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

Maria Helena Lopes Moreira da Veiga*
Universitat Autònoma de Barcelona
Edifici B, 08193 Bellaterra, Spain
Faculdade de Economia do Porto
Rua Dr. Roberto Frias, 4200 Porto, Portugal

September 2003

Abstract

This paper 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 in capturing the strong volatility clustering of data, caused by a possible change in the pattern of volatility in the last part of the sample. We use the Efficient Method of Moments (EMM) by Gallant and Tauchen (1996) to estimate logarithmic models with one and two stochastic volatility factors (with and without feedback) and to select among them.

JEL Classification: G13, C14, C52, C53
Keywords: Efficient Method of Moments, One (Two) Factor Volatility Logarithmic Model, Mean-Reversion, Persistent Volatility, Feedback, Projection, Seminonparametric (SNP), Reprojection.

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. Along these three last decades several models have been proposed with the aim of capturing these empirical facts. Stochastic volatility models, for instance, were

*Grant Sponsors: FEP and FCT (programa Praxis XXI). Email contact: mlopes@idea.uab.es.

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[1].

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 stochastic volatility models with only one volatility factor are not able to capture simultaneously the extra kurtosis and the volatility persistence. The introduction of a second factor of volatility allows that one might be slow mean reverting while the other might accomodate the fat tails.

This paper differs from previous ones since 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 structural change in volatility. The estimated models are direct extensions of the Gallant and Tauchen's (2001) by including the feedback feature[2]. The advantage of these modelizations compared to an affine is that they allow that the volatility be dependent on state, although there are not closed-form solutions[3]. 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[4].

This paper uses EMM (Efficient Method of Moments) by Gallant and Tauchen (1996). It is based on two compulsory phases: 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 asymptoti-

[1]See for example Chernov and Glysels (2000).

[2]We estimate logarithmic models that differ from the ones proposed in Chernov et al. (2001) by not allowing a stochastic instantaneous expected return.

[3]Pricing formulas may be computed by simulation.

[4]The advantage of two factors is that one can take care of persistent stochastic volatility while the other tries to deal with the tail behavior.

cally follows a chi-square distribution, it allows to apply 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 paper 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[5].

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.

The paper 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 paper.

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), etc..

This paper includes a feedback factor in the model[6] proposed in Gallant and Tauchen (2001), such as:

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

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

[5]However, the simulation of the evolution of the state vector is quite possible. The EMM is based on this.

[6]The specification differs from the one proposed in Chernov et al. (2001) because theirs include a stochastic drift in equation 1 and account for leverage effects.

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

where P_t is the daily price of a share of Microsoft and W_i with $i = 1, 2, 3$ are 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}[7]. This specification nests two groups of models: the first includes the logarithmic model with one volatility factor ($L1$), with $\beta_{13} = 0$ and $\beta_{22} = 0$, 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 is quite relevant especially when there seems to exist a change in the pattern of volatility. We can observe that structural change and feedback feature are imperfect substitutes, in the sense that the introduction of this feature allows one to capture the structural change[8]. Moreover, the volatility factors of equation 1 present drifts and volatilities that are linear functions of themselves, respectively and the drifts in equations 2 and 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 1.

2.0.1 Identification restrictions

To achieve identification it is necessary to impose some restrictions[9]. In this concrete case for the logarithmic specification we set

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

Therefore the previous specification becomes, for the first group:

[7]As Chernov, Gallant, Ghysels and Tauchen (2003) refer, the logarithmic models with feedback violate the standard regularity 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 equation1 with, as they say, "...the linear growth condition at the level of volatility so high that is unlike to be observed in the U.S. equity index returns." For more details please see the appendix A of their paper.

[8]We do not introduce jumps in the specifications because the change in the pattern of volatility occurs in the last part of the sample, which makes unfeasible the empirical application of such models.

[9]These restrictions are the minimum necessary to achieve identification.

$$\frac{dP_t}{P_t} = \alpha_{10}dt + \exp(\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t})dW_{1t} \quad (4)$$

$$dU_{2t} = \alpha_{22}U_{2t}dt + dW_{2t} \quad (5)$$

$$dU_{3t} = \alpha_{33}U_{3t}dt + dW_{3t} \quad (6)$$

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

$$\frac{dP_t}{P_t} = \alpha_{10}dt + \exp(\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t})dW_{1t} \quad (7)$$

$$dU_{2t} = \alpha_{22}U_{2t}dt + (1 + \beta_{22}U_{2t})dW_{2t} \quad (8)$$

$$dU_{3t} = \alpha_{33}U_{3t}dt + (1 + \beta_{33}U_{3t})dW_{3t} \quad (9)$$

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 Efficient Method of Moments

The models above are estimated using the Efficient Method of Moments (EMM).

Let $\{y_t\}_{t=-\infty}^{\infty}$, $y_t \in \Re^M$, 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 $p(y/\ x,\rho) = p(y_{-L}, ..., y_0|\rho)/p(y_{-L}, ..., y_{-1}/\rho)$. Tthe 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 as close as the efficiency of the ML estimator by making K large enough, Gallant and Long (1997).

3.0.2 Projection step

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[10]. The reshaped density is given by

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

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

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 $y = \sigma z + \mu$,

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

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 $P_K(z,x)$.

SNP Estimation Results In this subsection of the paper 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 previous section. 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 $L_r = L_p = K_z = K_x = 0$ till 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[11] has

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

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

and can be characterized as a Semiparametric GARCH.

3.1 The estimation step

In this section 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[12].

Defining the moment conditions by

[12]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.

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

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

$$\hat{\rho}_n = \arg\min m'(\rho, \tilde{\theta}_n)(\tilde{I}_n)^{-1}m(\rho, \tilde{\theta}_n). \tag{10}$$

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

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

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

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

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

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

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

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[13]. 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

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

[13] 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.

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.

4 EMM Empirical Results

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 4 and 7 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 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[14]. 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 and reports 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[15]. 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 - long memory property. If we also observe the graph of volatility we will see a possible structural change in the volatility pattern for the last period of the sample[16]. Recent studies, for instance: Beine and Laurent (2001), Granger and Hyung(1999) and Diebold and Inoue (1999) report that there is an increase in volatility persistence if we do not account for structural 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 ACF's (autocorrelation functions) of the squared returns of the absolute values of returns and we observe specially

[14] See tables 2 and 4.

[15] 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.

[16] See figure 3.

for the latter that the ACF decays slower towards zero[17]. 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 paper reports an empirical result for that parameter of -0.902 much smaller in absolute value than the previous one[18]. Both evidences 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 structural change 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 annually 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 modelize the data. We observe that one factor of volatility is extremely slow mean reversing while the other is very fast mean reversing[19].

Finally, the table 4 summarizes the EMM quasi-t-ratios diagnostics for $L1$, $L1F$ and $L2$. These statistics are asymptotically normal when evaluated at the true parameter values. Since they are evaluated at the point estimates they are asymptotically downward biased relative to 2.0. 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 paper 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

[17] See figures 6 and 7.

[18] See table 3.

[19] 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.

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.

4.1 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 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 $\{\hat{y}_\tau\}_{\tau=1}^{t}$[20]. 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:

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

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

With this procedure we obtain calibrated functions inside the simulation that gives predicted us values of U_{2t} and U_{3t} given $\{y_\tau\}_{\tau=1}^{t}$. 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 10.0,10.1, 11 and 12 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 sightly slower moving than $\tilde{U}_{3t}$ as we can verify by figures 10.0 and

[20] In fact, given the length of simulation, these regressions are as Gallant and Tauchen (2001) say analytic projections.

10.1[21]. 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 and at the same time it is computationally faster to implement.

5 Conclusion

This paper 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 data, $L1F$ 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 a possible structural change in volatility. 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 stochastic volatily model for this sample. The $L1F$ model is able to reproject volatility quite well. It even does not miss the crash of 1987.

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

[21] $\tilde{U}_{2t}$ 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 asymetry to the estimate.

Acknowledgements *I thank my advisor Michael Creel for introducing me to the idea of Efficient Method of Moments and for his advice and encouragement during the research. I also thank seminar participants at the Symposium of Economic Analysis in Salamanca 2002 and at the Universitat Autònoma de Barcelona for the helpful remarks.*

References

[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] Aït- Sahalia, Yacine (1999a): Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Methods, *Working paper*, Princeton University.

[4] Bates, D. (1996a): Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options, *Review of Financial Studies,* 9, 69-107.

[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*, University of Liège.

[7] 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.

[8] Campbell, J., A. Lo and C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.

[9] Chernov, M., Gallant, A. R., Ghysels, E. and G. Tauchen (2001): Alternative Models for Stock Price Dynamics, *Discussion paper*, University of North carolina at Chapel Hill.

[10] 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.

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

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

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

[13] Gallant, A.R., C.-T. Hsu and G. Tauchen (1999): Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance, *Review of Economics and Statistics,* (forthcoming).

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

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

[16] Gallant, A. R. and G. Tauchen (1989): Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications, *Econometrica*, 57, 1091-1120.

[17] Gallant, A.R. and G. Tauchen (1993): SNP: A Program for Nonparametric Time Series Analysis, version 8.3, User's Guide, *Discussion paper*, University of North carolina at Chapel Hill.

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

[19] Gallant, A. R. and G. Tauchen (1997): EMM: A Program for Efficient Method of Moments Estimation, Version 1.4, User's guide, *Discussion paper*, University of North carolina at Chapel Hill.

[20] 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.

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

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

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

[24] 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.

[25] Nelson, D. B. (1990): ARCH Models as Diffusion Approximations, *Journal of Econometrics* 45, 7-39.

[26] Sundaresan, S. M. (2000): Continuous-Time Methods in Finance: A Review and an Assessment,*The Journal of Finance*, VOL LV, NO.4, 1569-1622.

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

6 Figures and Tables

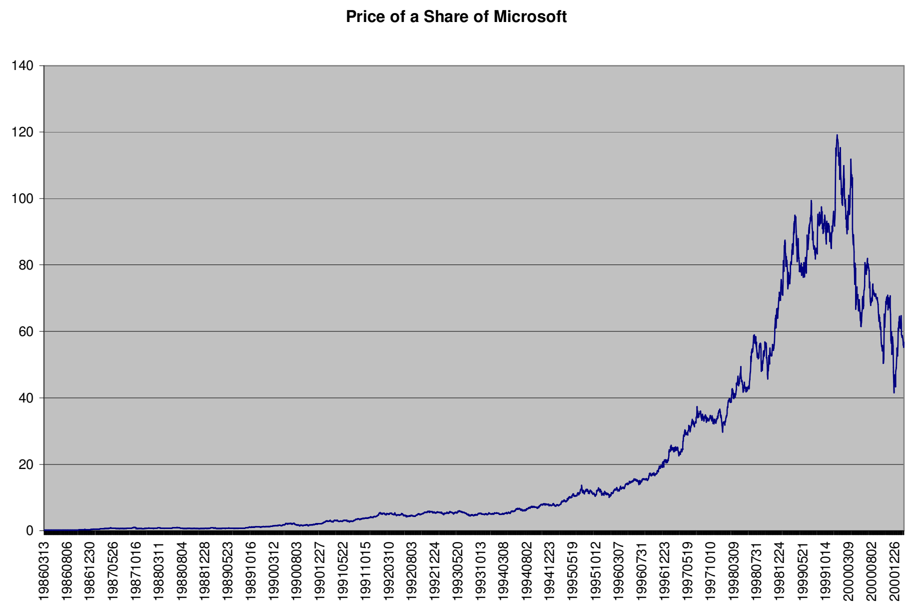


Figure 1

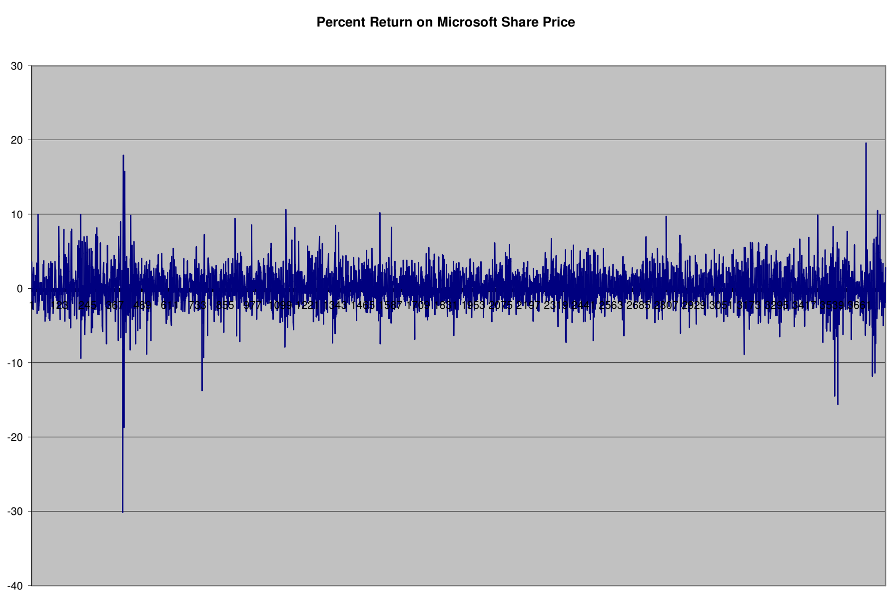


Figure 2

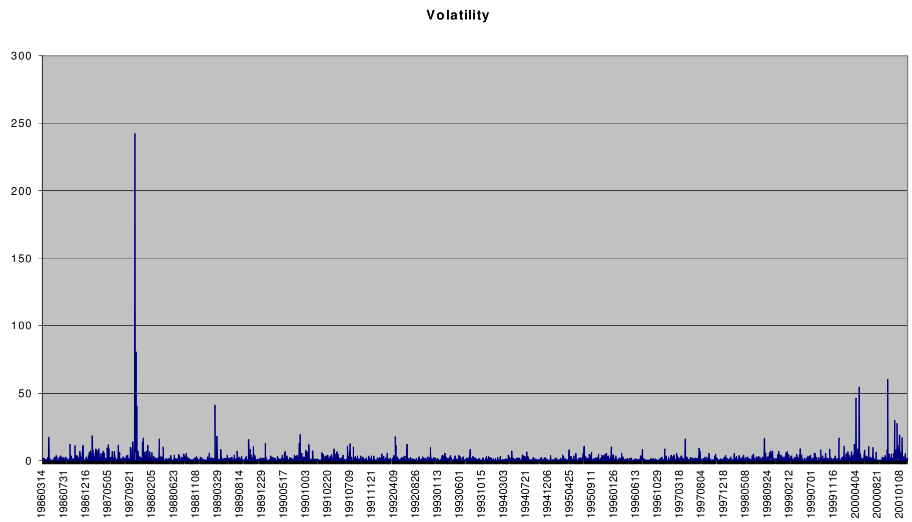


Figure 3

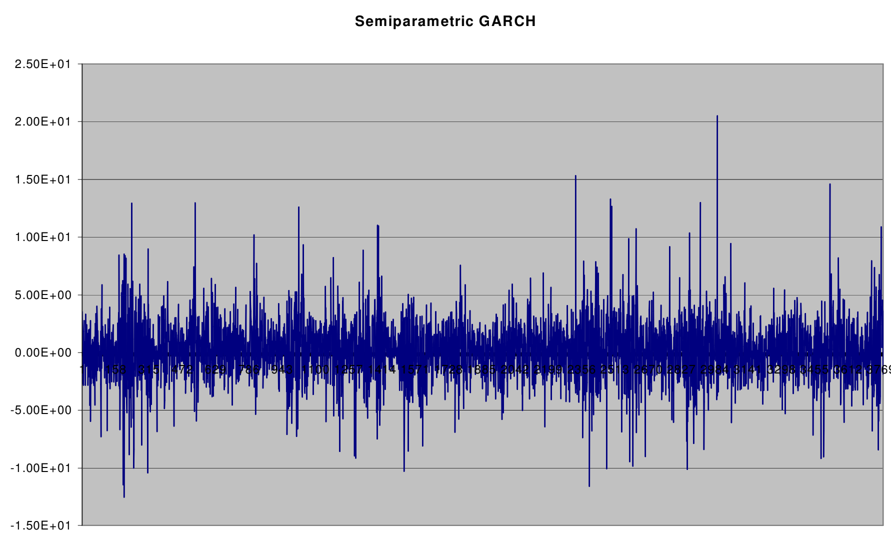


Figure 4

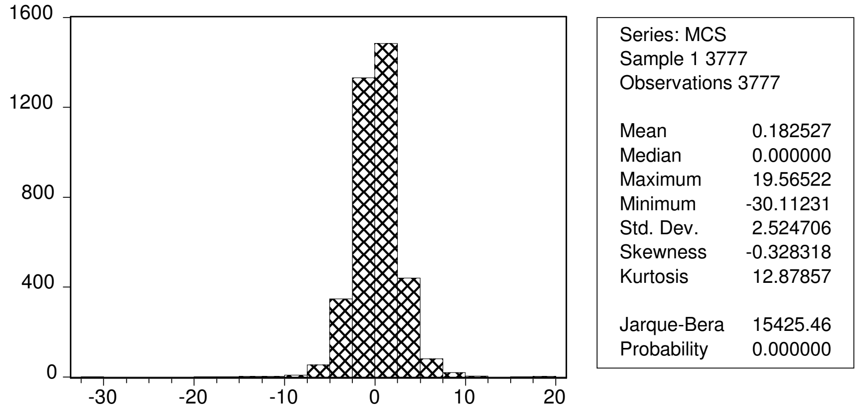


Figure 5: Sample Statistics- Percent Return on Microsoft, March 14, 1986 - February 23, 2001

			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).

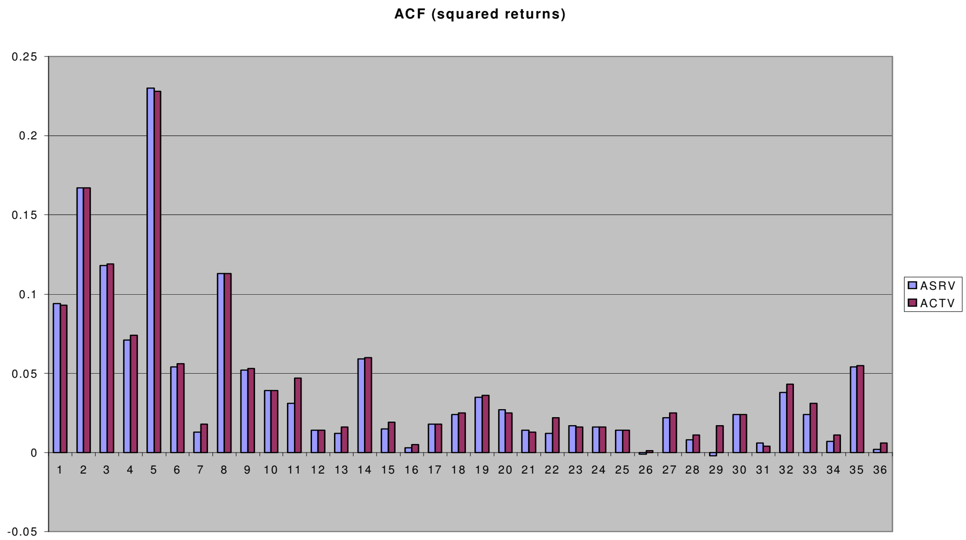


Figure 6: ACF's of the squared returns of the Gallant and Tauchen (2001) sample (ACRV) and of my sample (ACTV).

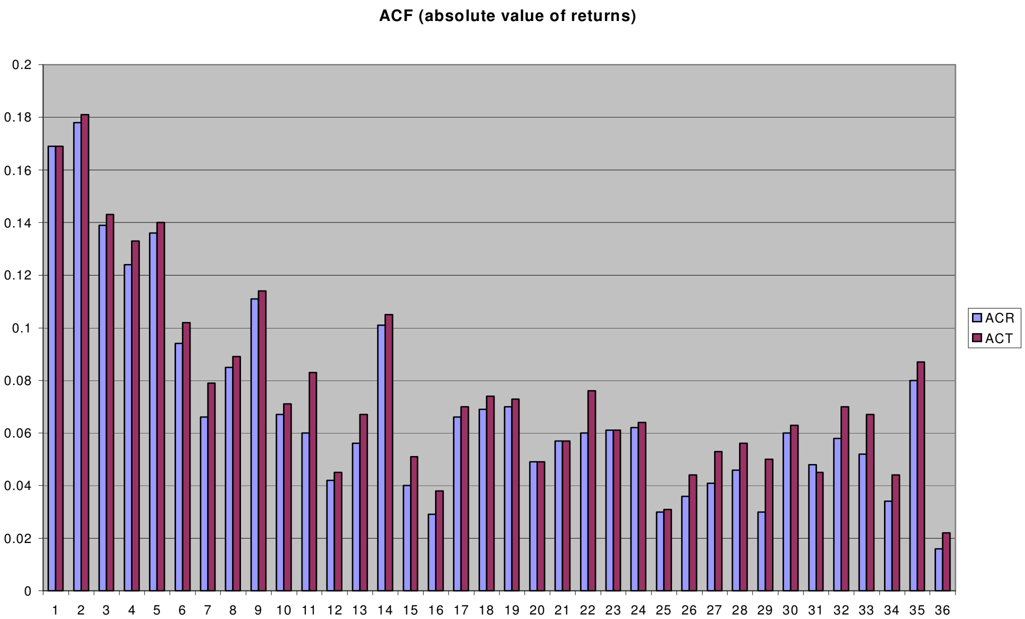


Figure 7: ACF's of the absolute value of returns. Gallant and Tauchen (2001) sample ACR and my sample ACT.

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
Sale 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

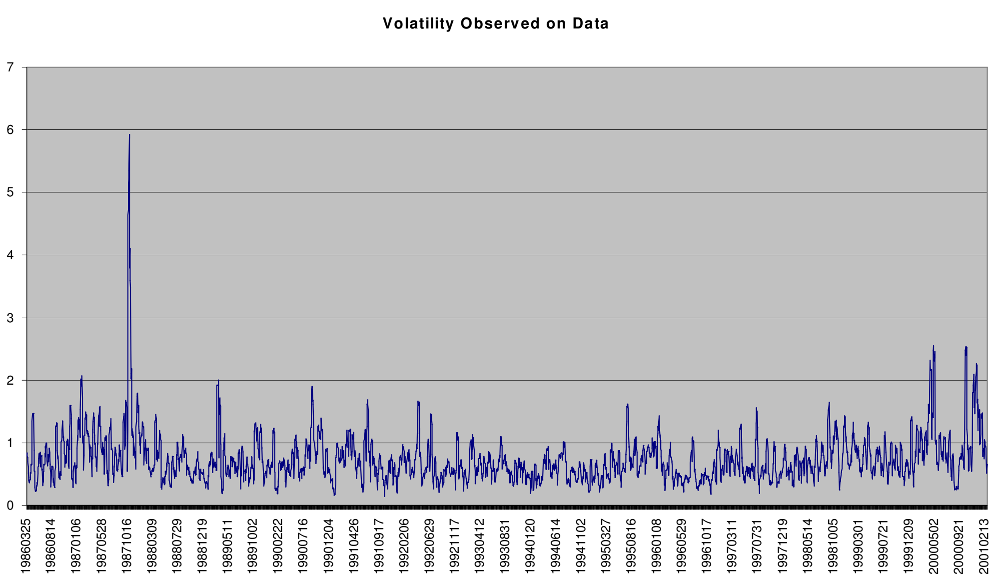


Figure 8

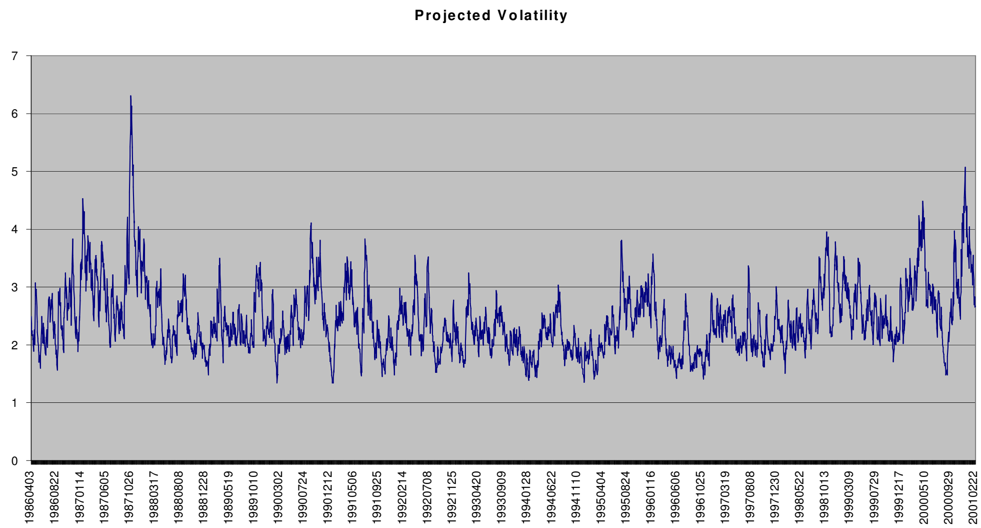


Figure 9

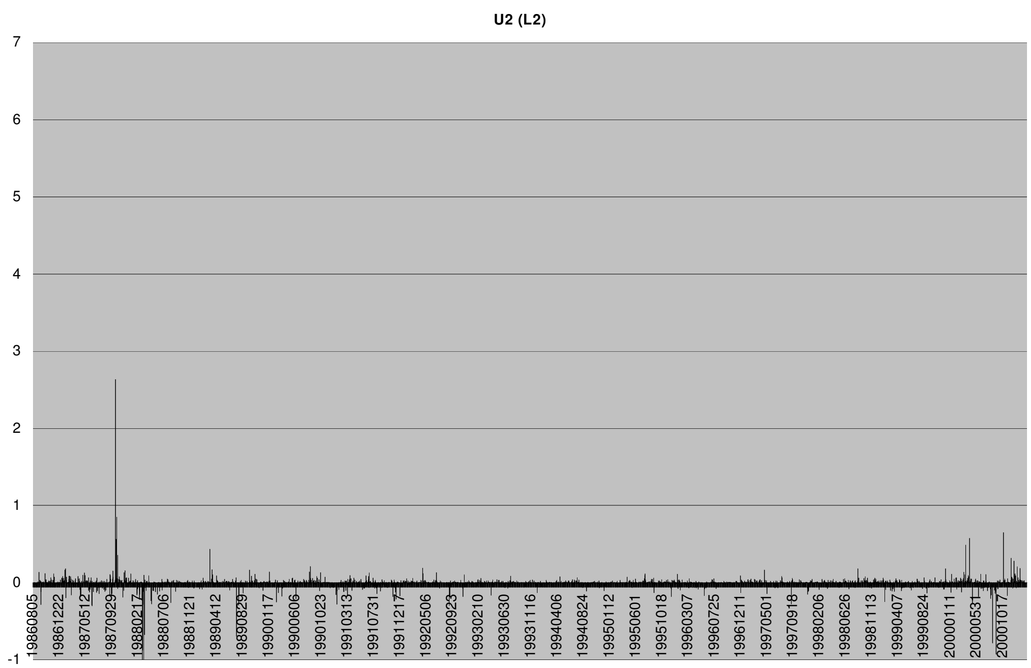


Figure 10.0

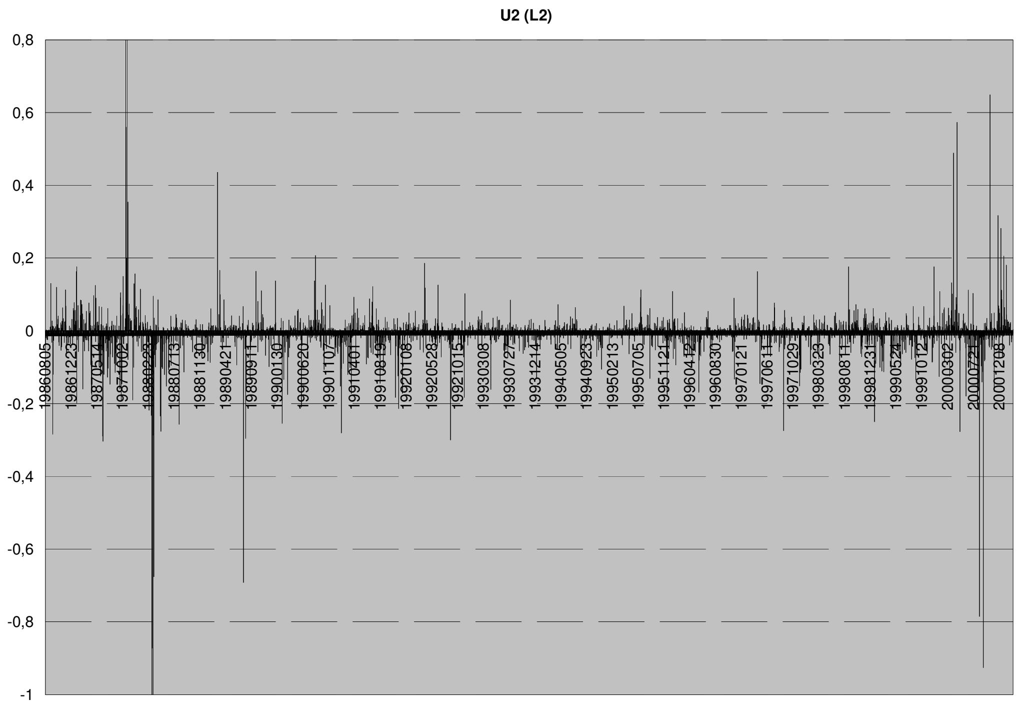
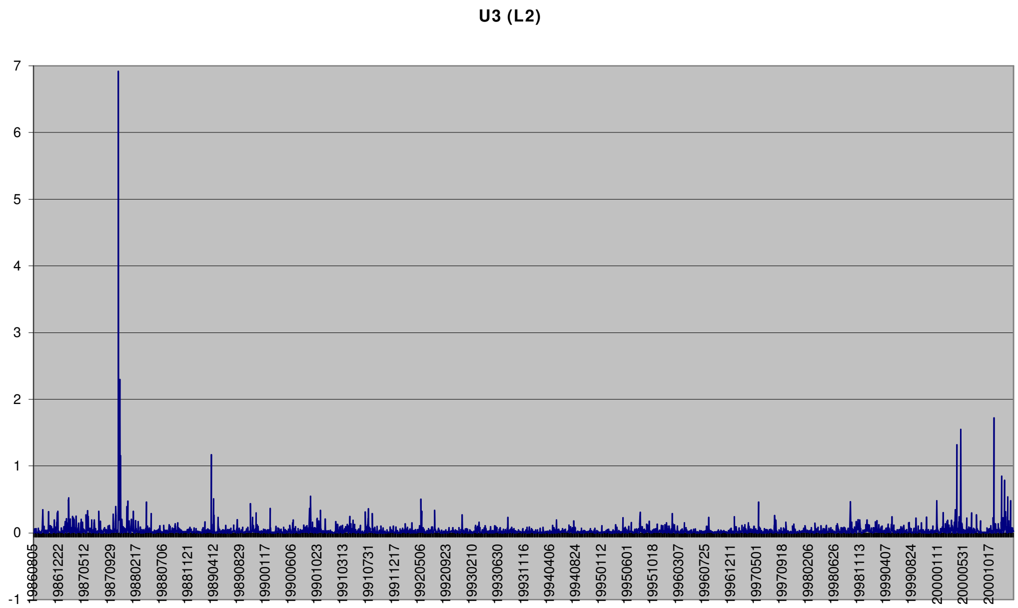


Figure 10.1 - Different scale



U3 (L2)
7
6
5
4
3
2
1
0
-1
19860805
19861222
19870512
19870929
19880217
19880706
19881121
19890412
19890829
19900117
19900606
19901023
19910313
19910731
19911217
19920506
19920923
19930210
19930630
19931116
19940406
19940824
19950112
19950601
19951018
19960307
19960725
19961211
19970501
19970918
19980206
19980626
19981113
19990407
19990824
20000111
20000531
20001017

Figure 11

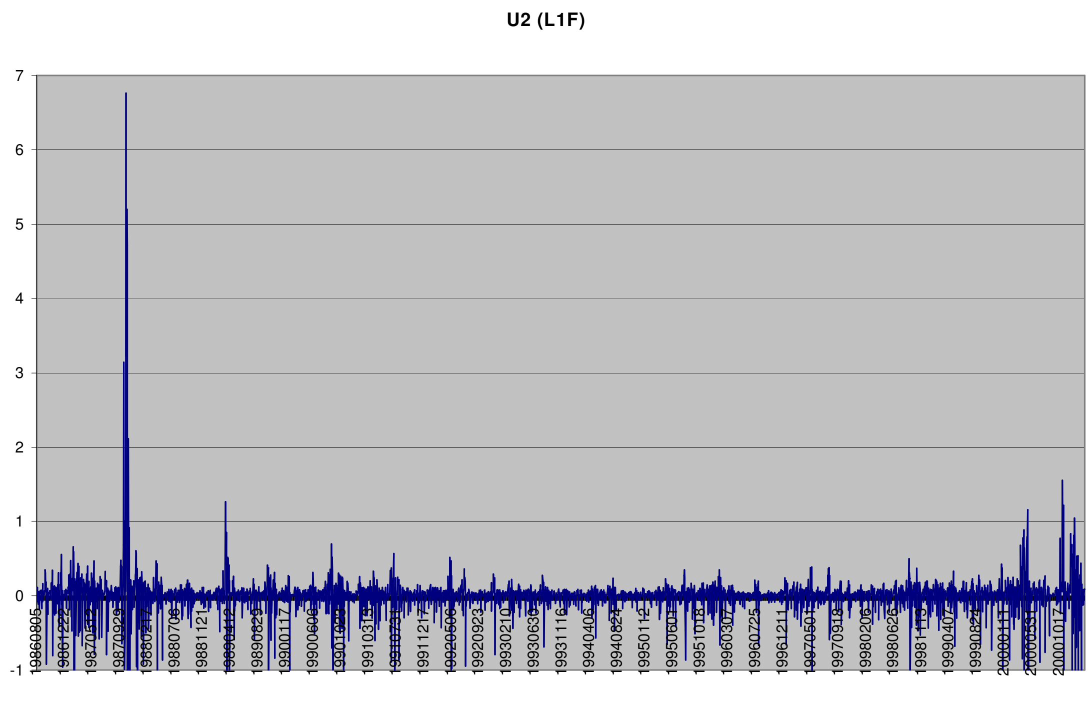


Figure 12