MAXIMUM LIKELIHOOD APPROACH FOR STOCHASTIC VOLATILITY MODELS

Jordi Camprodon i Masnou

Master Thesis supervisor: Josep Perelló i Palou

Departament de Física Fonamental, Universitat de Barcelona, Diagonal 647, 08028 Barcelona.

Master in Computational and Applied Physics (UB-UPC)

September 2011

Abstract

Volatility is a measure of the amplitude of price return fluctuations. Despite it is one of the most important quantities in finance, volatility is a hidden quantity because it is not directly observable. Here we apply a known maximum likelihood process which assumes that volatility is a time-dependent diffusions coefficient of the random walk of the price return and that it is a Markov process. We use this method using the *expOU*, the *OU* and the *Heston* models which are previously imposed. We find an estimator of the volatility for each model and we observe that it works reasonably well for the three models. Using these estimators, we reach a way of forecasting absolute values of future returns with current volatilities. During all the process, no-correlation is introduced and at the end, we see that volatility has non-zero autocorrelation for hundreds of days and we observe a significant correlation between volatility and price return called leverage effect. We finally apply this methodology to different market indexes and we conclude that its properties are universal.

Keyboards: *stochastic volatility, maximum likelihood, estimator of volatility*

1 Introduction

The first idea of modeling stock prices was given by Bachelier in 1900 with the arithmetic Brownian dynamics [1]. Its properties had some disagreements with empirical prices and, correcting that, Osborne proposed the Geometric Brownian Motion model (GBM) which was used for many years [12]. In this definition, volatility was associated with the diffusion coefficient of a random walk.

However, after the 1987 crash, it was noticed that Geometric Brownian Motion was unable to reproduce the behaviour of real markets [2]. In order to solve this huge problem, the stochastic volatility models were proposed. These models assume that stock price and volatility are random variables.

In the present work, we will focus our attention on volatility. It is a measure of the amplitude of return fluctuations and it is associated with the risk of holding an asset. Concretely, the higher the volatility the riskier the market index. In fact, volatility is such an important variable that investors pay sometimes more attention to volatility than to the direct price of the stocks [3] and eventually they trade through derivatives (futures and options).

One of the main properties of the volatility is that while returns themselves are uncorrelated, the absolute value of returns or their squares have a positive, significant and slowly decaying autocorrelation function. This issue can be observed in the market because, as Mandelbrot says, large changes tend to be followed by large changes, and small changes tend to be followed by small changes [7].

It is known that volatility has extremely useful properties [11, 8] but the biggest problem is that volatility itself is not observed. This fact implies some curiosities because, on one the hand, it is impossible to find the best estimator of volatility but, on the other hand, many models of volatility are not incorrect.

Among all the stochastic volatility models [5], we will study the exponential Ornstein-Uhlenbeck (expOU), the Ornstein-Uhlenbeck (OU), and the Heston models [6, 13, 9]. More concretely, we will try to apply the maximum likelihood methodology presented in [4] to all these models. Then, we will compare some properties in order to see if this methodology allows us to describe the market in a better way. We will do all of that using different indexes whose data has been downloaded from *Yahoo Finance*.

This work is divided into five sections. In Section 2 we present the stochastic volatility models and their main characteristics, while in Section 3 we show the
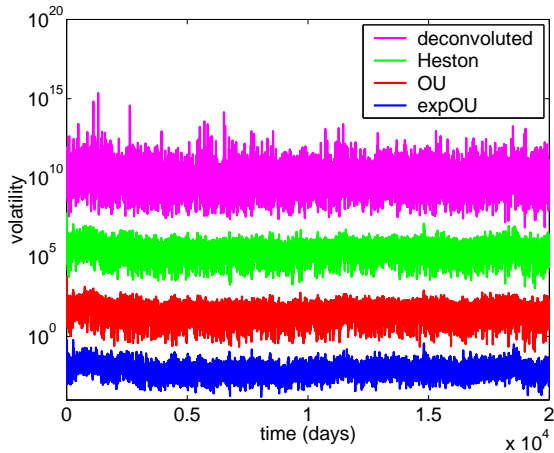
Figure 1: A comparison between different volatilities. On the top we observe the deconvoluted volatility computed using equation 11. The other ones are volatilities calculated using our maximum likelihood algorithm applied to different models. The Heston, the OU and the expOU models are represented in this order from above. All the plots are shifted in order to observe them more clearly

maximum likelihood approach. In Section 4 we explain the results we get using our algorithms. Conclusions are written in Section 5.

2 The stochastic volatility market models

The starting point of any stochastic volatility model is the GBM model:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma(t)dW_1(t) \qquad (1)$$

where $S(t)$ is a financial price or the value of an index, μ is the drift and $\sigma(t)$ is a random volatility.

If we define the zero-mean return $X(t)$ as

$$X(t) = \ln\left(\frac{S(t+t_0)}{S(t_0)}\right) - \left\langle \ln\left(\frac{S(t+t_0)}{S(t_0)}\right)\right\rangle \qquad (2)$$

with t_0 the initial time, we can rewrite the GBM as

$$dX(t) = \sigma(t)dW_1(t) \qquad (3)$$

Lots of models assume that volatility is a function of another random process, $\sigma(t) = f(Y(t))$, and $Y(t)$ is also a diffusion process [6, 13, 9]. Then, we can work with these stochastic differential equations:

$$dX(t) = f(Y(t))dW_1(t) \qquad (4)$$
$$dY(t) = -g(Y(t))dt + h(Y(t))dW_2(t) \qquad (5)$$

where $W_i(t)$ $(i = 1, 2)$ are Wiener processes.

As $f(y)$ is a monotonically increasing function, we could confuse Y with volatility. That is because σ and Y have similar behaviour in the market.
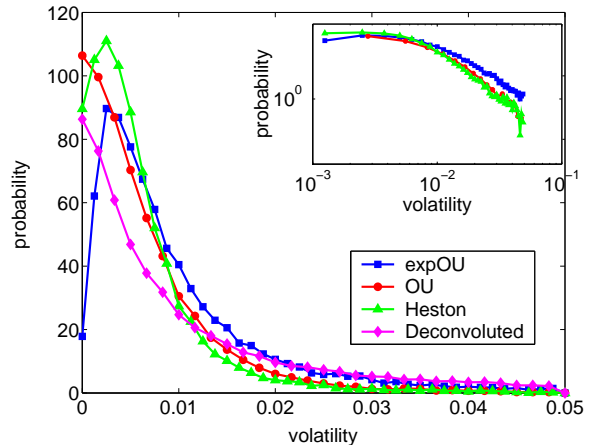

Figure 2: Probability distribution of the different volatilities. Inset figure shows the same representation in log-log scale.

	expOU	OU	Heston
$f(y)$	me^y	y	$y^{1/2}$
$g(y)$	αy	$\alpha(y-m)$	$\alpha(y-m)$
$h(y)$	k	k	$ky^{1/2}$

Table 1: Analytical expressions of $f(y)$, $g(y)$ and $h(y)$ that appear in equation 5. m, α and k are constants whose values are taken from [10].

Each stochastic volatility model has its own expressions of $f(y)$, $g(y)$ and $h(y)$. Table 1 summarizes the ones we are going to deal with.

Analyzing equation 5, $g(y)$ could be thought as a force that makes the volatility return to the normal level. $h(y)$ could be viewed as the volatility of the volatility.

3 Maximum likelihood approach

As we said in Section 1, we present a methodology proposed in [4] that allows us to have some criteria for choosing the best values of the volatility.

Let us imagine we have been able to get X and Y in the time interval $t - s \leq \tau \leq t$. Then, it can be demonstrated that the probability density of this realization is:

$$\ln P(X, Y) = -\frac{1}{2}\int_{t-s}^{t}\left(\frac{X'(\tau)}{f(Y(\tau))}\right)^2 d\tau$$

$$-\frac{1}{2}\int_{t-s}^{t}\left(\frac{Y'(\tau) + g(Y(\tau))}{h(Y(\tau))}\right)^2 d\tau + ... \qquad (6)$$

where X' and Y' are the derivatives of X and Y. Computationally, equation 6 is not feasible because X' and
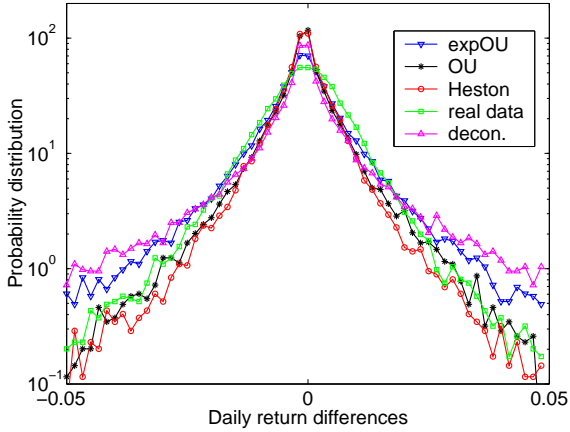
Figure 3: Comparison between the probability distributions of the return differences, dX, calculated using equation 3. Logarithmic scale is used in order to see more differences.
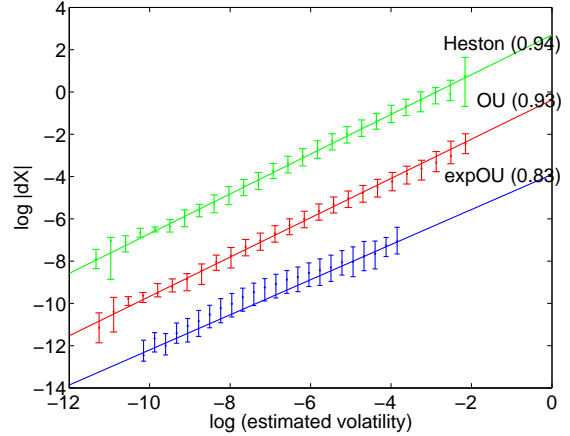

Figure 4: Logarithm of the return differences as a function of the logarithm of the volatility. All the models are shifted for better understanding. In brackets, we can find the value of the slope of the linear regression. The points represent the medians and the error bars are the first and third quartiles in the bins.

Y' do not exist as explicit functions and we only have discrete values of X and Y. For these reasons, we have to work with its discrete expression:

$$\ln P(X,Y) = -\frac{1}{2} \sum_{\tau=t-s}^{t} \left(\frac{X(\tau) - X(\tau - \Delta t)}{f(Y(\tau - \Delta t))\Delta t} \right)^2 \Delta t$$

$$-\frac{1}{2} \sum_{\tau=t-s}^{t} \left(\frac{Y(\tau) - Y(\tau - \Delta t)}{h(Y(\tau - \Delta t))\Delta t} + \frac{g(Y(\tau - \Delta t))}{h(Y(\tau - \Delta t))} \right)^2 \Delta t \quad (7)$$

The first term of equation 6 is a measure of the variations of X with respect to the volatility. We notice that the higher this fluctuations are, the lower the contribution to the probability is. The second part computes the fluctuations of the volatility with respect to the volatility of the volatility. Again, the bigger this term, the lower the contribution.

Our goal is to find a proper realization of Y given X. Then, we should consider the following conditional probability:

$$\ln P(Y|X) = \ln P(X,Y) - \ln P(X) \quad (8)$$

As we want to maximize this probability for a fixed set of X, the second term can be neglected. To sum up, if we compute different realizations of Y for the same X and we keep the one that makes bigger equation 6, we will be taking the realizations with smaller fluctuations. With this method, we can filter the Wiener noise $dW_1(t)$ and obtain our new estimation $Y_{est}(t)$ of the hidden volatility $Y(t)$.

One of the hard points of the project is the computational task. Specifically, we have implemented one algorithm which follows four steps:

1. Looking at equation 4, we generate a simple realization of Y as:

$$\bar{Y}_{est}(\tau) = f^{-1}\left(\left| \frac{dX(\tau)}{dW_1(\tau)} \right| \right) \quad (9)$$

where $t - s \leq \tau \leq t$.

2. We substitute $\bar{Y}_{est}$ and X into equation 7 and we compute the probability.

3. We iterate I times the points 1. and 2. and we keep the realization with higher probability. From this realization we get $Y_{est}(t) = \bar{Y}_{est}(t)$.

4. Finally, the estimator of the volatility at time t is $\sigma_{est}(t) = f(Y_{est}(t))$.

We observe that this procedure strongly depends on I and s. We also notice that the stochastic volatility model has to be chosen before starting the computation.

4 Simulation results and comparison between models

We have implemented the algorithm with $s = 10$ and $I = 100000$ and we have calculated our estimation of the volatility. We have used these values because if they were bigger our results would not be clearly improved. As we have taken daily data, we have worked with $\Delta t = 1$ day. In order to see if our methodology works, we have compared it with other ways of calculating the volatility.

As a first approximation, volatility can be viewed as something proportional to return differences:

$$\sigma_{prop}(t) = \frac{|dX(t)|}{\langle |dW_1(t)| \rangle} \quad (10)$$
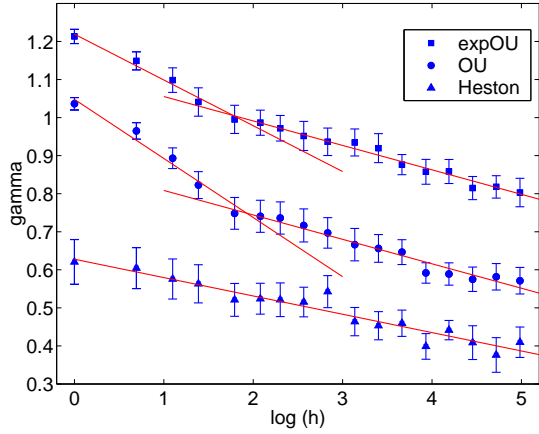
Figure 5: Representation of the magnitude $\gamma(h)$ that appears in equation 14. The errorbars correspond to the error on the slope of the regression of Figure 4. The data has been divided in two trams in the case of the expOU and the OU models. All the plots are also shifted.
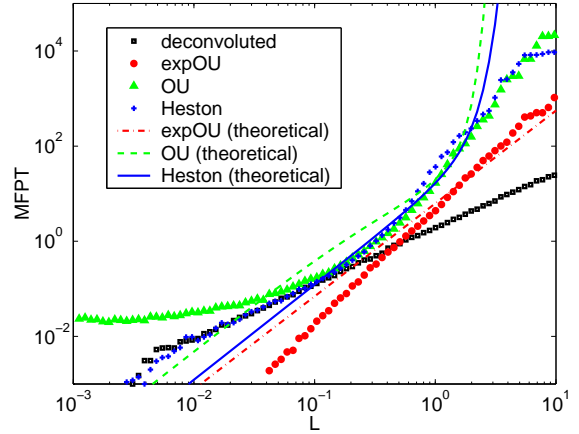

Figure 6: MFPT of the volatility of the three models and the deconvoluted. Lines represent the theoretical form that each curve should follow. Log-log scale is used in order to show the power-law behaviour.

This method might be too simple and we should add Wiener noise. Then, we can compute the deconvoluted volatility:

$$\sigma_{\text{decon}}(t) = \left| \frac{dX(t)}{dW_1(t)} \right| \qquad (11)$$

From equations 10 and 11, we can get $Y_{\text{prop}} = f^{-1}(\sigma_{\text{prop}})$ and $Y_{\text{decon}} = f^{-1}(\sigma_{\text{decon}})$. We observe that Y_{decon} is calculated with the first computed random value dW_1 while Y_{est} chooses an optimal value after I iterations.

In Sections 4.1 - 4.4 we have used Dow Jones daily data from October 1928 to July 2011.

4.1 Behaviour of our estimator

As we have seen, Y_{est} should be less noisy than Y_{decon}. We observe this fact in Figure 1 because the deconvoluted volatility oscillates much more than the estimated volatility calculated with all the models. In fact, the range of values of the deconvoluted is three or four orders of magnitude bigger than the others.

We know that our maximum likelihood approach chooses dW_1 but we do not know exactly if the different forms of $f(y)$, $g(y)$ and $h(y)$ clearly change the results. In order to compare how our algorithm works with each model, we have calculated the probability distribution of the different volatilities. We observe in Figure 2 that the position of the peaks and the height of them depend on the model. These differences show that the algorithm does not effect all the models in the same way.

In order to test the values given by each model, we should calculate $dX(t)$ multiplying $\sigma(t)$ with Wiener noise as it is written in equation 3. Doing that, we can compare our results with the real data of $dX(t)$. In Figure 3, we observe that the Gaussian shape of

the peak of the real $dX(t)$ is not reproduced in any model. On the other hand, we see that the tails of the real $dX(t)$ are very similar to the ones given by the OU model.

4.2 Predictive power of the method

Taking logarithms to equation 3, we get:

$$\ln |dX(t)| = \ln(\sigma(t)) + \ln |dW_1(t)| \qquad (12)$$

And if we use the conditional median of $\ln |dX(t)|$ given $\ln(\sigma(t))$ we should have

$$M[\ln |dX(t)| | \ln(\sigma(t))] = \ln(\sigma(t)) + c \qquad (13)$$

where c is a constant. In Figure 4 we plot this relationship using each estimator. We observe the slopes are near to 1 as they should be in the ideal case. We also notice that the Heston model is the one that reproduces better this dependence because its slope is equal to 0.94 while the other slopes are 0.93 and 0.83.

In order to see the prediction power of each model, we can forecast $dX(t+h)$ given $Y(t)$. From equation 13, we can give the following proposal [4]:

$$M[\ln |dX(t+h)| | \ln(\sigma(t))] = \gamma(h)\ln(\sigma(t)) + c \qquad (14)$$

If we calculate $\gamma(h)$ for some values of h, we can see the degree of predictability of each model. Figure 5 shows this issue and we observe that there is an acceptable linear relation between $\gamma(h)$ and $log(h)$. We also see that the expOU and the OU models have different slopes for small[1] and large[2] values of h. Using that relation we can write

$$M[\ln |dX(t+h)| | \ln(\sigma(t))] = (a\ln(h)+b)\ln(\sigma(t)) + c \qquad (15)$$

	expOU[1]	expOU[2]	OU[1]	OU[2]	Heston
a	-0.12	-0.064	-0.15	-0.064	-0.048
b	0.82	0.72	0.85	0.67	0.63

Table 2: Experimental values of the coefficients of equation 15. [1] is valid for $h \lesssim 7$ while [2] works for $h \gtrsim 7$.
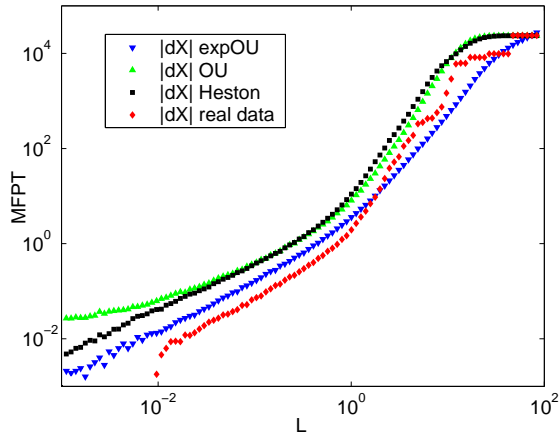

Figure 7: MFPT of the return differences calculated using equation 3. The expOU, the OU and the Heston models are compared with the well-known real data.

where a and b are the coefficients of the regression. Table 2 shows their experimental values.

We remark that, with equation 15, we have found a strategy that allows us to forecast $|dX|$ at time $t+h$ knowing information at t. We have also observed that the loss of information has a double time scale in the expOU and the OU models while the Heston model has a single time scale.

4.3 Mean First-Passage Time

In this section we are going to analyse the Mean First-Passage Time (MFPT) of the volatility. The MFPT of the volatility represents the mean time one has to wait in order to observe the volatility in one concrete value λ.

In order to compare different methodologies, we have to work with the dimensionless magnitude $L = \lambda/\sigma_s$ where σ_s is the volatility's normal level which depends on the volatility model and the stock data. Specifically, we have used the values in Table 3 that are calculated in [10].

	expOU	OU	Heston
σ_s	$me^{1/4\nu^2}$	m	$\gamma/\nu\sqrt{2}$

Table 3: Expressions of the volatility's normal level of the three models. The expressions of the constants are $\nu = \alpha^{1/2}/k$, $\gamma = \Gamma(\beta + 1/2)/\Gamma(\beta)$, $\beta = (1 + \rho\nu^2)/2$ and $\rho = m^2 - k^2/4\alpha$.

expOU	OU[3]	OU[4]	Heston[3]	Heston[4]	decon
2.4	0.5	3.3	1.5	3.1	1.1

Table 4: Scaling exponents of the MFPT of the volatility calculated with our maximum likelihood method. The exponent of the deconvoluted procedure is also written. The OU and the Heston models have two different exponents for $L \lesssim 0.2$ [3] and $L \gtrsim 0.2$ [4].
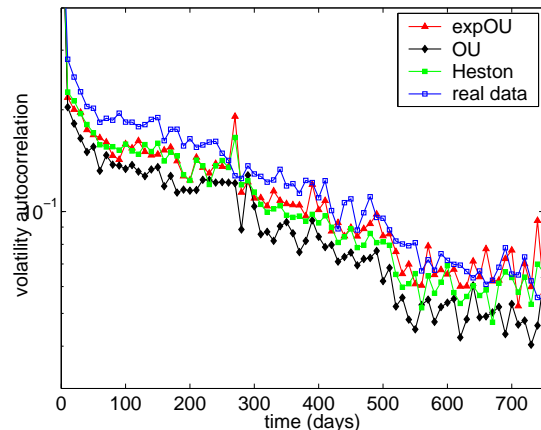

Figure 8: Comparison between the autocorrelation of our volatilities with the autocorrelation of the "real" one. The "real" volatility is calculated as the square of the return differences, $dX(t)^2$.

In Figure 6 we show the MFPT of the volatility calculated with our models. We observe that the MFPT of the volatility follows a power law in all the cases. The expOU model and the deconvoluted process have only one exponent for the whole range of L. The other ones are described with different exponents for small[3] and big[4] values of L. The experimental scaling exponents are summarized in Table 4.

In Figure 6 we also compare our experimental results with the theoretical curves given by [10]. We notice they differ significantly. In fact, big values of L in the expOU model and medium values in the Heston model are the only points that coincide with the theoretical curves. This low agreement could be attributed to the way of accepting the Wiener noise. In other words, our algorithm radically changes the range of values of the volatility and it strongly affects the behaviour of the MFPT.

Finally, the deconvoluted points that appear in the plot are lower for big L. As the deconvoluted volatility has bigger peaks and they appear more frequently, the time needed to get a huge value of L is significantly smaller.

As the volatility is a hidden variable and we only observe the price, we can calculate $|dX|$ multiplying each σ by $dW_1(t)$. Doing that, we can compare the MFPT of the experimental $|dX|$ with the MFPT of the real $|dX|$. In Figure 7 we see that each model has two

	expOU	OU	Heston	real data
$L \lesssim 1$	1.1	0.8	1.0	1.3
$L \gtrsim 1$	2.4	3.1	2.9	2.9

Table 5: Scaling exponents of the MFPT of dX. All the curves in Figure 7 have a characteristic exponent for $L \lesssim 1$ and another for $L \gtrsim 1$.
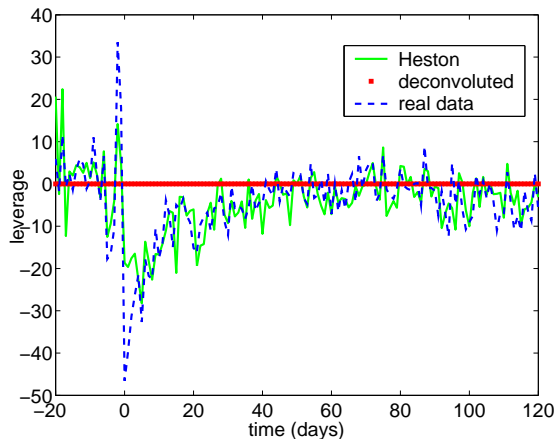

Figure 9: Leverage correlation given by equation 16 of the Heston model. It is compared with the deconvoluted procedure and the "real" volatility, $dX(t)^2$.

different exponents for $L \lesssim 1$ and $L \gtrsim 1$. In addition, we see that all the exponents are very similar. Their values are shown in Table 5.

4.4 Correlations

In this section we have studied how our maximum likelihood approach effects the main correlations of the volatility.

It is well-known that volatility fluctuations have long memory and that the autocorrelation function decays slowly. In Figure 8 we have plotted the volatility autocorrelation of each estimator. We observe that all of them preserve this property because they show significant and positive autocorrelation for hundreds of days. This fact clearly manifests the robustness of the proposed method.

The other important correlation that has to be studied is the leverage effect. The leverage correlation is defined by

$$\mathcal{L}(\tau) = \frac{\langle dX(t)\sigma(t+\tau)^2 \rangle}{\langle \sigma(t)^2 \rangle^2} \quad (16)$$

and it is a measure of the correlation between the variations of return and volatility.

In Figure 9 we have plotted the leverage correlation of the Heston model as an example. We observe there exist negative decaying correlation for some days. That fact is really important because in the analytical calculation of equation 6 the correlation between
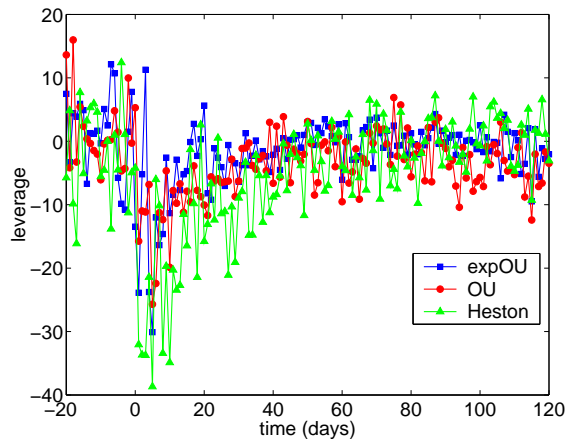

Figure 10: Leverage correlations of the expOU, the OU and the Heston models. As always, volatilities are calculated using the maximum likelihood method.

$dW_1(t)$ and $dW_2(t)$ is not imposed. It means that the anticorrelation of the volatility and the return appears without previous introduction. In some way, it could be a proof that our methodology is coherent because simply playing with the Wiener noise, the leverage effect is manifested.

In order to see if it happens in all the models, we have plotted the leverage effect of each estimator in Figure 10. It is clear that all the models show this negative correlation. In addition we see that the Heston model is the one whose anticorrelation is bigger.

It should be noticed again that although no previous relationship between $dW_1(t)$ and $dW_2(t)$ is introduced, we find an important negative correlation between return and volatility.

4.5 Different market indexes

We have studied how our maximum likelihood approach effects different stochastic volatility models. Here, we would also like to verify if there are any differences between working with one stock market or another.

Concretely, we have calculated our estimation of the volatility for the following indexes: Dow-Jones Industrial Average (DJ), Standard and Poor's-500 (S&P), German index DAX, Japanese index NIKKEI, American index NASDAQ, British index FTSE-100, Spanish index IBEX-35 and French index CAC-40.

The first thing we realize is that each volatility's stock has its own range of values. It means that we cannot compare directly the values of the volatility. What we observe is that in all the markets, the estimated volatility is considerably less noisy than the deconvoluted. In addition, we could say that the reduction of the oscillations is done in a similar way because the coefficient

$$\frac{\text{variance}(Y_{\text{est}})}{\text{variance}(Y_{\text{decon}})} \quad (17)$$
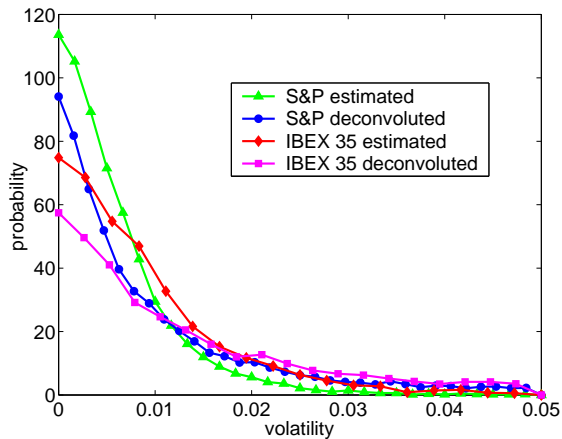
Figure 11: Probability distribution of the volatilities of the Standard and Poor's-500 (S&P) and the IBEX-35. For each one, it is plotted our estimated and the deconvoluted volatilities. The estimated is calculated using the Heston model.
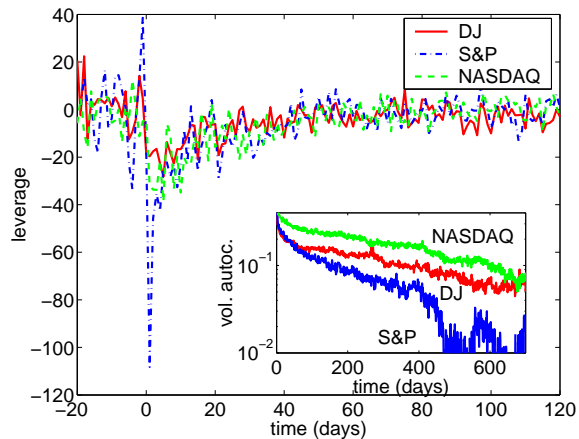

Figure 12: Comparison between the leverage effect of the Dow-Jones Industrial Average (DJ), Standard and Poor's-500 (S&P), and NASDAQ. The inset Figure shows the decay of the volatility autocorrelation of them. The Heston model has been used in order to calculate the volatility estimator of each stock.

is independent of the stock data and it only depends on the model.

In Figure 11 we plot the volatility given by the Heston model for two different indexes. As we said, we notice the different width of the probability distribution of two stocks because each one has a different range of values. We can also appreciate the reduction of the fluctuations achieved with our estimated volatility.

Finally, we have tried to deduce if the appearance of volatility's correlations is general for all the markets. For the Heston model, we show in Figure 12 that there are some stocks which manifest more leverage than others. As an example, the S&P has bigger anticorrelation than the Dow Jones. However the important fact is that we find leverage in all the stocks. The same happens with the volatility autocorrelation because although the NASDAQ decays more slowly, all the stocks manifest significant autocorrelation for hundreds of days.

5 Conclusions

It is fairly known that volatility is one of the main quantities in finance because it is a measure of price fluctuations and it gives information related to the risk of holding an asset. In this work, we have used a maximum likelihood method in order to obtain optimal values of the volatility.

We have applied it to the expOU, the OU and the Heston models and we have compared them with a deconvoluted method. We have realized that fluctuations of the estimated volatility are smaller in all the models than in the decovoluted calculation.

The Heston model has been the best one to show that there is a linear relationship between the logarithms of return and volatility. In addition, the nov-

elty is that we have found a strategy that allows us to forecast future returns with actual volatilities. We have also observed that the loss of information has a double time scale in the expOU and the OU models.

Speaking about extreme events, we have seen that our maximum likelihood approach does not reproduce the MFPT of the volatility as we expected. However, we have observed the nice concordance between the MFPT of our estimated return differences and the MFPT of the real ones.

We have also focused on volatility's correlations and we have realized that all the models show the existence of significant volatility autocorrelation for hundreds of days. In addition, correlations between variations of return and volatility have also been observed. This last issue, also called leverage effect, has been surprisingly found because we have not introduced previous relationship between returns and volatilities. All of that has allowed us to say that our methodology is really robust.

Finally, we have computed the same procedure using different stock indexes. We have seen that the reduction of volatility's noise follows the same pattern and we have corroborated that all the markets show volatility autocorrelation and leverage effect. In other words, we have observed that all this methodology shows universal properties.

References

[1] L. Bachelier. Theory of speculation. *The random character of stock market prices*, 1018:17–78, 1900.

[2] J. Bouchaud and M. Potters. *Theory of financial risk and derivative pricing: from statistical*

physics to risk management. Cambridge Univ Pr, 2003.

[3] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[4] Z. Eisler, J. Perelló, and J. Masoliver. Volatility: a hidden markov process in financial time series. *Physical Review E*, 76(5):056105, 2007.

[5] J. Fouque, G. Papanicolaou, and K. Sircar. *Derivatives in financial markets with stochastic volatility*. Cambridge Univ Pr, 2000.

[6] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327, 1993.

[7] B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[8] R. Mantegna and H. Stanley. *An introduction to econophysics: correlations and complexity in finance*. Cambridge Univ Pr, 2000.

[9] J. Masoliver and J. Perelló. Multiple time scales and the exponential Ornstein-Uhlenbeck stochastic volatility model. *Quantitative Finance*, 6(5):423–433, 2006.

[10] J. Masoliver and J. Perelló. Extreme times for volatility processes. *Physical Review E*, 75(4):046110, 2007.

[11] J. McCauley. *Dynamics of markets: econophysics and finance*. Cambridge University Press Cambridge, 2004.

[12] M. Osborne. Brownian motion in the stock market. *Operations research*, pages 145–173, 1959.

[13] J. Perelló, J. Masoliver, and N. Anento. A comparison between several correlated stochastic volatility models. *Physica A: Statistical Mechanics and its Applications*, 344(1-2):134–137, 2004.