular, they have shown that the introduction of several modifications to the models produce models that are quite elaborate but "...they still can not accommodate features that could be described as "nonlinear nonparametric".

Finally, we increase the order of the short run process (now ar(2)) and the improvement is almost inexistent. Another possible extension is to include in the model leverage effects. For this, the SNP polynomial part of the auxiliary model should include asymmetry.

5.5 Conclusion

In this chapter we reported some empirical facts of financial time series and we propose a two factor long memory stochastic volatility model as an alternative to the LMSV model of Breidt et al. (1994). We still model the volatility persistence by assuming that the volatility of the returns shows a long memory feature captured by a fractionally integrated process. The innovation is that we introduce a short run volatility factor that allows the model to generate extra kurtosis and simultaneously to accommodate better the volatility persistence.

The estimation method that we use in this chapter is the EMM (efficient method of moments) by Gallant and Tauchen (1996) because of its testing advantages. In fact, the minimized criterion function scaled by the number of observations follows asymptotically a chi-square distribution which allows us to test if the model is corrected specified. Finally, there is empirical evidence that the short run volatility factor seems to improve the EMM criterion as in Ming Liu (2000) and the long memory stochastic volatility model with two factors of volatility performs better than the two benchmark models in terms of the specification test but it is still rejected in its original form and its extensions as in Gallant et al. (1997). Since it is known that the χ^2 specification test tends to over reject the null of correct specification, we face a dilemma; Is the model appropriate to describe financial data or shall we doubt about the ability of stochastic volatility models (in discrete time) in modelling the stock returns? Gallant et al. (1997) emphasis the disability of stochastic volatility models to fit the main features of data. In particular, they have shown that the introduction to the models of several modifications produce models that are quite elaborate but "...they still can not accommodate features that could be described as "nonlinear nonparametric".

In future research, we think of computing a Monte Carlo experiment in order to infer if the specification test is over rejecting when we use the 2FLMSV model and in affirmative case we think of testing the model further by studying its forecasting performance.

Bibliography

[1] Baillie, R. T. (1996): Long Memory Processes and Fractional Integration in Econometrics, *Journal of Econometrics* 73, 5-59.

[2] Bollerslev, T. (1986): Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics,* 31, 307-327.

[3] Bollerslev, T., and H. Mikkelsen (1996): Modeling and Pricing Long Memory in Stock Market Volatility, *Journal of Econometrics*, 73/1, 151-184.

[4] Breidt, F., N. Crato and De Lima P. (1994): Modeling Long-Memory Stochastic Volatility, *Manuscript*, Iowa State University, Ames, IA.

[5] Carnero, M. A., Peña, D. and Ruiz, E. (2001): Is Stochastic Volatility More Flexible than GARCH?, *Working paper* 01-08, Universidad Carlos III de Madrid.

[6] Chernov, M., Gallant, A. R., Ghysels, E. and G. Tauchen (2003): Alternative Models for Stock Price Dynamics, *Journal of Econometrics*, 116, 225-257.

[7] Chumacero, R. A. (1997): Finite Sample Properties of the Efficient Method of Moments, *Working paper*, University of Chile.

[8] Dolado, J.,Gonzalo J. and Mayoral L. (2002): A Fractional Dickey-Fuller Test for Unit Roots, *Econometrica*, 70, 5, 2002.

[9] Engle, R. F. and Patton A. J. (2001): What Good Is a Volatility Model?, *Working paper,* University of California, San Diego.

[10] Gallant, A.R., Hsieh, D. and G. Tauchen (1997): Estimation of Stochatic Volatility Models with Diagnostics, *Journal of Econometrics*, 81, 159-192.

[11] Gallant, A. R. and G. Tauchen (1996): Which Moments to Match, *Econometric Theory*, 12, 657-681.

[12] Ghysels, E., A. Harvey and E. Renault (1995): Stochastic Volatility, in *Handbook of Statistics vol. 14.* Amsterdam: North Holland.

[13] Harvey, A. (1993): Long Memory and Stochastic Volatility, *Manuscript,* London School of Economics.

[14] Helms, B., Kaen, F., and Rosenman, R. (1984): Memory in Commodity Futures Contracts, *Journal of Futures Markets*, 4, 559-567.

[15] Hosking, J.R.M. (1981): Fractional Differencing, *Biometrika*, 68, 165-176.

[16] Liu, M. (2000): Modeling Long Memory in Stock Market Volatility, *Journal of Econometrics* 99,139-171.

[17] Lo, A.W. (1991): Long Term Memory in Stock Market Prices, *Econometrica* 59, 1279-1313.

[18] Mandelbrot, B.B. and Taqqu, M (1979): Robust R/S Analysis of Long-Run Serial Correlation, *Proceedings of the 42nd Session of the International Statistical Institute*, International Statistical Institute.

[19] Schwarz, G. (1978): Estimating the Dimension of a Model, *The Annals of Statistics*, vol 6, no. 2, pp 461-464.

[20] Taylor, S.J (1986): Modelling Financial Time Series, *John Wiley.*

[21] Zaffaroni, P. (2000): Time Series Models of Changing Volatility, *Lecture notes.*

5.6 Figures and Tables

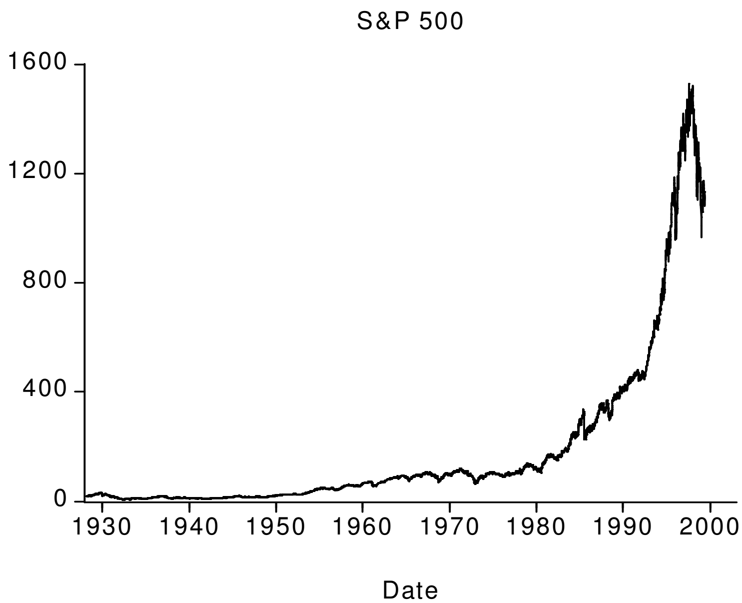

Figure 1

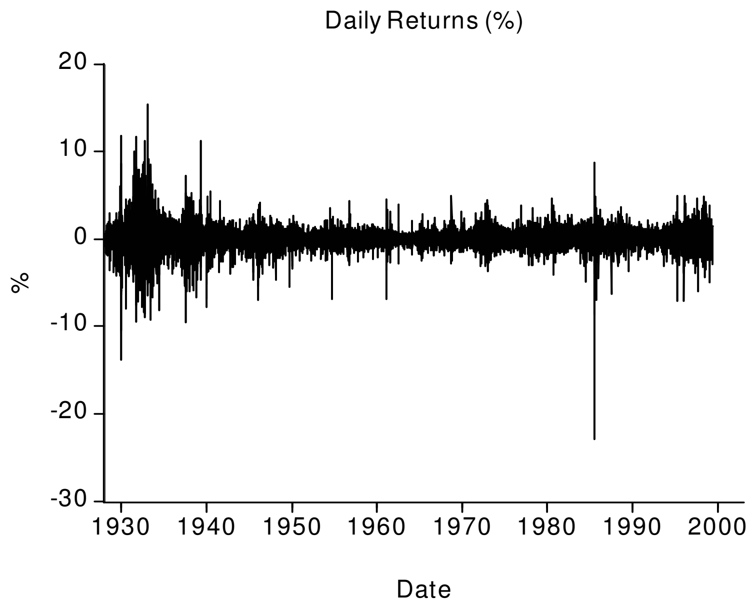

Figure 2

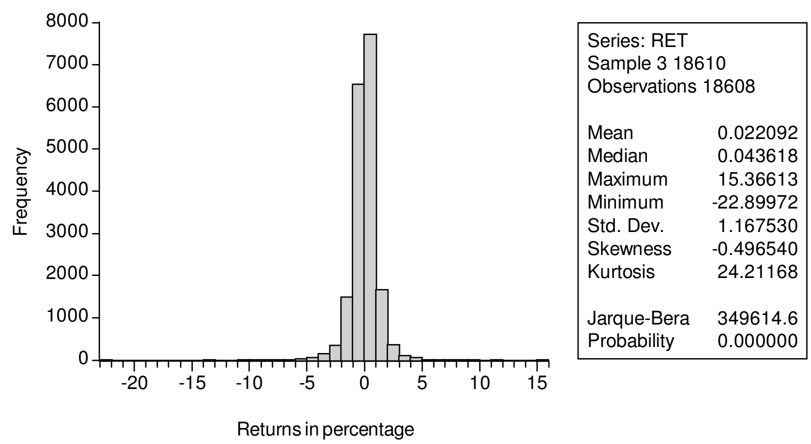

Figure 3: Histogram of Returns

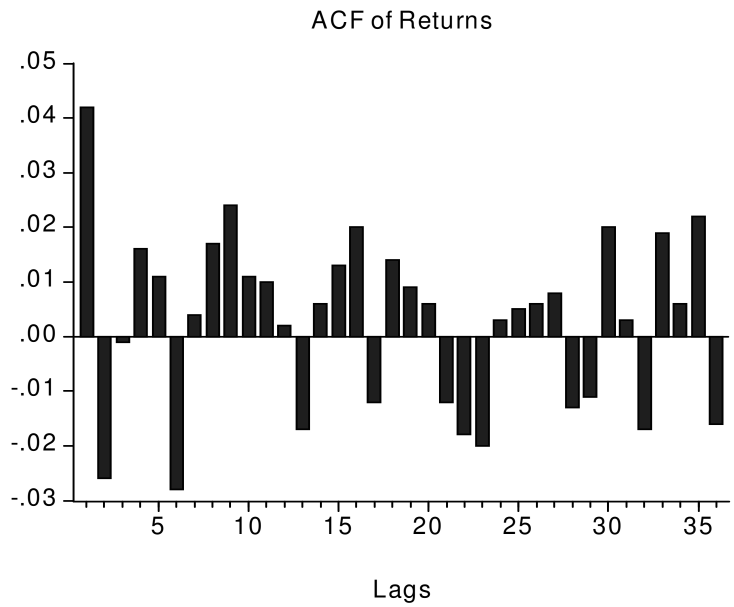

Figure 4

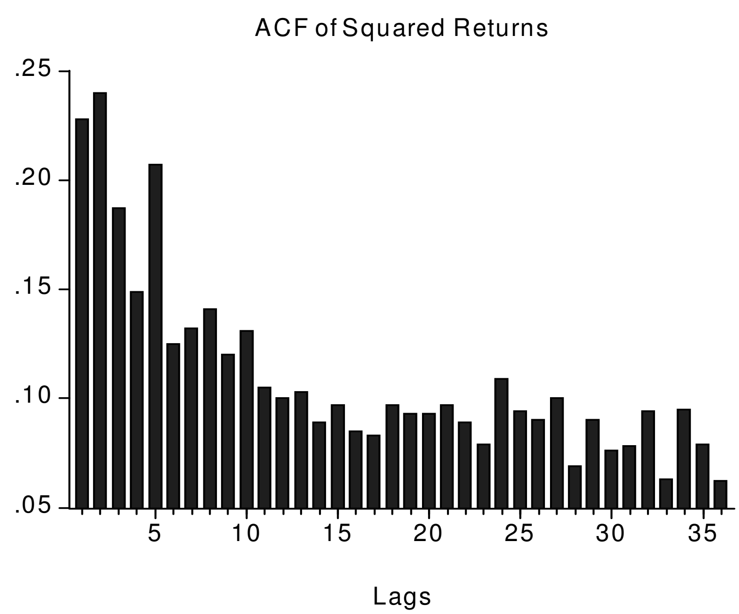

Figure 5

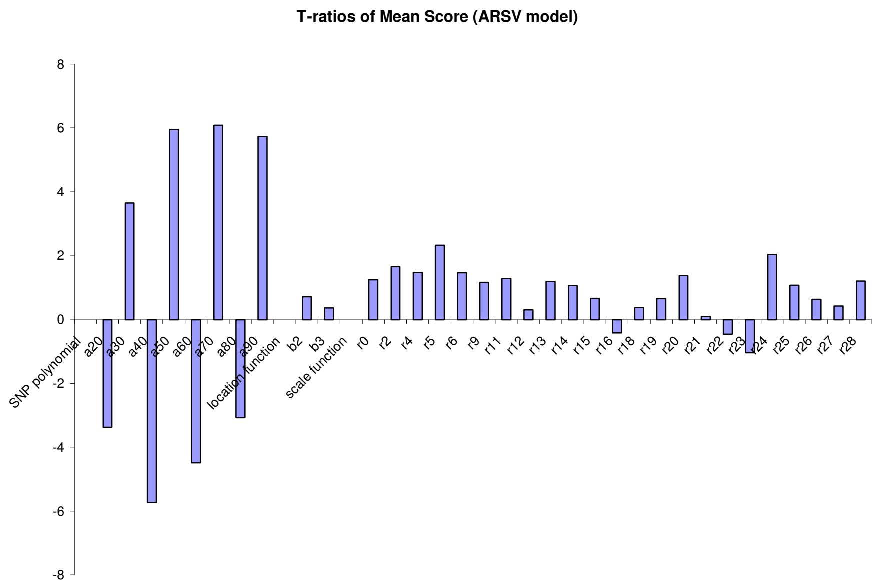

Figure 6

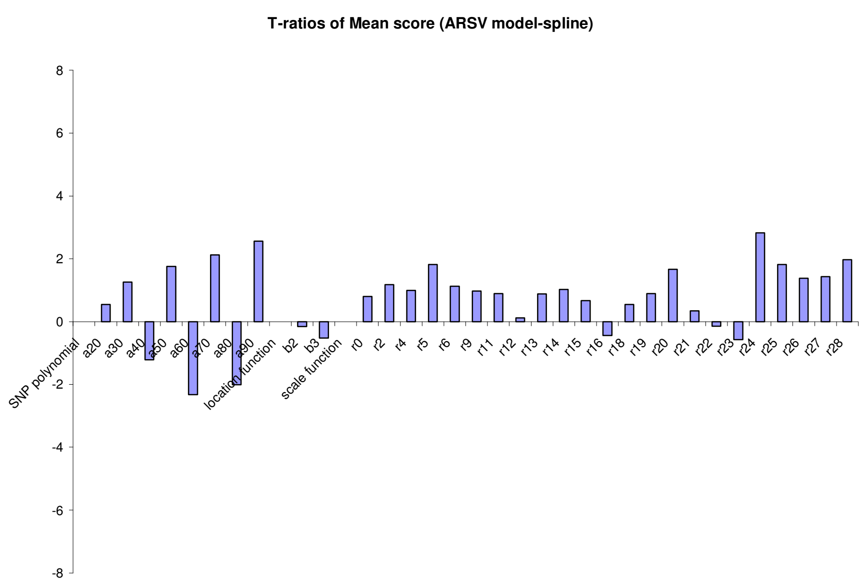

Figure 7

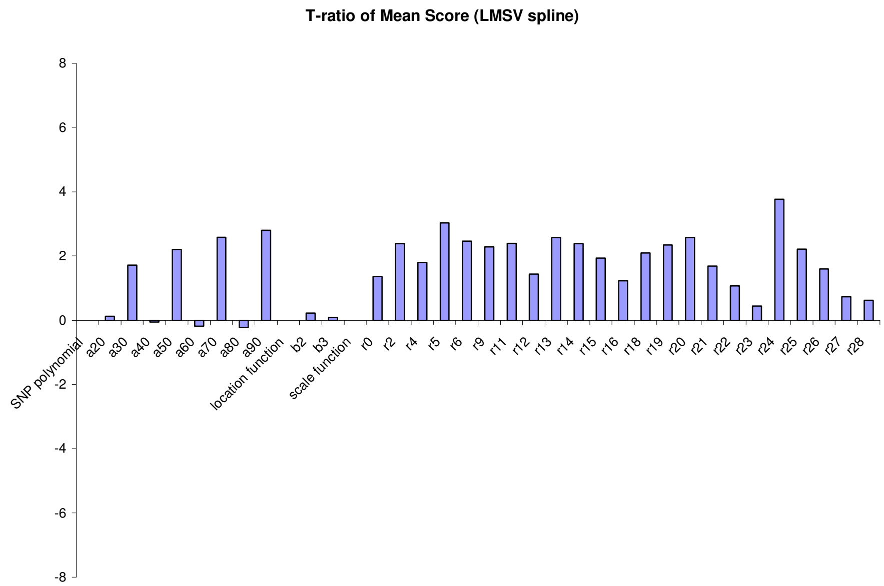

Figure 8

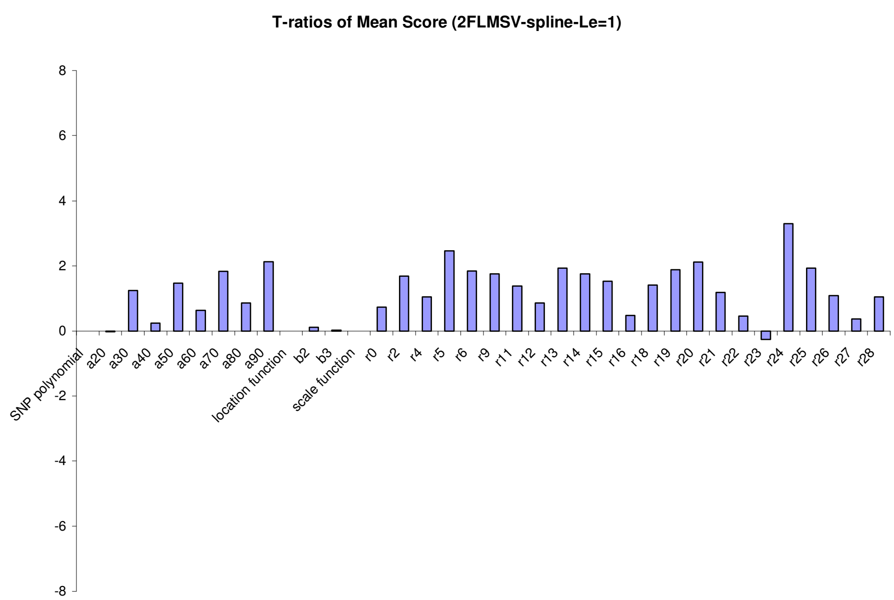

Figure 9

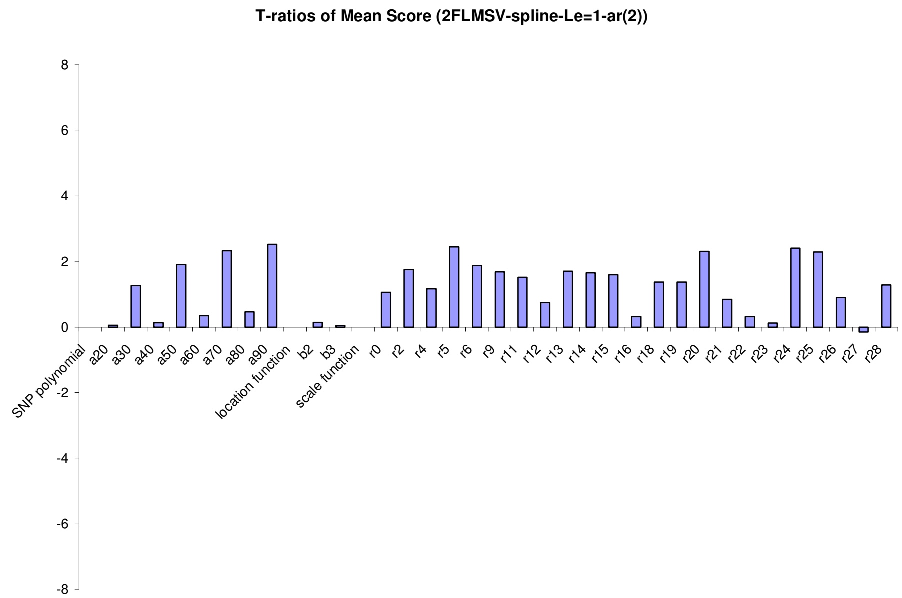

Figure 10

R/S	$\mathbf{q = 0}$	$\mathbf{q = q^*}$
Q	1432.63	592.93
$\hat{J}$	0.739	0.649
$\hat{d}$	0.239	0.149

Table 1

S&P 500	$t_{H_0:d_0=0.3}$	$t_{H_0:d_0=0.4}$	$\hat{d}_1$
Squared Returns	-1.649476	-4.049*	0.2632

Table 2: * means that the null hypotheses is rejected

Model	χ^2	df	$p-value$	$L\epsilon$
ARSV, Gaussian error	144.1972	27	<0.0001	0
LMSV, Gaussian error	122.1446	28	<0.0001	0
2FLMSV, Gaussian error	116.7419	26	<0.0001	0
2FLMSV, Gaussian error	110.4044	25	<0.0001	1

Table 3: χ^2 is the value of the EMM criterion, which follows a χ^2statistic with degree of freedom of df. $L\epsilon$ is the autocorrelation order of the error of the fractional integrated process for the volatility factor.

	μ_y	c_1	c_2	σ_y	σ_η	ϕ	a_1	σ_ϵ	d	$L\epsilon$
Gaussian										
ARSV	0.041	0.109	-0.054	0.756	0.130	0.990				0
LMSV	0.036	0.104	-0.051	0.652				0.427	0.541	0
2FLMSV	0.036	0.105	-0.051	0.661	0.121	-0.453		0.408	0.555	0
2FLMSV	0.037	0.106	-0.051	0.691	0.175	-0.421	0.571	0.191	0.562	1

Table 4: Fitted parameter values (Gaussian errors). Not all the parameters are free due to identification restrictions across parameters.

Model	χ^2	df	$p - value$	$L\epsilon$
ARSV, spline error	70.9361	23	<0.0001	0
LMSV, Spline error	62.7775	24	0.00026	0
2FLMSV, spline error	56.5525	22	<0.0001	0
2FLMSV, spline error	51.3086	21	0.00024	1
2FLMSV$_{ar(2)}$, spline error	45.5148	20	0.001	1

Table 5: χ^2 is the value of the EMM criterion, which follows a χ^2statistic with degree of freedom of df.

	μ_y	c_1	c_2	σ_y	σ_η	ϕ_1	ϕ_2	σ_ϵ
spline								
ARSV	0.109	0.113	-0.047	1.131	0.099	0.992		
LMSV	-0.218	0.09	-0.065	0.686				0.426
2FLMSV	-0.284	0.08	-0.07	0.816	0.219	-0.219		0.405
2FLMSV	-0.360	0.07	-0.087	0.754	0.309	-0.208		0.177
2FLMSV$_{ar(2)}$	-0.209	0.09	-0.062	0.708	0.111	-1.13	-0.697	0.19

Table 6: Fitted parameter values (spline errors). Not all the parameters are free due to identification restrictions across parameters.

	a_1	b_0	b_1	b_2	b_3	d	$L\epsilon$
spline							
ARSV		0.000	0.723	-0.120	0.156		0
LMSV		0.420	1.058	-0.073	0.0075	0.56	0
2FLMSV		0.449	0.882	-0.051	-0.025	0.572	0
2FLMSV	0.601	0.618	0.964	-0.056	-0.051	0.561	1
2FLMSV$_{ar(2)}$	0.567	0.396	1.062	-0.043	-0.039	0.576	1

Table 6 (cont.)

Chapter 6

Conclusions and Future Research

The main purpose of this thesis has been to model and forecast the volatility of the financial series using specifications either in continuous or discrete time.

In the third chapter, we study why stochastic volatility models with one factor of volatility in continuous time fail to fit the main features of data. The main reason, considering our sample, is that they are not able to accommodate all the persistence of data. In order to capture this strong volatility persistence, we introduce a feedback feature into the factor volatility specification that allows the volatility of the volatility factor to be high when itself is high and vice-versa. Under these conditions, the model with one factor and feedback is able to capture simultaneously the kurtosis and the persistence of data, without even missing the stock crash of October 1987.

Several authors as: Beine and Laurent (2000), Granger and Hyung (1999) and Diebold and Inoue (1999) report that there is a correlation between volatility persistence and changes of pattern, if we do not account for the latter. In our case the existence of this strong persistence can be perhaps explained, in some degree, by this evidence and consequently the feedback could be seen as an imperfect or indirect substitute of a jump.

In future research, we would like to investigate deeply this question by considering a sample with a visible volatility change. The idea would be to fit a stochastic volatility model with a jump and the stochastic volatility model with both, one factor of volatility and feedback. If the second performs similar to the first, the feedback factor could be seen as a substitute of a jump, with the advantage of avoiding some estimation difficulties. We would also compute a Monte Carlo experiment to generalize the conclusions.

In the fourth chapter we ask the question: Is there evidence that stochastic volatility models in continuous time forecast volatility, accurately? Are their forecasting performances better than the forecasting performances of other well known models? And the answer is yes for our sample.

The evaluation procedure adopted is the following: we regress the realized volatility, computed using 15-minutes data, on a constant and on the volatility forecasts of several models. If the volatility forecast is an unbiased estimator of realized volatility, considered a good measure of ex-post volatility, the constant is statistically insignificant and the coefficient of the volatility forecast is statistically equal to one. We observe for the out of sample periods that the benchmark models: GARCH and ARFIMA, face problems in tracking the growth pattern of the realized volatility. This occurrence is due to the volatility increasing at the very end of the sample. Remember we are using the same data of the third chapter. In order to avoid this phenomenon we consider the year before its occurrence and we observe that the continuous time model still performs better than the benchmarks. In future research, would be interesting to observe if these results are repeated for other financial series or if they are a particular feature of the Microsoft data.

Finally, in the fifth chapter we propose a model in discrete time with two factors of volatility. Our main purpose with this work is to ask the question: Is there evidence that two volatility factors in discrete time are enough to fit the features of data, as in continuous time? Theoretically we prove, by introducing a second factor of volatility whose aim is to capture the short-run dynamics, that the model is able to generate extra kurtosis and to accommodate better the volatility persistence. Furthermore, we are also able to prove that the autocorrelation of first order of the squared returns implied by our model is smaller than the one implied by the benchmark model: LMSV. Empirically, we test the performance of our model by fitting it to the returns of the S&P 500 composite index. The results show us that the long memory stochastic volatility model with two factors of volatility performs better than the two benchmark models. However, it is still rejected in its original form and extensions, as in Gallant et al. (1997).

Chumacero (1997) studies the small sample properties of EMM estimators of the ARSV model and he confirms the previous findings for the χ^2 specification test, using a Monte Carlo experiment. Inference based on the over identifying restrictions test as other χ^2 statistics shows important over rejections. So, if this is the case for the ARSV model, there is a huge probability that the same happens with the 2FLMSV model.

A doubt remains: Is the model corrected specified? or shall we disbelieve on the ability of stochastic volatility models (in discrete time) in modelling the stock returns as the specification test suggests?

In future research, we think of computing a Monte Carlo experiment in order to infer if the specification test is over rejecting when we use the 2FLMSV model and in affirmative case we think of testing the model further by studying its forecasting performance.

Bibliography

[1] Beine, M. and Laurent, S. (2000): Structural Change and Long Memory in Volatility: New Evidence from Daily Exchange Rates , *Discussion paper*, Université of Liège.

[2] Chumacero, R. A. (1997): Finite Sample Properties of the Efficient Method of Moments , *Working paper*, University of Chile.

[3] Diebold, F.X., Inoue,A. (1999): Long Memory and Structural Change , *Manuscript*, Departmant of Finance, Stern School, NYU, May 1999.

[4] Gallant, A.R., Hsieh, D. and G. Tauchen (1997): Estimation of Stochatic Volatility Models with Diagnostics, *Journal of Econometrics*, 81, 159-192.

[5] Granger, C.W.J., Hyung, N. (1999): Occasional Structural Breaks and Long Memory, UCSD, *Discussion paper* 99-14, June 1999.